Preference Learning in Terminology Extraction: A ROC-based approach

Jérôme Azé, Mathieu Roche, Yves Kodratoff, and Michèle Sebag

LRI – Laboratoire de Recherche en Informatique
UMR8623, CNRS, Université Paris Sud, 91405 Orsay Cedex, France
(e-mail: {aze,roche,yk,sebag}@lri.fr)

Abstract. A key data preparation step in Text Mining, Term Extraction selects the terms, or collocation of words, attached to specific concepts. In this paper, the task of extracting relevant collocations is achieved through a supervised learning algorithm, exploiting a few collocations manually labelled as relevant/irrelevant. The candidate terms are described along 13 standard statistical criteria measures. From these examples, an evolutionary learning algorithm termed ROGER, based on the optimization of the Area under the ROC curve criterion, extracts an order on the candidate terms. The robustness of the approach is demonstrated on two real-world domain applications, considering different domains (biology and human resources) and different languages (English and French).
Keywords: Text Mining, Terminology, Evolutionary algorithms, ROC Curve.

1 Introduction

Besides the known difficulties of Data Mining, Text Mining presents specific difficulties due to the structure of natural language. In particular, the polysemy and synonymy effects are dealt with by constructing ontologies or terminologies [Bourigault and Jacquemin, 1999], structuring the words and their meanings in the domain application. A preliminary step for ontology construction is to extract the terms, or word collocations, attached to the concepts defined by the expert [Bourigault and Jacquemin, 1999, Smadja, 1993]. Term Extraction actually involves two steps: the detection of the relevant collocations, and their classification according to the concepts.

This paper focuses on the detection of relevant collocations, and presents a learning algorithm for ranking collocations with respect to their relevance, in the spirit of [Cohen *et al.*, 1999]. An evolutionary algorithm termed ROGER, based on the optimization of the Receiver Operating Characteristics (ROC) curve [Ferri *et al.*, 2002, Rosset, 2004], and already described in previous works [Sebag *et al.*, 2003a, Sebag *et al.*, 2003b], is applied to a few collocations manually labelled as relevant/irrelevant by the expert. The optimization of the ROC curve is directly related to the recall-precision tradeoff in Term Extraction (TE).

The paper is organized as follows. Section 2 briefly reviews the main criteria used in TE. Section 3 presents the ROGER (ROc-based GEnetic learneR)

algorithm for the sake of self-containedness, and describes the bagging of the diverse hypotheses constructed along independent runs. Sections 4 et 5 report on the experimental validation of the approach on two real-world domain applications, and the paper ends with some perspectives for further research.

2 Measures for Term Extraction

The choice of a quality measure among the great many criteria used in Text Mining (see e.g., [Daille *et al.*, 1998, Xu *et al.*, 2002, Roche *et al.*, 2004b]) is currently viewed as a decision making process; the expert has to find the criterion most suited to his/her corpus and goals. The criteria considered in the rest of the paper are:

- Mutual Information (MI) [Church and Hanks, 1990]
- Mutual Information with cube (MI^3) [Daille *et al.*, 1998]
- Dice Coefficient ($Dice$) [Smadja *et al.*, 1996]
- Log-likelihood (L) [Dunning, 1993]
- Number of occurrences + Log-likelihood (Occ_L)[1] [Roche *et al.*, 2004a]
- Association Measure (Ass) [Jacquemin, 1997]
- Sebag-Schoenauer ($SeSc$) [Sebag and Schoenauer, 1988]
- J-measure (J) [Goodman and Smyth, 1988]
- Conviction ($Conv$) [Brin *et al.*, 1997]
- Least contradiction (LC) [Azé and Kodratoff, 2004]
- Cote multiplier (CM) [Lallich and Teytaud, 2004]
- Khi2 test used in text mining ($Khi2$) [Manning and Schütze, 1999]
- T-test used in text mining ($Ttest$) [Manning and Schütze, 1999]

Vivaldi *et al.* [Vivaldi *et al.*, 2001] have shown that the search for a quality measure can be formalized as a supervised learning problem. Considering a training set, where each candidate term is described from its value for a set of statistical criteria and labelled by the expert, they used Adaboost [Schapire, 1999] to automatically construct a classifier.

The approach presented in next section mostly differs from [Vivaldi *et al.*, 2001] as it learns an ordering function (term t_1 is more relevant than term t_2) instead of a boolean function (term t is relevant/irrelevant).

