TEMPORAL FUZZY COGNITIVE MAPS FOR CORPORATE CREDIT RATINGS

HAOMING ZHONG

NANYANG TECHNOLOGICAL UNIVERSITY

School of Computer Engineering

A thesis submitted to the Nanyang Technological University in fulfilment of the requirement for the degree of Doctor of Philosophy

2015
Abstract

Corporate credit ratings are one of the key problems of the credit risk management, which has attracted much research attention since the credit crisis in 2007. Methodologies and models based on finance, statistics, and expert knowledge have been developed in past decades. The essential problem of corporate credit rating is finding out the relationships between financial factors and associating them to generate the credit assessment. Artificial Intelligence (AI) technologies such as neural networks and SVMs have demonstrated their remarkable performance on automatic credit ratings. However, the relationships between input financial factors and their rating results is not interpretable, which brings difficulties to justify and revise models. FCM is a hybrid approach combining advantages of expert systems and AI approaches, which makes it a promising approach for the credit rating problem. The work described by the thesis investigates three topics of FCM on the corporate credit rating problem, including

(1) the FCM structure and the training algorithms,

(2) relationships among FCM, correlation, and causation, and

(3) temporalized FCMs.

To study the capacity of FCM on credit rating, an experimental comparison study over the effectiveness of five learning algorithms, i.e., BP, ELM, I-ELM, SVM, and FCM, is carried out targeting at comparing FCM with other AI approaches on the corporate credit rating problem. The effectiveness of the five algorithms is studied in terms of reliability and discrimination capacity. The experimental results show that Neural network based solutions, including BP, ELM, I-ELM, FCM, outperform SVM on reliability because they have better error distribution, while the SVM achieves a better performance on the discrimination testing. FCM, with a smaller structure than SLFN, obtains an equivalent performance as other approaches. It shows that the structure of FCM is more efficient than SLFN.

To further study the capacity of FCM on credit rating with small sample size, an individual rating case is studied. Nokia, as a telecom corporation, has been downgraded from A1 to B1 in past six years. Due to its wide rating variation in six years, Nokia is an
appropriate example to test the performance of different AI models on single corporation analysis with incomplete sample data. By exploiting correlations between financial factors as priori knowledge, the FCM outperforms BP and SVM on the reliability. The result also indicates that SVM cannot identify cases which are not covered by the training samples. In addition, another experimental study has been carried out to compare the difference brought by the correlation combination. The study shows that the combination improves the reliability of the FCM and the consistency of FCM training results as well. The high consistency between trained FCMs facilitates the interpretation and justification of training results. Finally, the study demonstrates that there are strong temporal correlation between different financial factors.

As one of the three causation rules, the temporal attributes play an essential role in the causality identification. FCM is a dynamic causal inferring tool. By temporalizing its relationships, FCM is able to support temporal causal inferring. The novel temporalized FCM with gradient descent learning algorithm is proposed to study the causal relationships between corporate financial factors. Three corporations, Nokia, Ericsson, and Google, are used to verify the proposed approach. The result shows that more than 50% of correlations have been eliminated by the training algorithm. The remained relationships, which are called “h-causations”, are different from the initial correlations. An analysis on some “h-causations” indicates that they can be justified based on corporation financial conditions, which gives a subjective proof that “h-causations” are closer to causations than correlations.

Based on the gradient descent algorithm, a temporalized FCM can be built by using machine learning. Meanwhile, FCM ca be built based on human expertise. A novel temporal FCM model, tFCM, is proposed to map human expertise to temporalized FCM with mapping patterns. An empirical study compares the difference between two kinds of causal relationships, the “if ... then ...” relationship and the “changes followed by changes” relationship, on error accumulations and sensitivities.

**Keywords:** Corporate credit rating, ELM, I-ELM, SLFN, Neural network, SVM, FCM, Causality inference, Temporal fuzzy knowledge.
Acknowledgments

First, I would like to express my deep gratefulness to my supervisor, Dr. Miao Chunyan, for her patient and systematic guidance, heuristic advice and encouragement in my research work and the writing up of this thesis.

Next, I would give my thankfulness to Dr. Shen Zhiqi and Dr. Huang Dong for their patience of listening and giving me valuable advice.

Last but not least, I am deeply grateful to my wife, Yuhong, for her heuristic advice and encouragement for my research work, and consistent support throughout my Ph.D. study.
## Contents

Abstract ................................................................................................................................. i  
Acknowledgments ................................................................................................................... iii  
List of Figures ......................................................................................................................... viii  
List of Tables .......................................................................................................................... x  

1 Introduction .......................................................................................................................... 1  
1.1 Background ....................................................................................................................... 1  
1.1.1 Corporate credit ratings .............................................................................................. 1  
1.1.2 Credit rating techniques ............................................................................................. 3  
1.2 Scope and objectives ......................................................................................................... 4  
1.3 Thesis organization .......................................................................................................... 5  

2 Literature review .................................................................................................................. 7  
2.1 Corporate credit rating and its methodologies ................................................................. 7  
2.2 Credit Rating Methodologies and performance evaluation ............................................. 7  
2.2.1 Subjective analysis and expert systems ..................................................................... 8  
2.2.2 Financial models ....................................................................................................... 8  
2.2.3 Statistical models ...................................................................................................... 10  
2.2.3.1 MDA models ....................................................................................................... 11  
2.2.3.2 Conditional probability models ......................................................................... 12  
2.2.4 Scorecard .................................................................................................................... 12  
2.2.5 Performance testing .................................................................................................... 14  
2.2.5.1 The error testing ................................................................................................. 14  
2.2.5.2 ROC and CAP curves ....................................................................................... 16  
2.3 Machine learning .............................................................................................................. 17  
2.3.1 Artificial neural networks .......................................................................................... 19  
2.3.2 Support vector machine ............................................................................................. 22  
2.3.3 Fuzzy cognitive maps ................................................................................................ 23  
2.3.3.1 Layered FCMs .................................................................................................. 29  
2.3.3.2 Training of FCMs .............................................................................................. 30
2.3.3.3 Causation and correlation in FCMs .............. 34
2.4 Summary ........................................... 36

3 A comparison work of machine learning based credit rating models 37
3.1 Data sets ............................................. 38
3.1.1 Raw financial data acquisition ......................... 39
3.1.2 Data pre-process .................................. 39
3.1.3 Data normalization ................................ 40
3.2 Experimental environments and parameters .............. 41
3.2.1 ELM parameters .................................. 42
3.2.2 I-ELM parameters ................................ 43
3.2.3 BP parameters .................................... 43
3.2.4 SVM parameters .................................. 44
3.2.5 FCM parameters .................................. 44
3.3 Experiment results ................................... 46
3.3.1 The reliability .................................... 46
3.3.2 The discrimination capacity ......................... 50
3.3.2.1 Rating distribution .............................. 50
3.3.2.2 ROC testing .................................... 54
3.3.2.3 Summary of the results ....................... 61
3.4 Conclusion .......................................... 62

4 FCM based credit rating model: a case study of Nokia 65
4.1 A FCM based methodology .............................. 66
4.1.1 The FCM structure ................................ 66
4.1.2 Longest path inferring ............................. 67
4.1.3 GA based learning algorithm ......................... 69
4.1.4 Priori knowledge combination ....................... 70
4.1.4.1 Concepts ...................................... 70
4.1.4.2 Relationships ................................. 71
4.2 The FCM credit rating model ......................... 71
4.2.1 Factor selection .................................. 71
4.2.2 Data preparation .................................. 73
4.2.2.1 Sample selection .............................. 73
4.2.2.2 Data normalization ............................. 75
4.2.3 Correlation coefficient ............................. 76
4.3 Experiment result .................................... 76
4.3.1 Learning error ................................... 77
4.3.2 Testing error ........................................ 78
4.3.3 The consistency .................................. 78
4.4 Summary and future work ............................ 81

5 Temporized fuzzy cognitive maps for corporate financial factor analysis 83
5.1 Temporal extensions of FCMs ...................... 84
  5.1.1 Discrete temporal extensions .................... 84
5.2 Temporalized FCM and learning algorithm ....... 86
  5.2.1 The definition of temporalized FCM .......... 86
  5.2.2 Learning algorithm ................................ 87
    5.2.2.1 Weight initialization ...................... 88
    5.2.2.2 Gradient descent learning ................. 88
    5.2.2.3 Learning speed and relationship refinery 89
5.3 Experiment result ................................... 90
  5.3.1 Data preparation ................................ 90
  5.3.2 Result ........................................... 91
    5.3.2.1 Learning accuracy ......................... 91
    5.3.2.2 H-causation and correlation .......... 91
    5.3.2.3 The h-causation and causation ..... 96
  5.3.3 Experiment summary ............................. 97
5.4 Conclusion .......................................... 98

6 The tFCM and tFCM studio 100
6.1 Temporal Fuzzy Cognitive Maps ................... 100
  6.1.1 The formal definition of tFCM ............... 101
    6.1.1.1 Temporal attributes ..................... 102
    6.1.1.2 The temporalization ...................... 103
  6.1.2 tFCM patterns .................................. 106
    6.1.2.1 Concept patterns ......................... 106
    6.1.2.2 Effect patterns .......................... 107
6.2 Design tFCMs with tFCM studio ................. 110
  6.2.1 Specify global attributes .................... 110
  6.2.2 Design concepts ................................ 111
  6.2.3 Design causal relationships .................. 113
6.3 Dynamic maps and balance maps .................. 113
6.4 Experiments ........................................ 115
  6.4.1 Evolution results .............................. 120
  6.4.2 Error analysis ................................ 120
6.5 Summary ............................................ 122
7 Conclusions

7.1 Summary of contributions .................................................. 125
7.2 Recommendations for future work ........................................ 131
7.3 Summary ............................................................................ 132
Publications ............................................................................. 133

A Corporate financial statement .................................................. 134
 A.1 Income statements ............................................................. 134
 A.2 Balance sheets ................................................................. 134
 A.3 Cash flow statements ......................................................... 136

B Financial ratios ....................................................................... 138
 B.1 Profitability ratios ............................................................. 138
 B.2 Liquidity ratios ................................................................. 139
 B.3 Activity ratios ................................................................. 139
 B.4 Debt ratios ........................................................................ 140
 Bibliography ............................................................................ 141
List of Figures

2.1 The illustration of KMV model ........................................ 9
2.2 Result distributions from different rating models .................. 18
2.4 A simple fuzzy cognitive map ......................................... 24
2.5 The FCM of the bad weather driving ............................... 26
2.6 The dynamics comparison with different initial states ............. 27
2.7 The dynamics comparison with different causal relations ......... 28

3.1 The testing error distribution of ELM ............................... 42
3.2 The testing errors of three I-ELM learning trials .................. 44
3.3 The testing error distribution of BP ................................ 45
3.4 The FCM structure for credit rating ................................. 46
3.5 The rating distributions for MIS and SnP testing data ............. 51
3.6 The output distributions for MIS testing data ..................... 52
3.7 The output distributions for SnP testing data ..................... 53
3.8 The ROC estimation of 1 year default ............................. 56
3.9 The ROC estimation of 5 years default ............................ 57
3.10 The ROC estimation of 10 years default .......................... 58

4.1 The sample of layered FCMs ........................................ 67
4.2 The converting from a weight matrix to a chromosome .......... 69
4.3 The FCM model for credit ratings .................................. 73
4.4 The normalized financial data of Nokia ........................... 74
4.5 Comparison of weight distributions ................................ 80

5.1 A hypothetical FTCM ................................................ 86
5.2 The correlation to correlation coefficient ........................ 92
5.3 The comparison between correlation and h-causation ............. 93
5.4 Top 3 most influencing factors in Nokia .......................... 94
5.5 Top 3 most influencing factors in Ericsson ....................... 94
5.6 Top 3 most influencing factors in Google ........................ 95
5.7 The \( h \)-causations of R&D-sales and R&D-profit for three corporations . 96
5.8 The h-causation between gross profit and assets ........................................... 98
6.1 The elements of FCM .................................................................................. 103
6.2 A sample effect function .............................................................................. 105
6.3 The semantic diagram of the tFCM ............................................................... 105
6.4 The definition of a base temporal condition ................................................ 109
6.5 The tFCM studio .......................................................................................... 109
6.6 The UI form of time domain setting .............................................................. 110
6.7 The UI form of concepts in tFCM studio .................................................... 112
6.8 The UI form of cause associations ................................................................. 113
6.9 The theoretical error rate against the number of iterations ......................... 116
6.10 The model of experiments .......................................................................... 117
6.11 The temporal domain and fuzzy time sets .................................................. 117
6.12 The evolving results of two different models ............................................. 121
6.13 The errors of two different models ............................................................. 123
## List of Tables

2.1 An example of the financial performance measurement. .......................... 12  
2.2 An example of the credit assessment scorecard. ................................. 13  
2.3 The different rating results .......................................................... 16  
3.1 Selected financial ratios. ........................................................................ 40  
3.2 Data sets for experiments ..................................................................... 40  
3.3 The normalization of credit ratings ....................................................... 41  
3.4 The learning errors, testing errors, and overfitness comparison. .......... 48  
3.5 The error distribution comparison. ....................................................... 50  
3.6 The conformation of rating distributions for ELM, I-ELM, BP, and SVM. 54  
3.7 Moody’s corporate idealized 1, 5, and 10 years cumulative probability of default (PD) rates .......................................................... 55  
3.8 The threshold points and distance to (0, 1) .......................................... 60  
3.9 The best two approaches on each measurement. ................................. 62  
4.1 The Nokia’s rates given by Moody’s Investor Services since 2001 ........ 74  
4.2 The 21-level rate map. ........................................................................ 76  
4.3 The correlation coefficients between financial factors and credit ratings 76  
4.4 Training errors for different algorithms ............................................... 77  
4.5 Testing errors for different algorithms ............................................... 78  
4.6 The consistency comparison between FCM (C) and FCM (N) ............. 79  
5.1 The error of models. ............................................................................. 91  
6.1 The decaying rate for concepts ........................................................... 118  
6.2 The temporal relationships between concepts ...................................... 119  
7.1 The comparison between tFCM and other temporal FCM extensions .... 130  
A.1 Sample income statements ............................................................... 135  
A.2 A sample cash flow statement ........................................................... 137
Chapter 1

Introduction

The corporate credit rating is one of the fundamental components of the modern financial engineering. It indicates the financial health of corporates. Corporate credit ratings are costly processes, which require both financial and industrial expertise. Previous work showed that artificial intelligent technologies can help to reduce the cost and increase the accuracy of corporate credit ratings. Fuzzy cognitive maps (FCMs) are soft computing tools, which naturally combine the advantages from Artificial Intelligence (AI) technologies (e.g., neural network) and the human participations. They have the potential to conduct corporate credit ratings. This thesis addresses three key issues in exploiting FCMs for corporate credit ratings: the model performance on credit ratings, the capability on combining priori knowledge with learning algorithm, and temporal relationship analysis.

1.1 Background

1.1.1 Corporate credit ratings

The credit risk management (CRM) has attracted much research attention from industries and academics since the credit crisis in 2007 [1,2]. The CRM provides accurate prediction, measurement, monitoring and reports on the possible credit risk exposure, which can help financial firms reduce financial loses. In general, the CRM includes corporate credit data access and aggregation, corporate credit ratings (in the rest of the thesis, "credit ratings" is used for short), and decision supports. Among them, credit ratings, which is also named as “credit scorings” in literature [3,4], plays a central role in CRM. A credit rating is an
CHAPTER 1. INTRODUCTION

alphanumeric symbol indicating the ability and willingness of a corporation to meet its financial liabilities in full and on time. It also indicates the financial health of the corporation and the relative likelihood that the corporation may default. A lower credit rating is an indicative of a lower likelihood \(^1\). In summary, credit ratings indicate the counterparties’ default probabilities for a financial firm and they are used to support business decisions on the customer relationship management, lending application assessment, product designing, pricing, and assets/liabilities management etc. In addition, they can also be used to calculate the regulator capital, run stress tests and scenario analysis, which are required by international capital management standards, i.e., the Basel II \([5]\) and the newly released Basel III \([6]\). The computation of credit ratings have the following features.

- **More objectives**: the credit ratings are not only used to support credit approval decisions, but also used to calculate customer’s credit limits, bond’s prices, and the regulatory capitals for financial firms;

- **More scale levels**: the credit ratings are scaled to more levels. Long term ratings from Moody’s and S&P are scaled to 21 levels and different levels represent different credit risks. That is, the values of credit ratings categorize corporations into 21 classes, which makes it a multi-class classification problem;

- **More attributes**: different regions and different industries may require different rating models to assess the credit risks; and

- **Temporal causal relations**: Credit ratings are long term inferring process based on time related knowledge, e.g., quarter revenue, long term risk, and short term risk. Temporal causal relations has been proven as important relations between corporates’ operation and financial factors \([7\text{-}10]\).

\(^1\)From Standard & Poor’s Rating Service website, credit ratings also cover the credit quality of an individual debt issue, a municipal bond, or a mortgage-backed security, etc. In this thesis, we focus on corporate credit ratings.
CHAPTER 1. INTRODUCTION

1.1.2 Credit rating techniques

There are two approaches to obtain credit ratings, i.e., the financial models and the statistical models [11–13]. Corporate credit scorecards (“scorecards” will be used in the remaining of the thesis for short) [14] are created to summarize results from different models into final credit ratings [15]. The logic and calculation process used in scorecards are straightforward and easy to be understood. In addition, factors and weights in scorecards can be adjusted to reflect the market changes to obtain more accurate results. However, the calculation of credit ratings in scorecards have to involve expertise of finance and industries as well. Meanwhile, the scorecard model itself does not capture any knowledge of the interconnection between the raw data and the listed factors. The credit score is given based on present financial factors rather than future predictions. That is, scorecard users have to undergo a high learning curve on the manual process in order to be proficient in the process, and they are also criticized for their objectivity on the factor and weight. Moreover, there is no systematic automation process for the result verification, adjustment, and justification.

Artificial Intelligent (AI) technologies have been studied for credit ratings for years, e.g., artificial neural networks (ANNs) [16,17] and support vector machines (SVMs) [18,19]. A comparison study has been carried out to compare the performance of SVMs and ANNs in credit ratings [20] and the experimental results show that ANNs achieve the accuracy comparable to those of SVMs. That is, both approaches demonstrate their capability on producing remarkable prediction on credit ratings.

ANNs are usually trained by gradient based algorithms, i.e. backpropagation (BP) algorithms, and the BP algorithms have been criticized on the difficulty to decide learning rates, ease to stuck on local minimums, overfit problems, and time consuming [21]. Extreme Learning Machine (ELM) and Incremental extreme learning machine (I-ELM) have been proposed to overcome these drawbacks [21–25]. A comparison study has been carried out on ELM, SVM, and K Nearest neighboring (KNN) to evaluate consumer credit scoring [26]. The experimental results show that ELM is faster than SVM and KNN while obtaining close accuracy rates.
Although ANN and SVM approaches demonstrate their capability on data classification and recognition, they have not been widely accepted by financial industry as major assessment tools. Their results are hard to interpret by human experts for further verifications and justifications. While Fuzzy cognitive maps (FCMs), as soft computing tools, combine the advantages from ANN and expert systems. They are promising for conducting corporate credit ratings.

A Fuzzy Cognitive Map (FCM) is a directed graph, where vertices represent concepts, directed edges represent the causal effect relationships between concepts, and the weights of edges represent the degree of the causal effect [27,28]. FCMs are important mathematical models representing the structured causality knowledge for quantitative inference. They have advantages such as simple representation, inconsistent knowledge and circle causalities supporting for knowledge modeling and inferring [29]. Such advantages make FCMs widely accepted as useful tools in decision making [30–32], prediction [33], and strategic planning [34] in social science, medicine, engineering, business, production systems, environment, agriculture, and information technologies [29]. Particularly, Zhai et al. [35] applied the FCMs on bankruptcy prediction and achieved an impressive average accuracy rate, i.e., 89.58%.

1.2 Scope and objectives

In order to exploit capability of FCM on credit ratings, following three problems are investigated: comparing the FCM with other AI approaches on credit rating, the performance of FCM on credit rating with small sample size, and modelling sophisticated temporal relations.

First, in order to compare FCM with other learning algorithms for credit ratings, an experimental comparison study over the performances of five learning algorithms, i.e., BP, ELM, I-ELM, SVM, and FCM, is carried out. The performances of the five algorithms are compared in terms of reliability and discrimination capacity.

Second, for the purpose of further improving the capability of FCMs on credit rating while retaining its understandability, we investigate how to combine priori knowledge
CHAPTER 1. INTRODUCTION

with FCMs to model credit ratings problem. Meanwhile, we study the inner learning and inferring mechanisms regarding the FCMs, and compare the FCMs with other learning algorithms with small sample set. In addition, a real case study on the application of the FCM for credit ratings is carried out to verify the models and the theories.

Third, as one of causation characteristics, temporal relationships exist between corporate financial factors inherently. The capacity to learn temporal relationships is desirable for FCM as a causality modeling tool for credit rating. We study how to learn temporal relationships from given data and compare the results with other approaches.

Finally, in order to introduce a full temporal extension to FCMs to model human temporal causation expertise, we have designed and constructed a temporal Fuzzy Cognitive Map (tFCM) using a discrete linear temporal domain. Our work focus on presenting a formal definition of the tFCM, designing and constructing corresponding software tools to capture temporal causation expertise. Theoretical and empirical studies of different causality models have been carried out regarding the error accumulation and the sensitivity of the inference.

1.3 Thesis organization

The rest of the thesis is organized as follows.

- Chapter 2 gives an introduction on credit ratings and the related technologies, such as financial models, statistical models, scorecard, and the model testing. Besides of that, AI approaches, such as SLFNs, SVMs, FCMs, and their learning algorithm, are also introduced. The relationship between correlation and causation is discussed under the section of FCM introduction.

- Chapter 3 presents an experimental comparison study, where five learning algorithms are compared in terms of reliability and discrimination capacity with real credit rating data.
Chapter 1. Introduction

- Chapter 4 proposes a methodology to train FCM with small sample data by combining priori knowledge. A layered FCM model with an original inferring approach is proposed and verified by financial statements and credit ratings took from Nokia. The result is compared with BP and SVM.

- Chapter 5 studies on temporal relationship learning. The FCM has been redesigned and trained by a gradient descent learning algorithm. The training results are compared with other models and justified with the backgrounds of different companies respectively.

- Chapter 6 introduces a novel FCM model, tFCM, to capture human temporal causation expertise. A formal definition of tFCM and related tools are carried out in the chapter. The discussion on cause selection is also studied in terms of error accumulation and sensitivity.

- Chapter 7 summarizes the achievements of our research and gives a brief outline on the unsolved issues regarding rating models and corresponding learning algorithms.
Chapter 2

Literature review

2.1 Corporate credit rating and its methodologies

Corporations play a central role in the financial industry. Their financial activities cover loans, bonds, shares, acquisition, and other finance products. In most of corporate financial activities, corporations deal with financial firms. A corporation is called as a counterparty by financial firms whom it deals with. It is called a default when a counterparty is fail to fulfill its liabilities in a deal. Defaults bring loss to financial firms. From historical records, the defaults of a small number of counterparties may result in a very large loss for the financial firm [36]. To understand the default risk of a counterparty, firms have to assess the counterparty on its probability of default (PD), which is indicated by its credit rating. In financial market, a corporation is backed by its corporate credit rating. It is used for customer relationship management, limits allocation, credit approval, portfolio risk analysis, regulation, regulatory/economic capital calculation, pricing, performance measurement, debt structuring, securitization, and risk reporting [11]. Studies show that a precise credit rating methodology will largely help financial firms reduce their risk and improve their profit [37]. It is one of the cornerstone of modern financial industry.

