View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Biblioteca Digital da Produção Intelectual da Universidade de São Paulo (BDPI/USP)

A comparison of the effect of feature selection and balancing strategies upon the sentiment classification of Portuguese News Stories.

Brett Drury
ICMC
Universidade de São Paulo
São Carlos, Brazil
Email: brett.drury@gmail.com

Alneu de Andrade Lopes
ICMC
Universidade de São Paulo
São Carlos, Brazil
Email: alneu@icmc.usp.br

***Abstract*—Sentiment classification of news stories using supervised learning is a mature task in the field of Natural Language Processing. Supervised learning strategies rely upon training data to induce a classifier. Training data can be imbalanced, with typically the neutral class being the majority class. This imbalance can bias the induced classifier towards the majority class. Balancing and feature selection can mitigate the effects of imbalanced data. This paper surveys a number of common balancing and feature selections techniques, and applies them to an imbalanced data set of manually labelled Brazilian agricultural news stories. The strategies were appraised with a 90:10 holdout evaluation and compared with a baseline strategy. We found that: 1. the feature selection strategies provided no identifiable advantage over a baseline method and 2. balancing produced an advantage over baseline with random oversampling producing the best results.**

I. Introduction

The sentiment classification of news stories using supervised learning strategies is a popular area of research which has recently produced a large number of research articles. Models produced with supervised learning techniques are influenced by the characteristics of the training data. A common characteristic of labelled data for sentiment analysis is that it can be imbalanced, i.e. the labelled classes have differing numbers of training examples. Typically the neutral class has a much larger number of training examples. A classifier which is biased towards a neutral majority class may impede the detection of ”interesting” stories which are typically contained within the ”positive”, ”negative” or related classes.

The research literature reveals that there are two common techniques which can be employed to mitigate the effects of imbalanced data. These techniques are: 1.feature selection and 2. balancing.

This paper evaluates five simple feature selection techniques which represent documents as: 1. bigrams, 2. trigrams, 3. nouns, 4. verbs and 5. adjectives and 5 statistical association feature selection techniques. These feature selection techniques were: 1. Chi-square, 2. Term Frequency - Inverse Document Frequency (TF-IDF), 3. Mutual Information (MI), 4. Probability Proportion Difference (PPD) and 5. Categorical Proportion Difference (CPD). The 5 balancing strategies evaluated were: 1. random over-sampling, 2. random over-sampling with artificial data generation , 3. random under-sampling, 4. majority class reduction with Tomek Links and 5. majority class reduction with Tomek Links combined with random oversampling with artificial data generation.

The remainder of this paper will present the following: 1. related work, 2. experiments with feature selection, 3. experiments with balancing and 4. conclusion and future work.

II. Related Work

The related work will cover: 1. sentiment classification 2. feature selection, 3. feature selection applied to sentiment classification and 4. balancing.

Sentiment analysis is the identification of opinions in text [12]. The classification of documents into sentiment categories is one method of sentiment identification [11]. A common strategy used to undertake sentiment classification is supervised learning. Supervised learning uses labelled data, i.e. data that has had a category, typically: positive, neutral or negative, assigned to it by an oracle. This data is used to induce a model from a classifier. This model is then used to classify unlabelled documents into the aforementioned predefined categories. The seminal paper which described the use of a supervised learning strategy for a sentiment classification task was produced by [16]. They classified movie reviews with three types of classifiers. They found that this method outperformed a human baseline. A similar approach was proposed by [21] who applied a supervised machine learning approach to classifying on-line reviews. Supervised sentiment classification has been used to classify Cantonese texts into sentiment categories [23].

There are a number of papers which describe strategies which classify Portuguese texts into sentiment categories. [9] classified Brazilian news into sentiment categories to support a stock trading strategy. [14] classified on-line Portuguese news. The Popstar tool[1] can analyze on-line Portuguese news as well as social media messages for sentiment content. The developers of Popstar use it to track the popularity of specific Portuguese politicians.