3 Learning ranking functions

This section first briefly recalls the ROGER (*ROc-based GEnetic learneR*) algorithm, used for learning a ranking hypothesis and first described in [Sebag *et al.*, 2003b, Sebag *et al.*, 2003a]. The N'ROGER variant used in this paper involves two extensions: i) the use of non-linear ranking hypotheses; ii) the

[1] Occ_L is defined by ranking collocations according to their number of occurrences, and breaking the ties based on the term Log-likelihood.

exploitation of the ensemble of hypotheses learned along independent runs of ROGER. Using the standard notations, the dataset $\mathcal{E} = \{(\mathbf{x}_i, y_i), i = 1..n, \mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1, +1\}\}$ includes n collocation examples, where each collocation $\mathbf{x}_i$ is described by the value of d statistical criteria, and its label y_i denotes whether collocation $\mathbf{x}_i$ is relevant.

3.1 ROGER

The learning criterion used in ROGER is the Wilcoxon rank test, measuring the probability that a hypothesis h ranks $\mathbf{x}_i$ higher than $\mathbf{x}_j$ when $\mathbf{x}_i$ is a positive and $\mathbf{x}_j$ is a negative example:

$$\mathcal{W}(h) = Pr(h(x_i) > h(x_j) \mid y_i > y_j) \tag{1}$$

This criterion, with quadratic complexity in the number n of examples[2] offers an increased stability compared to the misclassification rate ($Pr(h(x_i).y_i > 0$), with linear complexity in n); see [Rosset, 2004] and references therein. The Wilcoxon rank test is equivalent to the area under the ROC (Receiver Operating Characteristics) curve [Jin *et al.*, 2003]. This curve, intensively used in medical data analysis, shows the trade-off between the true positive rate (the fraction of positive examples that are correctly classified, aka recall) and the false positive rate (the fraction of negative examples that are misclassified) achieved by a given hypothesis/classifier/learning algorithm. Therefore, the area under the ROC curve (AUC) does not depend on the imbalance of the training set [Kolcz *et al.*, 2003], as opposed to other measures such as Fscore [Caruana and Niculescu-Mizil, 2004]. The ROC curve also shows the misclassification rates achieved depending on the error cost coefficients [Domingos, 1999]. For these reasons, [Bradley, 1997] argues the comparison of the ROC curves attached to two learning algorithms to be more fair and informative, than comparing their misclassification rates only. Accordingly, the area under the ROC curve defines a new learning criterion, used e.g. for the evolutionary optimization of neural nets [Fogel *et al.*, 1995], or the greedy search of decision trees [Ferri *et al.*, 2002].

In an earlier step [Sebag *et al.*, 2003b], the search space $\mathcal{H}$ considered is that of linear hypotheses ($\mathcal{H} = \mathbb{R}^d$). To each vector w in $\mathbb{R}^d$ is attached hypothesis h_w with $h_w(x) =< w, x >$, where $< w, x >$ denotes the scalar product of w and x. Hypothesis h defines an order on $\mathbb{R}^d$, which is evaluated from the Wilcoxon rank test on the training set $\mathcal{E}$ (Eq. 1), measured after cross-validation.

The combinatorial optimization problem defined by Eq. 1, thus mapped onto a numerical optimization problem, is tackled by Evolution Strategies (ES). ES are the Evolutionary Computation algorithms that are best suited to parameter optimization; the interested reader is referred to [Bäck, 1995]

[2] Actually, the computational complexity is in $\mathcal{O}(n \log n)$ since $\mathcal{W}(h)$ is proportional to the sum of ranks of the positive examples.

for an extensive presentation. In the rest of the paper, ROGER employs a $(\mu + \lambda)$-ES, involving the generation of λ offspring from μ parents through uniform crossover and self-adaptive mutation, and deterministically selecting the next μ parents from the best μ parents + λ offspring.

3.2 Extensions

An extension first presented in [Jong *et al.*, 2004] concerns the use of non-linear hypotheses. Exploiting the flexibility of Evolutionary Computation, the search space $\mathcal{H}$ is set to $\mathbb{R}^d \times \mathbb{R}^d$; each hypothesis $h = (w, c)$, composed of a weight vector w and a center c, associates to x the weighted L_1-distance of x and c:

$$h(x = (x_1, ..., x_d)) = \sum_{i=1}^{d} w_i |x_i - c_i|$$

It must be noted that this representation allows ROGER for searching (a limited kind of) non linear hypotheses, by (only) doubling the size of the linear search space. Previous work has shown that non-linear ROGER significantly outperforms linear ROGER for some text mining applications [Roche *et al.*, 2004a].

