Sindh Univ. Res. Jour. (Sci. Ser.) Vol.49(1) 87- 92 (2017)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Quantitative Prediction of Offensiveness using Text Mining of Twitter Data

A. IMRAN, W. ASLAM*, M. I. ULLAH++

Department of Computer Science and IT, University of Sargodha, Pakistan.

Received 8th March 2016 and Revised 29th December 2016

Abstract: Virtual communities reflect worldwide connectivity, and an enabler for real time information sharing and targeted advertising. Twitter has widely emerged as one of the extensively used micro blogging service. This is the platform to share ideas, feelings and views for any event. People have freedom to post Tweets for a particular event. The success of an event can be predicted by users' responses. Individual interaction patterns can strongly indicate personalities. Garbage or bosh replies can harm the fidelity of an event. To make it trustworthy, we have performed sentiment analysis for the prediction of offensiveness in Tweets. We have collected data from Twitter search and stream API. Text mining techniques (preprocessing, stemming, negation rule, tokenization and stop words removal) are used for cleaning data. Our approach can predict offensiveness in Tweets effectively. We also performed comparative analysis of different machine learning classifiers, i.e., Naïve Bays (NB), Support Vector Machine (SVM) and Logistic Regression (LR) to find sentiment polarity and found that SVM outperforms others. An in-house tool, 'Interaction Pattern Predictor', is developed using Python programming language. Our results are trustworthy as we have used three large data dictionaries to train our developed tool.

Keywords: Offensiveness, Virtual community, Sentiment Analysis, Text Mining, Twitter.

1. INTRODUCTION

Micro blogging services have gained strong attention in modern era. Currently, there are about 2.34 billion users connected via virtual communities (The Statistics Portal, 2016). These networks give the benefit of worldwide connectivity, information sharing, targeted advertising and faster communication. Different micro-blogging sites like Facebook, LinkedIn, Google+ and Twitter are the popular sources for connecting people from various zones of the world. These computer-mediated tools have, interestingly, become the need of different fields including education, business, sports, politics, etc.

With the proliferation in virtual community sites (SNSs), Twitter emerged as a widespread micro blogging platform (Ibrahim and Yusoff, 2015). It facilitates users to post messages of length up to 140 characters (called Tweets). Twitter is widely adopted by various people from different zones of world. With 645 million active users and 800 million Tweets daily, Twitter has attained worldwide admiration in a very short time (Chang *et al*, 2013). Twitter uses the concept of followings and followers. Followings are the persons who share their ideas, moods or feeling. These Tweets are public in nature and can be viewed by all followers. Followers are the persons who can see the Tweets posted by the followings. Followers can reply or re-tweet the followings. Twitter allows users to post or comment freely.

It is observed that some people carry out offensive activities due to liberal nature of Tweets. Tweets against the religious scholars of any religion can make people offensive and can trigger serious reactions. Therefore, it is a matter of prime interest to observe the nature of Tweets and be able to predict offensiveness. Due to an immense number of Tweets, this prediction is quite a time taking process. Digging information from Tweets calls for developing automated tools based on machine learning algorithms. As there is no clear definition of offensiveness, currently it is one of the challenging problems.

In this article, offensiveness in Tweets is predicted using text mining techniques for preprocessing of data. A comparative analysis of different machine learning classifiers, i.e., Support Vector Machine (SVM), Naïve Bays (NB) and Logistic Regression (LR) is performed.

This article is structured as follows. Section 2 describes the work related to offensiveness prediction. Research methods and related techniques are discussed in Section 3. Working of experimental tools and their results are presented in Section 4. A comparison of different machine learning classifiers is also elaborated in this section. In Section 5, conclusion and future directions of this research article are discussed.

2. RELATED WORK

Prediction of offensiveness is one of the major challenges for virtual communities. Several researchers

++Corresponding author Email: drikramullah@uos.edu.pk.
*Department of Computer Science & IT, The Islamia University of Bahawalpur, Pakistan.

have addressed this challenge. In this work we focus on Twitter data only, sentiment analysis on which can determine the types of Tweets: positive, negative or neutral. Nature of Tweets can be further used to predict positivity and negativity on Tweets. Four approaches are used to predict types of Tweets: dictionary based (Fei *et al*., 2012), statistical based (Pender and Karunarathna, 2013), semantics based (Ostrowski, 2015) and learning based (Khan *et al*., 2015). These approaches are used to develop a system for product reviews.