There is some evidence that feature selection (FS) can mitigate the effects of imbalanced data [24]. There is also evidence that feature selection is a suitable method for mitigating

[1] www.popstar.pt

the effects of imbalanced data for sentiment classification [4]. [10] evaluated 6 FS strategies for their suitability for a text classification task. The strategies evaluated were: 1. Document Frequency, 2. Mutual Information, 3. Information Gain, 4. CHI, 5. Bi-Normal Separation and 6. Weighted Log Likelihood Ratio. They were compared with a new technique, "weighed frequency and odds" (WFO). WFO was found to be the superior method. A similar study by [15] concluded that Document Frequency Difference produced the superior results. Finally, [18] conducted a survey of feature selection and found that Gain Ratio was the most effective feature selection for sentiment classification.

Balancing can mitigate the effects of imbalanced data. Balancing strategies attempt to equalize the numbers of training examples in each class. [5] states that there are three main types of balancing strategies: 1. over-sampling, 2. under-sampling and 3. majority class reduction. Oversampling is a strategy which balances the number of training examples by increasing the training candidates in the minority classes. Under-sampling balances the number of training candidates by removing training candidates from the majority classes. Majority class reduction reduces the number of training candidates in the majority class. A comprehensive survey of balancing techniques is provided by [5].

III. Experiments

The experiments for this paper evaluated a number of common : 1. feature selection (FS) and 2. balancing strategies against a baseline strategy which used neither balancing or FE.

A. Evaluation Set Construction

The data set for these experiments contained 500 news stories which were written in Brazilian Portuguese. These news stories where gathered from the Internet, and formed part of a larger collection. The 500 stories were randomly chosen from the larger collection.

The news stories were manually sorted into sentiment categories (neutral, negative and positive) by a single independent annotator who is a native speaker of Brazilian Portuguese. There were no tacit annotation guidelines, and therefore the annotator used his intuition in the annotation process. The annotator reported there were a small number of edge cases where the sentiment of a news story was unclear. The sentences in the news story were tagged with part of speech (POS) tags with the Aelius POS Tagger [2].

The evaluation set had one large majority class (neutral) which held 293 documents (58.60%) and two smaller classes, positive, which held 73 documents (14.50%) and negative, which held 134 documents (26.80%). The composition of the evaluation set was akin to similar training sets reported in the research literature.

B. Evaluation Methodology

The experiments for both the feature selection and balancing strategies used a 10 X 90:10 holdout evaluation. A 10 X 90:10 holdout evaluation reserves 90% of the data set to train a classifier (training set), and uses 10% for evaluating the induced model (test set). An accuracy figure is derived by: 1. counting the number of correct classifications and dividing the number of correct classifications by the total number documents in the test set. This process is repeated 10 times and an average accuracy figure is calculated. Multiple runs are made to reduce the bias in any single sample.

The classifier used in these experiments was a Naive Bayes (NB) classifier from the Python NLTK [13] library. A NB classifier was used because it is a common and simple generative classifier.

C. Simple feature selection Experiments

These set of experiments evaluated the effect of extracting features based upon POS tag information. POS tag information identifies the category of a word, for example, verb, noun, adjective, etc. The assumption behind these experiments was that certain groups of words, such as adjectives, may convey sentiment[19] better than other groups of words. The experiments extracted: verbs, nouns and adjectives, and compared the results with a baseline which used the full text of each document in the training set. A second experiment evaluated the effect of removing frequent words (stop words) from the extracted features.

The results for the experiment is documented in Table I.The results were poor. The simple feature selection strategies marginally outperformed the full-text baseline, but the difference was within the standard deviation, which suggests that the gain was not statistically significant. The experiments suggest that removing the stop words decreased the accuracy of the classifier, but again the difference was within the standard deviation and again no strategy gave an advantage.

