ISSN: 1984-3151

UTILIZAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NO RECONHECIMENTO DE ENTIDADES NOMEADAS NO PORTUGUÊS

USE OF MACHINE LEARNING TECHNIQUES IN RECOGNITION OF PORTUGUESE NAMED ENTITIES

Paulo Roberto Simões Pellucci; Renato Ribeiro de Paula; Walter Borges de Oliveira Silva; Ana Paula Ladeira
Centro Universitário de Belo Horizonte, Belo Horizonte, MG

paulo.pellucci@gmail.com; r3nato.rp@gmail.com; waltinho1979@gmail.com; aladeira.unibh@gmail.com

Recebido em: 23/05/2011 - Aprovado em: 30/06/2011 - Disponibilizado em: 24/07/2011

RESUMO: Técnicas de Aprendizado de Máquina são comumente utilizados em tarefas que envolvem a identificação de padrões. O reconhecimento de entidades nomeadas consiste em identificar todos os nomes de um documento e classificá-los em categorias prévias. No presente artigo foram analisadas três técnicas com aprendizado supervisionado, sendo que, os melhores resultados foram obtidos com o algoritmo NaiveBayes. Inúmeros experimentos foram realizados usando a ferramenta Weka, no entanto, os valores obtidos por meio da estatística Kappa ficaram abaixo do esperado.
PALAVRAS-CHAVE: Reconhecimento de Entidades Nomeadas. Aprendizagem de Máquina. Aprendizado supervisionado. Weka.

ABSTRACT: Machine Learning techniques are commonly used in tasks involving pattern recognition. The named entity recognition consists in identify all the names in a document and classify them into prior categories. The present paper analyzed three techniques with supervised learning and the best result was obtained with the NaiveBayes algorithm. Numerous experiments were performed using the Weka tool, however, the values obtained by the Kappa statistic were lower than expected.
KEYWORDS: Named Entity Recognition. Learning Machine. Supervised learning. Weka.

---

1 INTRODUÇÃO

Nas últimas décadas tem-se observado uma explosão na quantidade de informação armazenada e disponibilizada em documentos. Acredita-se que grande parte das informações, hoje, esteja no formato textual, tornando essencial o estudo de métodos de análise e processamento, focados nesse tipo de informação (BAEZA-YATES; RIBEIRO-NETO, 1999).

Observa-se ainda que as estratégias de busca e extração de informação não tem sido suficientes para resolver o problema, diante do volume de dados e informações gerados (WIVES, 2004).

A alternativa encontrada é utilizar as tecnologias de informação para tornar este acervo crescente de conhecimento mais acessível. Neste sentido, o estudo de técnicas de aprendizado de máquina para extração de informação em documentos textuais tem sido impulsionado pelo volume de textos produzidos diariamente.

Dentre as tarefas envolvidas no processamento da linguagem natural, o reconhecimento de entidades

nomeadas, do inglês *Named Entity Recognition* (NER), busca identificar e classificar elementos de um texto em categorias pré-definidas tais como pessoa, organização, lugar, quantidade, dentre outras (DUARTE; MILIDIÚ, 2007).

O NER provê características chave que auxiliam em tarefas mais elaboradas de gerenciamento de documentos e extração de informação (DUARTE; MILIDIÚ, 2007).

A identificação automática de entidades nomeadas pode basear-se no contexto o qual o termo está inserido, ou seja, no conjunto de termos ao redor do termo em questão (GOULART; STRUBE DE LIMA, 2009). Neste sentido, uma possibilidade é tentar reconhecer o padrão de termos no qual a entidade nomeada candidata se enquadra.

Diante disso, o presente trabalho tem como objetivo geral comparar as técnicas de aprendizado de máquina NaiveBayes, SVM e a árvore de decisão (*Decision Table)* no processo de classificação das entidades nomeadas de documentos em Português.

