

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Andre Sih

PREDIÇÃO DO GRAU DE RATINGS CORPORATIVOS

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para a obtenção do título de Mestre em Ciências da Engenharia Elétrica

Orientador: Carlos Kubrusly

Rio de Janeiro
Agosto de 2006



Andre Sih

Predição do Grau de Ratings Corporativos

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Carlos Kubrusly

Orientador

Departamento de Engenharia Elétrica - PUC-Rio

André D'Almeida Monteiro

Sócio, Gávea Investimentos

Prof. Carlos Eduardo Pedreira

Departamento de Engenharia Elétrica - COPPE/UFRJ

Prof. Carlton Anthony Taft

DMF – CBPF

Prof. Guilherme Caloba

Departamento de Engenharia de Produção – COPPE/UFRJ

Prof. Leonardo Macrini

Departamento de Engenharia Mecânica – PUC-Rio

Prof. Marcelo Medeiros

Departamento de Economia – PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Andre Sih

Graduou-se em Engenharia de Computação na PUC-Rio, em 1999. Iniciou sua carreira aos 19 anos na Xerox Corporation na área de Gerência de Suporte à Software e Engenharia, formando-se em *Software Specialist Engineer* pela Xerox University. Atualmente é sócio-consultor da PDP Consultoria em Gestão Empresarial, e pesquisador do Laboratório de Inteligência Computacional Aplicada no Departamento de Engenharia Elétrica da PUC-Rio, tendo como área de pesquisa, Business Intelligence, Redes Neurais, Estatística, Gestão de Riscos, Finanças e Investimentos.

Ficha Catalográfica

Sih, Andre

Predição do grau de ratings corporativos / Andre Sih ; orientador: Carlos Kubrusly . – Rio de Janeiro : PUC, Departamento de Engenharia Elétrica, 2006.

84 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui referências bibliográficas.

1. Engenharia elétrica – Teses. 2. Redes neurais. 3. Ratings. 4. Estatística. 5. Investimentos. 6. Predição de risco. 7. Standard & poor's. 8. Seleção de atributos. I. Kubrusly, Carlos. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

À minha mãe Marie Louise Sih Kwat Lie

Agradecimentos

Agradeço à todos aqueles que me apoiaram neste desafio e tornaram possível a realização deste trabalho, não podendo deixar de mencionar os meus:

Professores, em particular Carlos Eduardo Pedreira pela atenção, dedicação e orientação prática;

Conselheiros, em particular: Ricardo Weiss, membro do conselho deliberativo da FAPES e chefe da área de *ratings* do BNDES; Oswaldo Fonseca, membro do conselho deliberativo da FAPES; Peter Weil especialista em *ratings* do Credit Suisse; Nilson Teixeira, Economia-Chefe do Credit Suisse; e André D’Almeida Monteiro, Sócio-Gerente da Gávea Investimentos, pelas valiosas contribuições à este trabalho e por todo o conhecimento compartilhado ao longo deste estudo;

Todos os meus familiares;

Amigos, em particular à Deisy e Hiroko Kondo e aos inúmeros companheiros de estudos da PUC-Rio, em particular ao Diogo Duarte, por todo o apoio.

Agradeço também às instituições que fomentaram e viabilizaram os estudos: PUC-Rio, CAPES e CNPq.

Resumo

Sih, Andre; Kubrusly, Carlos. **Predição do Grau de Ratings Corporativos**. Rio de Janeiro, 2006, 84p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo desta dissertação é analisar a relevância de um conjunto inicial de 18 atributos tais como Despesas Financeiras, Receitas e Liquidez Corrente, dentre outros, em relação à classificação de risco (grau) de uma empresa: especulação ou investimento, conforme classificação realizada pela agência Standard & Poor's. Avaliou-se comparativamente a eficácia de métodos lineares e não-lineares de seleção de atributos tais como Análise de Componentes Principais (PCA), Informação Mútua (IM) e Informação Mútua para Seleção de Atributos com Distribuição Uniforme (MIFS-U) e métodos lineares e não-lineares de predição tais como Regressão Múltipla Linear, Discriminante Linear de Fisher e Redes Neurais. Identificou-se através destes métodos e de conhecimento a priori, um conjunto de cinco fatores (atributos) capaz de estimar com alto índice de eficácia se o grau de uma empresa é de investimento ou especulação, a saber: Lucro Líquido, EBIT, Receitas, Valor de Mercado e Setor.

Palavras-chave

Redes Neurais, Ratings, Estatística, Investimentos, Predição de Risco, Standard & Poor's, Seleção de Atributos.

Abstract

Sih, Andre; Kubrusly, Carlos. Corporate Ratings Grade Prediction. Rio de Janeiro, 2006, 84p. M.Sc. Thesis – Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro.

The purpose of this thesis is to analyze and rank the relevancy of 18 variables to S&P corporate ratings grades assignment. Beyond, we predict (classify) the Corporate Grades into two groups – Investment or Speculative. To achieve this goal, we applied and compared linear and non-linear Statistics models and Machine Learning Techniques (Multiple Linear Regression, Linear Fisher's Discriminant, Neural Networks MLP) and feature selection methods such as Principal Component Analysis (PCA), Correlation, Mutual Information (MI) and Mutual Information for Features Selection under Uniform Distribution MIFS-U). The 17 of the initial set of 18 variables are financial variables such as Net Income, Interest Expense and Market Capitalization but one was the corporation's Sector. Combining linear and non-linear models and a priori knowledge, we identified a subset of five features (Net Income, EBIT, Total Revenues, Market Capitalization and Sector) that together reached up to 94.32% of success rate for the S&P grade prediction.