2.2 Credit Rating Methodologies and performance evaluation

Due to the critical role played by credit ratings in financial activities, the related techniques has grown with the financial industry for decades.
2.2.1 Subjective analysis and expert systems

In the beginning, a method called “credit approval” was widely used by financial firms to evaluate the deal risks based on their counterparties’ credit. The evaluation process is a purely judgmental process focused on several aspects of their counterparties. A common aspect set is called “5 Cs” [11,13]. They are

- Character: reputation of the counterparty
- Capital: the difference between the liabilities and assets
- Collateral: the market value of the collateral
- Capacity: the ability to pay
- Condition: the position of the counterparty in its business.

The “5 Cs” system uses industry experts to evaluate their counterparties. Human experts are always criticized for their costs and psychological biases [11]. Some artificial techniques such as decision tree and rule based systems were used to capture human knowledge into computational models today [38].

2.2.2 Financial models

With the growth of the financial industry, the weak points of subjective analysis and expert systems were realized. The approach lacks transparency in the process, brings psychological biases, and does not have quantitative results [11].

Financial models use financial theoretical frameworks to assess corporate credit risks. The most popular credit risk theoretical framework is proposed by Merton in year 1973 based on Black-Scholes model [39,40]. The model uses stochastic Brownian motion process to assess the future uncertain corporate assert value. By comparing the asset value distribution and the certain debt value, the theory gives the probability of default, i.e., the probability that the asset value is lower than the total value of payable liabilities at a
future time $T$. Merton model was first proposed to calculate price of options in financial market. In year 1984, the Merton model was extended by Moody’s KVM on corporate credit rating [41]. The extension is called KMV model, which is de facto standard for the credit measurement in financial industry today. The Figure 2.1 illustrates the theory of KMV model. In KMV model, the distance to default at time $T$ is calculated by equation 2.1

$$ DD(T) = \frac{\log(A) - \log(L) + \mu T - \frac{1}{2} \sigma^2 T}{\sigma \sqrt{T}} $$

where $A$ is the current market value of the corporation’s assets, $L$ is the value of total payable liabilities at the time $T$, $\mu$ and $\sigma^2$ are the expected rate of return on assets and the variance.

The Merton theory uses asset volatility to assess the credit risk. It is based on a assumption, the corporation will default if its debt value exceeds its asset value. However, it is not always true in real market. From cash flow point of view, a corporation defaults
CHAPTER 2. LITERATURE REVIEW

if it does not have enough cash to pay matured liability, no matter if its debt value is larger than its asset value or not. Another theory, the gambler’s ruin, is proposed to rating corporate credits by predicting the cash flow volatilities. The theory applies a two state Markov process to model a corporation’s incoming and spending. The cash flow volatilities and the probability distribution of cash flow are calculated by the Markov process for the future time $T$. The risk is presented as the sum of booked equity and average cash flow, divided by the cash flow volatility.

2.2.3 Statistical models

Financial models present a credit risk rating with consistent theories explaining the reasons behind the obtained credit ratings. However, the financial models do not always match empirical observation [11]. Therefore, statistical models are proposed based on historical data analysis.

A typical empirical statistical framework is built based on following steps

- Factor selection: Selecting related factors as input for the model.

- Model selection: Finding a proper model to connect the factors and the credit ratings.

- Induction: Using the training data to compute the model parameters.

- Deduction: Applying the trained model to predict the credit ratings for corporations.

Usually, rating agencies and financial firms assess corporate credit risks based on their financial factors capturing the features of the corporate credit rating. Financial factors can be obtained by analyzing the corporate financial statements [42], which are part of the corporate financial reports. In most countries, corporations are required to release audited financial reports periodically. The basic three common financial statements, e.g., income statements, balance sheets, and cash flow statements, can be found in financial reports. A financial statement contains information to measure corporate business performance. A common financial statement includes total revenue or total sales, total debts, total liabilities,
current liabilities, total assets, current assets, net profit, earnings before interest and taxes (EBIT), etc. A more complete list of financial statements is listed in Appendix A. The financial statements from reports are absolute values, which have dependencies on the corporate size, industries, and other factors. A financial ratio is a relative magnitude of two selected financial data and the calculation of the financial ratio removes dependencies to other factors, such as size, economic situations, etc.. By removing these dependencies, financial ratios are more appropriate for evaluating corporate financial conditions than financial data. Frequently used financial ratios include profitable ratios, liquidity ratios, debt ratios, etc.. Appendix B lists most frequently used financial ratios and their definitions.

After financial statements are retrieved and financial ratios are calculated, an appropriate analysis model is selected and used for data analysis.

2.2.3.1 MDA models

In statistics, Multivariate Discriminant Analysis (MDA) is a common approach to classify a set of given data into different categories by a linear discriminant function. The MDA is proposed by Fisher in 1936 [43]. In year 1968, Altman applied the MDA on corporate credit ratings. 66 corporations were selected. 33 of them are bankrupted and another 33 corporations are randomly selected. Based on the sample data, 22 financial factors are calculated based on corporation's financial reports. By analyzing the correlations and comparing the accuracies, the Altman z-score formula was proposed to calculate the “credit score” for corporations based on 5 selected financial factors [44]. The z-score is calculated by a simple linear function defined in 2.2

\[
z = 0.012x_1 + 0.014x_2 + 0.033x_3 + 0.006x_4 + 0.999x_5 \tag{2.2}
\]

where \(x_1 = \) Working capital/Total assets
\(x_2 = \) Retained earnings/Total assets
\(x_3 = \) EBIT/Total assets
\(x_4 = \) Market value equity/Book value of total debt
\(x_5 = \) Sales/Total assets

After some revisions [45,46], the z-score based MDA analysis approaches are still widely used in the financial industry to assess the corporation credits today.
CHAPTER 2. LITERATURE REVIEW

2.2.3.2 Conditional probability models

The MDA approach assigns scores to corporations based on a formula rather than giving the probabilities of default directly. The financial firms have to map the “score” to the probability of default. To overcome the drawback, conditional probability models have been proposed. Conditional probability models assume the probability of default is following a certain distribution function of their credit “score”. With different distribution function, the conditional probability models are categorized into logit analysis (LA), probit analysis (PA), and linear probability modelling (LPM). Among them, the LA is most popular approach [12,47]

LA is based on logistic regression. The model combines selected factors $x_1, x_2, \ldots, x_n$ into a probability score via a logistic function as Equation 2.3

$$P = \frac{1}{1 + \exp(\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \ldots - \beta_n x_n)}$$

LA and MDA were studied on 105 bankrupt cases happened between 1970 and 1976 with 9 selected financial factors [48]. The results showed that the LA outperform MDA on corporation credit assessment.

2.2.4 Scorecard

Table 2.1: An example of the financial performance measurement.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross Operating Balance/Operating Revenue (%)</td>
<td>&gt;= 22</td>
<td>8 - 21.9</td>
<td>5 - 7.9</td>
<td>&lt; 5</td>
</tr>
<tr>
<td>Cash Financing Result/Total Revenue (%)</td>
<td>&gt;= 10</td>
<td>0 - 9.9</td>
<td>(10) - 0</td>
<td>&lt; (10)</td>
</tr>
<tr>
<td>Net Working Capital/Total Expenditures (%)</td>
<td>&gt;= 18</td>
<td>4 - 17.9</td>
<td>0 - 3.9</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>Interest Payments/Operating Revenue (%)</td>
<td>&lt;= 1.5</td>
<td>1.6 - 5.5</td>
<td>5.6 - 8.5</td>
<td>&gt; 8.5</td>
</tr>
</tbody>
</table>

Scorecards are widely used in the finance industry to combine expert systems, financial models, and statistical models. A credit rating scorecard will break the rating problem...
into a set of high level factors, which can be further divided into sub-factors. Each factor and sub-factor will be measured respectively. The overall credit score of a corporation will be calculated by summarizing all values of factors and sub-factors [14,49]. To design a scorecard based credit rating model, the first step is to list out all possible factors and find out the corresponding measurement to measure them in scores. Measurement tables are prepared by listing out all factors with their measurement scales calculated by sophisticated statistical processes, AI processes, financial models, and domain expertise as well. Table 2.1 is used as an example to illustrate the financial performance measurement table, which consists of sub-factors of a particular factor (in this example, the factor is financial performance), and corresponding threshold values.

Table 2.2: An example of the credit assessment scorecard.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Weight</th>
<th>Sub-factor Weight</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Environment</td>
<td>40%</td>
<td>50%</td>
<td>4</td>
</tr>
<tr>
<td>- Return per share</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>- Revenue volatility</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Financial Performance</td>
<td>40%</td>
<td>50%</td>
<td>3.8</td>
</tr>
<tr>
<td>- Gross Operating Balance/Operating Revenue</td>
<td>50%</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>- Cash Financing Result/Total Revenue</td>
<td>25%</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>- Net Working Capital/Total Expenditures</td>
<td>15%</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>- Interest Payments/Operating Revenue</td>
<td>10%</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Debt profile</td>
<td>20%</td>
<td>50%</td>
<td>4.5</td>
</tr>
<tr>
<td>- Debt / Revenue</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>- Short term debt / Debt</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>= 4.02 ≈ 4</td>
</tr>
</tbody>
</table>

After all factors and sub-factors are measured, the total score can be obtained by the summation of the time result of factor scores and their corresponding weights. Table 2.2 is an example to depict the content of a scorecard: factors and sub-factors with their risk scores, and the total risk score. The higher the score is, the higher the risk is. The logic and calculation process in scorecards are straightforward and easy to be understood. In
addition, the factors and weights can be adjusted to reflect the market changes. This makes the scorecard be widely accepted by financial firms and credit rating agencies [50].

The calculation of the credit rating in scorecards uses the results of different models and expert judgments as inputs rather than using the raw financial data directly. Therefore, scorecard users must have the expertise of financial data and the counterparty's industry data as well. In addition, the scorecard model itself does not capture any knowledge of the interconnection between the raw data and the listed factors. This raised the requirement that users have to rely on statistics analysis tools or their own experience to decide factor weights. This judgmental procedure weakens the objectivity of scorecard results as well. In summary, scorecard users have to take a long learning curve on the manual process and spend vast amounts of time to be proficient in the process. Moreover, there is no systematic automation process of result verification, adjustment, and justification on corresponding weight setting.

2.2.5 Performance testing

In order to choice and calibrate the proper model for different kinds of corporations, performance testing methodologies are desirable for financial firms and rating agencies.

2.2.5.1 The error testing

The primary objective of credit rating is to predict which counterparties may default in future. The rating process will assign corporations into different levels from high default probability to low default probability. A threshold rating will be specified to classify corporations into two credit categories, i.e., the “positive” category and the “negative” category, which identifies the high risky corporations from low risky corporations.

If a credit rating model classifies a counterpart into a wrong credit category, it is an error. The most straightforward accuracy measurement is to calculate the error rate. For corporations \( C = \{c_1, c_2, \ldots, c_N\} \), if \( M \) of them are misclassified into wrong credit categories, e.g., a corporation with high credit risk is classified into a “positive” credit category or the
other way round. The error rate is calculated by equation 2.4

\[ ER = \frac{M}{N} \] (2.4)

The overall error rate has been used other work to compare the accuracies of different models [51]. However, the credit ratings also provide the PD of the counterparty besides identifying the corporations with “negative” credits. The standard deviation is a common approach to test the accuracy of PDs, which is defined as equation 2.5

\[ ED = \sqrt{\frac{\sum (p_i - p_i^*)^2}{N}} \] (2.5)

where the \( p_i \) is the actual probability of default for the corporation \( c_i \), and the \( p_i^* \) is the result of the rating model.

Besides the number of errors and the standard deviation, the error type is another measurement on rating models. With a predefined threshold rating, the credit rating classifies corporations into “positive” and “negative” credit categories. There are four possible results for a rating result. They are

- a positive corporation has been classified into the positive credit category (True positive (TP));
- a negative corporation has been classified into the positive credit category (False positive (FP));
- a positive corporation has been classified into the negative credit category (False negative (FN));
- and a negative corporation has been classified into the negative credit category (True negative (TN)).

The result 1 and 4 are the expected results, and 2 and 3 are errors. The result 2 is called type I error, and result 3 is called type II error. The relationship is shown in Table 2.3
### Chapter 2. Literature Review

<table>
<thead>
<tr>
<th>Positive corporation</th>
<th>Negative corporation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive rate</td>
<td>Correct (TP)</td>
</tr>
<tr>
<td>Negative rate</td>
<td>Type II error (FN)</td>
</tr>
<tr>
<td></td>
<td>Correct (TN)</td>
</tr>
</tbody>
</table>

Table 2.3: The different rating results

The type I and type II errors have been discussed and compared separately because the costs of these two types of errors are different. Based on historical data, the costs of type I errors are much larger than the costs of type II errors [12, 36]. Comparing rating errors on type I and type II individually is a common evaluation approach. It was used by [12, 38, 44, 46–48] to test the performance of different credit models.

#### 2.2.5.2 ROC and CAP curves

Besides error rate, standard deviation, and the type I/type II errors, there are other statistical methods to test rating models on accuracy and the discrimination capacity. The most popular methods are Receiver Operating Characteristic (ROC) and Cumulative Accuracy Profile (CAP) [52–56].

ROC is an extension of type I and type II error comparison. ROC uses the relationship between two ratios to illustrate the accuracy of the model. They are *true positive rate* (TPR) and *false positive rate* (FPR), which are defined in equation 2.6 and 2.7 respectively.

\[
TPR = \frac{TP}{TP + FN} \quad (2.6)
\]
\[
FPR = \frac{FP}{FP + TN} \quad (2.7)
\]

When the threshold rating moves from the lowest credit rating to the highest credit rating, the TPR and FPR will change from 1 to 0 with different velocities. The ROC curve depicts the relationship between TPR and FPR. A rating model with certain discrimination capacity will distribute the negative corporations and positive corporations into different credit ratings. Figure 2.2 shows three possible rating distributions of positive and negative corporations respectively. The diagram A illustrate the perfect model, which totally separates the negative corporations from positive corporations. The distribution
CHAPTER 2. LITERATURE REVIEW

of B is better than the distribution of C, since the positive corporations distributions are further from the negative corporations in B than they are in C.

Therefore, when the threshold number is moving from 0 to 1, the relationships between the corresponding TPR and FPR are different from each other for the three rating models. Figure 2.3a illustrates the three different ROC curves for them in Figure 2.2.

The Figure 2.3a shows that the better rating model B has a ROC curve closer to the perfect curve A than the model C has.

In the financial industry, CAP is another important rating model validation tool [52–54]. Different from ROC, the CAP illustrate the relationship between the accumulated negative corporation percentage and the accumulated total corporation percentage while the threshold rating is moving from the lowest to the highest. In CAP, the x-axis is the percentage of total corporations whose credit ratings are equal or lower than the threshold rating. The y-axis is the percentage of negative corporations whose credit ratings is equal or lower than the threshold ratings.

For the three result distributions shown in Figure 2.2, Figure 2.3b illustrate their CAP curves accordingly.

Both the ROC and CAP provide the indices of rating accuracy. According to [54], that “The ROC area under the curve has a natural interpretation as the unbiased percentage of correct decision that does not apply to the accuracy ratio, and the ROC curve, unlike the CAP curve, is independent of the probability of default.

“In addition, ROC analysis has an established theory of optimal threshold setting.”

Based on these reason, the ROC is preferred to CAP for the discrimination capacity assessing.

2.3 Machine learning

Machine learning have been used in credit rating for years, including techniques such as artificial neural networks [17,57,58], support vector machines [59,60], k-Nearest-Neighbors
CHAPTER 2. LITERATURE REVIEW

Figure 2.2: Result distributions from different rating models
CHAPTER 2. LITERATURE REVIEW

![ROC curves for different rating model](image1)

(a) ROC curves for different rating model

![CAP curves for different rating model](image2)

(b) CAP curves for different rating model

[61,62], Bayesian inference and networks [63] etc. The credit rating problem can be formalized to a typical supervised machine learning problem as what section 2.2.3 describes. For a corporation $i$, its financial factors are denoted by $X_i = \{x_{i1}, x_{i2}, ..., x_{in}\}$. Its credit rating will be denoted by $t_i$. The training data $S = \langle X_i^*, t_i^* \rangle$ is selected from the historical data. Algorithms build up models by learning from the training data and the model will be able to classify any given $X_i$ to the corresponding credit rating $t_i$.

Some academic studies showed that machine learnings outperform the statistical models on accuracy and efficiency [20, 38, 64, 65]. Among various machine learning techniques, ANNs and SVMs are dominating others [51, 66, 67]. The following sections provide a brief introduction of these two techniques.

2.3.1 Artificial neural networks

Artificial neural networks (ANNs) were proposed to mimic the way, in which biologic neural networks work [17, 57, 58]. ANNs are flexible nonlinear mathematical tools. With the capability to learn complex relationships between inputs and outputs, ANNs have been widely used in many applications, e.g., accounting and finance [68], health and medicine [69, 70], engineering and manufacturing [71, 72], marketing [73]. Meanwhile, ANNs are efficient alternatives to traditional statistical techniques for addressing regression and classification.
CHAPTER 2. LITERATURE REVIEW

problems [74]. Especially, they have been proved as useful high performance analysis tools for credit ratings [20,75,76].

Single Hidden Layer Feedforward Neural Network (SLFN) is a kind of frequently used ANN with excellent approximation capabilities [77,78]. As its name depicts, an SLFN has one hidden layer and no feedback from the output. More formally, assuming that an SLFN consist of $N$ hidden nodes and let $o_j$ represent the $j$th output of the SLFN, then $o_j$ can be obtained by using equation 2.8.

$$o_j = f(X) = \sum_{i=1}^{N} \beta_{ij} g_i(X)$$

(2.8)

where $X$ is the input vector, $\beta_{ij}$ is the connection weight vector from the $i$th hidden node to the $j$th output node, and $g_i(X)$ is the hidden node activation function. There are a number of different types of hidden nodes with different activation functions, where the additive nodes and RBF nodes are two types of frequently used hidden nodes [24].

1. Additive hidden nodes:

$$g_i(X) = g(A_i \cdot X + b_i), A_i \in R^d, b_i \in R,$$

where $A_i$ is the weight vector connecting the input layer to the $i$th hidden node, and $b_i$ is the bias of the $i$th hidden node.

2. RBF hidden nodes:

$$g_i(X) = g\left(\frac{|X - A_i|}{b_i}\right), A_i \in R^d, b_i \in R^+$$

(2.10)

where $A_i$ and $b_i$ are the $i$th RBF node’s centre and impact factor.

For a given set of $k$ training samples, denoted as $\{<X_i, T_i> | X_i \in R^d, T_i \in R^m, i = 1,2,...,k\}$, the error function of the SLFN is defined by equation 2.11 to measure the distance between the anticipated output, denoted as $T = \{t_1, t_2, ..., t_k\}$, and the output of SLFN, denoted as $O = \{o_1, o_2, ..., o_k\}$.

$$E = 1/2 \sum_{i=1}^{k} ||o_i - t_i||^2$$

(2.11)
CHAPTER 2. LITERATURE REVIEW

The learning objective is to minimize the distance of the network by adjusting the network connection weights. As BP algorithm is a gradient based supervised learning algorithm, when it is applied, the error will be backpropagated to hidden nodes, and the weight \( w_{ij} \) will be adjusted iteratively by adding the partial derivation derived from equation 2.12 accordingly.

\[
\Delta w = -\gamma \times \frac{\partial E}{\partial w_{ij}} \tag{2.12}
\]

where \( \gamma \) is a constant specifying the step length of each iteration. If \( \gamma \) is small, SLFNs may be stuck at local minimums and require more time to converge to results. On the other hand, if \( \gamma \) is large, it will be difficult for the algorithm to converge. That is the reason why the BP algorithm has been criticized on the difficulties to decide learning rates, ease to be stuck on local minimums, and time consuming [21]

Recently, ELM algorithm has been proposed in [21, 22] to overcome the drawbacks of the BP algorithm. When ELM is applied, the output weight matrix, denoted as \( \beta \), can be learned from training samples using a randomly assigned input weight matrix, denoted as \( A \). For the given training input \( X \), the hidden node activation matrix, denoted as \( H \), can be derived by \( H = (g(X)) \). Based on this, the output matrix \( O \) can be obtained by \( O = H\beta \). Particularly, when \( \beta = H'T \), we can have:

\[
O = HH'T \approx T
\]

(2.13)

where \( H' \) is the MoorePenrose pseudoinverse of \( H \), meaning that the random input weight matrix \( A \) and the output weight matrix \( \beta = H'T \) compose the pattern for the given training samples. A parameter \( \lambda \) has been introduced to improve the stability of ELM, where the output weight matrix \( \beta \) can be achieved by equation 2.14 [79, 80].

\[
\beta = H^T(I - \frac{1}{\lambda} TT^T')T
\]

(2.14)

ELM is an extreme fast learning approach for SLFNs with fixed hidden node numbers and it has better generalization performance than BP [81]. Meanwhile, another extreme
CHAPTER 2. LITERATURE REVIEW

learning approach, incremental extreme learning machine (I-ELM), has been proposed targeting for more flexible neural networks [24]. When I-ELM is applied, the hidden nodes are added one by one with random input weight vectors and bounded piecewise continuous activation functions. Let a network with \( n \) hidden nodes be denoted as \( f_n \), the residual error function is defined as \( e_n = f - f_n \), where \( f \) denotes the target network. Particularly, for the \( i \)th hidden node, the error of the new network will be minimum if the weight vector from the hidden node to output, i.e., \( \beta_i \), equals to \( \frac{(e_{i-1}, g_i)}{\|g_i\|^2} \), which has been proved in [24]. Since the actual \( e_{i-1} \) is not available, and a consistent estimate of the \( \beta_i \) is used instead, which can be obtained by equation 2.15.

\[
\beta_i = \frac{EHT}{HH^T} = \frac{\sum_{p=1}^{N} e_{i-1}(p)h(p)}{\sum_{p=1}^{N} h^2(p)}
\]

(2.15)

where \( h(p) \) is the activation function of the \( i \)th new hidden node for the \( p \)th training input. Meanwhile, it has been approved that I-ELM increases the learning speed and keeps SLFNs as universal approximators as well [24].

2.3.2 Support vector machine

Support vector machine (SVM) is another popular machine learning technique [59, 60]. SVMs have been widely used in various applications such as gene selections [82, 83], time series forecasting [84, 85], and pattern recognitions [86, 87]. SVM is a supervised learning technique. For a given training set, denoted as \( S = \{< x_i, t_i > | i = 1, 2, ..., n \} \), where \( x_i \in \mathbb{R}^m \) and \( t_i \in \{-1, 1\} \), after mapping the input set \( x \) into a high dimensional feature space, the training data can be separated by a set of linear hyperplanes. The hyperplane with maximum margin to sample data is the desired result. The sample data that are closest to the maximum margin hyperplane are called support vectors. That is, SVM can be represented as the following optimization problem.

\[
\min \left( \frac{1}{2} \alpha^T Q \alpha - \sum_{i=1}^{n} \alpha_i \right)
\]

(2.16)

subjects to

\[
0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{n} t_i \alpha_i = 0
\]

(2.17)
where $\alpha_i$ is a Lagrange multiplier for training sample $i$, $C(C > 0)$ is the upper bound of $\alpha_i$, $Q$ is an $n$ by $n$ matrix with $Q_{ij} = t_i t_j k(x_i, x_j)$, and $k(x_i, x_j)$ is the kernel function [60]. Different kernel functions are used for different problems, e.g., the linear function, the polynomial function, the radial basis function (RBF). The RBF kernel, as defined in equation 2.18, is commonly used because of its nonlinearity, small number of hyperparameters, and fewer numerical difficulties [88].

$$k(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2) \quad (2.18)$$

As a binary classifier, the value of SVM outputs $t_i$ is in $\{-1, 1\}$. Multi-class Support Vector Machines (MSVMs) are used to address multi-class problem. MSVMs solve a multi-class problem by a set of binary classifiers. The binary classifiers are organized in One Against One (OAO) or One Against All (OAA) structures, and OAO has demonstrated better performance than OAA [89]. SVMs have been recognized as one of the best classification approaches. Previous studies showed that SVMs outperform ANNs and traditional statistical approaches [20,90].