As técnicas NaiveBayes e SVM foram escolhidas tendo como base os experimentos realizados em Duarte e Milidiú (2007). No entanto, a árvore de decisão foi escolhida para enriquecer estes experimentos, mesmo não tendo sido utilizada por Duarte e Milidiú (2007), a literatura mostra que ela tem conseguido desempenho acima de técnicas mais complexas, como indica Kohavi (1995).

Dentre os objetivos específicos tem-se: investigar o processo de treinamento de algoritmos de aprendizado supervisionado e não supervisionado; identificar os algoritmos de aprendizados implementados nas técnicas NaiveBayes, SVM e árvore de decisão (*Decision Table)*; comparar os índices de acertos obtidos pelas técnicas escolhidas no processo de reconhecimento de entidades nomeadas no Português. Para isso, utilizou-se a estatística Kappa (COHEN, 1960) retornada pelas técnicas como uma medida de confiabilidade para verificar a concordância entre as taxas de acerto alcançadas.

A estrutura do presente artigo é organizada da seguinte forma: na seção 2 são mostrados os processos de aprendizado supervisionado e não supervisionado. Na seção 3 são explorados os métodos de classificação usados no desenvolvimento do presente trabalho. Na 4.ª seção expõem-se a metodologia de criação dos experimentos e as ferramentas utilizadas. Na seção 5 os resultados obtidos e consolidados são apresentados e finalmente na seção 6 é apresentada a conclusão juntamente com as sugestões de continuidade do trabalho.

2 APRENDIZADO SUPERVISIONADO E NÃO SUPERVISIONADO

Conforme Lorena e Carvalho (2007), as técnicas de aprendizado de máquinas empregam um princípio de inferência denominado indução, no qual é possível obter conclusões genéricas a partir de um conjunto particular de exemplos. Estas técnicas de aprendizados indutivos podem ser divididas em dois principais tipos: os supervisionados e os não supervisionados.

No aprendizado supervisionado é fornecida uma referência do objetivo a ser alcançado, isto é, um treinamento com o conhecimento do ambiente. Este treinamento são conjuntos de exemplos com entradas e uma saída esperada. O algoritmo de aprendizado de máquina extrai a representação do conhecimento a partir desses exemplos. O objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas antes.

De acordo com Mitchell (1997), o erro de um aprendizado supervisionado pode ser calculado como a diferença entre a saída desejada e a saída gerada, conforme a Eq. (1):

$$e_k = d_k - y_k \quad (1)$$

onde: k = estímulo, e = sinal de erro, d = saída desejada apresentada durante o treinamento, y = saída real do algoritmo após a apresentação do estímulo de entrada.

Diferentemente do aprendizado supervisionado, o não supervisionado não se utiliza referências, ou seja, não ocorre um treinamento com o conhecimento do ambiente. Lorena e Carvalho (2007) destacam que "o algoritmo de aprendizado de máquina não supervisionado aprende a representar (ou agrupar) as entradas submetidas, segundo medidas de similaridade" (p. 44).

De acordo com Souto *et al*. (2003), as técnicas de aprendizado não supervisionado são mais utilizadas quando o entendimento dos dados é feito através de padrões ou tendências.

3 CLASSIFICADORES UTILIZADOS

As técnicas descritas, a seguir, foram utilizadas nos experimentos realizados no presente trabalho. Elas foram escolhidas por serem amplamente utilizadas em problemas que envolvem o reconhecimento de padrões. Além disso, tais técnicas apresentam interfaces similares, o que dispensou a criação de diferentes arquivos de treinamento para cada uma delas.

3.1 NAIVEBAYES

Segundo Oguri e Milidiú (2006), o NaiveBayes é "provavelmente o classificador mais utilizado em Machine Learning" (p. 25). É denominado ingênuo (*naive*) por assumir que os atributos são condicionalmente independentes. Em outras palavras, considera-se que as entradas são independentes entre si, o que não ocorre na maioria dos problemas práticos. Mesmo que esta premissa de ingenuidade seja mantida, o classificador reporta resultados que não comprometem a qualidade.

Existem, basicamente, dois tipos de modelos estatísticos para os classificadores NaiveBayes: modelo binário e modelo multinomial, que serão apresentados a seguir.