Feature	Avg. Accuracy	Avg. Accuracy (-SW)
Nouns	0.28$\pm$0.06	0.27$\pm$0.04
Adjectives	0.24$\pm$0.05	0.22 $\pm$0.07
Verbs	0.29$\pm$0.02	0.24$\pm$0.06
Baseline (Full Text)	0.22$\pm$0.04	0.22$\pm$0.04

TABLE I. A COMPARISON OF SIMPLE FEATURE SELECTION STRATEGIES. SW=STOPWORDS

D. NGram feature selection Experiments

The NGram feature selection experiments extracted sequences of unigrams. Unigrams have no context whereas sequences of unigrams retain some context. For example, the bigram "not good" when split into unigrams: "not" and "good" changes its sentiment orientation.

These experiments evaluated two common forms of NGrams: bigrams and trigrams. A bigram is a sequence of two unigrams and a trigram is a sequence of three unigrams. We conducted two iterations of the experiments, one which included stopwords and one that did not. The FS strategies were compared with a full text baseline. The results are displayed in Table II.

In common with the previous experiment the results were poor. The bigram and trigram features outperformed the baseline, but the difference was within the standard deviation, and therefore in common with the previous experiment, no strategy gave an advantage. The removal of stop words did not effect the performance of the classifier.

Feature	Mean Accuracy	Mean Accuracy (-SW)
Bigrams	0.25 ± 0.05	0.24 ± 0.06
Trigrams	0.24 ± 0.04	0.25 ± 0.05
Baseline (Full Text)	0.22 ± 0.04	0.22 ± 0.04

TABLE II. A COMPARISON OF BIGRAM AND TRIGRAM EXTRACTION STRATEGIES. SW=STOPWORDS

E. Feature Selection with Statistical Association Measures Experiments

These experiments selected features with using simple statistical association measures. The statistical measures score the association of features to a specific class. We selected three common general feature selection techniques: Chi Squared[6], 2. TF-IDF[8], 3. Mutual Information[17] and two techniques which were specially designed for sentiment analysis: 1. Probability Proportion Difference (PPD)[1] and 2. Categorical Proportion Difference (CPD)[1].

We did not exclude stop-words because stop words will be frequent and be present in all classes and therefore should not be selected by any feature selection technique.

Chi Square Feature Selection: Chi Square feature selection tested a feature's association with a class against the feature's association with the remaining classes. A chi-square value was computed for a feature and its association with each class.

The equation we used to calculate the "class association value" for a feature is described in Equation 1 where f = feature, (fa,c_1) is the feature association of feature f with class c_1, c_1 = class 1, c_2 = class 2 and c_1 = class 3.

$$(fa, c_1) = \chi(f, c_1) - (\chi(f, c_2) + \chi(f, c_3)) \tag{1}$$

The feature with the highest value for a specific class was selected as a feature for that class, consequently a feature will be unique to a class.

Term Frequency Inverse Document Frequency (TF-IDF) Feature Selection: The TF-IDF feature selection applies a information selection technique to weight features. In this experiment we use TF-IDF to select features by computing a frequency of a feature within a class and then comparing it with the term's Inverse Document Frequency (IDF) for the whole text collection.

The equation we used for calculating a TF-IDF score for a feature in a specific class is described in Equation 2 where (fa,c_1) is the feature association of feature f with class c_1, tf is the term frequency of a feature f in class c_1, idf is the inverse document frequency of a feature f in all of the classes present in the labelled data D.

$$(fa, c_1) = tf(f, c_1)idf(\dot{f}, D) \tag{2}$$

The experiments used a range of minimum TF-IDF values $> 0.5 <= 6.0$ which were incremented by 0.5 in each subsequent iteration. The value selected for the iteration preclude features which have a TF-IDF value less than the selected minimum.

Mutual Information feature selector: The mutual information feature selector uses mutual information to compute a score a feature's association with a specific class. The feature extractor in our experiment used Equation 3 where f = feature, (fa,c_1) is the feature association of feature f with class c_1, c_1 = class 1, c_2 = class 2, c_1 = class 3 and mi = mutual information.

$$(fa, c_1) = mi(f, c_1) - (mi(f, c_2) + mi(f, c_3)) \tag{3}$$

The feature is selected as a feature for a class if it had a higher mutual information score (MIS) than its MIS for the remaining classes.