2.3.3 Fuzzy cognitive maps

Fuzzy Cognitive Maps (FCMs) were introduced by Bart Kosko in 1986 based on Cognitive Maps (CMs) [27,91]. A cognitive map is a directed graph, in which the vertices represent concepts and edges represent the relationships between these concepts. There are two kinds of relationships between concepts in a cognitive map, the positive relationships and negative relationships. E.g., a CM shows the relationships between three concepts, E - “education investment”, J - “jobless rate”, and C - “crime rate”. The relationship from “E” to “J” is negative, i.e., increasing education investment will decrease the jobless rate. The relationship from “J” to “C” is positive, which indicates that a high jobless rate is a cause of a high crime rate. FCMs enhance CMs by fuzzifying the causal relationships between concepts. In a FCM, each relationship has a weight, which is a numeric value in $[-1, 1]$ to indicate the strength of the relationship. Each concept also has a state in $[-1, 1]$ indicate
CHAPTER 2. LITERATURE REVIEW

the concept’s status. Figure 2.4 shows a simple FCM. For a specific time, the states of concepts can be represented by a vector. For example, a possible state vector for the concepts in the map in Figure 2.4 is (0.3, 0.1, 0.1).

\[ \begin{bmatrix} 0.3 \\ 0.1 \\ 0.1 \end{bmatrix} \]

Figure 2.4: A simple fuzzy cognitive map

The relationship weights between concepts can be represented by a matrix, which is called the weight matrix. The weight matrix of the map shown in Figure 2.4 is

\[
\begin{pmatrix}
0 & 0.5 & 0 \\
0 & 0 & 0 \\
-0.1 & 0 & 0
\end{pmatrix}
\]

A FCM can be represented as \( G = (n, E, C, f) \) formally, where \( n \) is the number of nodes in the graph. \( E = (w_{ij}) \) is the weight matrix, where \( w_{ij} \in [-1, 1] \) represents the weight of the directed edge from concept \( N_i \) to concept \( N_j \). \( C = \{C_i\} \) is the set of concepts. \( f \) is the threshold function of concepts. FCM inference is an iterative process. During the inferring, concept states are calculated by equation 2.19

\[
C_i(t + 1) = f \left( C_i(t) + \sum_{i=1, j \neq i}^{n} w_{ji} C_j(t) \right), \forall i \in \{1, 2, \ldots, n\}
\]  

(2.19)

where \( C_i(t) \) denotes the fuzzy state of the \( i \)th concept at the \( t \)th iteration. By a given state vector \((C_1(0), C_2(0), \ldots, C_n(0))\), a FCM can predict concept state changes led by interactions between concepts via its weight matrix. The FCM will eventually converge to a steady state, or to a set of fixed concept state patterns, the so call hidden pattern.

Expression in equation 2.19 updates the state value of each concept over the successive iterations. The threshold function can be either discrete or continuous, linear or non-linear functions. Four threshold functions have been used in most studies. They are described in below:

- Bivalent:

\[
f(x) = \begin{cases} 
0 & x \leq 0 \\
1 & x > 0 
\end{cases}
\]

(2.20)
CHAPTER 2. LITERATURE REVIEW

- Trivalent:

\[ f(x) = \begin{cases} 
-1 & x \leq -0.5 \\
0 & -0.5 < x < 0.5 \\
1 & 0.5 \leq x 
\end{cases} \quad (2.21) \]

- Logistic:

\[ f(x) = \frac{1}{1 + e^{-Cx}} \quad (2.22) \]

- Hyperbolic tangent:

\[ f(x) = \tanh(\lambda x) = \frac{e^{\lambda x} - e^{-\lambda x}}{e^{\lambda x} + e^{-\lambda x}} \quad (2.23) \]

where \( C \) is a parameter used to determine proper shape of the function. In most applications, the value of \( C \) is set to be 5 [92,93]. Threshold functions are used to force concept state values to be a normalized range. Although they are designed for the same purpose, different threshold functions have different characteristics. Comparison studies of different threshold functions for FCMs have been carried out in [94,95], where the findings show that logistic functions took more iterations to stable state than other functions during inferring. The inferring results are also different when different threshold functions are used. Some functions, such as bivalent and trivalent functions, force concept values into discrete value sets [96]. During inferring, hyperbolic tangent functions allow concepts to reach to extreme values, such as \(-1\), \(0\), and \(1\), while logistic function will not [95]. Therefore, different applications use different threshold functions according to their domains.

Dynamics is another important feature of FCM. FCM inferring processes are iterative processes. Let \( C(t) \) denote the system state vector at the \( t \)th iteration, \( C(t) = [C_1(t), \ldots, C_n(t)] \). The inferring of a FCM generates a sequence of state vectors. The inferring results help people analyzing the dynamics of the system with different initial conditions. In order to put the discussion into perspective, an example, i.e., Bad Weather Driving [97], is used to show how bad weather affects driving speed on a California freeway in daytime. Figure 2.5 depicts the FCM model for it. Particularly, the dynamics are
CHAPTER 2. LITERATURE REVIEW

Figure 2.5: The FCM of the bad weather driving.

compared under two scenarios: different initial states under the same causal relations and the same initial states under different relations. The results shown in Figure 2.6 and 2.7 illustrate that the dynamics of FCM not only depends on initial states, but also depends on causal relations.

Due to the dynamics of FCM, with discrete threshold functions, such as functions defined in equation 2.20 and 2.21, the inferring process may converge to a fixed state or keep cycling between a number of fixed states [96], which is called hidden patterns. With continuous threshold functions, such as functions defined in equation 2.22 and 2.23, the process may converge to states or appear chaotic response and other uncertain behaviors.

Combining advantages of neural networks, expert systems, and non-linear dynamical systems, FCMs are potential efficient soft computing tools for credit ratings. Conventionally, a FCM is a dynamic system. All concepts are associated with each other via causal relations. The inferring will not stop until the map converges to a steady state or a hidden pattern, where the potential steady states of a FCM are decided by input values, connection weights, and threshold functions. Therefore, it brings difficulties for users to design FCMs, which can always converge to meaningful steady states with any initial states, especially for problems with large scale of concepts and relationships. In addition, the accuracies of the unsupervised learning are uncertain. The complexity limits the sizes of FCMs. A survey
Figure 2.6: The dynamics comparison with different initial states.

(a) Dynamics of the states with the first initial state vector value.

(b) Dynamics of the states with the second initial state vector value.
(a) Dynamics of the states with the first weight setting.

(b) Dynamics of the states with the second weight setting.

Figure 2.7: The dynamics comparison with different causal relations.
CHAPTER 2. LITERATURE REVIEW

on FCM size has been done by Stach et al [98]. 15 different FCM applications are studied. There are at most 10 concepts in each FCM. Especially, 8 of these 15 applications use only 5 or 6 concepts in their FCMs.

FCMs have been extended in two major directions: (i) incorporating rules to build rule based FCMs (RB-FCMs) [99–101]; and (ii) incorporating temporal concepts and causal relationships to build temporal FCMs, which are expected to support long term, dynamic problems [34,102,103]. Specifically, the extended FCM (EFCM) [102] and the fuzzy time cognitive map (FTCM) [103] were proposed in 1992 and in 1995 respectively to map FCMs with a discrete temporal domain. Based on the discrete temporal FCM, the effect decaying was introduced in 1997 as an independent attribute of concepts by Tsadiras et al. in 1997 [34]. The study on the relationship between relationship weights and the based time interval, called B-Time, was carried out by Carvalho et al. in 2000 [104]. In addition, some guidelines are also proposed in the same work to help modelers choose suitable B- Times for their applications.

2.3.3.1 Layered FCMs

The layering technology was introduced to FCMs by Satur et al. in 1995 [105] to construct large scale FCMs with small fragments. According to their work, the layered FCM architecture possesses following characteristics:

(1) The expressiveness in its representation of the human knowledge and hence of the data itself,

(2) The structure can be interpreted by the human expert in a way that allows the human expert to interact with the system.

A layered FCM has similar structure as Scorecard, which will facilitate the knowledge capturing from scorecard based rating models. Layered FCMs are used in chapter 3 and chapter 4 for the credit rating problem.
CHAPTER 2. LITERATURE REVIEW

2.3.3.2 Training of FCMs

A FCM can be created based on knowledge from domain expertise and historical data. Learning algorithms, by modifying the FCM causal relationships, has been proven to be essential components for FCM to improve its reliability [93]. The first learning algorithm of FCM, called DHL, was introduced with FCM by Ksoko et al. in 1986 [27]. From then on, a number of learning algorithms were carried out to refine causal relationships for FCM [106,107]. According to the technologies they are using, these learning algorithms are categorized into two types, the Hebbian learning and the evolutionary learning [93].

2.3.3.2.1 Hebbian learning. Hebbian learning is based on a theory in neuroscience introduced by Donald O. Hebb in 1949 [108]. It is used to train FCMs under the principal that the weight of a causal relation will be increased if the corresponding two connected concepts are changing simultaneously. To be more specific, assuming that the weight for a causal relation between concept $i$ and $j$ is $w_{ij}$. At step $k$, the changes of concept $i$ and $j$ can be calculated using the equations $\Delta v_i(k) = v_i(k) - v_i(k - 1)$ and $\Delta v_j(k) = v_j(k) - v_j(k - 1)$ respectively. The weight $w_{ij}$ will be adjusted using equation 2.24

$$w_{ij} = w_{ij} + \gamma(\Delta v_i \Delta v_j - w_{ij})$$

(2.24)

where $\gamma$ is the learning coefficient. The first Hebbian learning algorithm, i.e., differential Hebbian learning (DHL), was proposed by Dickerson and Ksoko [109] and then its enhancements such as differential Hebbian learning (DHL) and active Hebbian learning (AHL) were carried out by Papageorgiou et al. in 2003 and 2004 [106,107] respectively. Hebbian learning algorithms are unsupervised learning methods. They are used to train a predefined FCM to converge with minimal error to improve the efficiency and the reliability of FCM.

2.3.3.2.2 Evolutionary learning. The evolutionary learning approach is a supervised learning method. It trains the causal relations of FCMs with a given sample data based on the evolution theory. The evolutionary learning is a type of global search algorithms
inspired by biological evolution. Genetic Algorithm (GA) is the most popular type of evolutionary learning.

GAs are powerful algorithms for combinatorial optimization problems. Compared with other heuristic algorithms such as greedy algorithms and tabu search algorithms, GAs are global algorithms. Ideally, they can search the global optimal solutions. Therefore, GAs have been widely applied to combinatorial optimization problems such as the travelling salesman problems (TSP) and the minimal spanning tree problems.

GAs evolve solutions based on three evolutionary principles: crossover, mutation and selection. Followings are required steps when GA is applied to a problem:

1. a bijective mapping function between the solution set and chromosome set,

2. a fitness function to evaluate solutions.

For example, to apply the GA algorithm to a TSP with 8 cities, chromosomes are defined as 8-bit arrays. An array defined in equation 2.25 shows a path between the 8 cities.

\[ x = [7 \ 1 \ 3 \ 2 \ 5 \ 8 \ 6 \ 4]. \] (2.25)

The fitness functions are defined based on chromosomes to evaluate the correctness of solutions. In the 8-city TSP example, equation 2.26 shows a possible fitness function.

\[ \text{Fitness}(x) = \text{Distance}(x_8, x_1) + \sum_{i=1}^{7} \text{Distance}(x_i, x_{i+1}) \] (2.26)

where Distance\((x_i, x_j)\) is the function calculate the distance between two cities, \(x_i\) and \(x_j\).

With the chromosome mapping function and fitness function, the GA can be applied on target problem domain with its three operations, i.e., crossover, mutation, and selection.

Crossover is a genetic operation that combines two chromosomes to produce a new chromosome. According to how many crossover points are selected and how they are decided, there are many different kinds of crossover operations, where one-point, two-point and uniform crossover are the most commonly used crossover operators [110]. The one-point crossover operator randomly selects a crossover point to produce the new chromosome while
CHAPTER 2. LITERATURE REVIEW

the two-point crossover operator selects two crossover points. Particularly, for each gene value, a uniform crossover operator decides with some probability which parent chromosome will contribute in the offspring chromosome.

Among them, the one point crossover is the most widely used crossover operators. Given two chromosomes, $X$ and $Y$, where $X, Y \in \{0, 1\}^n$. The crossover point $p$ is randomly selected [111]. Then all data beyond the point in either bit string is swapped between the two parent bit strings, i.e., $X$ and $Y$. producing two solutions with $X'$ and $Y'$ with

$$
X'_i = \begin{cases} 
X_i & \text{if } i \leq p \\
Y_i & \text{if } i > p 
\end{cases} \quad \text{and} \quad
Y'_i = \begin{cases} 
X_i & \text{if } i > p \\
Y_i & \text{if } i \leq p 
\end{cases}
$$

More concretely, the following example illustrates the operation:

$$
\begin{align*}
X &= 0110|010 \\
Y &= 1011|001
\end{align*} \quad \rightarrow \quad \begin{align*}
X' &= 0110|001 \\
Y' &= 1011|010
\end{align*}
$$

Mutation operators are used to maintain genetic diversity from one generation of a population to the next by altering one or more gene values in a chromosome from its initial state. There are two category mutation operators proposed in [111]: bit-flipping mutation and inversion mutation. When the bit-flipping mutation is applied, a randomly selected single bit in the string is flipped to form a new offspring string, an example bit-flipping mutation is depicted as follows.

$$
X = 0110|010 \quad \rightarrow \quad X' = 0110|110.
$$

When the inversion mutation is applied, a subsequence in the bit string is reversed. To be more specific, for a given vector, it will generate two cutting points randomly. An example inversion mutation is depicted as follows. First, two cutting points are generated randomly and then cuts the solution vector into three parts. The bit string between the two cutting points is reversed:

$$
X = 01|100|10 \quad \rightarrow \quad X' = 01|011|10.
$$
CHAPTER 2. LITERATURE REVIEW

Generally, crossover and mutation operators improve the candidate solutions in GAs. Crossover operators allow GAs to find a local optimal solution quickly, and play an important role in GAs. On the other hand, crossover operators shrink the search space to local optimal solutions and hard to escape from them.

Contrary to crossover operators, mutation operators help to enlarge the total search space and help the GAs escape from the current search space. Therefore, mutation operators help GAs avoid local optimal solutions to find better solutions. In practice, the mutation rate is very low due to its diversity property.

After the crossover and mutation, a selection process is applied to select chromosomes to breed a new generation. Chromosomes are selected according their fitnesses calculated using fitness functions. Fitter chromosomes have more chance to be selected. For a given set of chromosomes $P = \{x_1, x_2, \ldots, x_n\}$, the probability for a chromosome $x_i \ (x_i \in P)$ to be selected is obtained using equation 2.27

$$p(x_i) = \frac{\text{Fitness}(x_i)}{\sum_{x_j \in P} \text{Fitness}(x_j)}.$$  (2.27)

Roulette wheel sampling [112], also called stochastic sampling with replacement, is a simplest selection scheme. When this method is applied, candidate solutions with lower fitness is less likely to be eliminated. In addition, some weaker solutions still have a chance to survive the selection process. Sometimes, weaker solutions may include an important component for the recombination process. Therefore, this selection scheme is relatively fair.

With genetic operators and selection, the pseudo code of genetic algorithm is defined in Algorithm 1. For many applications, it can find comparable good solutions within given running times. Therefore, GAs have been widely applied to many optimization problems, especially the NP-hard problems such as the TSPs and the quadratic assignment problems (QAP).

The genetic algorithm (GA) was first applied to train FCMs in 2001 [114] and it has been proven as an efficient approach to train FCMs from sample data. In GA, how to design the fitness function is the key research topic [93,98]. A complete methodology to apply GA
CHAPTER 2. LITERATURE REVIEW

Procedure. 1 The GA pseudo code [113]

Set $n = 0$;

Initialize population $P(0)$;

Evaluate $P(0)$;

repeat
    $S = \text{selectforvariation}(P(t))$;
    \hspace{0.5cm} \text{crossover}(S);
    \hspace{0.5cm} \text{mutation}(S);
    \hspace{0.5cm} \text{evaluate}(S);
    \hspace{0.5cm} P(t + 1) = \text{selectionforsurvival}(P(t), S);
    \hspace{0.5cm} t = t + 1;

until terminate=true

on FCM, called real-code genetic algorithm (RCGA), was carried out by Wojciech Stach et al. in 2005 [98]. The work converted a FCM weight matrix to a chromosome coded as real numbers. A function $f$ was designed to calculate the fitness for each chromosome based on normalized error summary. Chromosomes will be cross joined with each other based on its fitness. Chromosomes with higher fitness will have higher probability to cross join with other chromosomes. Mutation randomly happens on chromosomes. Comparing to unsupervised Hebbian learning algorithms, GA can find out a solution without pre-estimated causal relationships.

2.3.3.3 Causation and correlation in FCMs

The essentiality of credit rating is predicting the probability of default of a corporation based on its historical financial data. From statistical point of view, the credit rating is a classification problem, which is to classify corporations into different categories by their credit risks from low to high. From the financial point of view, the causal relationships
between the corporate financial factors lead the credit rating. Therefore, financial firms like to know the causation behind the credit rating of a counterparty.

FCM is a causality inferring tool, in which the concepts are associated by causal relationships. However, some FCM applications used word “correlation” instead of "causality" or “causation” [109,115–117] to describe the relationships between concepts. The relationship between causation and correlation has been discussed for more than centuries. Comparing to correlation, causality is difficult to be identified, because between two correlated characteristics, the other possible causes of variation are often seem to be beyond control. There are many studies on the relationships between causation and correlation. An early work by S. Wright in 1921 [118] used Cognitive Map (CM) to study the causalities behind the correlations.

A simple causation rule was given by Schutt et al. in 2011 [119]. It suggests that a causal relationship is recognized if the events of two characteristics

• have temporal ordering
• have correlation
• and not related to unobserved characteristics

The work also pointed out that the difficulty of establishing rule 1 and rule 3 is higher than that of rule 2, especially for the rule 3. Therefore, it is extremely difficult to establish causality between two correlated events or observances based on computational process. In contrast, there are many statistical tools to establish a statistically significant correlation [120]. Correlation does not imply causation, but it does give hints of causation. Comparing to traditional causation analysis, the statistical causation analysis is more emphasized on the effects of causes rather than on the causes of effects [121]. Therefore, correlations are always used for causality inferring when there is no clear cause-effect presents because of its accessibility [122]. Investigating on the relationship between correlation and causation is one of our objectives in this work.
CHAPTER 2. LITERATURE REVIEW

2.4 Summary

The chapter gives a brief introduction of related technologies. Following chapters describe how these technologies help improve credit ratings on accuracies, efficiencies, and flexibilities.
Chapter 3

A comparison work of machine learning based credit rating models

As described in last Chapter, four learning algorithms, BP, ELM, I-ELM, FCM, and SVM were studied for the credit rating problem. In order to compare their performances on the problem, an empirical comparison study is carried out over these algorithms on a data set consisting of real financial data, which are derived from the annual financial reports of public listed companies in United States. In general, the performance of AI credit rating approaches can be measured by the following metrics

- **The reliability:** This indicates how reliable a credit rating approach is. The reliability can be measured by three sub-measurements: (i) **The accuracy** is used to measure the overall error produced on testing data when an approach is applied [18–20, 26, 123]. The error on testing data gives an indication, on which we can estimate the error of real data. (ii) **The overfitness** measures the susceptibilities of the learning algorithms to over fit the training data sets. This also serves as a key step to test whether the patterns produced by learning algorithms can occur in a wider data set. The difference between testing error and learning error is usually used to measure overfitness [124]. (iii) **The error distribution** is used to measure the distribution of errors for each single data in a set of given credit data. Credit ratings are used to support decisions for corporate transactions. An error exceeding certain scope on single case will bring large credit risk to financial firms. The error
distribution has been used in credit rating industry to assess the reliability of rating prediction models for years [125].

- **The discrimination capacity:** Discrimination capacity measures the ability that a model differentiate “negative” and “positive” samples. The discrimination capacity are measured by two sub-measurements: (i) **Rating distribution** is used to measure the discrimination capability by evaluating its result distributions. During the 21 rating levels, cut-off ratings, such as “Ba1” for MIS and “BB+” for SnP, are used to differentiate “negative” and “positive” corporations. Corporations with ratings higher than “Ba1” or “BB+” rating are “prime”, while others are not. It has been approved that the discrimination capability is decided by credit rating distributions and cut-off ratings [53,126]. Rating agencies avoid assign the cut-off ratings to improve the discrimination capabilities of their credit ratings. Meanwhile, rating agencies do not assign too many corporations the ratings on the highest and lowest end. Therefore, the rating distribution is a critical measurement to evaluate the quality of a rating approach [127]. Without a proper rating distribution, a credit rating approach cannot be used as an independent approach, but just as a reference. (ii) **The ROC testing:** CAP and ROC are the most popular statistical testing approaches in the financial industry [52–56] to measure the discrimination capacity. The International Monetary Fund (IMF) preferred the ROC to the CAP to test the discrimination of credit rating models [54]. Therefore, the ROC is used in the work to benchmark the different models.

Thus, in this experimental study, four approaches are measured in terms of accuracy, overfitness, error distribution, rating distribution, and ROC.

### 3.1 Data sets

A data set consisting of real financial data is used in our experimental comparison study. In this sub-section, a simple description is given regarding how we obtain the real financial
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

data, what pre-processes and normalizations need to be carried out on transforming the financial data into the inputs and target outputs for the five approaches, i.e., BP, ELM, I-ELM, SVM, and FCM.

3.1.1 Raw financial data acquisition

The financial reports are downloaded with the corresponding ratings from two rating agencies, i.e., Moody’s Investor Services (MIS) and Standard and Poor’s rating service (SnP)\(^1\). Usually, rating agencies and financial firms assess corporates’ credit risks based on their financial factors capturing the features of the credit ratings. Such financial factors can be obtained by analyzing the corporate financial statements \([42]\), which are part of the downloaded reports. A financial statement contains financial data measuring the corporate business performance, where the frequently used financial data includes total revenue or total sales, total debts, total liabilities, current liabilities, total assets, current assets, net profit, earnings before interest and taxes (EBIT), etc.

3.1.2 Data pre-process

After obtaining the raw financial data, the following data pre-processes is carried out on data to prepare the inputs and target outputs for the five algorithms: (i) transforming financial data from absolute values into financial ratios; (ii) normalizing the value of selected financial ratios so as to guarantee that the inputs are between the range \([-1, 1]\); and (iii) dividing data sets into multiple data sets for training and testing purposes.

The financial data from reports are absolute values, which have dependency on the corporation size. A financial ratio is a relative magnitude of two selected financial data, which makes it more appropriate for evaluating corporate financial conditions. Frequently used financial ratios include profitable ratios, liquidity ratios, debt ratios, etc. Kumar and Bhattacharya \([76]\) listed 25 financial ratios to evaluate the credit ratings. Meanwhile, it has been proven that some of listed financial ratios are less correlated to credit ratings \([20,128]\).

\(^1\)The financial reports are available from EDGAR system (http://www.sec.gov/edgar.shtml) by search Form 10-K. The ratings for public listed companies are available on the website of rating agencies.
Combining the findings in these papers, high correlated financial ratios identified in [20,128] is chosen to train SLFNs in our experiment. To be specific, one debt ratio, two liquidity ratios, and five profitable ratios have been chosen as financial factors, which are listed in Table 3.1.