Segundo Oguri e Milidiú (2006), o modelo binário representa um documento através de um vetor binário. O valor 0 (zero) em uma posição *k* (onde *k* seria uma palavra da frase) representa a não ocorrência do termo, enquanto que o valor 1 (um) representa ao menos uma ocorrência do termo.

Esta técnica foi utilizada por Duarte e Milidiú (2007) no artigo utilizado como base no presente trabalho.

Já o modelo multinomial assume que o documento é representado por um vetor de valores inteiros, caracterizando o número de vezes que cada termo ocorre no documento (OGURI; MILIDIÚ, 2006).

3.2 SVM

O SVM é uma técnica recente (da década de 90) utilizada no processo de reconhecimento de padrões e regressão linear. Segundo Duarte e Milidiú (2007), o SVM é uma técnica de aprendizado linear baseada no uso de *kernels* (núcleo) e de regras não-lineares. A ideia central do SVM é o núcleo do produto interno entre o chamado vetor de suporte e um vetor retirado do espaço de entrada (HAYKIN, 1999). Os vetores de suporte são um subconjunto dos dados de treinamento.

O SVM usa propriedades geométricas para calcular o hiperplano que melhor separa um conjunto de exemplos de treinamento (OGURI; MILIDIÚ, 2006).

Assim como o NaiveBayes, o modelo SVM é utilizado para resolver problemas de classificação binária. No entanto, de acordo com Duarte (2007), o SVM utiliza de decomposição em vários problemas binários onde o problema original é uma classificação multi-classe.

3.3 DECISION TABLE

O *Decision table* é uma das técnicas mais simples de aprendizado supervisionado e uma das mais simples de se entender o funcionamento (KOHAVI, 1995).

Algumas tabelas utilizam valores verdadeiro ou falso (*true* ou *false)* para representar as alternativas condições (estilo *if-then-else*). Outras utilizam de alternativas numeradas (estilo *switch-case*) e, ainda existem aquelas, que utilizam de lógica *fuzzy* (lógica difusa que apresenta vários níveis entre verdadeiro e falso) ou representações probabilísticas para as alternativas condicionais (WETS *et al*., 1996).

4 METODOLOGIA

O presente artigo utiliza um corpus disponibilizado por Duarte e Milidiú (2007), com anotações de POS (do inglês *part-of-speech* – ou seja, classes gramaticais) em cada palavra e qual classe (entidade nomeada) a palavra pertence. O corpus disponibilizado tem também a informação de sintagma nominal.

O sintagma nominal é a unidade sintática que pode exercer a função de sujeito em uma oração. A metodologia escolhida pelo presente artigo necessitava que no corpus houvesse mais marcações deste atributo. Como eram poucas estas marcações, o sintagma nominal não foi utilizado.

No presente artigo são utilizadas três categorias (também consideradas como classes) de maior interesse normalmente, sendo elas: Pessoa, Organização e Lugar. Sendo assim, após analisar e modificar o corpus original, o estudo foi baseado em um corpus contendo 566 frases da entidade Pessoa (PES), 446 frases da entidade Organização (ORG) e 240 frases da entidade Lugar (LOC), totalizando 1.250 sentenças.

Utilizou-se a mesma metodologia do artigo de Duarte e Milidiú (2007) na técnica do SVM. É montada uma janela de cinco posições. Esta janela é composta de duas palavras que antecedem a entidade nomeada (vizinhos anteriores), a entidade nomeada e, duas palavras posteriores (vizinhos posteriores). Como atributo de treinamento é utilizado o POS das palavras pertencentes à janela.

Os mesmos autores definem POS como o processo de etiquetar palavras de acordo com suas classes gramaticais, como: artigo, adjetivo, substantivo, etc. Assim como no artigo base, palavras que não possuem classe gramatical (tais como vírgula, ponto, interrogação e outras) receberam um valor de “outro”.