A new extension, inspired from ensemble learning [Breiman, 1998], exploits the hypotheses $h_1, \ldots, h_T$ learned along T independent runs of ROGER. The aggregation of the (normalised) h_i, referred to as H, associates to each example x the median value of $\{h_1(x), \ldots, h_T(x)\}$.

4 Goals of Experiments and Experimental Setting

The goal of experiments is twofold. On one hand, the ranking efficiency of N'ROGER will be assessed and compared to that of state-of-the-art supervised learning algorithms, specifically Support Vector Machines with linear, quadratic and Gaussian kernels, using SVMTorch implementation[3] with default options. Due to space limitations, only ensemble-based non-linear ROGER, termed N'ROGER, will be considered.

On the other hand, the results provided by N'ROGER will be interpreted and discussed with respect to their intelligibility. The experimental setting is as follows. An experiment is a 5-fold stratified cross-validation process; on each fold, i) SVM learns a hypothesis h_{SVM}; ii) ROGER is launched 21 times, and the bagging of the 21 learned hypotheses constitutes the hypothesis $h_{n'Roger}$ learned by N'ROGER; iii) both hypotheses are evaluated on the fold test set and the associated ROC curve (True Positive Rate *vs* False Positive Rate) is constructed. The AUC curves are averaged over the 5 folds.

[3] http://www.idiap.ch/machine_learning.php?content=Torch/en_OldSVMTorch.txt

The overall results reported in the next section are averaged over 10 experiments (10 different splits of the dataset into 5 folds).

The ROGER parameters are as follows: $\mu = 20; \lambda = 100$; the self adaptative mutation rate is 1.; the uniform crossover rate is .6.

5 Empirical validation

After describing the datasets, this section reports on the comparative performances of the algorithms, and inspects the results actually provided by N'ROGER.

5.1 Datasets

In both domains, the data preparation step [Roche *et al.*, 2004b] allows for categorizing the word collocations depending on the grammatical tag of the words (e.g. Adjective, Noun).

A first corpus related to Molecular Biology involves 6119 paper abstracts in English (9,4 Mo) gathered from queries on Medline[4]. The 1028 Noun-Noun collocations occurring more than 4 times are labelled by the expert; the dataset includes a huge majority of relevant collocations (Table 1).

A second corpus related to Curriculum Vitae[5] involves 582 CVs in French (952 Ko). The "Frequent CV" dataset includes the 376 Noun-Adjective collocations with at least 3 occurrences (two hours labelling required), with a huge majority of relevant collocations. The "Infrequent CV" dataset includes the 2822 Noun-Adjective collocations occurring once or twice (two days labelling required), with a significantly different distribution of relevant/irrelevant collocations (Table 1). Examples of relevant *vs* irrelevant collocations are respectively *compétences informatiques* and *euros annuels*;

although both collocations make sense, only the first one conveys useful information for the management of human resources.

Collocations	# collocations	Relevant	Irrelevant
Molecular Biology	1028	90.9%	9.1%
CV, Frequent collocations	376	85.7%	14.3%
CV, Infrequent collocations	2822	56.6%	43.4%

Table 1. Relevant and irrelevant collocations.

5.2 Ranking accuracy

After the experimental setting described in section 4, Table 2 compares the average AUC achieved for N'ROGER and SVMTorch with linear, Gaussian

[4] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

[5] Courtesy of the VediorBis Foundation.

and quadratic kernels. On these domain applications, both supervised learning approaches significantly improve on the statistical criteria standalone (Table 3). Further, N'ROGER improves significantly on SVM using any kernel, excepted on the *Infrequent CV* dataset. A tentative interpretation for this result is based on the fact that this dataset is the most balanced one; SVM has some difficulties to cope with imbalanced datasets.

Corpus	N'ROGER ($\sim$ 17s/fold)	SVM ($\sim$ 1.5s/fold) Linear	Gaussian	Quadratic
Molecular Biology (MB)	0.73 ± 0.05	0.50 ± 0.08	0.46 ± 0.08	0.59 ± 0.08
Frequent CV (F-CV)	0.64 ± 0.08	0.48 ± 0.08	0.48 ± 0.08	0.50 ± 0.10
Infrequent CV (I-CV)	0.73 ± 0.01	0.72 ± 0.01	0.72 ± 0.02	0.71 ± 0.02

Table 2. Ranking accuracy (Area under the ROC curve) of learning algorithms.