Probability Proportion Difference (PPD): PPD is a feature selection strategy which was designed for sentiment classification. In the original paper the authors used the strategy for a two class (positive and negative) classification problem. For this paper we adapted PPD for a three class classification problem as per Equation 4 where: N is the number documents in class denoted in its subscript where feature F is present, W is the total number of features in a given class,F represents is a feature and (fa,c_x) is the feature association of feature f with a given class.

$$(fa, c_1) = \frac{N_{c_1}}{W_{c_1} + F} - (\frac{N_{c_2}}{W_{c_2} + F} + \frac{N_{c_3}}{W_{c_3} + F}) \tag{4}$$

The PPD feature selection allows the selection of N top features, for our experiments we use used a range of $>= 100 <= 600$ features per class in increments of 100.

Categorical Proportion Difference (CPD)

The last feature selection strategy was Categorical Proportion Difference (CPD). The CPD calculates a value for a feature based upon its frequency in a class. CPD was originally designed for a two class classification problem and we have adapted it for a three class classification problem. The CPD feature selection strategy we used followed Equation 5 where (fa,c_1) is the feature association of feature f with the class in the subscript, Nf is the frequency of a feature within the class denoted in the subscript.

$$(fa, c_1) = \frac{Nf_{c_1} - (Nf_{c_2} + Nf_{c_3})}{Nf_{c_1} + Nf_{c_2} + Nf_{c_3}} \tag{5}$$

The CPD features for each class were extracted subject to a minimum score. The minimum score was selected from a range $>= 0.1 <= 1.0$ in increments of 0.1. A feature which had a score lower than the selected minimum for a class was not included as a feature for that class.

Experiments: The experiments for the statistical FS strategies followed the same methodology as the previous experiments and were compared with a baseline of no feature selection (full text). The results are shown in Table III.

The results from the experiments show that there were no significant difference between any of the feature selection strategies. The strategies designed for sentiment analysis did not provide any significant advantage over baseline or the general feature selection strategies.

Feature Selection Strategy	Mean Accuracy
Chi-Squared	0.29 ±0.07
TF-IDF	0.18 ±0.04 - 0.23 ±0.05
Mutual Information	0.25 ±0.05
PPD	0.22 ±0.06 - 0.23 ±0.06
CPD	0.21 ±0.04 - 0.23 ±0.01
Baseline (Full Text)	0.22±0.04

TABLE III. A COMPARISON OF FEATURE SELECTION .

F. Balancing Experiments with POS Tag Feature Selection

The results for the feature selection experiments did not provide any advantage over a baseline strategy. It would be reasonable to assume that for the evaluation set FS strategies were inadequate. We therefore evaluated balancing strategies to see if they produced better results than FS method for mitigating the effects of imbalanced data.

We evaluated the following balancing strategies: 1. random oversampling, 2. random oversampling with artificial data generation, 3. random undersampling, 4. majority class reduction with Tomek Links[20] and 5. combination of majority class reduction with random oversampling

Random Oversampling: Random oversampling selects random training examples from minority classes and either: 1. duplicates the selected training example or 2. generates an artificial training example with data related to the selected training example. This process continues until the minority classes have the same number of training examples as the majority class[5].

We evaluated two types of artificial data generation: 1. synonyms and 2. word forms. The "synonym approach" randomly selects a training example and splits the text into unigrams. A series of synonyms for each unigram is collected from Onto.Pt[7]. From the series of synonyms one is chosen at random to replace the original unigram. If there are no synonyms for the unigram then the original unigram is retained. The document is then added to the minority class. The "word form" approach is similar to synonym approach except that it uses JSpell[3] to return related word forms to the original unigram rather than synonyms.

Random Undersampling: Random undersampling selects random training examples from the majority classes and removes them. This process continues until all classes have the same number of training examples[5].

Majority class Reduction: Majority class reduction seeks to remove training examples from the majority class which are similar to training examples in the minority class. These training examples are assumed to be at the classification borderline and may inhibit the accuracy of a classifier[5].