Cada linha do corpus disponibilizado é formada pela palavra, seguida do seu POS, o sintagma e a entidade nomeada, todos separados por um caractere “_” (FIG. 1).

```
O_ART_I_O
processo_N_I_O
eleitoral_ADJ_I_O
também_PDEN_O_O
preocupou_V_O_O
Itamar_NPROP_I_PER
```

FIGURA 1. Corpus disponibilizado com a palavra, seu POS, sintagma nominal e classe pertencente.

FONTE: Adaptado de Duarte e Milidiú (2007).

Na Tabela 1, dada a seguir, são listadas as classes gramaticais que foram marcadas para a construção do corpus.

Durante a realização dos experimentos, utilizou-se as classes da ferramenta Weka, por já apresentar implementados todos os algoritmos avaliados no presente trabalho. Foi criado um aplicativo escrito no *framework* .NET para o desenvolvimento dos arquivos utilizados no treinamento das técnicas.

O Weka recebe como entrada arquivos extensão *arff* e possui um layout que deve ser seguido para utilizar nos algoritmos. Este layout é dividido em duas seções: cabeçalho e dados. O cabeçalho contém o nome da relação (tag @RELATION), uma lista de atributos (tag

@ATTRIBUTE) e seus tipos. Esta lista de atributos são as colunas na parte de dados do arquivo.

Tabela 1

CLASSES GRAMATICAIS E SEUS RESPECTIVOS ACRÔNIMOS, UTILIZADOS PARA A CONSTRUÇÃO DO CORPUS

Acrônimo	Classe gramatical
ADJ	Adjetivo
ADV	Advérbio
ART	Artigo
KC	Conjunção coordenativa
KS	Conjunção subordinativa
N	Substantivo
NPROP	Substantivo próprio
NUM	Numeral
PCP	Particípio
PDEN	Palavra denotativa
PREP	Preposição
PROADJ	Pronome adjetivo
PROPESS	Pronome pessoal
PROSUB	Pronome substantivo
V	Verbo
VAUX	Verbo auxiliar
DUMMY	Classe não representada

FONTE: http://www.linguateca.pt/acesso/anotacao.html

O primeiro arquivo gerado para a ferramenta determinou valores diferentes para cada classe gramatical. Este arquivo não segue a metodologia do artigo base, foi apenas uma tentativa de verificar o funcionamento das técnicas. Um exemplo do arquivo pode ser visto na FIG. 2.

```
@RELATION ner

@ATTRIBUTE categtwo   INTEGER
@ATTRIBUTE categone   INTEGER
@ATTRIBUTE categpone   INTEGER
@ATTRIBUTE categptwo   INTEGER
@ATTRIBUTE class      {PES,ORG,LOC}

@DATA
14,6,99,99,ORG
14,6,99,14,LOC
99,6,20,9,ORG
6,9,14,16,ORG
99,20,99,9,PES
6,9,99,11,PES
6,9,20,9,PES
13,20,99,99,PES
9,14,99,11,PES
```

FIGURA 2. Atributos classificados como valores inteiros. A palavra (EN) foi omitida na imagem.

Pode-se observar que a palavra (entidade nomeada) não foi apresentada na figura, pois, sempre gera o mesmo valor, afinal sempre serão da mesma classe gramatical. Os atributos possuem valores inteiros diferentes para cada classe gramatical. Isso é possível de ser notado nos diferentes valores que foram gerados para uma mesma linha.