Corpus	MI	MI^3	$Dice$	L	Occ_L	Ass	J	$Conv$	$SeSc$	CM	LC	$Ttest$	$Khi2$
MB	0.30	0.35	0.31	0.42	0.57	0.31	**0.59**	0.35	0.43	0.31	0.46	0.31	0.31
F-CV	0.31	0.40	0.39	0.43	**0.58**	0.32	**0.58**	0.39	0.40	0.31	0.44	0.36	0.36
I-CV	0.29	0.30	0.33	0.30	0.37	0.29	**0.50**	0.40	0.39	0.30	0.45	0.30	0.30

Table 3. Ranking accuracy (Area under the ROC curve) of statistical criteria.

A more detailed picture is provided by Fig. 1, showing the ROC curve associated to SVM, N'ROGER and the Occ_L and J measures on the *Frequent CV* dataset on a representative fold (termed *RF* in this paper). Interestingly, the major differences between N'ROGER and the other measures are seen at the beginning of the curve, i.e. they concern the top ranked collocations. Typically, a recall (True Positive Rate) of 50% is obtained for 18% false positive with N'ROGER, against 23% with Occ_L, 31% with J measures and 68% for quadratic SVM[6].

In summary, N'ROGER improves the accuracy of the top-ranked collocations, and therefore the satisfaction and productivity of the expert if he/she only examines the top results. A proof of principle of the generality of the approach has been presented in [Roche *et al.*, 2004b], as the ranking function learned from one corpus, in one language, was found to outperform the standard statistical criteria when applied on the other corpus, in another language.

5.3 Analysis of a ranking function

As shown in [Jong *et al.*, 2004], the weights associated to distinct features by ROGER can provide some insights into the relevance of the features. Accordingly, the hypotheses constructed by N'ROGER are examined.

Fig. 2 displays the weights and center coordinates of all 13 features (section 2) for a representative ROGER hypothesis h (closest to the ensemble

[6] SVM ROC Curves is not significant as its AUC is lower than .5 on this test fold.

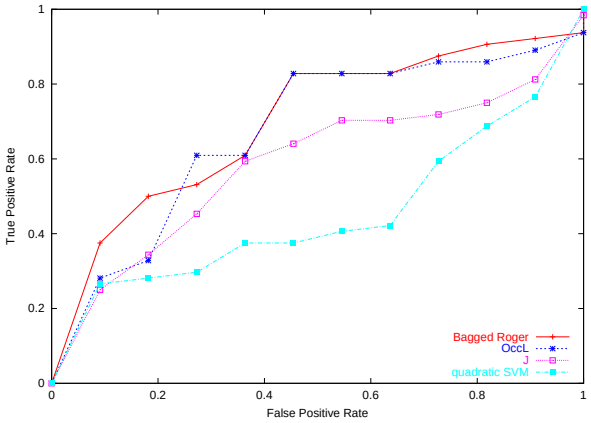

Fig. 1. ROC Curves on Frequent Collocations of CV corpus (for the test set of RF).

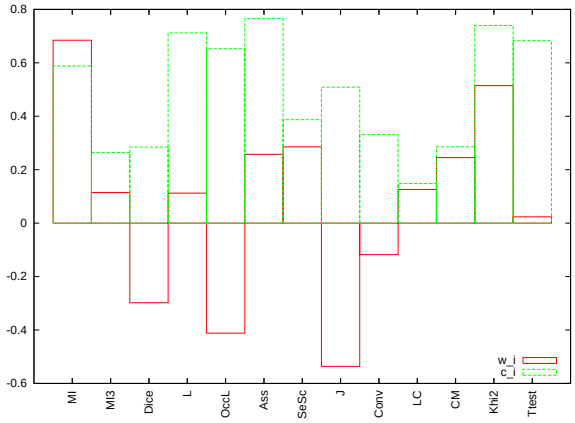

Fig. 2. Weights (w_j, c_j) on the frequent CVs (for the learn set of RF).

N'ROGER hypothesis H) learned on a fold of the *Frequent CV* dataset. Although AUC(h) is lower than that of H (.61 *vs* .64), it still outpasses that of standalone features (statistical criteria).

As could have been expected, ROGER detects that the mutual information (MI) criterion does badly (AUC(MI)= .31, Table 3), with a high center c_{MI} and weight w_{MI} values (collocations with high MI are less relevant, everything else being equal). Inversely, as the Occ_L criterion does well (AUC(Occ_L) = .58), the center c_{Occ_L} is high associated with a highly negative weight w_{Occ_L} (collocations with low Occ_L are less relevant, everything else being equal) (see Tab. 4).

