IJCSN International Journal of Computer Science and Network, Volume 5, Issue 2, April 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Pre-processing Techniques in Sentiment Analysis through FRN: A Review

[1] Ashwini M. Baikerikar, [2] P. C. Bhaskar

[1] Department of Computer Science and Technology, Department of Technology, Shivaji University, Kolhapur, Maharashtra, India.

[2] Department of Electronics and Communication Technology, Department of Technology, Shivaji University, Kolhapur, Maharashtra, India.

Abstract - The objective of the paper is to demonstrate the viability of analyzing online data. It displays a framework which after effects pattern investigation that will be shown as results with various segments introducing positive, negative and neutral. It is challenging task to summarize opinion about the products due to diversity and size. Mining online opinion mining is a difficult text classification task of sentiment analysis. Multivariate content technique called Feature Relation Network that considers semantic data, influencing the syntactic connections between n-gram features. FRN empowers the consideration of heterogeneous n-gram features for improved opinion classification, by joining syntactic data about n-gram relations. FRN selects the features in a more computationally effective way than numerous multivariate and hybrid methods. Appropriate feature selection and representation with sentiment analysis, accuracies using support vector mechanism sentiment analysis; the task of text pre-processing is to be explored.

Keywords - *Sentiment Analysis, Text pre-processing, Feature Relation Network (FRN), Support Vector Machine (SVM).*

1. Introduction

The Internet is rich in directional content (i.e., text containing opinions and emotions). The web gives volumes of content based information about shopper inclinations, put away in online audit sites, web forums, websites, and so forth. Sentiment analysis has emerged as a method for mining opinions from such text archives. It utilizes machine learning techniques joined with linguistic attributes/features keeping in mind the end goal to recognize in addition to other things the feeling extremity (e.g., positive, negative, and neutral) and polarity (e.g., low, medium, and high) for a specific content.

Sentiment analysis in reviews is the process of exploring reviews on the internet to determine the overall opinion about a product. Reviews represent the user-generated content, and this is to enhance the attention and a rich resource for marketing peoples, sociologists, psychologists and others who might be concerned with opinions, views, public moods and general or personal opinions [1].The number of reviews on the web represents the user's feedback. It is hard for humans or companies to summarize the state or general opinions about products due to the big diversity and size of social media data. This leads to automated and real time opinion extraction and mining. To decide about the sentiment of opinion is a challenge due to the subjectivity issue which is required of what people think. Sentiment analysis is considered as a classification problem as it classifies the orientation of a text into two things, either positive or negative.

Machine learning is one of technique that is widely used approach towards sentiment classification which leads to lexicon based methods and linguistic methods [2]. It has been stated that these methods do not perform as well in sentiment classification as done in topic categorization as the nature of an opinionated text requires more understanding of the text while the repetition of some keywords could be the key for a specific classification [2]. Machine learning classifiers Naïve Bayes, maximum entropy and (SVM) support vector machine are used in for sentiment classification to get accuracies that range from 75% to 83%, in comparison to a 90% accuracy or higher in topic based categorization[2],[3].

SVM classifiers are utilized for assessment investigation with a few univariate and multivariate strategies for highlight choice, achieving 85-88% correctness in the wake of utilizing chi-squared for selecting the pertinent characteristics in the writings [4]. A system/network based strategy that is Feature Relation Network (FRN) enhanced the execution of the classifier to 88-90% exact, which is

the most noteworthy precision accomplished in document level for sentiment analysis to the best of our insight [4] .

2. Literature Survey

Sentiment analysis involves several important tasks, including sentiment polarity and intensity [11]. Polarity assignment is concerned to analyze whether a text has a positive, negative, or neutral semantic orientation. Sentiment intensity orients whether the text are positive/negative sentiments that depicts mild or strong. Given the two phrases "I don't like you" and "I hate you," both would be considered as a negative semantic orientation but the latter would be considered more intense. Classifying the sentiment polarities and intensities entails the use of classification methods applied to linguistic features. While many other classification methods have been used for mining, Support vector Machine (SVM) has outperformed various techniques including Naïve Bayes, Decision Trees, Winnow, etc. [8], [14], [17], [19]. The most popular class of features used for opinion mining is n-grams [134]. Various n-gram categories have attained state-of-the-art results [4], [22]. Larger n-gram feature sets use the feature selection methods to extract correct attribute subsets. Next, the two areas: n-gram features and feature selection methods used for sentiment analysis. N-gram features for Sentiment Analysis: N-gram features can be classified into two groups: fixed and variable. Fixed n-grams are perfect sequences occurring at the character or token level. Variable n-grams are extraction patterns capable of representing more cleanly linguistic phenomena.