Dictionary based approach is based on a dictionary of words with sentiment polarities, e.g., WordNet (Fellbaum, 1998), SentiWordNet (Rosenthal *et al*., 2015) and SenticNet (Cambria, 2016). The dictionary is a repository of English language lexicons such as nouns, verbs, adjective and synonyms (Bhonde *et al*., 2015). Statistical based approach is used to assess the probability of sentiments from a set of documents called corpus using which the polarity (positive, negative or neutral) of opinions can be estimated. Moreover, this corpus is used as a model (Pender and Karunarathna, 2013). The concept of point wise mutual information (PMI) is used to find co-occurrence frequencies in Alta Vista search engine. Semantic orientation has been implemented using this PMI (Turney *et al*., 2003). Although, this is a good technique for sentiment analysis but it needs large corpus, which is a big problem.

Semantic based approach is similar to dictionary based approach and works on the basis of matching synonyms. However instead of words, synonyms are matched. Different web applications have been implemented using semantic based technique, e.g., WordNet Ostrowski, (2015). This approach has been used in to device a tool that extracts data from multiple sources into a single schema (Vlach *et al*., 2003). Learning based approach is based on machine learning, in which opinions are input and sentiments are output. (O'Hare *et al*., 2009) has used this approach to predict stock exchange using supervised learning. Their predictions were based on topic dependent sentiment analysis using Naïve Bays and Support Vector Machine classifiers to extract data.

Offensiveness has been predicted in three large online discussion communities, viz., CNN, Breitbart and IGN, by examining the profiles. Such interaction patterns in these online communities have been characterized by investigating the previous history of persons, i.e., when they joined these communities and when they get banned (Benevenuto *et al*, 2010). (DeAndrea *et al*., 2011) has worked to identify offensive words like spam, rumors, narcissism and self-marketing attitude in Facebook. Negativity has been predicted into three dimensions, i.e., reacting to negative Tweets, not posting comments to one's status and seeking communal support. In our work, we use a hybrid technique using dictionary based approach by combining three large online dictionaries (WordNet, SenticNet and SentiWordNet). We apply text mining approaches like text preprocessing, stop word removal, stemming, negation rule and tokenization. Finally a comparative analysis between different machine learning classifiers (Support Vector Machine, Naïve Bays and Logistic Regression) is made.

3. RESEARCH METHODOLOGY

The pragmatic demonstration of our work, from data collection using Twitter search and stream API to analysis and results has been distributed in the following main steps.

Data Collection

We collected data using 'sentiment viz' tool against the keyword 'USA PRESIDENTIAL ELECTION 2016' from both search and stream APIs of Twitter. Our dataset consists of 3946 unique Tweets for a period of 42 days, from May, 5 2016 to June 17, 2016, thus averaging 95 Tweets per day. **(Fig 2 and 3)** illustrate examples of positive and negative Tweets and comments.

Fig.1: An example of a positive Tweet.

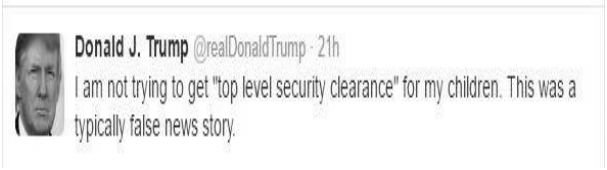

Fig.2: An example of a negative Tweet.

Preprocessing

Preprocessing is used for the classification of data, so that it can be analyzed. This task contains mainly tokenization, feature extraction and data cleaning. Tokenization is the process of segmenting the text data into tokens. These tokens consist of words, phrases, symbols and other meaningful text. Tokenization works as an input for future processing (Bird, 2006).