Although these tendencies could have been exploited by linear hypotheses, this is no longer the case for the J criterion (AUC(J) = .58): interestingly,

the center c_J takes on a medium value, with a high negative weight w_J. This is interpreted as collocations with either too low *or too high* values of J, are less relevant everything else being equal. The current limitation of the approach is to provide a "conjunctive" description of the region of relevant collocations[7].

	MI w_{MI} = **0.68** c_{MI} = **0.59**	Occ_L w_{Occ_L} = **-0.41** c_{Occ_L} = **0.65**	N'ROGER
Collocation	**Rank**	**Rank**	**Rank**
expérience commerciale	297	258	1
formation informatique	300	123	2
société informatique	298	299	3
gestion informatique	299	76	4
colonne morris	1	211	90
bouygue telecom	2	213	298
fromagerie riches-mont	3	212	297
sauveteur secouriste	4	151	296
expérience professionelle	146	1	300
ressource humaine	44	2	299
baccalauréat professionel	193	3	22
baccalauréat scientifique	148	4	58

Table 4. Rank of relevant collocations given with 2 measures (MI and Occ_L) and N'ROGER. For each measure the weights (w_i, c_i) used by N'ROGER are given (on the learn set of RF).

6 Discussion and Perspectives

The main claim of the paper is that supervised learning can significantly contribute to the Term Extraction task in Text Mining. Some empirical evidence supporting this claim have been presented, related to two corpora with different domain applications and languages. Based on a domain- and language-independent description of the collocations along a set of standard statistical criteria, and on a few collocations manually labelled as relevant/irrelevant by the expert, a ranking hypothesis is learned.

The ranking learner N'ROGER used in the experiments is based on the optimization of the combinatorial Wilcoxon rank test criterion, using an evolutionary computation algorithm. Two new features, the use of non-linear hypotheses and the exploitation of the ensemble of hypotheses learned along independent runs of ROGER, have been exploited in N'ROGER.

Further research is concerned with enriching the description of collocations, e.g. adding typography-related indications (e.g. distance to the closest typographic signs) or distance to the closest Noun, possibly providing additional cues on the role of relevant collocations. Another perspective is to

[7] In the sense that a single center c is considered, though the condition *far from* c_i actually corresponds to a disjunction.

extend ROGER using multi-modal and multi-objective evolutionary optimization [Deb, 2001], e.g. enabling to characterize several types of relevant collocations in a single run. A long-term goal is to study along a variety of domain applications and expert goals, the eventual regularities associated to i) the (domain and language independent) description of the relevant collocations; ii) the ranking hypotheses.

Acknowledgment: We thank Oriane Matte-Tailliez for her expertise and labelling of the Molecular Biology dataset and Mary Felkin who did her best to improve the readability of this paper. The authors are partially supported by the PASCAL Network of Excellence, IST-2002-506778.

References

[Azé and Kodratoff, 2004]J. Azé and Y. Kodratoff. Extraction de "pépites" de connaissance dans les données : une nouvelle approche et une étude de la sensibilité au bruit. *Revue RNTI*, E-1:247–270, 2004.

[Bäck, 1995]T. Bäck. *Evolutionary Algorithms in theory and practice.* New-York:Oxford University Press, 1995.

[Bourigault and Jacquemin, 1999]D. Bourigault and C. Jacquemin. Term extraction + term clustering: An integrated platform for computer-aided terminology. In *Proc. of EACL'99, Bergen.*, pages 15–22, 1999.

[Bradley, 1997]A.P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[Breiman, 1998]L. Breiman. Arcing classifiers. *Annals of Statistics*, 26(3):801–845, 1998.

[Brin *et al.*, 1997]S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: generalizing association rules to correlations. In *Proc. of ACM SIGMOD'97*, pages 265–276, 1997.

[Caruana and Niculescu-Mizil, 2004]R. Caruana and A. Niculescu-Mizil. Data mining in metric space: An empirical analysis of supervised learning performance criteria. In *Proc. of "ROC Analysis in AI" Workshop (ECAI)*, pages 9–18, 2004.

[Church and Hanks, 1990]K.W. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16:22–29, 1990.

[Cohen *et al.*, 1999]W. Cohen, R. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270, 1999.