A pattern of fixed and variable n-grams used for opinion mining, includes word, part-of-speech (POS), character, legomena, semantic and syntactic n-grams. Word n-grams involves bag-of-words (BOWs) and higher order word n-grams that are bi-grams, tri-grams etc. word n-grams have been used effectively in several studies [2]. Typically, unigrams to trigrams are used [4], [22], though 4-grams have also been employed. Word n-grams provide a feature set foundation, with other feature categories added to them [4], [13]. Feature Selection for Sentiment Analysis: Different sentiment classification studies have placed the limited emphasis on feature selection techniques, other than their benefits. Feature selection can improve classification accuracy [23], a key feature subset of sentiment discriminators, and provide greater emphasis into important class attributes. There are two types of feature selection methods [4], [24] both of which have been used in previous sentiment analysis work: univariate and multivariate.

Other Feature Selection Methods: In addition to prior sentiment feature selection methods, it is important to briefly discuss multivariate and hybrid methods used in related tasks. Principal component analysis (PCA) has been used considerably for dimensionality reduction in various text style classification problems.

Recently, number of dimensionality reduction techniques has also been applied to non-text feature selection problems. These involve conditional mutual information (CMIM), geometric mean, harmonic mean, general averaged divergence analysis, and discriminative locality alignment (DLA) [23], [24], [25], [26]. CMIM outperformed comparison techniques (including DTM) on image classification and biomedical prediction tasks [25]. DLA outperformed methods such as PCA and linear discriminant analysis on image classification tasks [26]. Hybrid methods that combine univariate measures with multivariate selection strategies can potentially improve the accuracy and convergence efficiency of otherwise slower multivariate methods. For instance, a hybrid GA utilizing the IG measure has been shown to converge faster than regular GA, when applied to feature sets spanning up to 26,000 features [22].

3. Methodology

Sentiment analysis that deals with different levels of the analyzed texts, including word or phrase [5], sentence [6], [7] and the document level [4],[8], in addition to some studies that are carried out on a user level [9],[10]. Word level sentiment analysis look into the orientation of the words or phrases in the text representing their effect on the overall sentiment, while sentence level look at sentences which express a single opinion and define its orientation. The document level opinion mining look at the overall sentiment of the whole document, and the user level sentiment searches for the possibility that associated users on the social network could have the same opinion [10].

There exist three approaches towards sentiment analysis that are machine learning based methods, lexicon based methods and linguistic analysis [2]. Machine learning techniques are based on training an algorithm, mostly classification done on a set of selected features for a specific task and then test on another set whether it is able to detect the correct features and give the correct classification. A lexicon based method depends on a predefined list or corpus of words having a certain polarity. An algorithm then searches for those words, counting them or estimating their weight and measuring the overall polarity of the text [2], [6], [7], [11]. Lastly the linguistic approach uses the syntactic structures of the words or phrases, the negation, and the pattern of the text to determine the text orientation. This approach is usually combined with a lexicon based method [7], [2].

3.1 Pre-processing

Pre-processing the data is the process of data cleaning and processing the text for classification.Online text data contains usually lots of noise and uninformative parts such as HTML tags, scripts, advertisements and other aspects. In addition, on words level, many words in the text do not have an impact on the general orientation of it. The dimensionality of the problem is high while keeping those words and hence making the classification harder as each word in the text is treated as one dimension. The hypothesis of having the data correctly pre-processed, to reduce the noise in the text so as to help to improve the performance of the classifier and fasten the classification process, thus aiding in real time sentiment analysis. The steps involves here are online data cleaning, blank space removal, abbreviation expansion, stemming, stopwords removal, handling the negation text and lastly feature selection called transformations, and the last step applying some functions to select the required patterns is called filtering [13].

Features in the context of opinion mining are the words, terms or phrases that extremely express the opinion as positive, negative or neutral. This means that they have a high impact on the orientation of the text than other words in the similar text. There are many other methods that are used in feature selection, where some are syntactic, where position of syntactic words such as adjectives, and some are univariate, based on each features relation to a unique category such as chi squared x^2.