<table>
<thead>
<tr>
<th>Ratios in [20] (p-values) a</th>
<th>Ratios in [128] (r) b</th>
<th>Type</th>
<th>Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debt ratio (0)</td>
<td>DR (0.1818)</td>
<td>Debt ratio</td>
<td>Total Assets</td>
</tr>
<tr>
<td>Current ratio (0.36)</td>
<td>CR (0.15071)</td>
<td>Liquidity ratio</td>
<td>Current Assets</td>
</tr>
<tr>
<td>Quick ratio (0.37)</td>
<td>QR (0.12926)</td>
<td>Liquidity ratio</td>
<td>Current Liabilities</td>
</tr>
<tr>
<td>Net profit margin (0)</td>
<td>NPM (0.27147)</td>
<td>Profitable ratio</td>
<td>Net profit</td>
</tr>
<tr>
<td>Gross profit margin (0.02)</td>
<td></td>
<td>Profitable ratio</td>
<td>Gross profit</td>
</tr>
<tr>
<td>Return on total assets (0.01)</td>
<td>EBIT/TA (0.50944)</td>
<td>Profitable ratio</td>
<td>Total assets</td>
</tr>
<tr>
<td></td>
<td>NP/TA (0.48495)</td>
<td>Profitable ratio</td>
<td>Net profit</td>
</tr>
<tr>
<td></td>
<td>EBIT/Sales (0.31458)</td>
<td>Profitable ratio</td>
<td>Sales</td>
</tr>
</tbody>
</table>

aHigher p-value indicates lower importance of the ratio.
bHigher |r| indicates higher importance of the ratio

Table 3.1: Selected financial ratios.

The rating data is downloaded from two rating agencies, i.e., MIS and SnP. They are calculated by different rating methodologies. In order to avoid the possible learning errors brought by the differences and separate training data and testing data, the data has been divided into four rating sets, and illustrated in Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>MIS ratings</th>
<th>SnP ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data</td>
<td>80 records</td>
<td>80 records</td>
</tr>
<tr>
<td>Testing data</td>
<td>338 records</td>
<td>700 records</td>
</tr>
</tbody>
</table>

Table 3.2: Data sets for experiments

3.1.3 Data normalization

To improve the accuracy of the learning and prediction, data need to be normalized into real numbers in [−1, 1]. In MIS and SnP ratings, corporations are categorized into 21 different levels according to their credit ratings from “excellent” to “poor”. Each class is labelled using alphanumeric symbols, where MIS and SnP use different labels for the same level. For example, the most “excellent” class is labelled as “Aaa” and “AAA” by MIS and
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

SnP respectively. Some researchers use other labels to classify credit ratings, e.g., integer numbers in [0, 12] and real numbers in [−6, 6] were used by [129] and [126] respectively. To normalize the values of credit rating labels for our experiments, real numbers in [−1, 1] are assigned to MIS and SnP ratings, as illustrated in Table 3.3.

<table>
<thead>
<tr>
<th>MIS labels</th>
<th>Aaa</th>
<th>Aa1</th>
<th>Aa2</th>
<th>Aa3</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>Baa1</th>
<th>Baa2</th>
<th>Baa3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SnP labels</td>
<td>AAA</td>
<td>AA+</td>
<td>AA</td>
<td>AA-</td>
<td>A+</td>
<td>A</td>
<td>A-</td>
<td>BBB+</td>
<td>BBB</td>
<td>BBB-</td>
</tr>
<tr>
<td>Real numbers</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS labels</th>
<th>Ba1</th>
<th>Ba2</th>
<th>Ba3</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>Caa1</th>
<th>Caa2</th>
<th>Caa3</th>
<th>Ca</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>SnP labels</td>
<td>BB+</td>
<td>BB</td>
<td>BB-</td>
<td>B+</td>
<td>B</td>
<td>B-</td>
<td>CCC+</td>
<td>CCC</td>
<td>CCC-</td>
<td>CC</td>
<td>C</td>
</tr>
<tr>
<td>Real numbers</td>
<td>0</td>
<td>-0.1</td>
<td>-0.2</td>
<td>-0.3</td>
<td>-0.4</td>
<td>-0.5</td>
<td>-0.6</td>
<td>-0.7</td>
<td>-0.8</td>
<td>-0.9</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 3.3: The normalization of credit ratings

There is no normalization needs to be carried out on some chosen financial ratios such as NPM, EBIT/TA, etc. While normalization is required for the DR and CR to make sure all DR and CR values fall in [−1, 1].

3.2 Experimental environments and parameters

In this section, we introduce our experimental environments and the parameters settings. First, the all approaches are implemented and tested under Java platform on a Intel Core 2 Duo 3.0GHz CPU PC with 3GB memory. The neuroph [130] and libsvm [131] are used to implement BP and SVM respectively. ELM and I-ELM are implemented independently based on an open source matrix lib, i.e., EJML [132].

There are two approaches to set SLFNs for $k$-class classification problems, the single output node approach and the multioutput node approach [79]. The single output node approach is used in our experiments to test BP, ELM, and I-ELM. To make sure all the approaches are set properly, the MIS rating data is used to calibrate them to find best settings for the different approaches. As discussed in previous section, the values of credit ratings categorize corporations into different classes, which have linear relationships among them. Therefore, credit ratings can be regarded as discrete regression problems. Actually,
regression models have been applied in k-class classification problems \cite{133} as well as credit ratings \cite{134}. Since the credit ratings have been normalized to be real numbers, for a given testing data set, i.e., \( S = \{ \langle x_i, t_i \rangle | i = 1, 2, ..., N \} \), the error generated by a credit rating approach can be measured by the average of the summation of the distance between the actual output and the desired output of every sample, as defined by the error function in equation 3.1.

\[
e(S) = \frac{\sum_{\langle x_i, t_i \rangle \in S} \| o_i - t_i \|^2}{N}.
\] (3.1)

### 3.2.1 ELM parameters

Based on equation 2.14, the hidden layer size and the parameter \( \lambda \) have to be decided before ELM can be applied. We test the hidden layer size and \( \lambda \) from range [5, 150] and \( \{2^{-18}, 2^{-17}, ..., 2^{17}, 2^{18}\} \) respectively with the MIS rating sets. Figure 3.1 depicts the testing error distribution against the hidden layer size and \( \lambda \).

**Figure 3.1:** The testing error distribution of ELM.
The results show that the testing error is very stable when the hidden layer is large enough, which is consistent with the result reported by [79]. Meanwhile, the lowest testing error appears at $\lambda = 2^2$ and $\text{hiddenlayersize} = 20$.

### 3.2.2 I-ELM parameters

The I-ELM is an incremental learning approach. During the learning process, hidden nodes are added one by one to increase the learning accuracy. To decide when to stop the learning process, the I-ELM learning algorithm has been executed for several times to find the best result. Figure 3.2 depicts the results of the three learning trials, denoted as trial A, B, and C. From Figure 3.2, we can observe that the values of the best hidden layer size of I-ELM for trial A and B are less than 20, and the one of trial A is 4. Meanwhile, we also notice that trial C obtains the highest accuracy when the hidden layer size is 40. In addition, when the hidden layer size of the I-ELM is large enough, the testing errors of all I-ELM learning trials are stable, meaning that adding new hidden nodes does not impact much the accuracy. Therefore, in order to make sure that we can present the best performance for I-ELM, we choose 40 as the I-ELM hidden layer size.

### 3.2.3 BP parameters

Based on equation 2.12, there are two parameters, i.e., the learning rate $\gamma$ and the size of the hidden layer, have to be decided before BP can be applied. We test the hidden layer size and $\lambda$ in ranges $[5, 150]$ and $\{0.001 \times 1.5^0, 0.001 \times 1.5^1, ..., 0.001 \times 1.5^{11}, 0.001 \times 1.5^{12}\}$ respectively with the MIS rating sets. Figure 3.3 shows the testing error distribution against $\lambda$ and the size of the hidden layer. The result shows that the testing error is stable with proper learning rate when the hidden layer size is large enough. Increasing the size of hidden layer will not improve performance. The best testing error appears when learning rate $\gamma = 0.001 \times 1.5^5$ and hidden layer size is 100.
Figure 3.2: The testing errors of three I-ELM learning trials.

3.2.4 SVM parameters

The One Against One MSVM with RBF kernel is used in our experiment. Based on equations 2.16, 2.17, and 2.18, two parameters, i.e., the $\gamma$ and $C$, play very critical role on performance. The cross validation based grid search was introduced in [88] to find best parameters for RBF kernel. By running the grid search against the MIS training data, the best point is found when $C = 128, \gamma = 0.125$.

3.2.5 FCM parameters

A three-layer FCM has been created for the testing. The first layer consists of 8 concepts for the 8 selected input factors, denoted by $I = c_i, i \in [1, 8]$. The second layer is the middle concept layer, which includes $k$ concepts, denoted by $M = c_m, m \in [1, k]$. The final layer has only one concept, the rating result concept, $c_0$. There is no expert knowledge used to build the relationships between concepts. Three kinds of relationships have been created between different layers. The input relationships link input concepts to middle concepts.
Figure 3.3: The testing error distribution of BP.

i.e., for any $c_i \in I$ and $c_m \in M$, there is a relationship from $c_i$ to $c_m$. Among the middle layer concepts $M$, for any two adjacent concepts, $c_m$ and $c_{m+1}$, a pair of relationships link them from $c_m$ to $c_{m+1}$ and from $c_{m+1}$ to $c_m$ as well. The last set of relationships link the middle concepts and the output. For each middle concept $c_m \in M$, two relationships created to link $c_m$ to $c_o$ and $c_o$ to $c_m$ as well. The FCM diagram is illustrated in Figure 3.4

The learning algorithm decides the weights of these relationships based on the training data. Before that, there are two parameters have to be decided. The size of middle layer $k$ and the inferring length $l$. Without expert knowledge, a set of experiments has been conducted to test the $k$ and $l$ in ranges $[2, 15]$ and $[2, 50]$ respectively. The best testing error appears around where the middle size is 4 and the inferring length is 6. Comparing to the hidden layer size of ELM, I-ELM, and BP, which are 20, 40, and 100 respectively, the size of FCM is small.
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

Figure 3.4: The FCM structure for credit rating

3.3 Experiment results

After all parameters have been decided, the four data sets are fed into approaches to evaluate their performance in terms of the reliability and the rating distributions.

3.3.1 The reliability

The reliability of the learning approaches can be measured from three aspects, i.e., the accuracy, the overfitness, and the error distribution. First, the error function defined in equation 3.1 is used to calculate the total error against different data sets. For training results, the function gives the training error of the approach. The testing error can be achieved by apply the function to the testing results. Testing error indicates the accuracy of each approach.

In previous work, the testing error has been used as indicator of overfitness [124,135]. Therefore, we measure the overfitness by comparing the training error and the testing error. equation 3.2 is used to measure the overfitness for each approach.
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

\[ O(T, T') = \frac{||e(T) - e(T')||}{e(T')} \]  \hspace{1cm} (3.2)

where \( T \) denotes the testing data, \( T' \) denotes the training data.
### Table 3.4: The learning errors, testing errors, and overfitness comparison.

<table>
<thead>
<tr>
<th>Method</th>
<th>Training Error</th>
<th>Testing Error</th>
<th>Overfitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELM</td>
<td>0.006</td>
<td>0.0122</td>
<td>0.1314</td>
</tr>
<tr>
<td>I-ELM</td>
<td>0.0056</td>
<td>0.01235</td>
<td>0.139</td>
</tr>
<tr>
<td>BP</td>
<td>0.0067</td>
<td>0.01279</td>
<td>0.0123</td>
</tr>
<tr>
<td>SVM</td>
<td>0.0083</td>
<td>0.0175</td>
<td>0.0152</td>
</tr>
<tr>
<td>FCM</td>
<td>0.0002</td>
<td>0.00270</td>
<td>0.0224</td>
</tr>
</tbody>
</table>
Table 3.4 gives the results of errors and the overfitnesses for each approach applied on different data sets. From the results, we have that the SVM gives best training error on both rating sets. However, regarding on testing error, the ELM outperforms other approaches on the MIS rating set, and the BP is the best on the SnP rating set. Regarding overfitness, all SLFN based approaches show the consistency between training error and testing error, while I-ELM is the best for both MIS and SnP rating sets. FCM shows its capability on the learning accuracy. Its testing errors outperform BP and I-ELM on MIS rating set, and ELM, I-ELM, and SVM on SnP rating set. There is also an interesting observation that the testing errors of ELM and FCM are lower than their training errors on both MIS and SnP rating sets.

The testing error and overfitness provide a summary of overall performance. For credit ratings, a reliable approach controls the error for each individual case in an acceptable level to make sure the error will not cause significant lose. In order to compare the error distribution, the error for a rating record \((x_i, t_i) \in S\) is defined using the distance between \(t_i\) and the output \(o_i\), as defined in equation 3.3.

\[
d_i = ||t_i - o_i||
\]  

(3.3)

To measure the error distribution for a rating approach \(R_k\), each single data \(x_i\) is classified to four categories according to the value of \(d_i\), i.e., \(0 < d_i \leq 0.2, 0.2 < d_i \leq 0.4, 0.4 < d_i \leq 0.6, 0.6 < d_i\). The four categories represent correct, acceptable, wrong, and opposite results respectively. Table 3.5 shows the results distributions on the four categories for different rating approaches applied to particular testing data respectively.

From Table 3.5, we can see that all outputs of the three SLFN based approaches are either correct or acceptable for MIS rating set, while one of the SVM outputs is wrong. ELM and BP produce the best error distribution on MIS rating set. For SnP rating set, the BP produces 2 wrong outputs, ELM produces 13 wrong outputs, and I-ELM produces 13 wrong outputs and 2 opposite outputs. FCM produces only 1 wrong result on MIS rating set, and 11 wrong results on SnP rating set. Although SVM has closed performance
### Chapter 3. A Comparison Work of Machine Learning Based Credit Rating Models

<table>
<thead>
<tr>
<th></th>
<th>0 &lt; ( d_i ) ≤ 0.2 Correct</th>
<th>0.2 &lt; ( d_i ) ≤ 0.4 Acceptable</th>
<th>0.4 &lt; ( d_i ) ≤ 0.6 Wrong</th>
<th>0.6 &lt; ( d_i ) Opposite</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIS rating set</td>
<td>ELM 335</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>I-ELM 329</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>BP 335</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SVM 330</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FCM 332</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SnP rating set</td>
<td>ELM 592</td>
<td>95</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>I-ELM 570</td>
<td>115</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>BP 632</td>
<td>66</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SVM 614</td>
<td>66</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>FCM 599</td>
<td>90</td>
<td>11</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.5: The error distribution comparison.

...on accuracy comparing to other approaches, it obtains the worst error distribution among the five approaches. SVM produces 20 *wrong* and *opposite* outputs on SnP rating data, which is 2.9% of test data.

Based on the aforementioned experimental results, we can see that SVM performs well on accuracy (the 2nd best approach on both MIS and SnP rating sets). However, it cannot compare with any SLFN based approach on overfitness and error distributions. Such results largely reduce the reliability of SVM for credit ratings.

#### 3.3.2 The discrimination capacity

##### 3.3.2.1 Rating distribution

Figure 3.5 shows the original rating distributions for MIS and SnP testing data respectively, i.e., the target rating distribution. The output distributions for the five approaches are depicted in Figure 3.6 and 3.7 respectively.

The conformation between the output distribution and the target distribution can be measured by comparing the number of corporations in each rating level. For a given rating result set \( S = \{s_1, s_2, ..., s_n\} \), let \( n_l(S) \) denote the number of corporations assigned to the rating level \( l \). For a rating level \( l \), assuming that \( O = \{o_1, o_2, ..., o_n\} \) represents the set of output and \( T = \{t_1, t_2, ..., t_n\} \) represents the set of target results, then the conformation
Figure 3.5: The rating distributions for MIS and SnP testing data.

between the output and the target can be measured by a distance function defined by equation 3.4
Based on equation 3.4, the overall conformation between the output distribution and

\[ m_t(O, T) = \frac{||n_t(O) - n_t(T)||}{n_t(O) + n_t(T)} \]  \hspace{1cm} (3.4)

Figure 3.6: The output distributions for MIS testing data.
the target distribution for all rating levels can be obtained by the summation of the conformation of every rating level. This is, the overall conformation can be derived by the equation 3.4. Particularly, the conformation between the output and the target results
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

3.6. 

\[ TP + FN = \sum_{x_i \in X} (1 - PD(x_i)) \]  

\[ M(O, T) = \sum_{i=1}^{N} m_i(O, T) = \sum_{i=1}^{N} \frac{|n_i(O) - n_i(T)|}{n_i(O) + n_i(T)} \]  

\[ M'(O, T) = m_{Ba2}(O, T) + m_{Ba1}(O, T) + m_{Baa3}(O, T) \]

Table 3.6: The conformation of rating distributions for ELM, I-ELM, BP, and SVM.

<table>
<thead>
<tr>
<th></th>
<th>ELM</th>
<th>I-ELM</th>
<th>BP</th>
<th>SVM</th>
<th>FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIS testing data</td>
<td>5.246</td>
<td>5.997</td>
<td>7.803</td>
<td>4.504</td>
<td>4.987</td>
</tr>
<tr>
<td>[ M(O, T) ]</td>
<td>0.764</td>
<td>0.776</td>
<td>0.580</td>
<td>0.264</td>
<td>0.6962</td>
</tr>
<tr>
<td>[ M(O, T) ]</td>
<td>0.964</td>
<td>0.698</td>
<td>0.865</td>
<td>0.660</td>
<td>0.879</td>
</tr>
</tbody>
</table>

Table 3.6 shows that SVM outperforms the other three approaches on threshold level distributions. That is, SVM can help users to largely differentiate “prime” corporations and “not prime” corporations. Regarding the overall distribution, SVM is the best approach on MIS rating set, while BP is the best approach on SnP rating set.

3.3.2.2 ROC testing

According to Section 2.2.5.2, the ROC testing is a recommended approach by IMF to test the discrimination capacity of rating models. To draw the ROC curve, the real default data is required. However, the default data is not available for the companies we collected. Therefore, the only way to generate the ROC curve is based on RD data published by rating agencies to estimate the errors. The PD data for Moody’s rating is available in [136]. The idealized PD of different corporate credit rates are listed in Table 3.7

For the given testing data set \( X = \{x_1, x_2, x_3, ..., x_k\} \), the \( PD(x_i) \) is the PD of \( x_i \in X \) given by the official credit rating mapping table. The number of positive corporations and negative corporations can be estimated by equation 3.7 and 3.8 respectively.

The results in Table 3.6 show that SVM outperforms the other three approaches on threshold level distributions. That is, SVM can help users to largely differentiate “prime” corporations and “not prime” corporations. Regarding the overall distribution, SVM is the best approach on MIS rating set, while BP is the best approach on SnP rating set.
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

\[ FP + TN = \sum_{x_i \in X} PD(x_i) \]  

(3.8)

For a rating model \( \mathcal{M} \), the \( \mathcal{M}(x_i) \) is the credit rate calculated by the model for a sample \( x_i \in X \). By a given threshold rate \( t \), the \( TP \) and \( FP \) can be estimated by the equation 3.9 and 3.10 respectively.

\[ TP = \sum_{\mathcal{M}(x_i) > t} (1 - PD(x_i)) \]  

(3.9)

\[ FP = \sum_{\mathcal{M}(x_i) > t} PD(x_i) \]  

(3.10)

Therefore, by a given threshold value \( t \) the \( TPR \) and \( FPR \) can be estimated by equation 3.11 and 3.12 respectively.

\[ TPR = \frac{TP}{TP + FN} = \frac{\sum_{\mathcal{M}(x_i) > t} (1 - PD(x_i))}{\sum_{x_i \in X} (1 - PD(x_i))} \]  

(3.11)

\[ FPR = \frac{FP}{FP + TN} = \frac{\sum_{\mathcal{M}(x_i) > t} PD(x_i)}{\sum_{x_i \in X} PD(x_i)} \]  

(3.12)

Based on the equation 3.11 and 3.12, the ROC estimations of different rating models for 1 year, 5 years, and 10 years default estimations have been calculated and shown in Figure 3.8, 3.9, 3.10 together with corresponding estimations of Moody’s rating for comparison.

<table>
<thead>
<tr>
<th>MIS labels</th>
<th>Aaa</th>
<th>Aa1</th>
<th>Aa2</th>
<th>Aa3</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>Baa1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-year PD (%)</td>
<td>0.0001</td>
<td>0.0006</td>
<td>0.0014</td>
<td>0.0030</td>
<td>0.0058</td>
<td>0.0109</td>
<td>0.0389</td>
<td>0.0900</td>
</tr>
<tr>
<td>5-years PD (%)</td>
<td>0.0029</td>
<td>0.0310</td>
<td>0.0680</td>
<td>0.1420</td>
<td>0.2610</td>
<td>0.4670</td>
<td>0.7300</td>
<td>1.1000</td>
</tr>
<tr>
<td>10-years PD (%)</td>
<td>0.0100</td>
<td>0.1000</td>
<td>0.2000</td>
<td>0.4000</td>
<td>0.7000</td>
<td>1.2000</td>
<td>1.8000</td>
<td>2.6000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS labels</th>
<th>Baa2</th>
<th>Baa3</th>
<th>Ba1</th>
<th>Ba2</th>
<th>Ba3</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-year PD (%)</td>
<td>0.1700</td>
<td>0.4200</td>
<td>0.8700</td>
<td>1.5600</td>
<td>2.8100</td>
<td>4.6800</td>
<td>7.1600</td>
<td>11.6200</td>
</tr>
<tr>
<td>5-years PD (%)</td>
<td>1.5800</td>
<td>3.0500</td>
<td>5.2800</td>
<td>8.4100</td>
<td>11.8600</td>
<td>16.1200</td>
<td>20.7100</td>
<td>27.0500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS labels</th>
<th>Caa1</th>
<th>Caa2</th>
<th>Caa3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-year PD (%)</td>
<td>17.3816</td>
<td>26.0000</td>
<td>50.9902</td>
</tr>
<tr>
<td>5-years PD (%)</td>
<td>36.3137</td>
<td>48.7500</td>
<td>69.8212</td>
</tr>
<tr>
<td>10-years PD (%)</td>
<td>47.7000</td>
<td>65.0000</td>
<td>80.7000</td>
</tr>
</tbody>
</table>

Table 3.7: Moody’s corporate idealized 1, 5, and 10 years cumulative probability of default (PD) rates
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

![1 year ROC estimation](image)

**Figure 3.8:** The ROC estimation of 1 year default
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

Figure 3.9: The ROC estimation of 5 years default
Figure 3.10: The ROC estimation of 10 years default
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

In a ROC curve, the point closest to the \((0, 1)\) is the best threshold point to discriminate the positive and negative samples. The distance from the \((0, 1)\) to the threshold point indicates the discrimination capacity of the model. When a model has a better discrimination capacity than others, its threshold point is closer to \((0, 1)\) than others. Table 3.8 shows the threshold points for different models and their distance to the \((0, 1)\).
<table>
<thead>
<tr>
<th></th>
<th>1 year</th>
<th></th>
<th>5 years</th>
<th></th>
<th>10 years</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Threshold point (t)</td>
<td>Distance</td>
<td>Threshold point (t)</td>
<td>Distance</td>
<td>Threshold point (t)</td>
<td>Distance</td>
</tr>
<tr>
<td>ELM</td>
<td>B1</td>
<td>0.3656</td>
<td>Ba3</td>
<td>0.4156</td>
<td>Ba3</td>
<td>0.4172</td>
</tr>
<tr>
<td>I-ELM</td>
<td>B2</td>
<td><strong>0.3216</strong></td>
<td>Ba3</td>
<td><strong>0.3799</strong></td>
<td>Ba3</td>
<td><strong>0.3927</strong></td>
</tr>
<tr>
<td>BP</td>
<td>B1</td>
<td>0.3252</td>
<td>Ba3</td>
<td>0.3916</td>
<td>Ba3</td>
<td>0.4028</td>
</tr>
<tr>
<td>SVM</td>
<td>B1</td>
<td>0.3442</td>
<td>Ba3</td>
<td>0.4290</td>
<td>Ba3</td>
<td>0.4363</td>
</tr>
<tr>
<td>FCM</td>
<td>Ba3</td>
<td>0.4121</td>
<td>Ba2</td>
<td>0.4618</td>
<td>Ba2</td>
<td>0.4766</td>
</tr>
</tbody>
</table>

Table 3.8: The threshold points and distance to (0, 1)
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

The result shows that I-ELM outperforms other models in all three ROC tests, which contrasts its performance in the rating distribution test. The SVM, which outperforms other models in the distribution test, only outperforms FCM in the 5 years and 10 years ROC tests. In the experiment, ROC curves are estimated based on the ideally PDs of Moody’s ratings, which are credit ratings for 1-year. In the 5 years and 10 years ROC estimations, I-ELM, ELM, and BP have better ROC estimations than the original Moody’s rating. It weaks the reliabilities of 5 years and 10 years ROC estimations as benchmarks of discrimination capacity. Therefore, only the result of 1 year ROC estimation is put into benchmarks.