Our majority class reduction technique used Tomek Links[20] to identify these borderline training examples. To identify Tomek Links we used a KNN cluster technique which used Levenshtein Distance[22] as a distance measure. If a nearest neighbour of a minority class training example was in the majority class then a Tomek Link was established. The majority class training example was removed. This process was continued until there were no Tomek Links detected.

Combination of majority class reduction with random oversampling: The final balancing strategy used a combination of majority class reduction with random oversampling. The random oversampling variant chosen was the "synonym approach". The "synonym approach" was chosen based upon the experimental results for the competing oversampling techniques.

The hybrid balancing strategy uses the majority class reduction strategy to reduce the majority class, and the random oversampling with replacement (synonyms) to balance the minority classes with the majority class.

The experiments where conducted with the feature selection methods described in section III-C. The results for each balancing strategy is documented in Tables IV and V.

FS.	R.OS	R.OS WF	R.OS Syn.
Trigrams	0.58 ±0.01	0.61±0.06	0.61±0.07
Bigrams	0.60 ±0.04	0.62±0.07	0.61±0.06
Nouns	0.59 ±0.08	0.60 ±0.06	0.60 ±0.06
Verbs	0.53 ±0.07	0.60 ±0.10	0.61 ±0.06
Adjectives	0.37 ±0.05	0.58 ±0.05	0.58 ±0.06
Full Text	0.59 ±0.06	0.61±0.05	0.62±0.05

TABLE IV. RESULTS FOR RANDOM OVERSAMPLING (R.OS) BALANCING TECHNIQUES WITH FEATURE SELECTION (FS) TECHNIQUES. SYN = SYNONYMS, WF = WORD FORMS.

FS.	R.US	TL.	R.OS Syn.TL.
Trigrams	0.34±0.07	0.25±0.03	0.59±0.10
Bigrams	0.36±0.06	0.23±0.05	0.61±0.08
Nouns	0.36±0.06	0.28±0.06	0.56±0.07
Verbs	0.40±0.07	0.28±0.04	0.58±0.06
Adjectives	0.26±0.07	0.22±0.04	0.56±0.06
Full Text	0.29 ±0.07	0.23 ±0.05	0.58±0.05

TABLE V. RESULTS FOR RANDOM UNDERSAMPLING (R.US), MAJORITY CLASS REDUCTION WITH TOMEK LINKS (TL), AND A HYBRID METHOD (R.OS SYN.TL.) BALANCING TECHNIQUES WITH FEATURE SELECTION (FS) TECHNIQUES. SYN = SYNONYMS.

The results demonstrate that the various oversampling techniques produced superior results to the undersampling and the majority class reduction techniques. The oversampling with artificial data generation produced marginally better results than the random oversampling. The gain was within the standard deviation and therefore we can't assume that these methods are superior. There was little difference between the feature selection techniques within each balancing technique. There was one exception, which was adjective feature selection in the random oversampling balancing technique. It performed significantly worse than the other feature selection techniques.

G. Balancing Experiments with Statistical Association Measures Feature Selection

We repeated the experiments with the balancing techniques described in section III-F with the statistical association feature selection techniques described in section III-E, we found that there was no statistical difference between the results described in section III-E and the results generated with balancing and feature selection.

IV. CONCLUSION

The experiments demonstrate that the evaluated feature selection techniques did not improve the accuracy of the classifier. The balancing techniques improved the classifier's accuracy, but the undersampling and the majority class reduction techniques produced inferior results to the various oversampling techniques. The random oversampling with artificial data generation produced marginally better results than the basic random oversampling.

A. Future Work

The results with simple artificial data generation suggest that this may be a possible research route. We will conduct experiments with SMOTE and its variants to evaluate their effectiveness for mitigating the effects of imbalanced training data.

We are releasing to the community the labelled data used in these experiments to encourage sentiment classification of Portuguese. The data can be obtained from http://goo.gl/74uEjN.