Feature Selection

Feature selection is the technique used for automatic selection of attributes relevant to predictive modeling of dataset. It reduces training time, improves accuracy and decreases redundancies (over-fitting). Feature selection

assigns a score to all features, which are based on an evaluation function (Ikonomakis *et al*., 2005). Information Gain (IG) is the technique used to assign scores based on their features. It is computed on term, *t*, as (Sui, 2013),

$$\mathbf{IG}(t) = \sum_{i=1}^{m} \mathbf{P}(c_i) \cdot \log \mathbf{P}(c_i) + \mathbf{P}(t) \sum_{i=1}^{m} \mathbf{P}(c_i|t) \cdot \log \mathbf{P}(c_i) + \mathbf{P}(t) \sum_{i=1}^{m} \mathbf{P}(c_i|t) \cdot \log \mathbf{P}(c_i|t),$$

Where P(c_i) denotes prior probabilities of categories set, i.e., (c_1, c_2, $c_{3,\ldots,}c_n$) and P(t) represents the prior probability of term *t*.

Classification through Supervised Learning

From the Tweet data, we determine offensiveness using Supervised learning in different machine learning classifiers, which are implemented and results compared. Brief introductions of Naïve Bays (NB) and Support Vector Machine (SVM) are given next.

Naïve Bays Classifier:

In Naïve Bays, patterns are matched by examining sets of categorized documents. It is a probabilistic classifier that matches the data with a bag of words. It streamlines the learning by classifying the features in an independent class. Accuracy of Naïve Bay is independent of feature dependencies on classes (Rish, 2001). Text categorization can be viewed in the context of subsequent documents probabilities, i.e.,$\mathrm{P}(c_i|d_j)$, where the probabilities of $j^{\text{th.}}$ document are represented in vectors. The weight vector is $d_j = < q1_j, q2_j, \ldots, q|T|_j >$, where qk_j is the weight of $k^{\text{th.}}$ feature in document belonging to class c_i. Naïve Bayes classifier is used to measure posterior probabilities, given as (Pak and Paroubek, 2010),

$$\mathbf{P}(c_i|d_j) = \frac{\mathbf{P}(d_j|c_i)\mathbf{P}(c_i)}{\mathbf{P}(d_j)}.$$

c_i denotes the posterior probability to select a random (arbitrary) document, P(d_j) is the probability of chosen arbitrary document that has the weight vector d_j and P($d_j|c_i$) is the conditional probability of the document d_j which is the member of class c_i. The estimation of P($d_j|c_i$) is complex and done as

$$\mathbf{P}(d_j|c_i) = \prod_{k=1}^{T} \mathbf{P}(W_{kj}|c_i).$$

Support Vector Machine:

SVM works to minimize structural risk. The major aim of this risk minimization is to determine a hypothesis to confirm minimum possible errors. According to the principles of risk minimization, training error and difficulty of hypothesis can be used for bounded true error. SVM is used to make resultant hypothesis free from true errors by maintaining its dimensions effectively. These dimensions can be represented in the context of Vapnik-Chervonekic or VC dimensions (Karpinski, *et al*., 1997).

Finding the maximum margins is formally represented as (Gunn, 1998),

$$\mathbf{minimize}_{w,b} < w.w >$$

$$y_i(< w.x_i > + b) \geq 1 \;\; ; \mathbf{i} = 1, \ldots, \mathbf{l}.$$

Here x_i is used as input vector and '1' exhibits training examples. Also y_j is the required output. For ease, the above mathematical problem can be reformulated as:

$$L(w, b, \alpha) = \frac{1}{2} < w.w > - \sum_{i=1}^{l} \alpha_i [y_i(< w_i.x_i > + b) - 1],$$

where $\alpha_{i>0}$ is Lagrange multiplier. This equation can be expressed in context of *w* and *b*. After substituting values, new equation can be formulated as

$$L(w, b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j < x_i x_j >.$$

The instance x_i represents the new feature space of x_j. The dimensional spaces x_i and x_j are used for dual formulation of product, which are the kernel functions.

4. EXPERIMENTS AND RESULTS

Response Volume over Time

The distribution of the collected data (3946 unique Tweets) over 42 days is shown in **(Fig-4)**, which indicates growing participation of people and decreasing infrequent peak levels. This data is categorized into three dimension, i.e., positive, neutral and negative and shown in **(Fig-5)**.