3.3.2.3 Summary of the results

In this experimental comparison study, we compare ELM, I-ELM, BP, and SVM on accuracy, overfitness, error distribution, rating distribution, and ROC estimation. Table 3.9 lists the two best approaches for each measurement. Considering ELM and I-ELM are extreme learning (EL) approaches, while ELM, I-ELM and BP are SLFN based approaches, the results can be summarized and discussed from the following three perspectives.

- SVM vs SLFN based approaches: When SVM and the three SLFN based approaches are concerned, SVM performs well on rating distribution and SLFN based approaches outperform SVM on reliability.

- BP vs EL: When BP and the two EL based approaches are concerned, these three SLFN based approaches achieve close performance on rating distributions while ELM obtains the best performance on reliability measurements. Meanwhile, we notice that the hidden node size of the BP trained SLFN, which is 100, is much larger than that of the ELM and the I-ELM. Therefore, we can have that EL approaches achieve more compact network architecture than BP.

- ELM vs I-ELM: When these two EL approaches are concerned, ELM outperforms I-ELM on accuracy and error distribution, which is consistent with results from previous
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

<table>
<thead>
<tr>
<th></th>
<th>Reliability</th>
<th>Discrimination capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measurements</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>ELM, SVM, FCM</td>
<td>BP, FCM</td>
</tr>
<tr>
<td>Overfitness</td>
<td>I-ELM, ELM</td>
<td>I-ELM, ELM</td>
</tr>
<tr>
<td>Error Distribution</td>
<td>ELM, BP</td>
<td>BP, FCM</td>
</tr>
<tr>
<td>Rating distribution</td>
<td>SVM, ELM</td>
<td>BP, SVM</td>
</tr>
<tr>
<td>Threshold distribution</td>
<td>SVM, BP</td>
<td>SVM, I-ELM</td>
</tr>
<tr>
<td>ROC testing</td>
<td>I-ELM, SVM</td>
<td>N/a</td>
</tr>
</tbody>
</table>

Table 3.9: The best two approaches on each measurement.

studies [80, 137]. Particularly, I-ELM has the lowest overfit on both MIS and SnP rating sets. This suggests a hybrid approach combining the advantages of these two EL approaches might be able to obtain high performances on all the three reliability measurements.

- **FCM**: with smaller size than SLFN based approaches, FCM achieves equivalent performance on reliability and distribution as other models. It indicates that FCM structure is more efficient than SLFN based approaches on credit rating problem. In addition, there is no domain knowledge applied on the FCM, which is a drawback of the FCM in the comparison.

3.4 Conclusion

The reliability of a credit rating approach is decided by factors such as the accuracy, the overfitness, and the error distribution. Previous work compared SVMs against BP on accuracies, which is not enough to evaluate credit rating approaches. Besides the reliability, discrimination capacity is also a critical measurement to evaluate credit rating approaches. This paper presents an experimental comparison study over these measurements on five learning algorithms, i.e., ELM, I-ELM, BP, SVM, and ELM with data extracted from the real financial reports and the corresponding ratings from two rating agencies, *Moody’s*
CHAPTER 3. A COMPARISON WORK OF MACHINE LEARNING BASED CREDIT RATING MODELS

Investor Services and Standard and Poor’s rating service. We study the effectiveness of the four algorithms in terms of reliabilities, and discrimination capacity. Our experimental results show that SLFN based approaches outperform SVMs on reliabilities, while SVMs are better on rating distributions than others, and I-ELM shows better performance on ROC testing than other approaches.

Regarding reliabilities, the overfitness and the error distribution indicate whether the output errors can be controlled in an expected scope. In real applications, learning error is the only basis for users to estimate the accuracy of output. Furthermore, the low variance of error distribution implies that the probability of “un-expected” cases is low. Therefore, an approach with lower overfitness and lower error distribution variance gives more reliable outputs than others. Based on these two measurements, the reliability of SVM is largely weakened, although it achieves good performances on accuracy and output distribution. The ELM, BP, and FCM are the most reliable approaches. ELM also achieves the best accuracy on MIS data. In addition, for the accuracy on SnP data, the performance of ELM is only 0.0005 less than what SVM achieves. Therefore, regardless of output distribution, ELM and BP outperform I-ELM and SVM on the credit rating problem. Considering the computational complexity, ELM is more attractive than BP.

The FCM’s performance in the experiment is impressive. With a 4-concept middle layer, the FCM achieve an equivalent results as BP, I-ELM, and ELM whose hidden layer sizes are 100, 40, and 20 respectively. In the experiment, FCM is trained by genetic algorithm without any priori knowledge. Its capability to combine the priori knowledge and machine learning need further investigation.

The work uses hundreds corporation samples to test the performance of AI models. However, a universal rating model for different corporations is not realistic. To make the result be more reliable, financial firms and rating agencies categorize corporations into different categories and study them separately to assess their credit risk. E.g., MIS categorize corporations into 22 market segments, 5 regions, and some political/economic groups. Each category has its own rating model, which has unique measure factors and
weights. Sometimes, exclusive rating models are also used for particular corporations. With the fine-grained categorization, the sample size has been reduced dramatically. Further investigation is necessary to study the performances of AI approaches under limited training samples.
Chapter 4

FCM based credit rating model: a case study of Nokia

Chapter 3 compares four prevalent AI models with FCMs on credit rating. The result shows that FCMs achieve equivalent performance on reliability and discrimination capacity as other models. In addition, the size of FCMs are smaller than SLFN based approaches in the experiment. It indicates that the structure of FCMs is more efficient than SLFN on credit ratings.

Beside to the reliability and discrimination, the ability to combine priori domain knowledge and understandable training results are other two important characters that a rating model should have. As discussed in Chapter 2, by combining advantages from neural networks and expert system, FCM has potential to be an AI credit rating models with these characters.

Another study carried out in this chapter is the ability of FCM to learn from small samples. In the financial industry, a universal rating model for all corporations is infeasible due to the variation of corporations. Corporations are categorized into segments according to their sizes, industries, regions, and company structures. Segments and some corporations are studied independently to improve the rating accuracies [138]. For each segment, the available data is limited. Being able to work with small sample size will improve the accuracy of credit models.

This chapter proposes an original FCM based methodology, which covers the FCM structure, the algorithm to calculate inferring steps, the GA training algorithm, and the
priori knowledge combination. The historical financial data and rating records of Nokia are selected to test the performance of the methodology on credit rating with small samples. The result shows that the methodology outperforms BP, SVM on accuracy, and by combining correlation coefficients as priori knowledge, the training results are more consistent than the original FCM models.

Section 4.1 gives the introduction of the methodology. The FCM credit rating model is described in section 4.2 together with the data selection and normalization. Section 4.3 compares the results of different models. In the end, the summary is given in Section 4.4

4.1 A FCM based methodology

4.1.1 The FCM structure

Considering that the layered FCM demonstrates an equivalent capacity on reliability and discrimination in chapter 3, a three-layer architecture for FCMs has been used in this chapter. The three layers are input layer, intermediate layer, and output layer.

(1) **Input layer**: This layer consists of input concepts, whose states are given by input data. There is no causal relations from any concepts to input concepts. Thus, the states of input concepts will not be changed during the inferring. For an input concept, there is only one causal relationship link it to a concept in the intermediate layer.

(2) **Intermediate layer**: The concepts in this layer play key roles in the inferring. They do not have input value. Their states are neutral in the beginning. During the inferring, their states are decided by effects propagated via causal relations from other concepts.

(3) **Output layer**: This layer consists of output concepts, which represent the inferring results of the FCM. There is no causal relations between output concepts, but only causal relations between intermediate concepts and output concepts.
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

Figure 4.1: The sample of layered FCMs.

Compared to the FCM structure used in chapter 3, each input concept in this model only associates with one intermediate concept. With this design, the credit rating expertise can be mapped to FCMs, which improves the interpretability of the model. Figure 4.1 depicts the FCM structure.

4.1.2 Longest path inferring

A FCM is usually an abstraction of the real world problems in most applications, especially for social and economic problems. In such applications, FCMs are obtained from domain experts by fuzzy logic theories. Particularly, errors are also brought by the inaccurate knowledge of the expert, such as incompleteness and the inaccuracy of the fuzzy relationships. FCMs are inferred via an iterative process. Errors will be accumulated iteratively and lead FCMs to unexpected results. Considering the error accumulation, an earlier state is more accurate than the later one during an inferring process\(^1\). Therefore, the inferring

---

\(^1\)The discussion of the error accumulation during the FCM inferring can be found in paper [139] and Chapter 6
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

should be stopped as early as possible in order to maintain the accuracy.

However, during the inferring of a FCM, inputs and their effects are propagated from input concepts to output concepts by the iterative inferring. The longer the inferring continues, the more causalities between concepts are calculated. In the early stage of a inferring, the output is not accurate because the incomplete coverage of input and causal relationships. If a FCM inferring is stopped once all causalities are covered for each concept, the result will retain the most accurate results with most wide coverage of causalities. In order to obtain the causality coverage as well as the accuracy, an algorithm called longest path algorithm is proposed to calculate the shortest inferring step length, which covers all causalities for a given FCM. The algorithm is defined in Procedure 2.

Procedure. 2 The longest path algorithm.

1: procedure CALCULATEINFERRINGSTEP($FCM(V, W)$) ▶ $V$ is the concept set, and $W$ is the weight matrix
2: for every $v_i \in V$, $l_i \leftarrow \emptyset$
3: $P \leftarrow 0$
4: repeat
5: newcausality $\leftarrow$ false
6: for $v_i \in V$ do
7: for $v_j \in V$ do
8: if $w_{ij} \neq 0$ and ($w_{ij} \not\in l_j$ or $l_i \not\subseteq l_j$) and ($l_i \neq \emptyset$ or $v_i$ is input concept) then
9: $l_j = l_j \cup \{w_{ij}\} \cup l_i$
10: newcausality $\leftarrow$ true
11: end if
12: end for
13: end for
14: $P \leftarrow P + 1$
15: until newcausality $= false$
16: end procedure

The algorithm maintains a causal relation set $l_i$ for each concept $v_i$, and simulate the inferring process to propagate relation sets to linked concepts. For a concept $i$, its relation set $l_i^k$ at step $k$ ($k > 0$) is calculated by equation 4.1

$$l_i^k = \{w_{ji} | w_{ji} \neq 0 \text{ and } l_j \neq \emptyset\} \cup \left( \bigcup_{w_{ji} \neq 0} l_j^{k-1} \right)$$  \hspace{1cm} (4.1)
CHAPTER 4. FCM based credit rating model: a case study of Nokia

If at the $k$th step, no concept receives any new relation to its relation set, i.e. $\forall l_j \in L, t_{j}^{k} = t_{j}^{k-1}$, according equation 4.1, for any following step $k'$, which $k' > k$, that $\forall l_i \in L, t_{i}^{k'} = t_{i}^{k'-1}$. In another words, all linked causal relations are fully captured by every concepts at the $k-1$th step. Because of the error accumulation during an inferring process, the inferring at step $k-1$ are the most accurate result with full causality coverage. Procedure 2 calculates the step number $P$, which $P = k - 1$. The number $P$ is the length of the longest path that the input data pass through from input concepts to other concepts. The inferring, which stops at step $P$, is called the **longest path inferring**.

![Diagram of weight matrix to chromosome conversion](image)

**Figure 4.2:** The converting from a weight matrix to a chromosome.

### 4.1.3 GA based learning algorithm

As discussed in section 2.3.3.2, Genetic algorithm (GA) has been proven as an efficient approach to train FCMs from sample data. A GA procedure consists of three parts, a chromosome structure, a fitness function, and GA operators. For a FCM training, the weight matrix of FCM is the learning object. A chromosomes is defined as an array of weights. Figure 4.2 shows how to transform the weight matrix into a chromosome.

The fitness function, denoted as $F$, is decided by the distance between the target output from training samples and the real output of the model. It is defined by equation 4.2

$$F = \sum \sum \max(0, \gamma - |Output_{i}^{target} - Output_{i}^{real}|) / \gamma$$  \hspace{1cm} (4.2)
where $\gamma$ is the distance constance.

GA operators defined in RCGA [98] are applied to train the three-layer FCM in the experiment.

## 4.1.4 Priori knowledge combination

In a FCM, knowledge is represented by concepts and causal relationships between concepts. The priori knowledge can be coded in FCMs as concepts and relationships.

### 4.1.4.1 Concepts

For the FCM structure shown in Figure 4.1, the output layer is decided by the output data. The priori knowledge is used to select concepts in the input and intermediate layers. In order to facilitate the priori knowledge mapping, the structure of public rating reports is used to decide concepts in the intermediate layer and the input layer.

(1) Intermediate layer:

In a rating report, high level measurements are used to measure objects in an abstract level. E.g., to measure a personal credit score, the high level measurements could be *job*, *health*, and *finance*. The value of a high level measurement is decided by its sub-measurements, which could be other high level measurements or financial factors. There is no direct data mapping between high level measurements and financial data. In the three-layer FCM, each high level measurement is represented by an intermediate concept.

(2) Input layer:

Under each high level measurement in a rating report, there is a set of sub-measurements, which are financial factors or other high level measurements. Each financial factor has a direct data source as its input. E.g., the *monthly salary* is a financial factor of *job*. Financial factors are represented in FCMs as input concepts. Each input concept will only connect to one intermediate concept, which is the high level measurement that the corresponding financial factor belongs to.
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

4.1.4.2 Relationships

In statistics, correlation coefficient is a prevalent tool to measure dependencies between variables. A correlation coefficient \( r \) is a numerical number in \([-1, 1]\) to indicate the relationship between two variables \( X \) and \( Y \). Pearson product-moment correlation is the most popular approach to calculate correlation coefficient.

In the proposed FCM, the association between an input concept and the output concept is not representable by a numerical number directly, but a set of relationships, which link them through intermediate concepts. In the methodology, correlation coefficients between financial factors and the credit rating are applied on relationships between inputs and intermediate concepts as weights to combine the priori knowledge from correlation analysis with the FCM. Equation 4.3 shows the combination formally.

\[
    w(c_i, c_m) = r(c_i, c_o)
\]  

where \( c_i \) is an input concept linked with intermediate concept \( c_m \), \( w(c_i, c_m) \) is the relationship weight between \( c_i \) and \( c_m \), \( c_o \) is the output concept, and \( r(c_i, c_o) \) is the correlation coefficient between \( c_i \) and \( c_o \) calculated based on sample data. Therefore, the relationship weights between input concepts and the corresponding intermediate concepts are decided by the correlation between the input data and the output data of training samples. The GA will only learning relationships between intermediate concepts and the output concept.

4.2 The FCM credit rating model

Based on the methodology given in Section 4.1, following steps are given to create the FCM based credit rating model, (i) selecting high level measurements and financial factors, (ii) preparing training data and testing data, and (iii) calculating correlation coefficients.

4.2.1 Factor selection

The financial factors, which capture the features of the corporate credit rating, can be obtained from corporate financial statements. Financial statements are alphanumeric doc-
ments which consist of the results of the corporate financial accounting [42]. They contain information about the corporate business performance. Thus financial statement analysis and data mining have attracted much research interest in the past decades, including performance indicator extraction [42], anomalies detection [140], and bankruptcy prediction [141]. Considering that the main objective of this work is to demonstrate the feasibility and advantages of FCM on credit rating, the model is simplified to include only key factors. The most frequently used factors identified by public rating reports can be classified into the following three high level measurements.

(1) **Market performance**: Factors of this measurement are used to measure the successfulness of a corporation regarding its products as well as markets. The market performance of a corporation can be measured using its *net sales*, *gross profit*, and *gross profit margin*. Net sales represent the market share of a corporation, which indicates its strength in the market and satisfactory by its clients. The gross profit implies the corporation’s earning ability. If the gross profit is lower than a certain level, it will not have sufficient fund to support its operations. Finally, the gross profit margin is used to reveal the proportion of money left over from gross profit after accounting for the expense.

(2) **Financial structure**: This measurement is used to indicate the financial status of a corporation. Factors in this measurement are used to evaluate if the corporation has enough financial flexibility to support its operational strategies and product innovations. The most frequently used financial structure factors are *assets*, *liabilities*, and *liabilities/ assets*.

(3) **Company operation**: This measurement is used to evaluate the activities of a corporation’s operation. The operation will impact the future of a company. Ideally, a company with high *research & development expense* will have a positive future. The selling & market expense will increase the market shares by the company. In general, operation factors includes *research & development expense* and *selling & market expense*.
CHAPTER 4. FCM based credit rating model: a case study of Nokia

In summary, the three measurements, *market performance measurement*, *financial structure*, and *company operation* are selected for the investigation. The output is the credit rating.

According to the methodology carried out in Section 4.1, three kinds of connections are specified to build up the FCM. As shown in Figure 4.3, the first kind of connections connects input concepts to intermediate concepts. Every input concept is connected to the intermediate concept which it belongs to. The second kind of connections is used to direct intermediate concepts to the output concept and other intermediate concepts. And the third kind of connections feeds the value of the output concept back to their intermediate concepts. Based on the longest path algorithm defined in Procedure 2, the stop number of longest path inferring is 4 for the FCM depicted by Figure 4.3.

![Figure 4.3: The FCM model for credit ratings.](image)

4.2.2 Data preparation

4.2.2.1 Sample selection

Testing the performance of different rating models in limited data set is one of our objectives. Nokia is selected as the sample data for the testing because (i) The quarterly financial
reports from 2003 to 2013 are available in Nokia website, which can be used as the source of input factors; and (ii) The public rating reports for Nokia have been published by Moody’s Investors Service Inc. (Moody’s) varies from A1 to B1 in the past decade, which provides a comparative coverage as a sample.

Table 4.1: The Nokia’s rates given by Moody’s Investor Services since 2001.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>Baa2</td>
<td>Baa3</td>
<td>Bal</td>
<td>Ba3</td>
<td>B1</td>
</tr>
</tbody>
</table>

Table 4.1 depicts the rating changes for Nokia by Moody’s since September 2001. We notice that the Nokia’s credit rating have been downgraded by multiple rating agencies recently. Therefore, the data from 2008 Q1 to 2014 Q2, are chosen for our study.

Figure 4.4: The normalized financial data of Nokia.
The financial reports are captured from Nokia website from year 2008 Q1 to 2014 Q2. Data from 2008 Q1 to 2009 Q4 has been removed because the analysis methodology had been changed by Moody's in 2009 due to financial crisis. Data from 2014 Q1 to 2014 Q2 has been removed because the financial reports structure have been changed. After the data clearance, the whole set of data has been separated into two sets, the training data includes data from 2010 Q1 to 2012 Q4. The data of year 2013 is used for the testing data.

4.2.2.2 Data normalization

The financial factors can be directly obtained from corporate financial reports and rating reports, e.g., revenues and gross profit margins. However, the original financial factors cannot be applied to a FCM model, further refinement is required in order to: (i) eliminate periodical fluctuations, which commonly exist in most corporate financial reports. For example, the quarter revenues change from Q1 to Q4 with same patterns annually. The periodical fluctuations mislead the learning processes. (ii) Normalize account numbers into numerical values in [-1,1] since most AI algorithms only accept numbers in [-1,1] as input and output. (iii) Normalize credit ratings into numerical values when the ratings from reports are qualitative assessments.

To be more specific, regarding to the normalization of account numbers, the yearly growth rates are used instead of account numbers as the input since the ratios and relative changes on accounts reflect the corporation’s situation better than account numbers. In addition, annual fluctuations have been eliminated in yearly growth rates. The following equation is used to calculate the growth rate for each account.

\[ \Delta C_t = (C_t - C_{t-1\text{year}})/C_{t-1\text{year}} \] (4.4)

The credit ratings from rating agencies are qualitative assessments. A fuzzify process is carried out to normalize them into numerical values. Credit ratings are classified into 21 levels by credit rating agencies from the lowest risk to the highest risk. Rating agencies denote these levels by symbols, e.g., Moody’s Investor Services denotations are from Aaa...
Table 4.2: The 21-level rate map.

<table>
<thead>
<tr>
<th>Rates</th>
<th>Aaa</th>
<th>Aa1</th>
<th>Aa2</th>
<th>Aa3</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>Baa1</th>
<th>Baa2</th>
<th>Baa3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Rates</td>
<td>Ba1</td>
<td>Ba2</td>
<td>Ba3</td>
<td>B1</td>
<td>B2</td>
<td>B3</td>
<td>Caa1</td>
<td>Caa2</td>
<td>Caa3</td>
<td>Ca</td>
</tr>
<tr>
<td>Values</td>
<td>0</td>
<td>-0.1</td>
<td>-0.2</td>
<td>-0.3</td>
<td>-0.4</td>
<td>-0.5</td>
<td>-0.6</td>
<td>-0.7</td>
<td>-0.8</td>
<td>-0.9</td>
</tr>
</tbody>
</table>

Table 4.3: The correlation coefficients between financial factors and credit ratings

<table>
<thead>
<tr>
<th>Financial factors</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research and development expense</td>
<td>0.6894</td>
</tr>
<tr>
<td>Sales</td>
<td>0.5245</td>
</tr>
<tr>
<td>Assets</td>
<td>0.6722</td>
</tr>
<tr>
<td>Liabilities</td>
<td>-0.4802</td>
</tr>
<tr>
<td>(\frac{\text{Liabilities}}{\text{Assets}})</td>
<td>-0.9138</td>
</tr>
<tr>
<td>Gross profit</td>
<td>0.1590</td>
</tr>
<tr>
<td>Gross profit margin</td>
<td>-0.0110</td>
</tr>
<tr>
<td>Marketing expense</td>
<td>0.6970</td>
</tr>
</tbody>
</table>

4.2.3 Correlation coefficient

Pearson product-moment correlation coefficient has been widely used for more than a century in statistical analysis to analyze the dependency between two variables [122]. For the given \(X\) and \(Y\), \(X = \{x_1, x_2, ..., x_n\}\) and \(Y = \{y_1, y_2, ..., y_n\}\), the coefficient \(r\) is defined in Equation 4.5

\[
r = \frac{\sum(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum(x_i - \overline{X})^2 \sum(y_i - \overline{Y})^2}} \tag{4.5}
\]

9 correlation coefficients are calculated based on normalized training data and assigned to the FCM as relationship weights between input concepts and intermediate concepts.

