Opinion Mining: A Comparison of Hybrid Approaches

Alex M. G. Almeida
Sylvio Barbon Jr.
Rodrigo A. Igawa

Londrina State University
PR 445 Km 380
Londrina-PR, Brazil
Email: alex.marino@gmail.com
sbarbonjr@uel.br
igawa.rodrigo@gmail.com

Emerson C. Paraiso

Pontifícia Universidade Católica do Paraná
Rua Imaculada Conceição, 1155
Curitiba-PR, Brazil
Email: paraiso@ppgia.pucbr.br

Stela N. Moriguchi

Uberlândia Federal University
Av. João Naves de Avila, 2121
Santa Mônica, Uberlândia-MG, Brazil
Email: stellanm@ufu.br

***Abstract*—Applications based on Opinion Mining and Sentiment Analysis are critical tools for information-gathering to find out what people are thinking. It is one of the most active research areas in Natural Language Processing, and a diversity of strategies and approaches have been published. We evaluate two strategies - Cognitive-based Polarity Identification and Crowd Explicit Sentiment Analysis - and combine them with emoticons and lexicon analysis in a four hybrid models cascade framework. These four approaches were compared to evaluate a suitable method to improve performance over different datasets. We performed experimental tests using the SentiStrength database, which is composed of five public datasets. Results show that emoticons attribution can improve accuracy while combined with Crowd Explicit Sentiment Analysis and Cognitive-based Polarity Identification approaches. In addition, hybrid approaches achieve better precision in case of neutral sentences. Datasets that provide a more informal use of language are suitable for hybrid approaches.**

Keywords–Opinion Mining; Sentiment Analysis; Machine Learning.

I. INTRODUCTION

Opinion Mining (OM) and Sentiment Analysis (SA) are fields of studies which analyze opinions, sentiments, and mood, based on textual data. Currently, OM can be related to Information Retrieval (IR) and Machine Learning (ML), which match classification indicators of semantic weight. Although emotions are conceptualized as subjective experiences which are hard to evaluate, for SA they are susceptible to a kind of computing effort accessible from text [1].

Some research fields have arisen due to the increasing use of Web environments, such as forums, blogs, and online social networks (OSNs). By providing ubiquitous information, these environments have encouraged studies concerning people's sentiments, opinions, and mood [2]. Knowledge is obtained from people's written thoughts: in this way, a company can improve its products and services by discovering people's wishes [3].

The OSN, typically, contains a huge volume of opinion in textual format. Additionally, it is a valuable source of information about people's sentiment, but it is not easy for humans to read it due to the large amount of data and the diversity of multimedia characteristics (slangs, emoticons, lack of grammatical accuracy, etc.). Thus, we studied several hybrid combinations such as in [4] associated with ML and lexicon-based approaches. While [4] evaluated only one dataset, our experiments covered five datasets with distinct features: user profile, written profile (varying frequency use of slang, emoticons, writing correction). Furthermore, based on the ML approach, it was possible to achieve accurate results and to identify a high precision combination.

The present paper presents a comparative study of two approaches to OM and SA in the recent literature. Both approaches have been chosen because of their simplicity. The first approach, Crowd Explicit Sentiment Analysis (CESA), is lexicon-based, while Cognitive-based Polarity Identification (CBPI) is a machine learning algorithm. The experiments were performed on five different datasets of web social media content available at the CyberEmotions Consortium [5].

In summary, this paper presents a comprehensive and in-depth critical assessment of both approaches, aiming to highlight the limitations and advantages of each one for predicting opinion polarity in many different datasets. To guide the comparative study, we aim to answer the following questions: a) Which approach is more accurate? b) How do emoticons aid the classification? c) Is there an approach more suitable for a single polarity? d) Does the dataset influence the selection of the approach?

The rest of this paper is organized as follows. Section II gives a brief review of the literature on OM and SA. Then, Section III describes the SentiStrengh Dataset. Next, Section IV describes the experimental settings, followed by Section V, which supplies a detailed descriptions of the approaches used in this paper. The following Section VI describes the results, and Section VII presents some conclusions and suggestions for future work.