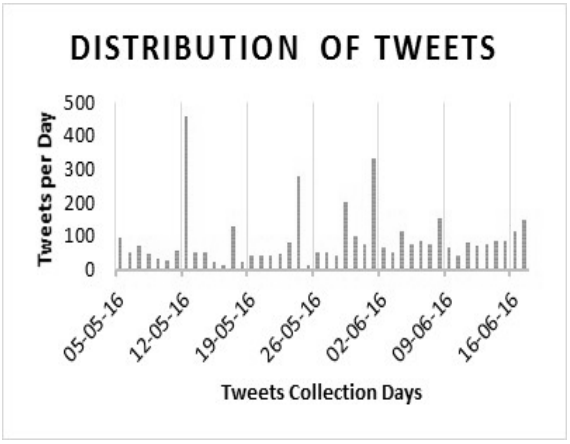

Fig. 1: The distribution of Tweets over considered days.

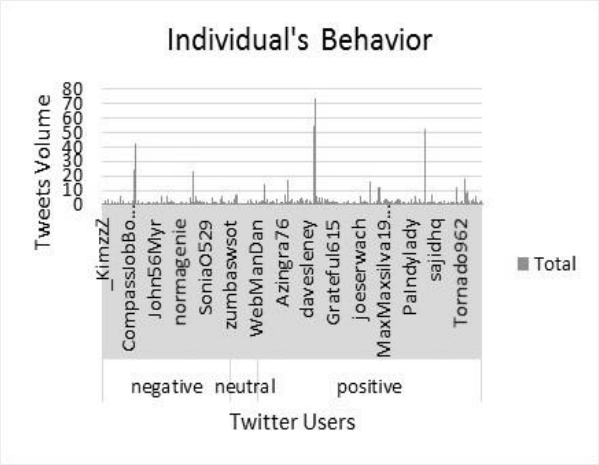

Fig.4: Interaction trends on Twitter.

Experimental Tool

To predict offensiveness in Tweets, we developed an in-house tool using Python language, namely 'Interaction Pattern Predictor' (IPP), which matches words specific to their categories, each one of which have weights assigned to them. Our tool uses a dictionary (a combination of WordNet, SenticNet and SentiWordNet) with 26112 words. This dictionary is distributed into six other source dictionaries containing positive (11078), negative (9097), neutral (373), ignore (578) prefixes (203) and offensive (4783) words. We focus on four sentiment categories, viz., positive, negative, neutral and offensive – weights of these categories are shown by result analyzer of our tool.

Dataset View

For searching data, 'sentiment viz' is invoked with multiple keywords that are separated by spaces (Kumar and Teeja, 2012). A dataset of Tweets along with additional information such as screen name, date and time is shown in **(Fig-6)**.

Date	Screen Name	Tweets Data
05-16-16 00:37	pashaterri	@USAelection @onalshale2 @realDonaldTrump @nytimes @HElectionToday same way she took on Benghazi let others die for her Remember Benghazi
05-16-16 00:46	nylek	@USAelection @realDonaldTrump @nytimes @HElectionToday now I know why ur a supporter of the orange rapist, ur dumb as a bag of rocks
05-16-16 00:46	nylek	@USAelection @realDonaldTrump @nytimes @HElectionToday an online poll? HAHAHHAHAHA, that's as meaningful as a Drudge poll...
05-16-16 01:30	IsmaelIromani68	@USAelection @HillaryClinton @HElectionToday I don't like any of them, this country is going down!
05-16-16 01:38	TheBernUnit	@USAelection @LRD90 @HElectionToday Do you think people actually fall for the communism fallacy? Have YOU actually fallen for it?
05-16-16 01:40	LRD90	@TheBernUnit @USAelection @HElectionToday - Yep. Some have fallen and can't get up. https://t.co/Wa4tq4QiQN
05-16-16 01:41	flaco_usa	@USAelection @realDonaldTrump @nytimes @HElectionToday the reason many white Democrats were directed to vote for you insuring her easy win
05-16-16 01:47	LRD90	@USAelection @TheBernUnit - No one in this country is going to put down their Chianti, Starbucks, or Mountain Dew long enough to rise up.
05-16-16 01:53	DonaldoTrumo	@USAelection @HillaryClinton @HElectionToday Trump is better to than Hillary
05-16-16 02:39	jaxvb13	@USAelection @realDonaldTrump ..truth will prevail ..at some point

Fig.5: Tweets dataset view.