4.3 Experiment result

To compare the performance of FCM with other AI models, BP based SLFN and SVM are tested with the same data. According to Section 3.2.3 and 3.2.4, the parameters of BP and
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

Table 4.4: Training errors for different algorithms

<table>
<thead>
<tr>
<th>Date</th>
<th>Moody’s</th>
<th>SVM</th>
<th>BP (Round)</th>
<th>FCM (C) (Round)</th>
<th>FCM (N) (Round)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 Q1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3545(0.4)</td>
<td>0.425(0.4)</td>
<td>0.46 (0.5)</td>
</tr>
<tr>
<td>2010 Q2</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4471(0.4)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>2010 Q3</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4919(0.5)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>2010 Q4</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4883(0.5)</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>2011 Q1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4932(0.5)</td>
<td>0.5</td>
<td>0.44 (0.4)</td>
</tr>
<tr>
<td>2011 Q2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2911(0.3)</td>
<td>0.35(0.4)</td>
<td>0.2</td>
</tr>
<tr>
<td>2011 Q3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2772(0.3)</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>2011 Q4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1466(0.1)</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>2012 Q1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0417(0)</td>
<td>0.125(0.1)</td>
<td>0.16 (0.2)</td>
</tr>
<tr>
<td>2012 Q2</td>
<td>0</td>
<td>0</td>
<td>0.0241(0)</td>
<td>-0.01(0)</td>
<td>0.12 (0.1)</td>
</tr>
<tr>
<td>2012 Q3</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.1620(-0.2)</td>
<td>-0.1875(-0.2)</td>
<td>-0.2</td>
</tr>
<tr>
<td>2012 Q4</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.1944(-0.2)</td>
<td>-0.16(-0.2)</td>
<td>-0.2</td>
</tr>
<tr>
<td>Error</td>
<td>N/a</td>
<td>0</td>
<td>0.00462</td>
<td>0.00338</td>
<td>0.00177</td>
</tr>
</tbody>
</table>

SVM are optimized to get the best performance. Meanwhile, to compare the improvements brought by the correlation coefficient combination, two different FCMs are tested in the experiment. The FCM with correlation coefficients is denoted by FCM (C), and the FCM without any preseted weight is denoted by FCM (N).

4.3.1 Learning error

Four models are trained with the data from 2010 Q1 to 2012 Q4. Table 4.4 shows the results for all three models. Errors are calculated by Equation 3.1. The result shows that with the small sample size, the BP’s training error is larger than other algorithm. Comparing the training error of FCM (N), the training error of FCM (C) is about 90% higher. It means without the constraints from the priori knowledge, the learning algorithm is easier to find a better match for the given samples. Since the training data only has 12 samples, a better match may bring problem of overfitness.

Because of the small sample size, the SVM gives a perfect match with the sample data. Its training error outperforms BP and FCM.
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

<table>
<thead>
<tr>
<th>Date</th>
<th>Moody's</th>
<th>SVM</th>
<th>BP (Round)</th>
<th>FCM (C) (Round)</th>
<th>FCM (N) (Round)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 Q1</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.2516(-0.3)</td>
<td>-0.25(-0.3)</td>
<td>-0.2</td>
</tr>
<tr>
<td>2013 Q2</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.3115(-0.3)</td>
<td>-0.25(-0.3)</td>
<td>-0.14 (-0.1)</td>
</tr>
<tr>
<td>2013 Q3</td>
<td>-0.3</td>
<td>-0.2</td>
<td>-0.1974(-0.2)</td>
<td>-0.22(-0.2)</td>
<td>0.04 (0.0)</td>
</tr>
<tr>
<td>2013 Q4</td>
<td>-0.3</td>
<td>-0.2</td>
<td>-0.2597(-0.3)</td>
<td>-0.2975(-0.3)</td>
<td>0.08 (0.1)</td>
</tr>
<tr>
<td>Error</td>
<td>N/a</td>
<td>0.005</td>
<td>0.00681</td>
<td>0.00285</td>
<td>0.0659</td>
</tr>
</tbody>
</table>

Table 4.5: Testing errors for different algorithms

4.3.2 Testing error

To verify the reliabilities of the different rating models, the data of year 2013 has been used for testing. Moody's downgraded the Nokia's Ba3 rating to B1 in August 2013, which is not included by the training sample. Therefore, the testing result tells if the model can identify the pattern, which is not included in the sample data, but implied by it. Table 4.5 shows the testing result.

The result shows that after rounding to the nearest credit rate, the BP and FCM (C) generate the same result. The FCM (C) achieves the lowest testing error than others before the rounding.

The result of FCM (N) shows that the FCM has overfit problem with small sample size if it does not combine any expert/priori knowledge.

The SVM successfully classifies the credit ratings of the first two quarters of 2013. However, since the training samples do not cover any sample for B1 rating, it fails to provide correct ratings for 2013 Q3 and Q4.

4.3.3 The consistency

GA is a random search algorithm. It produces different results in each learning process with the same training data. The consistency of trained models reveals how the training results are similar with each other. Consistent learning results are able to be understood, adjusted, and merged into an optimal one. The consistency of FCMs with the same structure can be measured by the standard deviations of causal relation weights.

78
TABLE 4.6: The consistency comparison between FCM (C) and FCM (N)

<table>
<thead>
<tr>
<th></th>
<th>Financial structure</th>
<th>Marketing performance</th>
<th>Company operation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FCM (C)</strong></td>
<td>Average</td>
<td>0.3604</td>
<td>0.1704</td>
</tr>
<tr>
<td></td>
<td>Deviation</td>
<td>0.017841</td>
<td>0.043025</td>
</tr>
<tr>
<td><strong>FCM (N)</strong></td>
<td>Average</td>
<td>0.2353</td>
<td>-0.0242</td>
</tr>
<tr>
<td></td>
<td>Deviation</td>
<td>0.39833</td>
<td>0.5328</td>
</tr>
</tbody>
</table>

The three causal relation weights from the intermediate concepts to the output concept are chosen to carry out the comparison study regarding the consistency. They are the relationships from company operation, market performance, and financial structure to ratings respectively. Each model is trained for 500 times. Figure 4.5 depicts the weight distribution for FCM (C) and FCM (N) respectively. The standard deviations and the average values are list in Table 4.6. The result indicates the standard deviation of FCM (C) causal weights is between 0.0018 and 0.043. In contrast, the corresponding standard deviation of FCM (N) causal weights is between 0.1791 and 0.5328. That is, the FCM (C) has much smaller standard deviations than those of the FCM (N). In another words, after combining with the priori knowledge, the GA training produces consistent FCM structures, which bring statistical meanings of the weights.

Considering relationship weights are numerical number in \([-1, 1]\), the standard deviations of FCM (N) results indicate that FCMs trained with the same training samples are very different from each other. With a wide range of variation, the average weights of FCM (N) are less statistically significant than FCM(C).

The average weights of trained FCM(C)s tell that the financial structure has the most positive impact to the rating result among the three intermediate concepts, and the expense of company operation is a negative factor to the rating result. The result is consistent with the past credit risk researches, in which the assets and liabilities structure are always the major factor to effect the credit rating and the operation expense does reduce the cash flow of a corporation [20, 128, 142].
CHAPTER 4. FCM based credit rating model: a case study of Nokia

(a) The weight distributions of FCM (C).

(b) The weight distributions of FCM (N).

Figure 4.5: Comparison of weight distributions
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

4.4 Summary and future work

Due to the corporate diversities on industries, scales, histories, and cultures, a universal credit rating model is infeasible. Corporations are separated into different segments for credit assessment. In rating agencies, some public corporations are studied and traced individually to increase the reliability of their credit ratings. In order to assess a corporation individually by using only its own historical data, the capacity to work with small sample size is essential for a credit rating model. For machine learning algorithms, their performance will be impacted by small sample size [143]. An experiment to test the performance for different AI models with small sample size is carried out by this chapter. A FCM based rating methodology is used to compare with BP and SVM. The methodology includes the layered structure of FCM, the longest path inferring, and the priori knowledge combination. The comparison shows that FCM outperforms SVM and SLFN on reliability with a small sample of Nokia.

Correlation coefficient, as part of the priori knowledge, is combined into the FCM as first layer weight in our methodology. Particularly, in order to evaluate the performance improvement of the correlation combination, an experimental study has been carried out to compare the consistencies between FCMs with correlation combination and normal FCMs. The results show that the correlation combination improves the consistency and reliability as well.

Essentially, a credit rating is a prediction of future financial status for a corporation. From financial point of view, credit rating is not just a classification problem, but also a corporate financial analysis problem. Financial factors are influencing each other via different causal relationships. High correlations are found between different financial factors of Nokia. Temporal characters are found in correlations between financial factors. E.g., the correlation coefficient between the gross profit margin and the rating in 15 months later is near 0.8, and the correlations between rating and R&D expense in 6 months later is more than 0.9. According to the causation rule given in Section 2.3.3.3, with the temporal relationship, correlation is a strong hint of a real causal relationship. Investigations on
CHAPTER 4. FCM BASED CREDIT RATING MODEL: A CASE STUDY OF NOKIA

temporal causal relationships between corporate financial factors will further improve the reliability of rating models.
Chapter 5

Temporized fuzzy cognitive maps for corporate financial factor analysis

Three AI technologies, i.e., SLFNs, SVM, and FCMs, have been studied on credit rating in Chapter 3 and 4 to compare their performance of reliability, discrimination, small sample size, and consistencies. FCMs show equivalent performance on reliability and discrimination with other technologies and outperform BP and SVM on small sample learning. Furthermore, by combining correlation coefficients, the consistency and reliability of FCM have been improved. During the investigation, the correlation between financial factors has been found. Substantial temporal correlations are identified between Nokia’s financial factors.

Corporations, as well as their businesses and products, have their life cycles. They start their businesses, develop their products and markets, grow up on size and market, mature at a stable level, revive with new products and markets, and eventually decline [7]. Research works revealed that corporate factors effect each other with time. Temporal causal relations between corporate factors have been studied, such as the ownership structure and the financial performance [10], the corporate social performance and the market risk [8], and the innovation investigation and the competitive advantages [9]. The capability to inferring temporal causal relationships between financial factors will help researchers and analyzers build the quantified cognitive model for corporate analysis. FCMs, as dynamic soft programming tools, are natural to be used for temporal causal relationship analysis.
CHAPTER 5. TEMPORIZED FUZZY COGNITIVE MAPS FOR CORPORATE FINANCIAL FACTOR ANALYSIS

They have been applied on discrete temporal causal problems [34,102-104] and continuous temporal causal problems [144] as well.

By combining methodologies from former chapters, a novel temporalized FCM and its learning algorithm have been proposed to analysis the temporal causal relationships between financial factors in this chapter. The temporalized FCM is studied with financial data from different corporations and compared with SLFNs. The comparison shows that temporalized FCM outperforms SLFN on accuracy. The differences between corporations are also identified by the learning algorithm.

Section 5.1 gives an introduction of other FCM based temporal causality models. The temporalization methodology is introduced in Section 5.2 with the learning algorithm. Section 5.3 carries out the experiment to compare FCMs with BP trained SLFNs and justify the learned relationships. Finally, Section 5.4 gives the summary of the work.

5.1 Temporal extensions of FCMs

Although FCMs are dynamic systems, they are used for one time point problems in the beginning. In one time point problems, for each given input data, a FCM will converge to a steady state or a circulated hidden pattern as the result. The transitions of the states during the inferring process are insignificant to users. By mapping inference iterations to a discrete time dimension, FCMs are extended to capture temporal relations and the inference processes indicate how concepts affect each other with time. Based on the time mapping, FCMs are be able to support time related dynamic problems [34,102-104].

5.1.1 Discrete temporal extensions

The extended FCM (EFCM) [102] and the fuzzy time cognitive map (FTCM) [103] were proposed in 1992 and in 1995 respectively to map FCMs to discrete temporal domains. One of the temporal attributes, the delay of effects, was studied accordingly. Moving forward, another temporal attribute, the decay of concept, was introduced in 1997 [34].

84
The Rule based FCM (RB-FCM) [99–101] is another important extension of FCMs. In order to support temporal causal inferring, temporal domains were mapped by the interval time between adjacent inferring iterations. The interval time is called the based time (B-Time) [104]. A B-Time is a parameter which is explicitly determined by modelers. Within a RB-FCM, relationship weights will be adjusted according to different B-Times. The study also presented some guidelines to help modelers to choose appropriate B-Times for their applications.

The formal definition of discrete temporalized FCMs was given by FTCMs in [103]. Temporal attributes, such as decay and delay, were carried out with the definition. Similar to FCMs, a FTCM is a directed graph represented by a 2-tuples \((X, E)\), where \(X\) is the concept set defining all relevant factors, and \(E\) is the causality set defining the causalities between concepts.

\[
X = \{i\}_{i=1}^{N} \\
E = \{e_{i,j} | i, j \in X\}
\]

The causal relations in FTCMs are specified by two parameters, the strength and time lag, while FCMs only capture the strength of causal relations. The strength and time lag for causality \(e_{ij}\) are represented by \(s_{ij}\) and \(t_{ij}\) respectively. The causal relation strength, \(s_{ij}\), is a real number between \(-1\) and \(1\). A time lag is an integral multiple of the unified time unit, which specifies the delay of a causal effect. Figure 5.1 depicts a hypothetical FTCM with five factors and six causal relations. Each causality is specified by \(s(t) : s = strength, t = timelag\).

A FTCM maps each inferring iteration to a time slot. The distance between two adjacent slots is one time unit. For the causalities whose time lag is \(N\), the cause will impact the effected concepts after \(N\) inferring iterations. Thus, the conventional inferring approach of FCMs is not applicable to FTCMs. A transform algorithm was carried out to solve this issue [103]. The algorithm transform lagged causalities to causality chains by inserting dummy nodes. For a causality \(e_{ij} : s_{ij}(t_{ij})\), if \(t_{ij} = N\) and \(N > 1\), then
Based on the correlation analysis on Nokia’s financial data, it is found that between two concepts \( C_i \) and \( C_j \), they may have correlations cross different time intervals. Because of its complexities, this kind of relationships cannot be represented by FCM as weights and delays. DCN used continuous temporal functions to represent causal relationships. By

### 5.2 Temporalized FCM and learning algorithm

#### 5.2.1 The definition of temporalized FCM

Based on the correlation analysis on Nokia’s financial data, it is found that between two concepts \( C_i \) and \( C_j \), they may have correlations cross different time intervals. Because of its complexities, this kind of relationships cannot be represented by FCM as weights and delays. DCN used continuous temporal functions to represent causal relationships. By

---

**Figure 5.1: A hypothetical FTCM**

\( N - 1 \) dummy nodes will be added between the nodes \( x_i \) and \( x_j \), which are denoted as \( x_{i1}, x_{i2}, \ldots, x_{i(N-1)} \). The dummy nodes break the causality \( e_{ij} \) to be a causality chain, denoted as \( (e_{i1,i1}, e_{i1,i2}, \ldots, e_{i(N-1),j}) \). Each causality among the causality chain is a conventional FCM causality. After the transformation, the FCM inferring approach can be applied to FTCMs.

However, the aforementioned approaches [34, 102-104] do not constitute a complete temporal attributes of causal realtionships. Some attributes such as long term effects, short term effects, and effect lasting, have not been studied.

More recently, another important work on temporalizing FCMs, the Dynamical Cognitive Network (DCN), was proposed to extend FCM theories to continuous time domains [144]. DCNs use the dynamic theory to build a continuous dynamic system, which supports FCMs, CMs, and non-linear systems. Theoretically, DCNs can support a full set of time related features. However, the complexity of DCN prevents it from discrete temporal problems.
CHAPTER 5. TEMPLORIZED FUZZY COGNITIVE MAPS FOR CORPORATE FINANCIAL FACTOR ANALYSIS

porting the temporal function to discrete time domains, a discrete function denoted by $f_{ij}(t)$ is used in this work to specify the temporal relationship between concepts $C_i$ and $C_j$. The value of $f_{ij}(t)$ specifies the effect to the concept $C_j$ caused by the concept $C_i$ after the time interval $t$. By introducing the causal relationship function, the FCM model is changed to be $FCM =< n, F, C, T >$, where $n$ is the number of nodes in the graph. $F = \{f_{ij}(t)\}$ represents the set relationship functions. $C = \{C_1, C_2, ..., C_n\}$ is the set of concepts, and $T$ represents a bounded linear discrete temporal domain, which the FCM will run inside.

For any time $t + 1 \in T$, the value of a concept $C_i \in C$ is decided by equation 5.1

$$C_i(t + 1) = \text{sigmoid}\left(\sum C_j(t) \otimes f_{ji}(t)\right) \quad (5.1)$$

In time $t + 1$, all concept values later than $t$ is not available. According to temporal order of causations, an effect is always later than its cause. Therefore, the equation is revised as

$$C_i(t + 1) = \text{sigmoid}\left(\sum_{k=0}^{t} C_j(k) f_{ji}(t-k)\right) \quad (5.2)$$

When study is only focused on temporal relationships within a given period $s > 0$, i.e., $f_{ji}(t) = 0$ when $t > s$, the equation is revisited again as

$$C_i(t + 1) = \text{sigmoid}\left(\sum_{k=t-s}^{t} C_j(k) f_{ji}(t-k)\right) \quad (5.3)$$

After mapping a FCM inferring process to a discrete time domain, intermediate nodes will introduce delays [102]. To avoid unintended delays, the layered structure is not applicable to temporalyzed FCMs. Concepts are connected with each other directly.

5.2.2 Learning algorithm

A FCM can be built from domain knowledge or trained by samples. After temporализing its relationships from a numeric weight matrix to a temporal function sets, the value space of a FCM has been increased from $R^n$ to $R^{sn}$. To increase the learning efficiency, a hybrid learning algorithm based on correlation coefficient and gradient descent learning is proposed.
5.2.2.1 Weight initialization

The discussion in section 2.3.3.3 shows that the correlation coefficients are always used as hints for possible causations [122]. The last chapter demonstrates notable improvement on the performance when correlation coefficients are combined into the FCM model as priori knowledge. The correlations between concepts narrow down the search space of learning algorithms to a meaningful scope. Considering the layered FCM structure is not applicable for temporalized FCM, the temporal correlation coefficients are used as the initial temporal relationship functions. Equation 5.4 gives the definition of the temporal correlation coefficient $Corr(X_j, X_i, t')$.

$$Corr(X_j, X_i, t') = \frac{\sum_{t=1}^{k} (X_j(t) - \overline{X_j})(X_i(t + t') - \overline{X_i(t + t')})}{\sqrt{\sum_{t=1}^{k} (X_j(t) - \overline{X_j})^2} \sqrt{\sum_{t=1}^{k} (X_i(t + t') - \overline{X_i(t + t')})^2}}$$  \hspace{1cm} (5.4)

where $X_j$ and $X_i$ are the temporal value sets of concept $C_j$ and $C_i$ respectively, $t' \in [1, s]$ is the time interval. Therefore, in the beginning, the initial relationship function links concept $C_j$ to $C_i$, $f_{ji}'$, is defined in equation 5.5

$$f_{ji}'(t') = Corr(X_j, X_i, t')$$  \hspace{1cm} (5.5)

5.2.2.2 Gradient descent learning

The correlation is not causation. It provides a reasonable start point for the learning algorithm to learn possible causalities between concepts. According to the three causation rules defined in section 2.3.3.3, the temporal correlation matches with the temporal rule and the correlation rule. When correlations are used as temporal causal relationships, there will be errors in two circumstances. The first circumstance is for concepts $C_i$ and $C_j$. If both of them have a common cause, $C_z$, effecting them, there will be a correlation between $C_i$ and $C_j$. It is called a propagated correlation. The second circumstance is a coincident correlation between two independent concepts, which is called a illusive correlation. To reduce the error, the learning algorithm must be able to adjust the initial relationships...
to eliminate propagated correlations and illusive correlations. A gradient descent base algorithm is proposed for the objective.

In a gradient descent learning algorithm, a parameter $k$ of model will be adjusted by its partial derivation of the error $E$. For a given training sample $\{< X_i, Y_i > | i = i, 2, ..., k \}$ with the time domain $T = \{1, 2, ..., t \}$. $X_i$ is the value of input concepts in different time, which can be represented as a matrix

$$X_i = \begin{pmatrix} x_i^0(0) & x_i^1(0) & \cdots & x_i^n(0) \\ x_i^0(1) & x_i^1(1) & \cdots & x_i^n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_i^0(t) & x_i^1(t) & \cdots & x_i^n(t) \end{pmatrix}$$

and $Y_i = \{(y_1, y_2, ..., y_m)\}$ denotes the anticipated outputs at time $t + 1$. For each training sample $< X_i, Y_i >$, with the error function defined in equation 3.1, the partial derivative of the $E^t$ to a temporal weight $f_{ji}(t')$ at time $t$ can be obtained by equation 5.6

$$\frac{\partial E^t_i}{\partial f_{ji}(t')} = 2C_i(t + 1)(1 - C_i(t + 1))(C_i(t + 1) - y_i)C_j(t - t') \quad (5.6)$$

For each round of learning, the adjustment of $f_{ji}(t)$ is decided by the sum of partial derivatives and the learning coefficient $\gamma$. The formula is defined in equation 5.7.

$$\Delta f_{ji}(t') = ||f_{ji}(t')||\gamma \sum_{i=1}^{k} \frac{\partial E^t_i}{\partial f_{ji}(t')} \quad (5.7)$$

5.2.2.3 **Learning speed and relationship refinery**

A gradient descent algorithm will converge to a result with an iterative learning process. The speed of convergence is controlled by learning coefficient $\gamma$. To make sure that the learning algorithm will not be blocked by local minimal points and also have the capability to arrive the nearest best solution, a three-stage learning is used. The learning speeds are different in each stage from large to small. In our experiment, the learning speeds for three stages are 1, 0.01, and 0.001 respectively.

In order to concentrate on substantial relationships, the model will be refined after the first stage to remove insubstantial relationships. The relationships $f_{ij}(t)$, which does not
have any \( t' \in [1, s] \) that \( \|f_y(t')\| \) is larger than a threshold value \( p \), will be removed from the model.

The learning result is a set of relationships, which are refined by the learning algorithm from correlations. Based on the causation rules, these relationships are not causations, because a learned relationship may relates to an unobserved characteristics. However, relationships learned from sample data does conform to the other two causation rules, i.e, temporal ordering and correlation. Therefore, we call them “\( h\)-causations”, which means the hypothesis of causations.

5.3 Experiment result

The quarterly financial reports of Nokia, Ericsson, and Google between 2008 to 2013 are used for study.

5.3.1 Data preparation

To study temporal relationships between different financial factors, following factors are selected. They are \( \text{total sales}, \ \text{gross profit}, \ \text{gross profit margin}, \ \text{R\&D expense}, \ \text{marketing expense}, \ \text{assets}, \ \text{liabilities}, \ \text{liabilities/assets}, \ \text{and credit rating}. \)

The original financial data are normalized by the sample approach as what chapter 4 uses.

The temporal correlation test shows that most correlation are insubstantial when the time interval is large than two years. Therefore, the experiment only studies the temporal relationships within two years.

Each training sample has all 9 factors as its outputs, and the historical values of the 9 factors from past two years are inputs. For each corporation, there are 12 training samples, which consists of data from 2008 to 2012 as inputs and data from 2010 to 2012 as outputs, and 4 testing samples, which consists of data from 2011 to 2013 as inputs and data from 2013 as outputs.
5.3.2 Result

5.3.2.1 Learning accuracy

The training errors and testing errors are key measurements to measure the accuracy of a learning algorithm. To verify the reliability of proposed methodology, the result of FCM is compared with BP trained SLFN.

The testing errors and training errors of FCM and SLFN for different corporations are listed in table 5.1 respectively.

The result shows that by combining the temporal correlation and gradient descent learning algorithm, the FCM has a better performance than SLFN on the training accuracy. The accuracy test verifies the reliability. The result will be further analyzed to compare the relationships between $h$-causation, correlation, and causation in the rest of this section.

5.3.2.2 H-causation and correlation

Correlation coefficients are used in the algorithm as initial temporal weight matrices. The relationship between correlations and $h$-causations is carried out in this section.

To compare the difference between the correlation and the $h$-causations, another correlation analysis is conducted between the two values first.

Considering the difference between corporations and different time interval, $h$-causations and correlations is separated into small groups according to their corporations and time intervals, and tested on correlation independently. Figure 5.2 depicts the correlation between correlation coefficients and $h$-causations in terms of different time intervals and corporations.

The result shows that correlations between the temporal correlation and the $h$-causation are all in $[-0.3, 0.3]$, which is not significant related. It implies that relationships are
adjusted largely by the learning algorithm from their initial weights. Figure 5.3 compares the correlation and *h-causation* between four different concept pairs taken from the model of Nokia.