B. Acknowledgments

This work was supported by FAPESP grant 11/20451-1.

References

[1] Basant Agarwal and Namita Mittal. Categorical probability proportion difference (cppd): A feature selection method for sentiment classification. In *Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology*, pages 17–26, Mumbai, India, December 2012. The COLING 2012 Organizing Committee.

[2] L. F. de. ALENCAR. Novos recursos do aelius para o processamento computacional raso do português. In E.LAPORTE, A. SMARSARO, and O. VALE, editors, *Dialogar é preciso: linguística para o processamento de línguas.*, pages 7–20, 2013.

[3] José João Almeida and Alberto Simões. Jspellando nas morfolimpíadas: Sobre a participação do jspell nas morfolimpíadas. In Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa.* IST Press, 2007.

[4] A. R. Balamurali, Aditya Joshi, and Pushpak Bhattacharyya. Harnessing wordnet senses for supervised sentiment classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1081–1091, 2011.

[5] Gustavo E. A. P. A. Batista, Ronaldo C. Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explorations*, 6(1):20–29, 2004.

[6] C.S. Dhir, N. Iqbal, and Soo-Young Lee. Efficient feature selection based on information gain criterion for face recognition. In *Information Acquisition, 2007. ICIA '07. International Conference on*, pages 523–527, July 2007.

[7] Hugo Gonçalo Oliveira. *Onto.PT: Towards the Automatic Construction of a Lexical Ontology for Portuguese*. PhD thesis, University of Coimbra, 2013.

[8] Li-Ping Jing, Hou kuan Huang, and Hong-Bo Shi. Improved feature selection approach tfidf in text mining. In *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on*, volume 2, pages 944–946 vol.2, 2002.

[9] Eduardo Motta Leandro Alvim, Paula Vilela. Sentiment of financial news: a natural language processing approach. In *1st Workshop on Natural Language Processing Tools Applied to Discourse Analysis in Psychology.*, 2010.

[10] Shoushan Li, Rui Xia, Chengqing Zong, and Chu-Ren Huang. A framework of feature selection methods for text categorization. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 692–700, 2009.

[11] Bing Liu. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[12] Bing Liu. Sentiment analysis and subjectivity. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL, 2010. ISBN 978-1420085921.

[13] Edward Loper and Steven Bird. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, pages 63–70, 2002.

[14] Inês Coimbra Morgado. Classification of sentiment polarity of portuguese on-line news. In *Proceedings of the 7th Doctoral Symposium in Informatics Engineering*, pages 139–150, 2012.

[15] Chris Nicholls and Fei Song. Comparison of feature selection methods for sentiment analysis. In *Proceedings of the 23rd Canadian Conference on Advances in Artificial Intelligence*, AI'10, pages 286–289, 2010.

[16] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, 2002.

[17] H. Peng, Fulmi Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, Aug 2005.

[18] Anuj Sharma and Shubhamoy Dey. A comparative study of feature selection and machine learning techniques for sentiment analysis. In *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, RACS '12, pages 1–7, 2012.

[19] Tom De Smedt and Walter Daelemans. A subjectivity lexicon of dutch adjectives. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *LREC*, pages 3568–3572. European Language Resources Association (ELRA), 2012.

[20] I. Tomek. Two Modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(2):679–772, 1976.

[21] Qiang Ye, Ziqiong Zhang, and Rob Law. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Syst. Appl.*, 36(3):6527–6535, April 2009.

[22] Li Yujian and Liu Bo. A Normalized Levenshtein Distance Metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1091–1095, June 2007.

[23] Ziqiong Zhang, Qiang Ye, Yijun Li, and Rob Law. Sentiment classification of online cantonese reviews by supervised machine learning approaches. *Int. J. Web Eng. Technol.*, 5(4):382–397, March 2009.

[24] Zhaohui Zheng, Xiaoyun Wu, and Rohini Srihari. Feature selection for text categorization on imbalanced data. *SIGKDD Explor. Newsl.*, 6(1):80–89, June 2004.