There are many ways to assess the importance of each feature by assigning a certain weight in the text. The most popular ones are Feature Frequency (FF), Term Frequency Inverse Document Frequency (TF-IDF), and Feature Presence (FP). FF is the number of occurrences in the document. TF-IDF is given by

$$TF - IDF = FF * \log (N/DF) \qquad (1)$$

where N indicates the number of documents, and DF is the number of documents that contains this feature [14]. FP takes the value 0 or 1 based on the feature absent or presence in the document.

3.2 Support Vector Machine

SVM has become a famous method of classification and regression for linear and nonlinear problems [14], [16]. This method tries to find the optimal linear separation between the data with a maximum margin that allows positive values above the margin and negative values below it. This problem is described as a "quadratic programming optimization problem" [17].

Let $\{(x_{11}, y_1),(x_{12},y_2),\ldots.(m_n,y_m)\}$ denote the set of training data, where x_{ij} denotes the occurrences of the events j in time i, and $y_i \in \{-1, 1\}$. A support vector machine algorithm is solving the following quadratic problem:

$$\min_{w,b(1/2)} w^2 + c* \sum_{i-1}^{n} \varepsilon i \text{ st}, \forall_{i:} y_i (<w,x_{ij}> +b) \geq 1- \varepsilon_i \ \varepsilon \geq 0 \qquad (2)$$

where ε_i are the slack variables in which there are non-separable case and C>0 is the soft margin which controls the differences between margin b and the sum of errors. It performs a penalty for the data in the incorrect side of classification (misclassified); this penalty rises as the distance to the margin rises, w is the slope of the hyper plane which separates the data [18]. The uniqueness of SVM comes from the ability to apply a linear separation on the high level dimension non-linear input data, and this is obtained by using an appropriate kernel function [19]. SVM effectiveness is affected by the various types of kernel functions that are selected and tuned based on the characteristics of the data.

4. Review Analysis Framework

We anticipate a computational form for sentiment analysis that comprises of three key stages. In the first place, to begin with most appropriate parts that will be extracted by using excessive data transformation, and filtering. Second, the classifiers will be created utilizing SVM on each of the features developed in the initial step and the accuracies coming about because of the forecast will be figured, and third the classifier's execution will be analyzed against various approaches. The difficult part of the framework is feature selection and here we analyze it in some profundity. Start by applying change on the data, which involves HTML names clean up, abbreviated structure improvement, stop words removal, invalidation dealing with, and stemming, by using natural language processing techniques, taking care of frameworks to perform them. Three distinctive component systems are figured in light of different feature weighting methods (FF, TF-IDF and FP). Moving to the filtering process where we analyze the chi-squared theory for each feature for every document and select the related features, trailed by the improvement of various features in perspective of the same past weighting procedures.

The data involve two data sets of movies reviews, where one was at first used as a piece of [3] containing 1400 documents (700positive and 700 negative)(Dat-1400), and the other was produced in [4],[20] with 2000 documents (1000 positive, 1000negative)(Dat-2000). Both sets are freely open. The first set is joined into the second set, as they were accessed differently so that could affect the clear comparison. Other than this allotment allows a reasonable

examination with different studies that used them autonomously. The feature used as a part of this study is unigrams.

4.1 Data Transformation

The content was at that point cleaned from any HTML tags. The abbreviated forms were extended utilizing pattern recognition and regular expression methods, and afterward the content was cleaned from non-alphabetic signs. Concerning stopwords, development of a stoplist from a few accessible standard stoplists, with a few changes identified with the particular attributes of the information. For example the words film, on-screen character, performer, movie, scene, director, are non-informative in cinema review data.

They were considered as stopwords since they are movable picture space particular words. Concerning negation, firstly taking after [3] by labeling the negative word with the accompanying words till the main punctuation mark occurrence. This tag was utilized as a unigram as a part of the classifier. By looking at the outcomes prior and then afterward adding the tagged negation to the classifier there was a small difference in the outcomes. This outcome is predictable with the findings of [21]. The reason is that it is difficult to describe a match between the labeled negation phrases among the entire set of documents. Hence, decreasing the labeled words after the negation to three and after that to two words taking in account the syntactic position and this led to more negation. To reduce the redundancy, stemming was performed.

4.2 Filtering

The technique utilizing for filtering is the univariate method called chi-squared. It is an analysis that is utilized as a part of text categorization to calculate the dependency between the word and the classification of the document it is specified in. In the word is occurring frequently in numerous classes, chi-squared value is low, while if the word is continuously occurring in couple of classifications then chi-squared value is high.