O arquivo usado no segundo experimento foi o mesmo utilizado em Duarte e Milidiú (2007) para a técnica do SVM. Um arquivo foi gerado contendo um vetor V = [0,1], onde 0 indica a não presença do atributo na janela e, 1 representa a presença do atributo na janela. Um exemplo do arquivo pode ser visto na FIG. 3.

```
@RELATION ner

@ATTRIBUTE ADJ   INTEGER
@ATTRIBUTE ADV   INTEGER
@ATTRIBUTE ADV-KS   INTEGER
@ATTRIBUTE ADV-KS-REL   INTEGER
@ATTRIBUTE ART   INTEGER
@ATTRIBUTE KC   INTEGER
@ATTRIBUTE KS   INTEGER
@ATTRIBUTE N   INTEGER
@ATTRIBUTE NPROP   INTEGER
@ATTRIBUTE NUM   INTEGER
@ATTRIBUTE PCP   INTEGER
@ATTRIBUTE PDEN   INTEGER
@ATTRIBUTE PREP   INTEGER
@ATTRIBUTE PROADJ   INTEGER
@ATTRIBUTE PROPESS   INTEGER
@ATTRIBUTE PROSUB   INTEGER
@ATTRIBUTE PRO-KS-REL   INTEGER
@ATTRIBUTE PRO-KS   INTEGER
@ATTRIBUTE V   INTEGER
@ATTRIBUTE VAUX   INTEGER
@ATTRIBUTE DUMMY   INTEGER
@ATTRIBUTE class      {PES,ORG,LOC}

@DATA
0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,ORG
0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,LOC
0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,1,ORG
0,0,0,0,1,0,0,1,0,0,0,0,1,0,1,0,0,0,0,0,0,0,ORG
0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,1,PES
0,0,0,0,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,1,PES
0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,PES
```

FIGURA 3. Atributos classificados em 0 e 1 para indicar não presença e presença respectivamente.

Até o momento, ou seja, para os dois primeiros experimentos, o tipo dos atributos foram todos valores numéricos. Para o terceiro, avaliou-se a possibilidade de se utilizar variáveis nominais. Neste caso, todas as marcações do POS foram utilizadas como dados do arquivo. A FIG. 4 mostra um trecho do arquivo gerado. Apesar da figura não listar todas as marcações do POS, todas foram utilizadas.

```
@RELATION ner

@ATTRIBUTE minusculo   {ADJ,ADV,ADV-KS,ADV-KS-REL,ART,KC,KS,N,
@ATTRIBUTE minusculo   {ADJ,ADV,ADV-KS,ADV-KS-REL,ART,KC,KS,N,
@ATTRIBUTE word        {ADJ,ADV,ADV-KS,ADV-KS-REL,ART,KC,KS,N,
@ATTRIBUTE plusculo    {ADJ,ADV,ADV-KS,ADV-KS-REL,ART,KC,KS,N,
@ATTRIBUTE plusculo    {ADJ,ADV,ADV-KS,ADV-KS-REL,ART,KC,KS,N,
@ATTRIBUTE class       {PES,ORG,LOC}

@DATA
PREP,ART,NPROP,DUMMY,DUMMY,ORG
PREP,ART,NPROP,DUMMY,PREP,LOC
DUMMY,ART,NPROP,V,N,ORG
ART,N,NPROP,PREP,PROPESS,ORG
DUMMY,V,NPROP,DUMMY,N,PES
ART,N,NPROP,DUMMY,NUM,PES
ART,N,NPROP,V,N,PES
PDEN,V,NPROP,DUMMY,DUMMY,PES
N,PREP,NPROP,DUMMY,NUM,PES
N,DUMMY,NPROP,DUMMY,DUMMY,PES
N,PREP,NPROP,KC,NPROP,PES
DUMMY,ART,NPROP,V,PREP,LOC
ADV,ART,NPROP,ADV,V,ORG
PREP,ART,NPROP,KC,PREP,LOC
```

FIGURA 4. Atributos sendo utilizados como multi-valores.

Todos os resultados obtidos serão apresentados na próxima seção.

5 RESULTADOS

Nesta seção são apresentados os resultados obtidos em cada experimento realizado. Além da porcentagem de acertos, foram investigados os valores retornados pela estatística Kappa.

A estatística Kappa fornece uma diferença que é medida entre a concordância observada na precisão da técnica e os valores “por acaso” (COHEN, 1960). Os valores da estatística Kappa ficam entre [0,1], onde o mais próximo de 0 (zero) significa o acerto “no chute” e o mais próximo de 1 (um) indica concordância exata da inferência dos valores pela técnica (CARLETTA, 1996).