Analysis and Results

After training our tool, the sentiments in Tweets are observed and categorized into four dimensions, viz., negative, positive, neutral and reclusive. **(Fig-7)** illustrates the results, in which about 48% Tweets have positive sentiments.

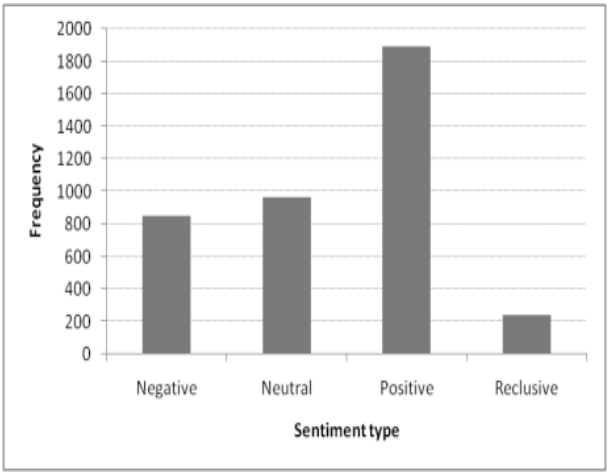

Fig.6: Tweet frequencies of sentiment categories.

The results of the three classifiers (SVM, NB and LR) are evaluated in Weka tool using 10 folds cross validation technique. The performance of these classifiers is measured on precision (*P*), recall (*R*) and F-measure (*F*). *P*, *R* and *F* are defined next:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F = \frac{2PR}{P + R}$$

Where *TP* is true positive (predicted and actual value both true), FP is false positive (predicted and actual both false) and FN is false negative (predicted false and actual true). The results of SVM, NB and LR are listed in **(Table 2)**.

Table 1: Performance of classifiers

Classifier	Precision	Recall	F-measure
SVM	**0.886**	**0.898**	**0.898**
NB	**0.811**	**0.826**	**0.816**
LR	**0.762**	**0.858**	**0.754**

Our results are based on streaming API, which uses library 'movie reviews' of natural language toolkit (NLTK). Different classifiers use these reviews and tested against them. Streaming API provides informative features and polarity values: for our case this information is presented in **(Fig-8)**.

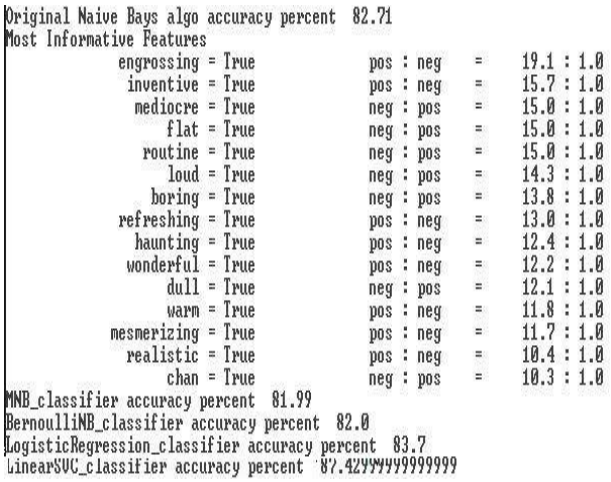

Fig.8: A comparative analysis of classifiers used.

The accuracy of the tested classifiers, SVM, LR and NB is 87.43%, 83.70% and 82.71% respectively.

5. CONCLUSIONS

In this article, we have predicted offensiveness in Tweets using text mining techniques. As compared to the current state-of-the-art works, we focus on Tweets that have enabled strong connectivity between virtual communities though with limitations of expression lengths. These limitations reflect core ideas of individuals but under stress of being misunderstood.

For our purpose, we have developed an in-house tool, IPP (in Python language), which can analyze sentiments in Tweets. In case of negative sentiments, it can further discriminate between being offensive or not. For interaction trends prediction, our tool uses streaming API, and uses a feature set of 5000 most popular words for training and testing our algorithm. We have also compared performance of three machine learning classifiers, viz., Naïve Bays, Support Vector Machine and Logistic Regression.

