



Universidade Federal de Pernambuco
Centro de Informática

Graduação em Engenharia da Computação

**Análise comparativa de técnicas de
seleção de protótipos**

Dayvid Victor Rodrigues de Oliveira

Trabalho de Graduação

Recife
22 de novembro de 2011

Universidade Federal de Pernambuco
Centro de Informática

Dayvid Victor Rodrigues de Oliveira

Análise comparativa de técnicas de seleção de protótipos

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: *Prof. Dr. George Darmiton*

Recife
22 de novembro de 2011

*Eu dedico este trabalho a João Rodrigues de Silva, meu
avô.*

Agradecimentos

Senhor, obrigado por ter me dado a fora e a confiana para completar esta jornada,
obrigado por ter me guiado atravs de todos os obstculos no meu caminho,
e por me manter
obrigado pela tua proteo em todo caminho,
obrigado pelos amigos que eu fiz,
que o Senhor cuide deles, assim como cuidou de mim,
obrigado por me permitir (...)

What can I give back to God, for the blessings You pour out on me?
—BONO (Boston, 2001)

Resumo

RESUMO

Palavras-chave: PORTUGUES

Abstract

ABSTRACT

Keywords: INGLES

Sumário

1	Técnicas de Seleção de Protótipos	1
1.1	Motivação e Contextualização	1
1.1.1	Seleção de Protótipos	1
1.1.2	Bases Desbalanceadas	1
1.2	Objetivo	1
1.3	Estrutura do Trabalho	1
1.3.1	Sessões	1
1.3.2	Metodologia Utilizada	1
1.3.3	Bases de dados	1
2	Técnicas de Seleção de Protótipos	2
2.1	ENN	2
2.2	Tomek Links	2
2.3	CNN	2
2.4	LVQ	2
2.4.1	LVQ 1	2
2.4.2	LVQ 2.1	2
2.4.3	LVQ 3	2
2.5	SGP	2
2.6	SGP 2	2
2.7	CCNN	2

Lista de Figuras

Lista de Tabelas

Técnicas de Seleção de Protótipos

1.1 Motivação e Contextualização

Classificadores são (...)

*"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web, the content, links, and transactions between people and computers. A **Semantic Web** which should make this possible has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The **intelligent agents** people have touted for ages will finally materialize."*Tim Berners-Lee

Tradução literal: *"Eu tenho um sonho para a Web [em que os computadores] tornam-se capazes de analisar todos os dados na Web, o conteúdo, links, e as transações entre pessoas e computadores. A **Web Semântica** que deve tornar isso possível ainda está para surgir; mas quando isso acontecer, os mecanismos dia-a-dia da burocracia do comércio e nossas vidas diárias serão tratados por máquinas falando com máquinas. Os **agentes inteligentes** que as pessoas têm falado por anos vão finalmente se concretizar."*Tim Berners-Lee

1.1.1 Seleção de Protótipos

1.1.2 Bases Desbalanceadas

1.2 Objetivo

1.3 Estrutura do Trabalho

1.3.1 Sessões

1.3.2 Metodologia Utilizada

1.3.3 Bases de dados

CAPÍTULO 2

Técnicas de Seleção de Protótipos

2.1 ENN

2.2 Tomek Links

2.3 CNN

2.4 LVQ

2.4.1 LVQ 1

2.4.2 LVQ 2.1

2.4.3 LVQ 3

2.5 SGP

2.6 SGP 2

2.7 CCNN

Referências Bibliográficas

- [CPZ11] Ruiqin Chang, Zheng Pei, and Chao Zhang. A modified editing k-nearest neighbor rule. *JCP*, 6(7):1493–1500, 2011.
- [dSPC08] Cristiano de Santana Pereira and George D. C. Cavalcanti. Prototype selection: Combining self-generating prototypes and gaussian mixtures for pattern classification. In *IJCNN*, pages 3505–3510. IEEE, 2008.
- [EJJ04] Andrew Estabrooks, Taeho Jo, and Nathalie Japkowicz. A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1):18–36, 2004.
- [FHA07] Hatem A. Fayed, Sherif Hashem, and Amir F. Atiya. Self-generating prototypes for pattern classification. *Pattern Recognition*, 40(5):1498–1509, 2007.
- [Har68] P. E. Hart. The condensed nearest neighbor rule. *IEEE Transactions on Information Theory*, 14:515–516, 1968.
- [HKN07] Jason Van Hulse, Taghi M. Khoshgoftaar, and Amri Napolitano. Experimental perspectives on learning from imbalanced data. In Zoubin Ghahramani, editor, *ICML*, volume 227 of *ACM International Conference Proceeding Series*, pages 935–942. ACM, 2007.
- [Koh86] Teuvo Kohonen. Learning vector quantization for pattern recognition. Report TKK-F-A601, Laboratory of Computer and Information Science, Department of Technical Physics, Helsinki University of Technology, Helsinki, Finland, 1986.
- [Koh88] Teuvo Kohonen. Learning vector quantization. *Neural Networks*, 1, Supplement 1:3–16, 1988.
- [PI69] E. A. Patrick and F. P. Fischer II. A generalization of the k-nearest neighbor rule. In *IJCAI*, pages 63–64, 1969.
- [TC67] P.E. Hart T.M. Cover. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IT-13:21 – 27, 1967.
- [Tom76] I. Tomek. Two Modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(2):679–772, 1976.
- [WP01] G. Weiss and F. Provost. The effect of class distribution on classifier learning, 2001.