Tabela 2

TAXAS DE ACERTO OBTIDAS NOS EXPERIMENTOS REALIZADOS

	Exp. 1	Exp. 2	Exp. 3
NaiveBayes	50.9%	65.1%	**76.9%**
SVM	53.5%	**70.1%**	76.0%
Decision Table	**72.7%**	69.3%	75.7%

Tabela 3

ESTATÍSTICA KAPPA OBTIDA NOS EXPERIMENTOS REALIZADOS

	Exp. 1	Exp. 2	Exp. 3
NaiveBayes	0.226	0.423	0.626
SVM	0.248	0.493	0.601
Decision Table	0.538	0.480	0.597

A técnica do NaiveBayes foi a que teve a maior taxa de acerto dentre todos os experimentos realizados (76,9% de acerto no terceiro experimento).

Para o primeiro experimento, as técnicas que utilizam de modelos binários (SVM e NaiveBayes) apresentaram porcentagens de acerto baixas. O *Decision Table* apresentou os melhores resultados.

Para o segundo experimento, as técnicas de modelo binário obtiveram uma porcentagem de acerto superior ao *Decision Table*. Ainda assim, a técnica *Decision Table* ficou bem próxima ao SVM, quem teve a maior porcentagem.

O terceiro experimento apresenta valores acima de 75% para todas as técnicas, o que torna dedutível que foi o melhor resultado. Neste experimento o NaiveBayes obteve a melhor porcentagem de acerto e uma melhora excelente em suas porcentagens, considerando que, nos anteriores, as porcentagens foram abaixo do esperado.

Na Tabela 3, dada a seguir, é possível verificar o valor da estatística Kappa para cada uma das técnicas nos experimentos realizados.

Observa-se que, diante dos valores apresentados para a estatística Kappa, muitos acertos não foram inferências exatas das técnicas. Segundo Carletta (1996), os valores recomendados para a estatística Kappa devem ser entre 0.8 e 1.

6 Conclusão

A maior dificuldade do estudo realizado foi tentar alcançar os mesmos valores expostos no artigo base. Em Duarte e Milidiú (2007), a técnica SVM apresentou um acerto de 84.8%, o que não se alcançou nesta pesquisa.

Um ponto importante que a investigação trouxe é que, nenhum dos três experimentos teve um valor considerável para a estatística Kappa. Os valores obtidos para a estatística não invalidam os resultados deste trabalho, eles demonstram que as técnicas utilizadas não foram adequadas para chegar ao acerto de Duarte e Milidiú (2007).

É possível deduzir que a metodologia adotada para definir os atributos usados no processo de classificação e, consequentemente, a maneira como foram montados os experimentos não foram adequadas. Conforme descrito no presente artigo, nem todos os atributos do corpus original foram usados e isto pode ter sido o caráter da distinção entre os valores deste trabalho e de Duarte e Milidiú (2007). Mais um ponto a se considerar é que neste último o algoritmo SVM foi desenvolvido, enquanto que no Weka, o algoritmo do SVM disponibilizado não é a mesma implementação usada no artigo supracitado.

O terceiro experimento foi o que trouxe a menor distinção entre as porcentagens de acerto das técnicas.

Como sugestão para futuros trabalhos, um foco maior neste terceiro experimento e uma mudança nos atributos no processo de classificação podem criar a possibilidade de maiores porcentagens nos resultados destas técnicas.

AGRADECIMENTOS

Os autores gostariam de agradecer a todos os envolvidos no desenvolvimento deste trabalho. O tema escolhido foi baseado em uma monografia e que despertou bastante interesse por não existir literatura publicada similar para o Português.

Assim, os agradecimentos ao UniBH, pelo ambiente acadêmico, pelo seu corpo docente e pelos recursos disponíveis, o que tornaram possível a realização deste trabalho. Por fim, agradecimentos especiais à enorme ajuda concedida pelo Júlio Cesar Duarte, autor do trabalho usado como referência.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B., **Modern Information Retrieval**. Addison-Wesley, 1999.

CARLETTA, J. **Assessing Agreement of Classification Tasks**: the Kappa Statistic, Computational Linguistics. 1996. 22:2.249-254. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.3107&rep=rep1&type=pdf>. Acesso em: nov. 2010.