Another comparison between the *h-causation* and the correlation is based on the remained relationships. As the initial relationships, the correlation coefficients are calculated for any two concepts. There are 81 relationships created for each corporations as initial models. After the training, only 35, 37, and 46 relationships remained for Nokia, Ericsson, and Google respectively. More than 50% of correlations have been removed by the learning algorithm.

The distribution of remained relationships can give more information about the reliability of the learning algorithm. In a FCM, a concept influence other concepts by its relationships. Figure 5.4 – 5.6 depict the top 3 most influencing concepts in the trained models for the three corporations respectively. By removing duplicated relationships, the 3 most influencing concepts associate 24, 23, and 23 relationships in their FCMs respectively.
CHAPTER 5. TEMPORIZED FUZZY COGNITIVE MAPS FOR CORPORATE FINANCIAL FACTOR ANALYSIS

Figure 5.3: The comparison between correlation and h-causation.
CHAPTER 5. TEMPORIZED FUZZY COGNITIVE MAPS FOR CORPORATE FINANCIAL FACTOR ANALYSIS

Figure 5.4: Top 3 most influencing factors in Nokia

Figure 5.5: Top 3 most influencing factors in Ericsson

94
The number nears 60% of the total remained relationship number. The 3 most influencing concepts are Rating, Gross profit, and Liabilities/Assets for Nokia, Sales, liabilities/assets, and gross profit for Ericsson, and R&D expense, liabilities/assets, and liabilities for Google. The liabilities/assets is one of most influencing concepts for all three corporations. It conform to the essential role that the liability asset ratio plays in corporate financial management [145]. Besides the liabilities/assets, the financial statements of Nokia and Ericsson are more influenced by sales, rating and gross profit than other factors, while the Google’s financial statements are driven by the R&D expense and the liabilities. That shows some difference between these two kinds of corporations. The next section will discuss these relationships.

The comparison shows that by removing propagated and illusive correlations, the h-causation is very different from the correlation.

The further study indicates that the algorithm eliminates more than 50% of correlation to improve the reliability of the result. Among the remained relationships, 60% of relationships associate with three most influencing concepts. Therefore, the learning algorithm
Figure 5.7: The \( h \)-causations of R&D-sales and R&D-profit for three corporations carries out results with statistical significant from the correlation relationships and sample data.

5.3.2.3 The \( h \)-causation and causation

Last section compares the \( h \)-causation and correlation. The empirical comparison indicates that \( h \)-causation is different from correlation. As another hint of causation, how close is the \( h \)-causation to the causation?

To further investigate the question, some \( h \)-causations are selected to compare with the background of the three corporations. The first is the \( h \)-causation between R&D expense and market factors, i.e., sales and gross profit.

According to Moody’s classification, Nokia and Ericsson belong to telecommunication equipment industry, while Google is an IT service corporation. All of these three corporations belong to technology industry. For technology corporations, the efficiency of R&D expense is one of the essential measurement of their performances [146–148]. The effects on sales and gross profit caused by R&D expense were widely used to assess the R&D efficiency of corporations [149, 150]. Figure 5.7 compares the two \( h \)-causations between different corporations.

The \( h \)-causations indicate that the effects on both the sales and gross profit caused by R&D are insubstantial in Nokia. Ericsson’s \( h \)-causations indicate that the R&D expense
brings positive effect to sales and profit within a year. However, the long term effect is significant negative. The two year effect to gross profit is $-1$ and to sales is $-0.92$. Considering the financial performance of the corporation, especially the Sony-Ericsson, a previous joint venture with Sony, which has been acquired by Sony in February 2012, the \textit{h-causation} indicates that the long term R&D strategy of Ericsson may not match with the market. The \textit{h-causation} of Google indicates that the R&D expense causes immediate effects on sales followed by insubstantial long-term effects. Its effects on gross profit are substantial after a year, which is a reasonable delay for effects of R&D investments.

The \textit{h-causation} between R&D expense and market factors indicates that Google and Ericsson have better R&D efficiency than that Nokia has. Ericsson’s long term strategy on R&D may not match with market.

A further verification is taken on the relationship between gross profit and assets. The two factors are selected because they are causally associated in finance. Gross profit is a factor, which causes the increasing of assets directly. The \textit{h-causations} of gross profit and assets are depicted in Figure 5.8.

The \textit{h-causation} tells that for Google and Nokia, the gross profit bring positive effects on assets within a year, while Ericsson is not. For Google, the effects brought by gross profit will become insubstantial after a year, which is justified by the financial disciplines. For Ericsson and Nokia, their \textit{h-causations} show significant negative effects on the assets in 15 and 18 month later. After comparing the financial histories of these three corporations, it is found that Ericsson and Nokia have paid dividends to their shareholders every year, while Google never do it in its history. Considering the dividend will reduce the assets of the corporation, it explains the \textit{h-causations} of Nokia and Ericsson.

5.3.3 Experiment summary

The empirical study of the novel methodology indicates that it outperforms BP on accuracy. The further analysis indicates that \textit{h-causation} is different from correlation. The comparison between \textit{h-causations} from different corporations with their background shows that learned \textit{h-causations} are justifiable.
5.4 Conclusion

The temporalization of relationship makes the FCM be able to support temporal causal relationships, which cross different time interval. Meanwhile, it also increase the value space of the FCM, which bring challenges for model construction.

The work proposed a novel learning methodology, which combines the correlation calculation and the gradient descent learning to train temporalized FCMs.

The experiment result verifies the accuracy of the learning methodology. The learned relationship functions, called as h-causations, are studied as another result of the learning methodology. The study reveals that h-causations are different from correlations, and close to causations. It makes h-causations be a potential tool for corporation analysis.

FCMs can be built with two different approaches. Machine learning is one of them. Besides that, FCMs are able to be built from human expertise. The temporalization brings challenges to the human expertise based approach by introducing the $R^m$ value space. FCM is a popular soft computing tool for people to build models based on their knowledge. As
one of three causation rules, temporal attributes are inherent parts of causal relationships. Tools to map temporal expert knowledge to temporalized FCM is desirable for temporalized FCMs.
Chapter 6

The tFCM and tFCM studio

As a combination of neural network and expert system, a FCM can be built based on human expertise. FCM’s Fuzzy cause-effect relationships facilitate the knowledge acquisition from human experts [109]. As one of the causation rules, temporal attributes are essential parts of causal relationship. It is common to define a cause-effect relationship with words like “immediate”, “long-term”, “later”, etc. The temporalization of FCM increases the value space of the model from $R^n$ to $R^m$. I.e., more values have to be decided than before. Former chapter shows how to train temporal relationships, called $h$-causations, based on sample data. This chapter is focused on how to map temporal cause-effect relationships with human expertise.

A formal definition of a temporalized FCM, called tFCM, is given in section 6.1 with patterns and methodologies to map the human expertise to the tFCM. The corresponding software, tFCM studio is introduced by section 6.2. Two tFCM models, the dynamic map and the balance map are given in section 6.3 to represent different opinions on what effects are caused by. An experiment is conducted to compare different tFCM models. The result and discussion are represented in section 6.4. The final section summaries the chapter.

6.1 Temporal Fuzzy Cognitive Maps

In a literal description of a cause-effect relationship, the causality is not the only element who has temporal attributes. The temporal attributes can come with the cause, effect, and related objects. Therefore, other elements in FCM need to be temporalized with the
causalities to capture these temporal attributes in a complete form. By analyzing possible elements of FCM and their temporal attributes, the section gives a formal definition of tFCM. In addition, element patterns are proposed to map tFCM with human cause-effect expertise in this section.

### 6.1.1 The formal definition of tFCM

The elements of FCM have to be identified out before the temporalization. FCM is defined as $G = (n, E, C, f)$, where $n$ is the number of concepts, $E$ is the causality set, and $C$ is the set of concepts, $f$ is the threshold function of concepts.

Concepts play a central role in FCM. FCM applications used concepts to represent two kinds of objects. They are events [109,116,151]) and state measurements [116,152–154]. In a cause-effect relationship, events are common sources of effects, while states are common objects effected by causes. E.g., for a causation that “increasing base interest rate will increase mortgage interest payment”, the cause, “increasing base interest rate”, is an event. It causes the “mortage interest payment”, which is a state, to be increased. Therefore, it is a dilemma to decide whether to map the state or the event to a concept when it is a cause of effects and effected by other causes as well.

To out of the dilemma, some practices represent states as concepts and use causal relationships to capture the events, which effect other concepts [152,155]. It separates the cause from the concept, and make concepts represent only states.

Input is another comment element of FCM. Some applications use input as initial values of their concepts [155], while some applications use dedicated input concepts [116,151,152].

For a concept in FCM, if there is no input to it, the state of the concept will be decided by all effects it received from cause-effect relationships and the threshold function. For a concept, the total effect it received is another element of the concept.

In summary, there are four value sets existed in FCM. $V$ denotes the states of concepts, $C$ denotes causes, $E$ denotes effects, and $I$ denotes inputs. A FCM is a DAG, in which concepts are linked by causal relationships. The value of concept at time $t$ is calculated
based on all received effects via a threshold function $f$. By separating the concept set into four different value sets, functions are needed to link these value sets together. Five functions are proposed to connect the four value sets together to be a tFCM. They are

1. $\psi$: $\psi$ stands for the effect response function specifying how a state is changed by the effects from other concepts; $\psi$ links the summary of effects to a state.
2. $\phi$: $\phi$ stands for the self-response function specifying how a concept affects itself;
3. $\varphi$: $\varphi$ stands for the input response function specifying how a concept is changed by the inputs;
4. $\varepsilon$: $\varepsilon$ stands for the cause function specifying how a cause is generated by a concept; and
5. $h$: $h$ stands for the effect function specifying how effects are generated by causes.

The relationships between these elements are illustrated in Figure 6.1. The further study on the temporal attributes of these elements gives a scope that the tFCM definition covers.

6.1.1.1 Temporal attributes

With the separation of FCM elements, all possible temporal attributes are listed for each elements below.

1. States ($V$): When a concept is isolated from others, its state may be maintained or continuously moving toward to a normal point with time. This phenomenon is called decaying.
2. Causes ($C$): A cause can be generated either by a change of the concept state, or the concept state itself. According to the way it is generated, a cause can be categorized into two types: Difference causality or equality causality respectively;
CHAPTER 6. THE TFCM AND TFCM STUDIO

Figure 6.1: The elements of FCM

(3) Effects(E): Delay, lasting are common attributes to describe temporal effects, where the delay is the response time to causes, and the lasting describes how long the effect lasts.

(4) Inputs(I): In a dynamic environment, an input can change with the time. That is called a dynamic input.

6.1.1.2 The temporalization

FCM is a dynamic system. The iterative inferring steps of FCM can be mapped to a time axis. A temporal domain is a time axis to map the value sets. In tFCM, a bounded linear ordered set \( T = \{t, t+1, ..., t+n\} \) is used as the time domain. Based on the time domain and the defined elements, the calculation process in the tFCM is given in below:

\[
\begin{align*}
V_i(t) &= \varphi_i(I_i, t) \\
C_i(t') &= \varepsilon_i(V_i(t')) \\
E_i(t') &= \sum h_{ji}(C_j, t') \\
V_i(t' + 1) &= f_i(\phi(V_i(t')), \psi(E_i(t')), \varphi_i(I_i(t' + 1)))
\end{align*}
\]  

(6.1)

where \( f_i \) is the threshold function of concept \( i \).

According to the work in Chapter 5, the effect function can be rewritten as 6.2
With the effect $h$, time related effects are supported by the tFCM. The temporal attributes are able to be represented by the effect function. Figure 6.2 shows how to encode those attributes into an effect function. A fuzzy logic based encoding method to design effect functions is carried out by Section 6.2. The rest temporal attributes can also be represented by different functions. They are listed in below.

- **Decaying**

  $$\phi_i(V_i(t)) = \lambda V_i(t) + (1 - \lambda)p$$

  where $\lambda$ is the decaying constance and $p$ is the normal point that the concept decays to.

- **Difference causality**

  $$\varepsilon_i(V_i, t) = V_i(t) - V_i(t - 1)$$

- **Equal causality**

  $$\varepsilon_i(V_i, t) = V_i(t)$$

- **Dynamic input**

  $$\varphi_i(I_i(t)) = \mu I_i(t) + \alpha$$

  where $\mu$ is the input coefficient and $\alpha$ is the input constance.

Based on the element separation and temporalized calculation, the tFCM can be defined by equation 6.7

$$tFCM =< T, N, E, I >$$

where $T$ is the time domain, $E = \{h_{ji}\}$ is the set of effect functions, $I$ is the input, and $N$ is the concept set, $N = \{< f_i, \psi_i, \phi_i, \varphi_i, \varepsilon_i >\}$.

Equation 6.7 gives a formal definition of the tFCM. Figure 6.3 shows the semantical structure of the tFCM.
CHAPTER 6. THE tFCM AND tFCM Studio

Figure 6.2: A sample effect function

Figure 6.3: The semantic diagram of the tFCM
6.1.2 tFCM patterns

Section 6.1.1 defines a full temporalized extension of FCMs based on a linear discrete time domain. This section gives some patterns to capture the temporal causality expertise to the tFCM. These patterns include two concept patterns and a fuzzy temporal effect pattern.

6.1.2.1 Concept patterns

In the formal definition of tFCM, a concept consists of five temporal functions. They are three response functions \( (\psi_i, \phi_i, \varphi_i) \), a threshold function \( (f_i) \), and a cause function \( (c_i) \). Since most concepts have common attributes, concept patterns are useful to facilitate the building process by predefined functions templates, and linking meaningful attributes to function parameters. Linear decaying concept is the basic concept pattern proposed by us.

**Definition 1. Linear decaying concept**

For a concept \( i \), if the three response functions are defined as

\[
\psi_i(E, t) = \sum_{j \neq i} E_{ij}(t),
\]

\[
\phi(V_i, t) = \lambda V_i(t) + (1 - \lambda)p,
\]

\[
\varphi(I_i(t)) = I_i(t)
\]

and the threshold function is defined as

\[
f(\phi_i(V_i(t), \varphi_i(I_i(t+1)), \psi_i(E_i, t)) = \text{sigmoid}(\phi_i(V_i, t) + \varphi_i(I_i(t+1)) + \psi_i(E, t))
\]

the concept \( i \) is called a linear decaying concept.

Decaying is a common phenomenon, which represents whether a concept tends to move toward a normal point, and the degree of the tendency. The linear decaying concept is a simple concept pattern in the tFCM and is common in applications. For each linear decaying concept, the decaying constance \( \lambda \) and the normal point \( p \) are the two attributes that need to be specified.
CHAPTER 6. THE TFCM AND TFCM STUDIO

In addition, another concept pattern is proposed to support the association between causes, e.g. and and or relationships. The concepts in this pattern are called association concepts. The formal definition of association concepts is given by definition 2.

**Definition 2.** Association concept

For a concept, if its state is only decided by causes from associated concepts as what equation 6.12 defines.

\[ V_i(t + 1) = \psi(C(t)), \]  

the concept is called the association concept. It is OR node if \( \psi(C(t)) = \max(C(t)) \), whereas it is AND node if \( \psi(C(t)) = \min(C(t)) \).

6.1.2.2 Effect patterns

Effects in conventional FCMs are numeric values between \(-1\) to \(1\). The value of an effect indicates how strong the concept is associated and effected by other concepts. In tFCMs, temporal effect functions defined by equation 6.2 specifies cause-effect relationships between concepts. For a well investigated effect relationship, the corresponding effect function can be defined precisely based on domain knowledge. However, FCM is widely used because they are able to capture “fuzzy” human knowledge, which does not have precisely defined functions. A fuzzy rule set based effect pattern is proposed in this section to map fuzzy human knowledge to temporal effect functions.

The effect pattern is called *Fuzzy Temporal Effects* (FTEs). A FTE is a fuzzy rule set defined on a serial of fuzzy time sets as conditions. These fuzzy time sets are called *Base Time Sets*.

**Definition 3.** Base Time Set

A base time set \( s \) is a fuzzy set on the time domain \( T \). The set and the membership function \( \mu \) are defined by Equation 6.13 and 6.14 respectively.

\[ s = (T, \mu) \]  

\[ \mu : T \to [0, 1] \]
CHAPTER 6. THE tFCM AND tFCM STUDIO

A set of base time sets consists the base temporal conditions of FTEs and is denoted by $S$. Figure 6.4 shows membership functions of three fuzzy time sets, Short Term ($\mu^{ST}$), Middle Term ($\mu^{MT}$), and Long Term ($\mu^{LT}$).

With fuzzy time sets as conditions, the following step is to specify effects in every conditions. Using the fuzzy time sets given by Figure 6.4 as an example, three numbers are assigned to indicate the effect in short term, middle term, and long term. They are denoted by $\gamma^{ST}$, $\gamma^{MT}$, and $\gamma^{LT}$ respectively. These numbers can be interpreted as:

When there is an event happened as a cause in any time point, as a result,

- the short term effect to associated concept is $\gamma^{ST}$
- the middle term effect to associated concept is $\gamma^{MT}$
- the long term effect to associated concept is $\gamma^{LT}$

All the three fuzzy rules compose a FTE. Definition 4 gives the formal definition of the FTE.

Definition 4. Fuzzy Temporal Effects

A FTE is a set of fuzzy rules on base time sets.

\[ FTE = \{(s, \gamma) | s \in S, \gamma \in [-1, 1]\} \]  \hspace{1cm} (6.15)

Using Center of Gravity (COG) based defuzzification, an effect function $h_{ij}(t)$ can be achieved from a FTE as:

\[ h_{ij}(t) = \frac{\sum_{s \in S} \gamma_s \mu_s(t)}{\sum_{s \in S} \mu_s} \]  \hspace{1cm} (6.16)

Based on the tFCM, Definitions 1–4 define patterns to represent human knowledge by tFCM. An inferring tools, tFCM studio, are built based on these patterns. The following section will introduce how to use the tFCM studio to design the tFCM for real world problems.
CHAPTER 6. THE tFCM AND tFCM STUDIO

Figure 6.4: The definition of a base temporal condition

Figure 6.5: The tFCM studio
CHAPTER 6. THE TFCM AND TFCM STUDIO

6.2 Design tFCMs with tFCM studio

tFCM studio is a java based software to help modelers design, debug, and infer tFCMs. Figure 6.5 depicts the user interface of the tFCM studio.

In the tFCM studio, a tFCM is built by three parts. They are global attributions, concepts, cause associations, and effects. They will be introduced one by one in this section.

6.2.1 Specify global attributes

For a tFCM, global attributes are the time domain and the base temporal conditions. A time domain is specified by the time slot size and the inferring length. The time slot size is the base time unit for all other map settings, especially for the effect functions. The inferring length specifies how many iterations an inferring will be executed. For example, the time slot length in Figure 6.6 is one month and the inferring length is “24”, which means that the tFCM will infer the interaction between concepts within two years based on inputs and cause-effect relationships.

The other kind of map attributes is a set of base time sets, which are named as base temporal conditions. The base temporal conditions are base time sets used as conditions to specify FTEs between concepts. According to definition 3 and 4, in a temporal fuzzy effect function, each time set is used as a rule condition for one of the effect rules. A unique name will be given for each time set to help modelers identify and understand these conditions. For example, ST, MT, and LT may be used to name three time sets, which represent short term, middle term, and long term respectively.

When the global setting is completed, the tFCM is ready to add concepts and causal effects between concepts.
6.2.2 Design concepts

Concept design phase consists of two steps: i) Identify concepts, and ii) Specify concepts. According to the tFCM definition, tFCM concepts need to cover two major concerns:

1. The inputs from external environment to the tFCM; and
2. The interactive concepts modeled by the tFCM.

The concepts are separated into two categories to represent these two concerns respectively. For each category, only the corresponding concerns will be catered. The two categories are environment concepts and inferring concepts. Environment concepts represent inputs from environment, and inferring concepts represent the interactive factors.

Environment concepts and inference concepts can be identified from the environment and domain knowledge.

Environment concepts have simple structures. Their states are only decided by the inputs from the environment, i.e. for a concept $i$, $v_i(t) = I_i(t)$. For other concepts, they do not have any input besides the initial states. The environment concepts are specified by the input functions.

To simplify the design, the linear decaying concept pattern is used to describe Inferring concepts. With Definition 1, each concept is specified by two required parameters, i.e., the decaying constant and the cause function.

Decaying constants are natural attributes for concepts. It can be found in related domain knowledge, e.g. the half-life of the concept quantity and the degree of the concept’s polarization.

Cause functions specify the way how concepts effect others. Equality causality and difference causality are most common cause function patterns used in tFCMs. The difference of inference results for the two cause functions will be discussed in the following section.

In tFCM studio, inference concepts and environment concepts are merged together. Each concept can be specified by corresponding decaying constants, cause functions, and input functions, as illustrated in Figure 6.7. In practices, they are defined separately.
CHAPTER 6. THE tFCM AND tFCM STUDIO

Figure 6.7: The UI form of concepts in tFCM studio
6.2.3 Design causal relationships

After all concepts are identified and specified, causal effects between inference concepts and environment concepts can be specified. In tFCM studio causal relationships are separated into two parts, i.e., the co-occurrent cause conditions, named as “cause associations”, and the effect functions to specify how causes effect.

In a cause-effect model, it is common to have rules that composite co-occurrence of multiple causes to generate new causes, such as: “C₁ and C₂ and C₃”. In tFCM studio, it is called cause association. Two association types are supported by the software, the AND and OR associations, which are defined in Definition 2. A cause association can associate concepts and other cause associations as well. The Figure 6.8 shows the cause association form in tFCM studio.

After a cause was generated from a concept or a cause association, it will impact related concepts. Causal effects are impacts that concepts received from causes. In tFCM theory, causal effects are specified by the effect functions. Designing effect functions is the key task for tFCM designers. The tFCM studio uses FTE model to specify effect functions. For a cause function, a set of fuzzy numbers is given to specify the effect under each base temporal conditions respectively. The cause function will be generated correspondingly;

6.3 Dynamic maps and balance maps

The former section defined two types of causality: the equal causality and the difference causality (equation 6.5 and 6.4). The equal causality and difference causality represent
different opinions regarding what a causal relationship between two concepts is. The equal causality represents an “if-then” relationship between concepts, while in a difference causality, a causal relationship always involve changes [156].

For a tFCM, if all cause functions are equal causalities, the tFCM is called a *dynamic map*. If all cause functions of concepts are difference causalities, the map is called a *balance map*. They are named after their different behaviors during inferring processes. For a dynamic map, as other dynamic systems, it can evolve to a steady state or a hidden pattern from any initial states. The behavior of balance map is different from dynamic map. Since concepts are effected by only changes from other concepts in a balanced map. Without input from the environment, the map will stay at its initial states as a equilibrium state, in which all competing influences are balanced. A balance map will be a dynamic system when there are some inputs from outside to the system, i.e., for any input from outside, it can evolve to a steady state or a hidden pattern.

A survey indicates that some FCMs used changes as causes [152, 154, 157–159], and some other FCMs used absolute value as causes [116, 153, 160, 161]. Further comparison shows that all the FCMs, which used the absolution values as causes, are one time point problems, i.e. they do not map the dynamic inferring process to any time domain.

[27, 156, 162] gave subjective explanations why causal relationship should associate relative variations rather than absolute value. This section gives an objective comparison between the dynamic map and balance map on error accumulation.

Errors are unavoidable in a tFCM/FCM because

- The model cannot cover all related concepts and relationships completely;
- The human expertise is fuzzy.

For a tFCM/FCM model, let $V$ denote the initial states of the concepts, $R$ denote the accurate relationship matrix, and $\Delta R$ denote the error of the relationship. Therefore, $(R + \Delta R)$ is the relationship of the model. The error of a dynamic map at the $n$th iteration can be estimated by equation 6.17.
\[ E_n = V'_n - V_n = (R + \Delta R + 1)^n V_0 - R^n V_0 \]
\[ = V_0 \sum_{i=0}^{n} C_i^n \sum_{j=0}^{i-1} C_j^i R^j \Delta R^{i-j} \]  

where \( C_i^n \) is the binomial coefficient, \( V_0 \) is the initial states, \( V'_n \) is the model result at the step \( n \), \( V_n \) is the real value at the iteration \( n \).