[Daille *et al.*, 1998]B. Daille, E. Gaussier, and J.M. Langé. An evaluation of statistical scores for word association. In *Proc. of The Tbilisi Symposium on Logic, Language and Computation, CSLI Publications*, pages 177–188, 1998.

[Deb, 2001]K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, Chichester, 2001.

[Domingos, 1999]P. Domingos. Meta-cost: A general method for making classifiers cost sensitive. In *Knowledge Discovery from Databases*, pages 155–164, 1999.

[Dunning, 1993]T. E. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, 1993.

[Ferri *et al.*, 2002]C. Ferri, P. Flach, and J. Hernandez-Orallo. Learning decision trees using the area under the ROC curve. In *Proc. of ICML'02*, pages 139–146, 2002.

[Fogel *et al.*, 1995]D.B. Fogel, E.C. Wasson, and E.M. Boughton. Evolving neural networks for detecting breast cancer. *Cancer Letters*, 96:49–53, 1995.

[Goodman and Smyth, 1988]M.F.R. Goodman and P. Smyth. Information-theoretic rule induction. In *Proc. of ECAI'88*, pages 357–362, 1988.

[Jacquemin, 1997]C. Jacquemin. Variation terminologique : Reconnaissance et acquisition automatiques de termes et de leurs variantes en corpus. In *Mémoire d'Habilitation à Diriger des Recherches, Université de Nantes*, 1997.

[Jin *et al.*, 2003]R. Jin, Y. Liu, L. Si, J. Carbonell, and A. Hauptmann. A New Boosting Algorithm Using Input-Dependent Regularizer. In *ICML 2003*. AAAI Press, 2003.

[Jong *et al.*, 2004]K. Jong, J. Mary, A. Cornuéjols, E. Marchiori, and M. Sebag. Ensemble feature ranking. *Proc. of PKDD-2004*, pages 20–24, 2004.

[Kolcz *et al.*, 2003]A. Kolcz, A. Chowdhury, and J. Alspector. Data duplication: An imbalance problem ? In *Workshop on Learning from Imbalanced Data Sets II (ICML)*, 2003.

[Lallich and Teytaud, 2004]S. Lallich and O. Teytaud. évaluation et validation de l'intérêt des règles d'association. *Revue RNTI*, E-1:193–217, 2004.

[Manning and Schütze, 1999]C. Manning and H. Schütze. *Collocations*, pages 165–184. Cambridge, MA: MIT Press, 1999.

[Roche *et al.*, 2004a]M. Roche, J. Azé, Y. Kodratoff, and M. Sebag. Learning interestingness measures in terminology extraction. a roc-based approach. In *Proc. of "ROC Analysis in AI" Workshop (ECAI)*, pages 81–88, 2004.

[Roche *et al.*, 2004b]M. Roche, J. Azé, O. Matte-Tailliez, and Y. Kodratoff. Mining texts by association rules discovery in a technical corpus. In *Proc. of IIPWM'04, Springer Verlag*, pages 89–98, 2004.

[Rosset, 2004]S. Rosset. Model Selection via the AUC. In *Proc. of the Twenty-First International Conference on Machine Learning (ICML'04)*, 2004.

[Schapire, 1999]R.E. Schapire. Theoretical views of boosting. In *Proc. of EuroCOLT-99*, pages 1–10, 1999.

[Sebag and Schoenauer, 1988]M. Sebag and M. Schoenauer. Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In *Proc. of EKAW'88*, 1988.

[Sebag *et al.*, 2003a]M. Sebag, J. Azé, and N. Lucas. Impact studies and sensitivity analysis in medical data mining with ROC-based genetic learning. In *Proc. of ICDM 2003*, pages 637–640, 2003.

[Sebag *et al.*, 2003b]M. Sebag, N. Lucas, and J. Azé. ROC-based Evolutionary Learning: Application to Medical Data Mining. In *Proc. of EA 2003*, pages 384–396, 2003.

[Smadja *et al.*, 1996]F. Smadja, K. R. McKeown, and V. Hatzivassiloglou. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics*, 22(1):1–38, 1996.

[Smadja, 1993]F. Smadja. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1):143–177, 1993.

[Vivaldi *et al.*, 2001]J. Vivaldi, L. Màrquez, and H. Rodríguez. Improving term extraction by system combination using boosting. *Proc. of ECML*, 2167:515–526, 2001.

[Xu *et al.*, 2002]F. Xu, D. Kurz, J. Piskorski, and S. Schmeier. A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and their Relations with Bootstrapping. In *Proc. of LREC*, 2002.