COHEN, J. **A Coefficient of Agreement for Nominal Scales**. Journal of Educational and Psychological Measurement. 1960. 20.37-46. Disponível em: <http://www.garfield.library.upenn.edu/classics1986/A1986AXF2600001.pdf>. Acesso em: out.2010.

DUARTE, Julio Cesar, MILIDIÚ, Ruy Luiz. **Machine Learning Algorithms for Portuguese Named Entity Recognition**. 2007. 10 p. Monografia (Monografias em Ciência da Computação) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007. Disponível em: <http://www.dbd.puc-rio.br/depto_informatica/07_09_duarte.pdf>. Acesso em: 1 out. 2010.

GOULART, R. R. V.; STRUBE DE LIMA, V. L. **O Contexto no Reconhecimento de Entidades Nomeadas em Textos de Biomedicina.** Simpósio de Tecnologias da Informação e da Língua (STIL), 2009, São Carlos. Anais do Simpósio de Tecnologias da Informação e da Língua (STIL), 2009. p. 1-10.

HAYKIN, S. S., **Neural Networks**: A Comprehensive Foundation, Editora Prentice-Hall, 2. ed., 842 p., 1999.

KOHAVI, R. **The power of decision tables**. European Conference on Machine Learning (ECML). 1995. 16 p. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.4576>. Acesso em: 1 nov. 2010.

LORENA, A. C.; CARVALHO, A. C. P. L. F. **Uma Introdução às Support Vector Machines.** Revista de Informática Teórica e Aplicada, vol.14, no2, pp 43-67, 2007. Disponível em: <http://seer.ufrgs.br/index.php/rita/article/viewPDFInterstitial/rita_v14_n2_p43-67/3543>. Acesso em: 16 out. 2010.

MITCHELL, Tom M. **Machine Learning**: McGraw-Hill, 1997. Cap 1, 10 f. Disponível em: <http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>. Acesso em: 6 out. 2010.

OGURI, Pedro, MILIDIÚ, Ruy Luiz, Renteria, Raúl. **Aprendizado de Máquina para o Problema de Sentiment Classification**. 2006. 54 p. Dissertação de Mestrado (Pós-graduação em Mestrado em Informática) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/acessoConteudo.php?nrseqoco=31636>. Acesso em: 8 nov. 2010.

SOUTO, M. C. P.; LORENA, A. C.; DELMBEM, A. C. B.; CARVALHO, A. C. P. L. F. **Técnicas de Aprendizado de Máquina para problemas de Biologia Molecular**. Minicursos de Inteligência Artificial, Jornada de Atualização Científica em Inteligência Artificial, XXIII Congresso da Sociedade Brasileira de Computação, 2003. 103–152 p. Disponível em: <http://www.dimap.ufrn.br/~marcilio/DAAD/BIB/GENEEXPRESSION/jaia2003-14-03-08.pdf>. Acesso em: 4 nov. 2010.

VAPNIK, V.; BOSER, B.; GUYON, I. **A Training Algorithm for Optimal Margin Classifiers**. Proc. of the 5th Annual ACM Workshop on Computational Learning Theory, pp.144-152, ACM Press, 1992. Disponível em: <http://portal.acm.org/citation.cfm?id=130401>. Acesso em: 1 nov. 2010.

WEKA, Weka 3: Data Mining Software in Java, Disponível em www.cs.waikato.ac.nz/ml/weka/, Acesso em: mar. 2010.

WETS, G.; WITLOX, F.; TIMMERMANS, H.; VANTHIENEN, J. **Locational choice modelling using fuzzy decision tables**, Biennial Conference of the North American Fuzzy Information Processing Society, Berkeley, CA, (1996). IEEE, pp. 80–84. Disponível em <http://alexandria.tue.nl/repository/freearticles/589700.pdf>. Acesso em: 15 nov. 2010.

WIVES, L. K. **Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (clustering) de documentos**. 2004, 131 f. Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.