For our tool, we have used a larger combination of three online available dictionary sources; hence our results are more trust worthy and comprehensive than existing works. As a future work, we intend to improve validation of our results by incorporating data from various virtual community networks such as Facebook, Twitter, YouTube and LinkedIn. Also our tool will be extended by embedding in it, the psychological disorder model, FFM.

REFERENCES:

Benevenuto, F., G. Magno, T. Rodrigues, V. Almeida, (2010). Detecting spammers on twitter. In Collaboration, electronic messaging, anti-abuse and spam conference (CEAS) Vol. 6, 12-16.

Bhonde, R., B. Bhagwat, S. Ingulkar, A. Pande, (2015). Sentiment Analysis Based on Dictionary Approach. International Journal of Emerging Engineering Research and Technology, 3(1).

Bird, S., (2006). NLTK: the natural language toolkit. In Proceedings of the COLING/ACL on Interactive Presentation Sessions (69-72). Association for Computational Linguistics.

Cambria, E., (2016). Affective computing and sentiment analysis. IEEE Intelligent Systems, 31(2), 102-107.

Chang, Y., X. Wang, Q. Mei, Y. Liu. (2013). Towards Twitter context summarization with user influence models. In Proceedings of the sixth ACM international conference on Web search and data mining (WSDM '13). ACM, New York, NY, USA, 527-536. DOI=http://dx.doi.org/10.1145/2433396.2433464.

DeAndrea, D. C., S. T. Tong, J. B. Walther, (2011). "Dark sides of computer-mediated communication", The dark side of close relationships II, 95-118.

Fei, G., B. Liu, M. Hsu, M. Castellanos, R. Ghosh, (2012). A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In 24th international conference on computational linguistics 309.

Fellbaum, C., (1998). Word Net. Blackwell Publishing Ltd.

Gunn, S. R., (1998). Support vector machines for classification and regression. ISIS technical report, 14.

Ibrahim, M. N. M., M. Z. M. Yusoff, (2015). Twitter sentiment classification using Naive Bayes based on trainer perception. In 2015 IEEE Conference on e-Learning, e-Management and e-Services (IC3e) (187-189). IEEE. (Naïve Bay reference 1).

Ikonomakis, M., S. Kotsiantis, V. Tampakas, (2005). Text Classification using Machine Learning Techniques. WSEAS Transaction on Computers, Vol. 8(4), 966-974.

Karpinski, M., A. Macintyre, (1997). Polynomial Bounds for VC Dimension of Sigmoidal and General Pfaffian Neural Networks, Journal of Computer and System Sciences, Volume 54, Issue 1, 169-176, ISSN 0022-0000, http://dx.doi.org/10.1006/jcss.1997.1477.

Khan, A. Z., M. Atique, V. M. Thakare, (2015). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. International Journal of

Electronics, Communication and Soft Computing Science and Engineering (IJECSCSE), 89.

Kumar, A., M. S. Teeja, (2012). Sentiment analysis: A perspective on its past, present and future. International Journal of Intelligent Systems and Applications, 4(10), 1-6.

O'Hare, N., M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, A. F. Smeaton, (2009). Topic-dependent sentiment analysis of financial blogs. In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion 9-16. ACM.

Ostrowski, D. A., (2015). "Using latent dirichlet allocation for topic modelling in twitter," Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), Anaheim, CA, 493-497.

Pak, A., P. Paroubek, (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In LREc Vol. 10, 1320-1326.

Pender, D., H. Karunarathna, (2013). A statistical-process based approach for modelling beach profile variability. Coastal Engineering, 81, 19-29.

Rish, I., (2001). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence Vol. 3, No. 22, 41-46. IBM New York.

Rosenthal, S., P. Nakov, S. Kiritchenko, S. M. Mohammad, A. Ritter, V. Stoyanov, (2015). Semeval-task 10: Sentiment analysis in twitter. Proceedings of SemEval-

Sui, B., (2013). Information Gain Feature Selection Based on Feature Interactions (Doctoral dissertation, University of Houston).

The Statistics Portal (2016). Retrieved Dec. 25, 2016, http://www.statista.com.

Turney, P. D., M. L. Littman, (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS), 21(4), 315-346.

Vlach, R., W. Kazakaos, (2003). Using Common Schemas for Information Extraction for Heterogeneous Web Catalogs. ADBIS 2003, LNCS 279, 18, 118-132.