In this stage the value of chi-squared test was processed for every feature that came about the output feature from the first stage. After that, 95% of significance level of the value of chi-squared insights, a last arrangement of features was chosen in both datasets, bringing about 776 out of 7614 elements in Dat-1400, and 1222 out of 9058 elements in Dat-2000. The two sets were used to build the features networks on which the grouping was directed. At this stage data set has three feature networks: FF, TF-IDF, and FP.

4.3. Classification Process

Subsequent to developing therefore mentioned networks applying SVM classifier on every stage. Selecting the Gaussian radial basis kernel function having the parameter y˝ that controls for the area in which the support vector has an impact in the data space. SVM was connected by utilizing the machine learning package "e1071" in R. The SVM is connected with various blends of C and y˝, due to the affectability of SVM execution to their values. For the classification, prepare every set that was isolated into two sections one for preparing and the other for testing, by proportion 4:1, that is 4/5 sections were utilized for preparing and 1/5 for testing. At that point preparing was performed with 10 folds cross approval for classification.

4.4. Performance Evaluation

The performance metrics used to evaluate the classification results are precision, recall and F-measure. These metrics are computed based on the values of true positive (tp), false positive (fp), true negative (tn) and false negative (fn) allocated classes. Precision is the number of true positive out of all positively assigned documents, and it is given by

$$\text{precision} = tp / tp+fp \qquad (3)$$

Recall is the number of true positive out of the actual positive documents, and it is given by

$$\text{recall} = tp / tp+fn \qquad (4)$$

Finally F-measure is a weighted method of precision and recall, and it is computed as

$$\text{F} - \text{measure} = \frac{2*\text{precision}*\text{recall}}{\text{precision}+\text{recall}} \qquad (5)$$

Where, its value ranges from 0 to 1 and indicates better results the closer it is to 1.

5. Results Analysis

In this segment, we analyze the outcomes by a few tests furthermore the performance of the classifier. We run the classifier on each of the features frameworks, coming about because of every data transformation, filtering and contrast the performance with the one accomplished by running the classifier on non-processed data in view of correctness.

It is encountered that "standard machine learning classification procedures", for example, support vector machines (SVMs) [20], can be connected to the entire

documents themselves and this is the reason to apply the classifier on the entire content with no pre-processing or feature selection strategies [3], [20]. Hence, to permit a reasonable comparison with different results in view of the tuned kernel stage parameters we are utilizing as a part of this stage gama=0.001 and C=10, here documents are ordered with no pre-processing. The classifier was applied on the Dat-1400 features network coming about because of the main stage of pre-processing. Table 1: defines about the classifier that exhibits the performances on both not pre-processed and pre-processed data for each of the features network (TF-IDF, FF, and FP).

Besides it thinks about these results to those that are achieved in performance evaluation for both TF-IDF and FF frameworks. The correlation depends on the refined correctness and the metrics calculated in Equations 3, 4, 5. Table 1: The classification correctness in rates on Dat-1400, the segment no pre-processing advert to the outcomes reported in [3], no pre-proc2 adverts to our outcomes with no pre-processing, and pre-processing adverting to the outcomes after pre-processing, with ideal parameters y= 10-3 and C=10.

Table 1: Pre-processing

	TF-IDF		**FF**			**FP**		
	no pre-proc	pre-proc	no pre-proc1	no pre-proc2	pre-proc	no pre-proc1	no pre-proc2	pre-proc
Accuracy	78.33	81.5	72.7	76.33	83	82.7	82.33	83
Precision	76.66	83	NA	77.33	80	NA	80	82
Recall	79.31	80.58	NA	76.31	85.86	NA	83.9	83.67
F-measure	77.96	81.77	NA	76.82	82.83	NA	81.9	82.82

Table 1: demonstrates that for the information that was not a subject to pre-processing, a great change happened on the correctness of the FF network, from 72.8% to 76.33%, while the exactness's of the FP network were slightly different, we accomplished 82.33% while reported 82.7%[3]. What's more we got 78.33% exactness in TF-IDF framework where [3] did not utilize TF-IDF. By examining further in the outcomes we see the expansion in the exactness while applying the classifier on the pre-processed information after the data transformation, with a most elevated precision of 83% for both network FF and FP. Table 1 demonstrates that despite the fact that the precision fulfilled in the FP network is near the one accomplished before and in [3], there is a major correction in the classifier execution on the TF-IDF and FF frameworks, and this demonstrates the significance of stemming and removing stopwords in accomplishing higher accuracy in classification of sentiments. We stress to that so it should be capable to use the SVM classifier on the whole document; one should design and utilize a kernel for that specific issue [22].After that we order the three unique network that were constructed after the filtering (chi-squared feature selection).The achievements (see Table 2) of the classifier were high contrasting with what was accomplished in past analysis and in [3]. Selecting the features taking into account their chi squared insights esteem helped diminishing the dimensionality and the noise in the content, permitting a performance of the classifier that could be practically identical to topic categorization. Table 2 introduces the correctness and evaluation of the classifier execution prior and then afterward chi squared was applied. Table 2: The classification in rates previously, then after the fact utilizing chi-squared on Dat-1400, with ideal parameters y=-5, C=10.