II. RELATED WORK

Prior to classifying opinions, extracting opinions is also a concern of OM and SA. In [6], the authors report experiments addressing feature-based opinion extraction and achieved good results limited to 40 pre-defined examples. In [7], classic Text Mining pre-processing techniques based on OM were studied and it was concluded that this kind of technique should be adapted in order to correctly clean the noisy text from platforms such as Twitter.

 ISBN: 978-1-61208-452-7

A more recent paper about Opinion Retrieval, a subarea of OM, used YouTube, where the goal was to obtain the target of an opinion as being either the video or the product [8].

SA is an important research area for audio and video. Recently [9] showed the importance of several feature extraction methods in experiments with EmoDB Dataset [10], which Suport Vector Machines (SVM) classification improved 11% in accuracy. In [11], J.G. Ellis et al. focus on predicting the evoked emotion at particular times within a movie trailer. The main idea is to learn a mid-level feature representation for multimedia content that is grounded in machine detectable concepts and then model human emotions based on this representation.

Regarding SA, the literature can be divided into two main approaches: Lexicon and Machine Learning. The first one mainly relies on a sentiment lexicon, i.e., a set of known and predefined terms or phrases that represent emotions, e.g., Opinion Finder available on [12].

On the other hand, Machine Learning approaches rely on an initial set of pre-labeled documents, opinions, or terms, to automatically extract features for further classification [13].

Within Lexicon-based SA, there are two more subdivisions of approaches, according to [13], [14]: Dictionary and Corpus-based. The second usually concerns a more dynamic set of words rather than fixed dictionaries to represent emotions. For example, in [15], the goal was to retrieve a new and adapted lexicon from a specific domain.

Also, in [16], it was reported that only one lexicon in a reference language should be necessary to perform a multi-language SA. In [17], the authors successfully analyzed the behavior of sports fans during FIFA 2014 World Cup on Twitter.

Dictionary-based approaches have also used a set of terms to be updated according to context as shown in [18], [19], the analysis of stock prices, or of the emergence of political topics, all based on the news. One last example of a dictionary-based solution is CESA, exposited in [20], where a dictionary was shown to be useful in combination with other tools.

Unlike Lexicon-based approaches, Machine Learning applied to OM and SA mainly depends on its labeled corpus to extract features in order to classify opinions. Examples are shown in [21], [22], in which SVM was applied to monitor not only the products, but also features that would describe or classify opinions in multiple domains.

Another case of the use of Machine Learning is to classify opinion polarity, based on sentence features [23]. Probabilistic classifiers (e.g., Naive Bayes) in [2], [24] also performed well at inferring the polarity of tweets, in a simple way. In this paper, experiments as with [24], have been used as an ML approach for comparison of lexicon-based approaches.

III. The SentiStrength Dataset

The SentiStrength [25] database consists of six datasets: Digg, BBC forum, Runners World forum, YouTube, Twitter, and MySpace. The dataset in its original form includes three fields:

1) positive strength
2) negative strength
3) message

TABLE I. Sentiment Strength x Sentiment Label

PS	NS	Message	SL
3	-1	I am so happy, #SherlockHolmes incredible soundtrack! Glad to see others appreciated. Now if it wins, crossing my fingers.	positive
1	-1	Grow on Twitter with Tweet Automator http://bit.ly/dwxFHu	neutral
1	-2	#4WordsOnObamasHand Don't Say The N-Word	negative

To make it possible to use datasets for sentiment classification, we reassigned all messages in datasets with sentiment labels (positive, negative, neutral) rather than sentiment strengths (positive and negative strengths). In Table I, the negative strength (NS) is a number between -1 and -5, and the positive strength (PS) is a number between 1 and 5.

To obtain sentiment label (SL) we followed two rules: neutral if the difference between the negative absolute value and the positive value is 0, and positive if the ratio of positive values to the negative values is bigger than 1.5; if not, it is negative [26].

IV. Experimental Settings

To perform this experiment, we used a reassigned dataset, as described in Section III, called the sample. The sample was composed of labeled messages (positives, neutrals, and negatives). Table II presents some sample characteristics: number of negative and positive emoticons, number positive and negative sentiment terms, and the number of messages for each dataset. Figure 1 shows the algorithm of the experimental procedure. We partitioned it into four tasks: Acquisition, Pre-processing, Training and Classification.