Keywords

Neural Networks, Ratings, Statistics, Investments, Risk Forecast, Standard & Poor's, Feature Selection.

Sumário

1 Introdução	14
1.1 Sobre o conceito de 'rating'	15
1.2 Atributos utilizados na especificação de um rating corporativo	20
2 Banco de Dados	22
2.1 Tratamento de dados faltantes	23
2.2 Formatação de dados	29
3 Metodologia	32
3.1 Seleção de atributos	32
3.1.1 Correlação	33
3.1.2 Stepwise	33
3.1.3 Análise de Componentes Principais (PCA)	34
3.1.4 Informação Mútua (IM)	35
3.1.5 MIFS-U	36
3.2 Predição do desfecho	38
3.2.1 Métodos de classificação lineares	39
3.2.1.1 Regressão Múltipla Linear	39
3.2.1.2 Discriminante Linear de Fisher	40
3.2.2 Métodos de classificação não-lineares	41
3.2.2.1 Redes Neurais Artificiais	41
4 Resultados	43
4.1 Seleção de atributos	43
4.2 Predição do grau da empresa	50
4.2.1 Predição através de Redes Neurais	51
4.2.1.1 Predição após selecionar os atributos com a Informação Mútua	51
4.2.1.2 Predição após selecionar os atributos com o MIFS-U	53
4.2.1.3 Predição após selecionar os atributos com o PCA	57
4.2.1.4 Predição após selecionar os atributos com o Stepwise	59

4.2.1.5 Predição após selecionar os atributos com conhecimento apriori	60
4.2.2 Predição através de métodos lineares	63
4.3 Análise comparativa dos resultados	64
5.Conclusões	68
6.Referências	70
Apêndice I Nome e rating das empresas utilizadas	75

Lista de figuras

Figura 1 – Agrupamento de <i>Ratings</i> em dois graus: Investimento e Especulação.	19
Figura 2 – Comportamento da evolução da taxa de acerto médio através dos atributos inseridos baseados na ordenação resultante do método de IM	53
Figura 3 - Comportamento da evolução da taxa de acerto médio através dos atributos inseridos baseados na ordenação resultante do método de MIFS-U	55
Figura 4 – Comportamento da evolução da taxa de acerto em 5 simulações para os graus de especulação e investimento utilizando aos atributos 8, 17, 1, 3 e 7 selecionados baseado na ordenação resultante do método de MIFS-U.	57
Figura 5 - Comportamento da evolução da taxa de acerto médio através dos atributos inseridos baseados na ordenação resultante do método de PCA ponderado	59
Figura 6 - Comportamento da evolução da taxa de acerto para os graus de especulação e investimento utilizando os atributos 1, 9, 4, 3, 18	62
Figura 7 – Comportamento da variação das taxas de acerto médio com a utilização dos métodos de seleção de atributos e predição de desfecho (grau de investimento ou especulação)	67

Lista de tabelas

Tabela 1 - Os dezoito atributos utilizados e seus códigos	25
Tabela 2 - Classificação de empresas quanto ao valor de mercado	29
Tabela 3 - Distribuição das 318 empresas do BD resultante nas 17 classes de <i>ratings</i> e classes de graus	30
Tabela 4 - Média e Desvios-padrão dos 18 atributos por grau	31
Tabela 5 - Resultados de ordenação dos 18 atributos de acordo com a relevância em relação ao grau da empresa.	45
Tabela 6 - Resultados da correlação e da IM em relação ao grau de uma empresa em ordem decrescente de Informação Mútua. * Os valores da Informação Mútua e da correlação não são passíveis de comparação direta, devido à natureza de cada método	46
Tabela 7 – Matriz de correlações entre todos os dezoito atributos e o grau de uma empresa.	47
Tabela 8 - Matriz de <i>p-values</i> entre todos os dezoito atributos e o grau de uma empresa.	48
Tabela 9 - Importância percentual e acumulada das componentes principais.	49
Tabela 10 - Relevância dos atributos em relação ao desfecho ponderados pela importância percentual das componentes principais	50
Tabela 11 - Resultados de predição do grau de empresas utilizando os atributos ordenados através do método de IM.	52
Tabela 12 - Resultados de predição do grau de empresas utilizando os atributos ordenados através do método de MIFS-U.	54
Tabela 13 - Resultados provenientes de 5 simulações da RN para a predição do grau de empresas utilizando os atributos 8, 17, 1, 3 e 7 selecionados baseado na ordenação através do método de MIFS-U.	56
Tabela 14 - Resultados de predição do grau de empresas utilizando os atributos ordenados através do método PCA ponderado.	58

Tabela 15 – Resultados provenientes de 5 simulações da RN para a predição do grau de empresas utilizando os atributos 1, 9, 4, 3, 18.	62
Tabela 16 – Testes estatísticos da regressão linear	63
Tabela 17 – Coeficientes da regressão linear	64
Tabela 18 – Matriz comparativa entre métodos de seleção lineares e não lineares com a utilização posterior de métodos de classificação lineares e não-lineares.	66

“No momento da determinação de um rating, a história, uma grande parcela da estatística e suas relações que refletem àquela história, são de uma importância apenas para, em um certo grau, nos ajudar a antever o futuro”

Standard & Poor's Corporation