Table 2: Chi-squared Classification

	TF-IDF		FF		FP	
	no chi	Chi	no chi	Chi	no chi	Chi
Accuracy	81.5	92.3	83	90	83	93
Precision	83	92.3	80	92	82	94
Recall	80.58	91.5	85.86	88.5	83.67	92.16
F-measure	81.77	92.4	82.83	90.2	82.82	93.06

Table 2 demonstrates a critical increment in the nature of the classification, with the most noteworthy exactness of 93% accomplished in the FP network, trailed by 92.3% in TF-IDF and 90.%in FF networks, in like manner the F-measure results is extremely close to 1, and that shows a superior of the classification. To the best of our insight, those outcomes were not reported in document level sentiment analysis utilizing chi-squared as a part of past studies. Henceforth, the utilization of transformation and then filtering on the writings information decreases the clamor in the writings and enhances the execution of the order.

It demonstrates how the forecast correctness of SVM gets higher the less the quantity of components is. A FRN selection technique (FRN) was proposed in [4] to choose relative features from Dat-2000 and improve the sentiment forecast utilizing SVM. The precision accomplished utilizing FRN 89.65%, in contrast with an exactness of85.5% they accomplished by utilizing chi-squared strategy among some other univariate and multivariate feature selection technique. We pre-processed Dat-2000, then ran SVM classifier, and we convey a high precision of 93.5% in TF-IDF followed by 93% in FP and 90.5% in FF (see Table 3), and that is too higher than what was found in [4].

Table 3: Correctness using Chi-squared

	TF-IDF	FF	FP
Accuracy	93.5	90.5	93
Precision	94	89.5	91
Recall	93.06	91.3	94.79
F-measure	93.53	90.4	92.87

Table 3: Best correctness in rates came about because of using chi-squared on 2000 documents with ideal parameters gama = 10-6 and C=10. The features that were utilized as a part of [4] are of various sorts including distinctive N-grams classifications, for example, words, POS tags, legomena etc, while we are utilizing unigrams as it were. We have shown that utilizing unigrams as a part of the classification has a better impact on the order results in correlation with other features. Figure 1: The relation between accuracies and the number of features, no pre-processing refers to the results in [3], pre-processing & X2(FP) refer our results.

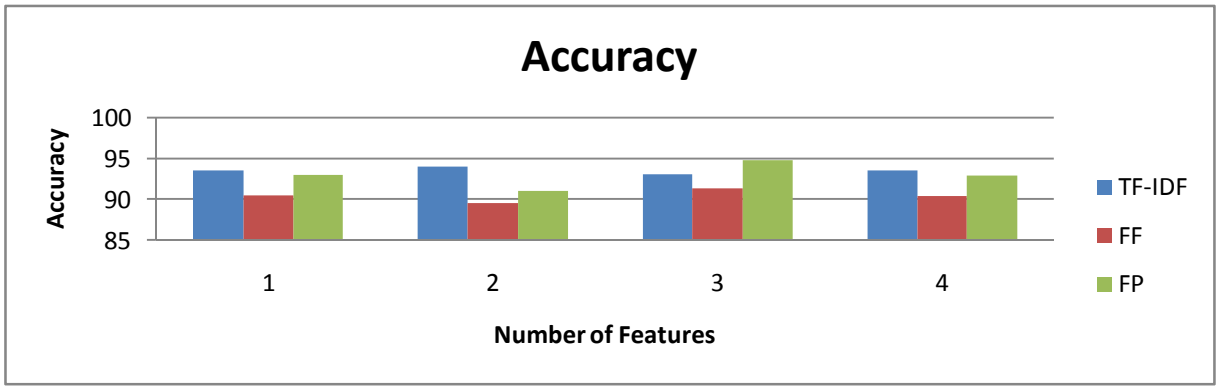

Figure 1: Relation between Accuracies and Features

6. Conclusion

In sentiment analysis feature selection, that emerges as a challenging area with lots of obstacles as it involves natural language processing. The challenge of this field is to develop the machines ability to understand text as human readers do. In this paper, we analyzed the part of text pre-processing in sentiment analysis, experimental results that demonstrate with appropriate feature selection and representation, sentiment analysis correctness using SVM in this area may be increased up to the level achieved in topic classification. Various pre-processing

methods are used to reduce the noise in the text in addition to using chi-squared method to remove unwanted features that does not affect its orientation. The level of accuracy achieved on the two data sets is comparable to the sort of accuracy that can be achieved in topic categorizing. Concluding that hybrid method for feature selection can be the future direction in the field of feature selection in sentiment analysis.

References

[1] H. Tang, S. Tan, X. Cheng, "A survey on sentiment detection of reviews, Expert Systems with Applications" 36 (7) (2009) 10760 10773.

[2] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? "Sentiment classification using machine learning techniques", in: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002.

[3] M. Thelwall, K. Buckley, G. Paltoglou, "Sentiment in twitter events", Journal of the American Society for Information Science and Technology 62 (2) (2011) 406 418.

[4] A. Abbasi, S. France, Z. Zhang, H. Chen, "Selecting attributes for sentiment classification using feature relation networks", Knowledge and Data Engineering, IEEE Transactions on 23 (3) (2011) 447 462.

[5] T. Wilson, J. Wiebe, P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis", in: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), 2005, pp. 347 354.

[6] H. Yu, V. Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences", in: Proceedings of the conference on Empirical methods in natural language processing, EMNLP-2003, 2003, pp. 129 136.

[7] L. Tan, J. Na, Y. Theng, K. Chang, "Sentence-level sentiment polarity classification using a linguistic approach", Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation (2011) 77 87.

[8] S. R. Das, "News Analytics: Framework, Techniques and Metrics", Wiley Finance, 2010, Ch. 2, the Handbook of News Analytics in Finance.

[9] P. Melville, W. Gryc, R. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification", in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009, pp. 1275 1284.

[10] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, P. Li, "User-level sentiment analysis incorporating social networks",Arxiv preprint arXiv:1109.6018.

[11] X. Ding, B. Liu, P. Yu, "A holistic lexicon-based approach to opinion mining", in: Proceedings of the international conference on Web search and web data mining, ACM, 2008, pp. 231 240.

[12] I. Feinerer, K. Hornik, D. Meyer, "Text mining infrastructure", Journal of Statistical Software 25 (5) (2008) 1 54.

[13] J.-C. Na, H. Sui, C. Khoo, S. Chan, Y. Zhou, "Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews", in: Conference of the International Society for Knowledge Organization (ISKO), 2004, pp. 49 54.

[14] V. Vapnik, "The nature of statistical learning theory", springer, 1999.

[15] C. Lee, G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization", Information processing & management 42 (1) (2006) 155 165.

[16] S. Russell, P. Norving, "Artificial Intelligence: A Modern Approach", second edition, Prentice Hall Artificial Intelligence Series, Pearson Education Inc., 2003.

[17] J. Wang, P. Neskovic, L. N. Cooper, "Training data selection for support vector machines", in: ICNC 2005. LNCS, International Conference on Neural Computation, 2005, pp. 554 564.

[18] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, "Estimating the support of a high-dimensional distribution", Neural computation 13(7) (2001) 1443 1471.

[19] B. Pang, L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts", in: Proceedings of the ACL, 2004.

[20] K. Dave, S. Lawrence, D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", in: Proceedings of WWW, 2003, p. 519 528.

[21] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers, Signal Processing", IEEE Transactions on 45 (11) (1997) 2758 2765.

[22] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Trans. Information Systems, vol. 26, no. 3, article no. 12, 2008.

[23] M. Hall and L.A. Smith, "Feature Subset Selection: A Correlation Based Filter Approach," Proc. Fourth Int'l Conf. Neural Information Processing and Intelligent Information Systems, pp. 855-858, 1997.

[24] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," J. Machine Learning Research, vol. 3, pp. 1157- 1182, 2003.

[25] F. Fleuret, "Fast Binary Feature Selection with Conditional Mutual Information," J. Machine Learning Research, vol. 5, pp. 1531-1555, 2004.

[26] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch Alignment for Dimensionality Reduction," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 9, pp. 1299-1313, Sept. 2009.