With the same approach, the error of balance model can be estimated by Equation 6.18
\[ E_n = V'_n - V_n = (R + \Delta R)^n \Delta V_0 - R^n \Delta V_0 \]
\[ = \Delta V_0 \sum_{i=0}^{n} C_i^n R^i \Delta R^{n-i} \]  

Diagram 6.9 illustrates the error rate of the two different models with the inferring iterations. Error rate is defined by Equation 6.19
\[ \text{Error Rate} = \frac{E_n}{V_n} \]

### 6.4 Experiments

To test and verify the model and theories of the tFCM, a dummy map of virtual social concepts is created, as shown in Figure 6.10. It consists of six virtual concepts: education budget, education, creativity, industry, jobless rate, and population. Concept states are numbers between -1 to 1 to indicate the prosperities of concepts.

**Education Budget** is an environment concept. In the experiment, it is grown 0.05 per year from its initial states. How other concepts will be affected by the stable increasing of education budget will be inferred from the map. All other concepts are decaying concepts.

The threshold function is defined as Equation 6.20, where decaying rate \( \lambda_i \) and initial states of concepts \( v_i(0) \) are specified in Table 6.1.

\[ v_i(t + 1) = \min(1, \max(-1, \lambda_i v_i(t) + \sum_{j \neq i} e_{ij}(t))) \]  

Three fuzzy time sets, short term (ST), middle term (MT), and long term (LT) are defined to specify the temporalized causal effect relationships, as depicted in Figure 6.11. Based on the fuzzy time sets, the effect degrees between concepts are given in Table 6.2.
Figure 6.9: The theoretical error rate against the number of iterations
CHAPTER 6. THE TFCM AND TFCM STUDIO

Figure 6.10: The model of experiments

Figure 6.11: The temporal domain and fuzzy time sets
### Table 6.1: The decaying rate for concepts

<table>
<thead>
<tr>
<th>Concept</th>
<th>Decaying rate $\lambda$</th>
<th>Initial state $v(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education budget</td>
<td>n/a</td>
<td>0.3</td>
</tr>
<tr>
<td>Education</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>Creativity</td>
<td>0.98</td>
<td>0.3</td>
</tr>
<tr>
<td>Industry</td>
<td>0.98</td>
<td>0.05</td>
</tr>
<tr>
<td>Jobless Rate</td>
<td>0.998</td>
<td>0.03</td>
</tr>
<tr>
<td>Population</td>
<td>0.92</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 6.2: The temporal relationships between concepts

<table>
<thead>
<tr>
<th>Concept</th>
<th>Education budget</th>
<th>Creativity</th>
<th>Industry</th>
<th>Jobless rate</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Creativity</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Industry</td>
<td>0</td>
<td>-0.2</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jobless Rate</td>
<td>0</td>
<td>0.1</td>
<td>-0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Population</td>
<td>0</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Chapter 6. The tFCM and tFCM Studio

The base time unit is one year for all effect degrees and decaying rates listed in Table 6.2 and Table 6.1. For example, the decaying rate of concept creativity is 0.98, which means that the creativity will decrease 2% per year without consideration of effects from other concepts. However, the base time unit for the inference is \( \frac{1}{20} \) year. The decaying rates and the effect degrees will be adjusted to be \( \sqrt[20]{\lambda} \) and \( E/20 \) respectively during the inferring.

Based on the model, four experiments have been conducted to compare differences between results of the balance model and the dynamic model. The first two experiments study differences between the results of two models. The evolving results for the balance model and the dynamic model are shown in Figure 6.12

6.4.1 Evolution results

Figure 6.12 shows the inference of the dynamic model falls into turbulence rapidly, which is common in dynamic systems. Dramatic increasing and decreasing are not realistic. Therefore, the result of dynamic model is difficult to be interpreted to users. Designers have to carefully adjust the effect degrees and decaying rates to control the inference to present a meaningful result. However the technical adjustment will also reduce the reliability of the result.

Compared to the dynamic model, the balance model gives an interpretable result, which shows smooth changes on concept states. The results show that “education”, “creativity”, and “jobless rate” increase, while “industry” and “population” decrease.

Two models infer the results based on the given concepts and relationships. If the knowledge contains errors, the result will also be wrong. In Section 6.4.2, another two experiments are described to show how the knowledge errors affect the results.

6.4.2 Error analysis

The error accumulation is another important concern for the tFCM. According to the theory in Section 6.3, the balance model reduces the error accumulation dramatically. A distance function is defined to verify the theory and compare the error accumulation between two
different models. For any given two state functions, \( v(t) \) and \( v'(t) \), the distance between them is denoted by \( D(v, v') \). \( D(v, v') \) is defined by equation 6.21

\[
D(v, v') = \sum_{t \in T} (v(t) - v'(t))^2
\]  

Figure 6.12: The evolving results of two different models

Two experiments are designed to find out the distance between a given concept in evolution results for different causal relationships based on different models (the balance
model and the dynamic model). The Industry concept is chosen as the testing concept. \( v_0(t) \) denotes the state of Industry in original model. The FCM is run several times. For each running, the causal relationships will be increased by 1%. \( v_n(t) \) denotes the state of Industry in the \( n \)th running, in which the relationship deviation from the original map is \( n\% \). Figure 6.13 depicts the \( D(v_n, v_0) \) against the relationship deviation from 0% to 50% for the dynamic model and the balance model respectively.

The result shows that the distance in the dynamic model is thousand times larger than that in the balance model when the relationship is 50% higher than the original one. However, it is still far lower than the distance it should be according to Equation 6.17. It is caused by the scope of concept states, which is not considered in Equation 6.17. For a tFCM, all states are in \([-1, 1]\). After the distance increases to a threshold value (here is around 6), the distance will not increase according to the equation 6.17, but a linear relationship with the relationship deviation. For the balance model, the experiment result matched with the theory well, since the distance in the model is small enough and no value hit the boundary of the value.

Based on the theories and experiment results, the dynamic model is sensitive to the relationships. Minor differences in maps will lead major differences in results, which makes the dynamic model fall into the circulations (hidden patterns) or stable states rapidly.

For the balance model, then the deviation of relationship is 50%, the distance of the concept state is around 0.1. That means the model is stable and error tolerant, i.e. it is robust. However the changing of the concept states is slow in the balance model, which reduces the dynamic capacity of the model.

### 6.5 Summary

FCMs are dynamic mathematical models in natural. Previous works mapped FCM inference processes to time domains with temporal attributes. Based on their work, tFCM is proposed to systematically temporalize FCM. With the model, some design patterns and
a software tool called tFCM studio are introduced to reduce the complexities brought by the temporalization.

Meanwhile, to further discussion on dynamics of FCMs, two causality models are carried out, the dynamic map and the balance map. The theoretical analysis and experimental studies indicate that the dynamic map is sensitive to the changes of relationships, while the balance map is more stable and error-tolerate than the dynamic map. Therefore, they
can be applied on different problems.
Chapter 7

Conclusions

7.1 Summary of contributions

Credit ratings are time-consuming data-intensive analysis tasks involving domain expert knowledge. Due to the raising requirements on regulatory capital calculation, stress tests and scenario analysis, the AI based corporate credit rating approaches attracted much research attention from industries and academics. A fast, accurate, and consistent credit rating technique will largely reduce the cost and improve the profit margin for financial firms. However, the training results of AI based approaches are not able to be justified, which reduces the reliability of AI based approaches. As a combination of neural network and expert system, FCM has potential to be an acceptable tool for credit rating.

The potential of FCM in credit rating problem is studied by the work. The contributions include:

- **The comparison between AI credit rating models**

  As first part of our work, an experimental study is done to compare the performances of five learning algorithms, i.e., BP, ELM, I-ELM, SVM, and FCM with a data extracted from financial reports and credit rating reports from two rating agencies, *Moody’s Investor Services* (MIS) and *Standard and Poor’s rating service* (SnP). The performances of the five algorithms are studied in terms of reliability and discrimination capacity. The experimental results show that neural network based approaches,
CHAPTER 7. CONCLUSIONS

including FCM, outperforms SVM on reliability, and SVM achieves better performance on discrimination. FCM achieves equivalent performances in the experiment with a smaller structure than other SLFN models.

- FCMs as soft computing tools for individual corporation credit analysis

Although AI approaches for credit ratings attracted much attention from both financial industry and academics, it still cannot replace the conventional credit rating approaches, because of two reasons, the understandability and the ability to work with small sample set. As soft computing tools, FCM can combine the advantages of expert system and AI technologies. The performance of FCM in corporate credit ratings has been demonstrated in comparison with other AI approaches. Based on that, a novel FCM methodology to assess credit risk with small sample size is proposed. The methodology includes

  - The 3-layers FCM architecture

    For the purpose of further improving the capability of FCM on modelling sophisticated causal relations while retaining its understandability, a 3-layers FCM architecture is used for credit rating. The architecture separate concepts into three layers, the input layer, the intermediate layer, and the output layer. The 3-layers architecture increases the understandability of FCMs. In addition, the architecture gives guidance for financial analysts to design a FCM for their problems.

  - The longest path inferring approach

    In order to avoid unnecessary interferences from the dynamics of FCM, an original inferring approach, the longest path inferring is introduced. The inferring process will stop FCMs when all necessary causal relations are calculated. An algorithm is proposed to calculate the number of required inferring steps.

  - The combination of correlation as priori knowledge
CHAPTER 7. CONCLUSIONS

By applying the correlation between input and the output concepts on the first layer relationships as the cause effect weight, the methodology improves the reliability of the FCM, and narrow down the search space into a meaningful scope, which increased the efficiency of GA algorithm. The combination also improves the interpretability of the trained FCM weight matrices.

- **An experimental comparison study**
  A case study with financial data from Nokia is carried out to compare proposed methodology with conventional FCM, SVM and BP trained SLFN. The result shows that the proposed methodology outperforms other approaches on testing errors. It also shows that SVM cannot predict credit ratings, which are not covered sample data, but implied by it. The consistency of trained FCM weight matrices has been largely improved after the correlation coefficient has been combined with the FCM. It makes the trained FCMs be more understandable than before.

- **Temporalized FCM**
  To further improved the capacity of credit assessment, a tool, which can forecast the financial factors of a corporation, is valuable. A FCM based approach is proposed for corporate financial factor forecasting. The work consists of following parts.

  - **Temporalized relationships**
    A temporalized relationship associates the future with past, and predict the future changes based on historical data.

  - **Gradient descent based training algorithm**
    An novel algorithm is proposed to refine the correlations to *h-causations* by eliminating propagated and illusive correlations.

An experiment is carried out to compare the approach with BP trained SLFNs. Financial data of Nokia, Ericsson, and Google from 2008 to 2013 is used as sample
data. The results show that the temporalized FCM outperforms BP on accuracy. In addition, the learning algorithm eliminates more than 50% correlations, which decreases the error of \textit{h-causations}. Some trained relationships are justifiable based on the backgrounds of different corporations.

- tFCM

The temporalization of FCM increases the value space of FCM, which brings difficulties to build FCMs from human expertise. A systematical model, called tFCM, is proposed with patterns and tools to build temporalized FCM from human expertise. The contributions of the work are listed below.

- A complete temporalized FCM model
  
  By systematical approach, the FCM model has been fully temporalized on four value sets, i.e., \textit{effects}, \textit{causes}, \textit{states}, and \textit{inputs}, and five functions, i.e., \textit{the effect response function}, \textit{the self-response function}, \textit{the input response function}, \textit{cause function}, and \textit{the threshold function}. The formal definition of tFCM is obtained by introducing the time domain, \( T \), to the value sets and functions.

- tFCM patterns
  
  In order to map fuzzy human expertise to tFCMs, patterns for concepts, and effects are defined. They are the linear decaying concept, association concept, and fuzzy rule set based FTE.

- Dynamics of tFCM
  
  Two causality models, \textit{dynamic map} and \textit{balance map}, are defined in the work. The two models are used for different scenarios. The \textit{dynamic map} makes tFCMs be sensitive to the relationships. The \textit{balance map} is concentrated on effects of changes. Theoretical analysis shows the balance map is more stable and error tolerable than the dynamic map.
CHAPTER 7. CONCLUSIONS

- **tFCM studio**
  
  In addition, a design and inferring tool, the tFCM studio, is designed to help users create, maintain, and execute tFCMs.

- **Experimental study for tFCMs**
  
  An experimental study has been carried out to verify the model and theories of the tFCM. The experiment results show the differences between the dynamic map and the balance map. After a 200 iteration inference, the error of balance map is lower than 0.1% of the error of dynamic map. The result of the balance model is also more realistic than the dynamic model’s result. However, the balance model is not suitable to simulate dynamic scenarios, such as stock markets, where dramatic changes in short period are often. In view of their complementary strengths, a hybrid model integrating the dynamic model and the balance model can be desirable for users according to their applications.

A comparison study on five temporal extensions on FCMs and the native FCMs in features shown in table 7.1 depicts that tFCMs cover more temporal features than others.
### Table 6.2: The temporal relationships between concepts

<table>
<thead>
<tr>
<th></th>
<th>Education</th>
<th>Creativity</th>
<th>Industry</th>
<th>Jobless rate</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>0.7</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Creativity</td>
<td>0.3</td>
<td>0.3</td>
<td>0</td>
<td>-0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Industry</td>
<td>0.0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>Jobless rate</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>Population</td>
<td>0.0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Chapter 6. The tFCM and tFCM Studio
7.2 Recommendations for future work

In the work described by the thesis, FCMs have been studied on the credit rating problem. The capacities of FCM on credit rating and temporal causal inferring are demonstrated by the work. The FCM could be further developed in a number of ways:

1. The scenario analysis problem

The scenario analysis is a financial analysis to assess the impact to an economic entity of various events based on historical or hypothesis scenarios. It attracts much attention since it was integrated into the Comprehensive Capital Analysis and Review (CCAR), a supervisory assessment conducted by United States Federal Reserve in 2011. There are two major problems in scenario analysis, how to define a hypothesis scenario, and how to analysis the impact of a scenario. Structural and non-structural models are proposed for the two problems recently. However, AI based approaches are not mentioned yet. With the capability of causal inferring demonstrated in credit rating assessment, FCM has potential to be a proper tool for the problems.

2. Macroeconomic problems

Comparing to corporation analysis, the macroeconomic problems, such as industry analysis, growth model, financial policy analysis, have more stable causal relationships. It makes the macroeconomic problems a good platform to test and improve the FCM theory.

3. Learning algorithms

   - Dynamic learning

   Causal relationships between financial factors are effected by corporate operations, such as acquisition, recapitalization, strategic reorganization, etc. Therefore, the time window of steady causal relationships is very limited. The causal relationships are part of the dynamic system. A learning algorithm which can
continuously update/adjust the relationship to follow the relationship changes is desirable

• Theoretical investigation on the illusive/propagated correlations

Based on an hypothesis that errors are brought by illusive/propagated correlations when they are used as causal relationships in FCM, the gradient descent algorithm to be used by the work to learn \textit{h-causations} from sample data. More theoretical analysis is desirable to improve the efficiency of the learning algorithm.

7.3 Summary

This thesis presents our investigation work on FCM as an AI based tool for corporate analysis. It compares different learning algorithms, proposes different FCM based approaches to analyze corporations. Particularly, two related topics are studied, combining FCM training with correlation as priori knowledge, and temporalizing FCM to support temporal causalities. For each work, experimental studies are carried out to verify theories and methodologies in the thesis.
Publications

Conference


Journal


Appendix A

Corporate financial statement

A corporate financial statement is a formal financial report released by the corporate to reveal its business activities from financial viewpoints. The format of corporate financial statements are different in different regions according to different regional regulatory. In United States, all corporates listed in public exchange markets must issue their financial statement reports as 10-K form and 10-Q form. 10-K forms are filled for yearly financial statement reports, while 10-Q forms are filled for quarterly financial statement reports. No matter how different these forms in format, they all include some important sections, such as income statements, balance sheets, and cash flow statements. The details are list in following sections.

A.1 Income statements

Income statements, or called revenue statements or earning statements, indicate how much the company earned from its business operation, and how much the company expend on sales, operation, interests, and tax etc..

A.2 Balance sheets

Balance sheets were named by the basic accounting equation.

\[ \text{Assets} = \text{Liabilities} + \text{Equity} \]  \hspace{1cm} (A.1)

Balance sheets represent the details of these three balanced elements. They are
## CHAPTER A. CORPORATE FINANCIAL STATEMENT

<table>
<thead>
<tr>
<th>Accounts</th>
<th>Numbers (example)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>100,000,000</td>
<td>The total cash earned by the company in the report period.</td>
</tr>
<tr>
<td>Cost of sales</td>
<td>(200,000)</td>
<td>The direct expense on producing and delivering the goods. The Cost of sales are calculated based on different formulas in different industries.</td>
</tr>
<tr>
<td>Gross profit</td>
<td>99,800,000</td>
<td>Gross profit = revenue - cost of sales</td>
</tr>
<tr>
<td>Operating expenses</td>
<td>(1,000,000)</td>
<td>The operating expenses cover all costs, which are not relevant with producing. Common operating expenses include marketing expenses, R&amp;D expenses, and administrative expenses.</td>
</tr>
<tr>
<td>Earnings before interest and taxes (EBIT)</td>
<td>98,800,000</td>
<td>Operating profit = gross profit - operating expenses</td>
</tr>
<tr>
<td>Net interest expenses/incomes</td>
<td>(200,000)</td>
<td>The interests the company paid or earned from its cash and loans</td>
</tr>
<tr>
<td>Earnings before incoming taxes</td>
<td>98,600,000</td>
<td>Earnings before taxes = EBIT - net interest expenses</td>
</tr>
<tr>
<td>Incoming taxes</td>
<td>30,000,000</td>
<td>The incoming taxes paid by the company</td>
</tr>
<tr>
<td>Net profit</td>
<td>68,600,000</td>
<td>Net profit = Earning before incoming taxes - Incoming taxes</td>
</tr>
</tbody>
</table>

Table A.1: Sample income statements
CHAPTER A. CORPORATE FINANCIAL STATEMENT

(1) **Assets**

   (I) **Current assets** refer to total amount of all cash and assets which can be converted into cash within 12 months.

   (II) **Non-current assets** represent total amount of all assets that cannot or will not be converted into cash within next 12 months.

(2) **Liabilities**

   (I) **Current liabilities** are obligations and debts will be paid within next 12 months.

   (II) **Non-current liabilities** are long-term obligations and debts, whose maturity date are 12 months later.

(3) **Equity** is net assets that owned by shareholders. Equity is calculated by following equation

\[
Equity = Assets - Liabilities
\]

A.3  **Cash flow statements**

The cash flow statements represent the changes of cash and cash equivalents during the report period. The cash flow statements are used to judge whether a corporate has enough financial flexibility to support its operation, investment, and financial activities. Therefore, a cash flow statement will be separated into three parts, e.g., the cash flow from operating activities, the cash flow from investing activities, and the cash flow from financing activities. Beside of these three parts, the cash flow statement will present two summary number: the cash and cash equivalent at beginning of the reporting period, and the cash and cash equivalent at end of the reporting period.

Table A.2 shows a sample cash flow statement taken from Wikipedia.
## Chapter A. Corporate Financial Statement

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receipts from customers</td>
<td>9,500</td>
</tr>
<tr>
<td>Cash paid to suppliers and employees</td>
<td>(2,000)</td>
</tr>
<tr>
<td>Cash generated from operations (sum)</td>
<td>7,500</td>
</tr>
<tr>
<td>Interest paid</td>
<td>(2,000)</td>
</tr>
<tr>
<td>Income taxes paid</td>
<td>(3,000)</td>
</tr>
<tr>
<td><strong>Net cash flows from operating activities</strong></td>
<td>2,500</td>
</tr>
<tr>
<td>Proceeds from the sale of equipment</td>
<td>7,500</td>
</tr>
<tr>
<td>Dividends received</td>
<td>3,000</td>
</tr>
<tr>
<td><strong>Net cash flows from investing activities</strong></td>
<td>10,500</td>
</tr>
<tr>
<td>Dividends paid</td>
<td>(2,500)</td>
</tr>
<tr>
<td><strong>Net cash flows used in financing activities</strong></td>
<td>(2,500)</td>
</tr>
<tr>
<td><strong>Net increase in cash and cash equivalents</strong></td>
<td>10,500</td>
</tr>
<tr>
<td>Cash and cash equivalents, beginning of year</td>
<td>1,000</td>
</tr>
<tr>
<td>Cash and cash equivalents, end of year</td>
<td>11,500</td>
</tr>
</tbody>
</table>

Table A.2: A sample cash flow statement
Appendix B

Financial ratios

Usually, financial ratios are used as indicators for judge a firm’s performance and financial situation. Normally, financial ratios are compared with benchmarks. However, since ratios change with industries and regions, the benchmarks are important to financial ratios. Different regions and industries have different financial ratio benchmarks. This appendix collected most frequently used financial ratio types and ratios.

B.1 Profitability ratios

Profitability ratios are used to measure the overall efficiency and performance for a corporate.

Gross profit margin:

\[ Gross \ profit \ margin = \frac{Gross \ profit}{Revenue} \]  
(B.1)

Operating margin

\[ Operating \ margin = \frac{EBIT}{Revenue} \]  
(B.2)

Profit margin

\[ Profit \ margin = \frac{Net \ profit}{Revenue} \]  
(B.3)

Return on assets (ROA)

\[ Return \ on \ assets = \frac{EBIT}{Assets} \]  
(B.4)
CHAPTER B. FINANCIAL RATIOS

B.2 Liquidity ratios

Liquidity ratios indicate whether the corporate can pay its short-term liabilities.

Current ratio

\[
Current \ ratio = \frac{Current \ assets}{Current \ liabilities}
\]  
\hspace{1cm} (B.5)

Quick ratio

\[
Quick \ ratio = \frac{Total \ assets}{Total \ liabilities}
\]  
\hspace{1cm} (B.6)

Current asset ratio

\[
Current \ asset \ ratio = \frac{Current \ assets}{Total \ assets}
\]  
\hspace{1cm} (B.7)

Operation cash flow ratio

\[
Operation \ cash \ flow = \frac{Operating \ cash \ flow}{Total \ liabilities}
\]  
\hspace{1cm} (B.8)

Cash to total assets

\[
Cash \ to \ total \ assets = \frac{Total \ cash \ flow}{Total \ assets}
\]  
\hspace{1cm} (B.9)

B.3 Activity ratios

Activity ratios measure how efficient a corporate can turn its assets into cash or revenue.

Asset turnover

\[
Asset \ turnover = \frac{Revenue}{Total \ assets}
\]  
\hspace{1cm} (B.10)

Average asset turnover

\[
Average \ asset \ turnover = \frac{Revenue}{Average \ assets \ for \ period}
\]  
\hspace{1cm} (B.11)

Degree of operating leverage (DOL)

\[
DOL = \frac{\Delta EBIT}{\Delta Revenue}
\]  
\hspace{1cm} (B.12)
CHAPTER B. FINANCIAL RATIOS

B.4 Debt ratios

Debt ratios are used to indicate how much liabilities a corporate owned. Debt ratios vary with different industries. Therefore, different debt ratios are designed to measure firms in different industries.

Debt ratio

\[
Debt \ ratio = \frac{Total \ liabilities}{Total \ assets} \quad \text{(B.13)}
\]

Debt to equity

\[
Debt \ to \ equity = \frac{Total \ liabilities}{Equity} \quad \text{(B.14)}
\]

Long-term debt to equity

\[
Long - term \ debt \ to \ equity = \frac{Long - term \ debt}{equity} \quad \text{(B.15)}
\]

Times interest earned ratio

\[
Times \ interest \ earned \ ratio = \frac{EBIT}{Annual\text{interest}\text{expense}} \quad \text{(B.16)}
\]
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


146
BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY