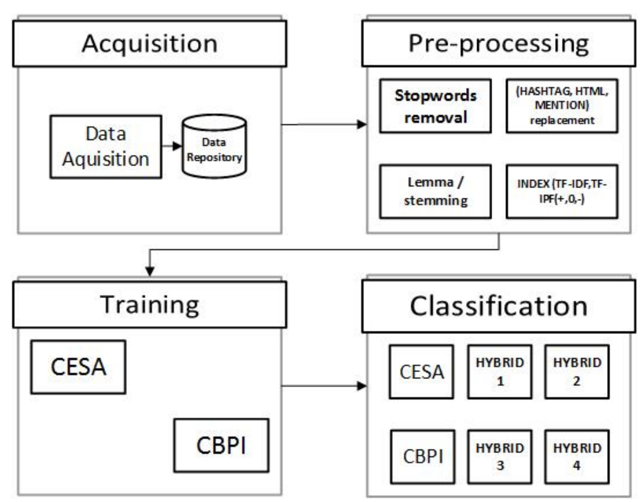


Figure 1. Complete Experimental Process Diagram

For the Acquisition task, we used SentiStrength Database where messages serve as input to the pre-processing task. The MySpace Dataset was ignored because is a deprecated OSN.

The pre-processing task performs intensive processing steps for each message and then sends it to the subsequent task. The pre-processing task consists of the following steps:

1) Stop words removal
2) Tags replacement (“@”, “#”)

TABLE II. DATASETS SPECIFICATION

Dataset	Emoticons		SWN Terms		Messages
	NEG	POS	NEG	POS	
YOUTUBE	46	277	595	2012	3407
TWITTER	130	427	556	1450	4127
RW1046	41	181	678	1758	1042
DIGG	10	28	471	742	1042
BBC	7	13	753	1062	1042

3) Stemming
4) Indexing by Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Inverse Polarity Frequency (TF-IPF)

For the Training task, we focused on building a classifier based on the input vector plus the answer vector to generalize the knowledge and finally predict the messages. In the main CESA and CBPI, we used a 10-fold cross-validation. With regard to the Classification task, accuracy refers to the ability of the classifier to predict the class label correctly for a new set of data.

V. THE COMPARED APPROACHES

This article includes different approaches of previous research. In these studies, the main issues are classification accuracy, data sparsity with insufficient data or very few useful labels in the training set, and a high percentage of sentences incorrectly classified as neutral [4]. For this research, we chose CESA due to the fact that it is simple to implement [20] and highly accurate for negative and positive strengths.

The CBPI approach was chosen because it is a simple solution and has one of the best neutral classification results. The challenge of hybrid approaches is to combine these techniques with emoticon interpretation in order to improve the accuracy of the results. Moreover, a hybrid approach can reduce the dependency of ML [27] on SentiWordNet (SWN) [28] because a prior polarity attribution diminishes the influence of the training set on an ML solution, which is good for low-quality datasets.

A. Emoticon Attribution

Emoticon attribution is used in a simple way. Attribution is done based on a list of emoticons. It uses regular expressions, based on Table III, to detect the presence of emoticons which are then classified as positive or negative and reduces the dependency on machine learning [27].

TABLE III. EMOTICONS POLARITY

Emoticon	Polarity
:-) :) :o) :] :3 :c) :N =] 8) =) :} :^) :)	positive
:-D :D 8-D 8D x-D xD X-D XD =-D =D=-3 =3 B^D	positive
>:[:-(:(:-c :c :-<:<:-[:[:{	negative
>:\>:/ :-/ :-. :/ :\=/ =\:L =L :S >.<	negative

B. SentiWordNet

SWN uses a list of positive and negative words, as shown in Table IV, to check the sentiments for each term in the message. After pre-processing, for each word, the sentiment score is found and the final polarity is given by the sum of each sentiment score.

TABLE IV. POSITIVE AND NEGATIVE POLARITY OF WORDS

Sentiment Words	Polarity
better good well great happy free best lucky safe fine ready big strong special normal	positive
bad guilty sick sad lost tired ill stupid weird horrible wrong terrible hurt worse empty uncomfortable	negative

Let a document *D* be defined as a set of messages, as follows:
$D = \{m_1, m_2, m_3, ..., m_i\}$.

Let *m* defined as a set of words for each message, as follows:
$m = \{w_1, w_2, w_3, ..., w_j\}$.

Let *PW* be defined as a set of positive words and *NW* defined as negative words as
PW={set of positive words} and
NW={set of negative words}.

For each word in a message the score is calculated by (1).

$$S(w_j) = \begin{cases} 1 & (w_j \in m_i) \wedge (m_i \in D) \wedge (w_j \in PW); \\ -1 & (w_j \in m_i) \wedge (m_i \in D) \wedge (w_j \in NW); \\ 0 & (w_j \in m_i) \wedge (m_i \in D) \wedge (w_j \notin (NW \cup PW)). \end{cases} \tag{1}$$

The final message polarity of a message obtained by SWN is given (2).

$$P(m_i) = \begin{cases} positive & \sum_{k=1}^{j} S(w_k) > 0; \\ negative & \sum_{k=1}^{j} S(w_k) < 0. \end{cases} \tag{2}$$

C. Crowd Explicit Sentiment Analysis

The lexicon-based OM used in this research is CESA [20]. Based on Explicit Semantic Analysis (ESA), the main idea consists in using a set of documents as a matrix to represent the meaning of the concepts obtained from a *corpus* [29]. The main improvement of CESA over ESA is grounded in forming a vector of sentiments.

As with any other lexicon-based approach, CESA requires a *Lexicon v*; in this paper, we use WeFeelFine [30]. The first step of CESA consists in acquiring the dataset, followed by pre-processing. These pre-processing techniques result in a sentiment vector—to be manually labeled—along with the initial *corpus* and the TF-IDF.

The matrix M_{mn}, the key part of the CESA, results from a combination of the sentiment vector and the *corpus*. m represents the size of the *corpus* and n the number of sentiments obtained by the pre-processing phase. We perform a modified version of the CESA: in our version, the part of speech tagger was not employed, due to the results shown in [7], which concluded that the classical part of speech tagger should be adapted when applied to an informal *corpus* like an OSN.

D. Cognitive-based Polarity Identification

The Machine Learning approach used in our experiments is found in [24]. As specified in its own presentation, the Cognitive-based Polarity Identification prioritizes simplicity and uses a Bayesian Classifier along with a TF-IPF for the feature extraction. By this time, there should be three sets: a set of positive posts, another of negative posts, and a set of neutral posts.

From there, the procedure takes into consideration an adapted TF-IDF which the authors call TF-IPF. The TF measures how frequently a term t occurs in the ith polarity document, while the Inverse Polarity Frequency measures how important t is for that polarity class. Then, the polarity score $S_i(t)$ (where $i = 1, 2, 3$) considering positive, negative, and neutral is associated to the term t for the polarity i is $S_{class_i}(t) = TF_i(t) * IPF(t)$.

As a consequence, the terms containing higher values of $S_{class_i}(t)$ are considered features. Following the same threshold used in [24], we considered as document feature any term with a $S_{class_i}(t)$ greater than 0.5.

By applying Naive Bayes, the aim is to predict the class c_i from the probability that the document D, represented as a set of m features f_k where $(k = 1, 2, 3, .., m)$ belongs to that class $p(c_i|D)$, given by (3) where $p(f_k|c_i)$ is the probability that the features belong jointly to the class c_i (A is a normalization factor). Lastly, the classifier predicts the class i of the document D presenting the highest probability.

$$p(c_i|f_1, f_2, .., f_m) = \frac{1}{A} p(c_i) \prod_{k=1}^{m} p(f_k|c_i) \tag{3}$$

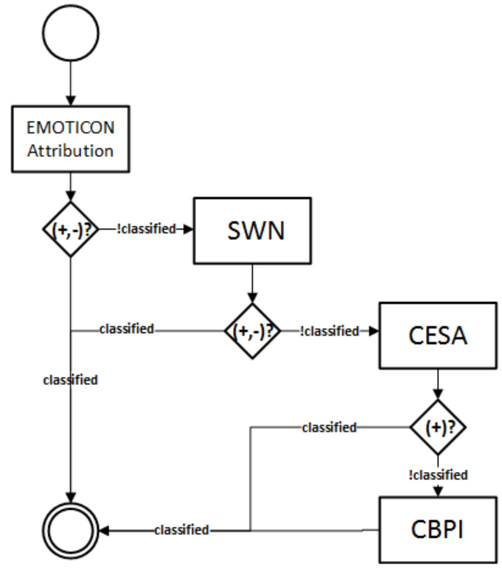
E. Hybrid Schema 1

Hybrid Classification Schema 1, as shown in Figure 2(a), is composed of four steps:

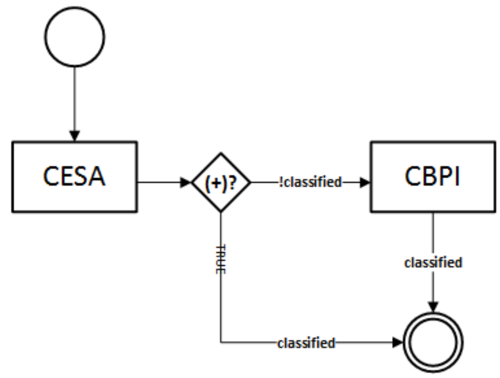
1) Emoticon Attribution
2) SWN
3) CESA
4) CBPI

Emoticon Attribution and SWN are baseline filters, and the CESA and CBPI are machine learning methods. To define Emoticon Attribution as the first step on baseline filters, we considered the fact that the presence of an emoticon in a microblog message represents a sentiment that extends to the whole message [2]. For ML methods, we defined CESA as the first step because it is good for negative and positive classification while CBPI is one of the best methods for neutral rating. To perform the classification, each message is subjected to the four indicated steps in order. The step Emoticon Attribution classifies a particular message as positive or negative verifying the existence or not of emoticons. When subjected message does not contain any emoticon, then the message is processed by SWN which classifies it only as positive or negative, as shown in V-B. Once all filtering steps are performed and the message is not classified yet (as positive or negative), then it is processed by CESA.

When a subjected message comes to CESA step it emphasizes positive messages, in other words, it is being verified if the message is positive or not. If the message is not classified as positive, then it is finally treated by CBPI, which will classify it as positive, neutral, or negative. Hybrid 1 is done using the three scores as given in (4).



(a) Hybrid Classification Schema 1



(b) Hybrid Classification Schema 3

Figure 2. Hybrid classifications schemas

$$P(m) = \begin{cases} \textit{positive} & (S_E > 0) \vee (S_E = 0 \wedge S_S > 0) \vee \\ & (S_E = 0 \wedge S_S = 0 \wedge S_D > 0) \vee \\ & (S_E = 0 \wedge S_S = 0 \wedge S_D \leq 0 \wedge S_M > 0); \\ \textit{negative} & (S_E < 0) \vee (S_E = 0 \wedge S_S < 0) \vee \\ & (S_E = 0 \wedge S_S = 0 \wedge S_D \leq 0 \wedge S_M < 0); \\ \textit{neutral} & (S_E = 0 \wedge S_S = 0 \wedge S_D = 0 \wedge S_M = 0). \end{cases} \tag{4}$$

where S_E, S_S, S_D and S_M are Emoticon Attribution, SWN, CESA and CBPI respectively.

F. Hybrid Schema 2

Hybrid 2 performs the same baseline filters as Hybrid 1. Hybrid 2 differs from Hybrid 1 in the CESA step, which emphasizes negative messages. The next step, CBPI, performs the same way as Hybrid 1 and the message is classified as positive, neutral or negative. Hybrid 2 is done using the three scores as given in (5).

$$P(m) = \begin{cases} \textit{positive} & (S_E > 0) \vee (S_E = 0 \wedge S_S > 0) \vee \\ & (S_E = 0 \wedge S_S = 0 \wedge S_D \geq 0 \wedge S_M > 0); \\ \textit{negative} & (S_E < 0) \vee (S_E = 0 \wedge S_S < 0) \vee \\ & (S_E = 0 \wedge S_S = 0 \wedge S_D < 0) \vee \\ & (S_E = 0 \wedge S_S = 0 \wedge S_D \geq 0 \wedge S_M < 0); \\ \textit{neutral} & (S_E = 0 \wedge S_S = 0 \wedge S_D = 0 \wedge S_M = 0). \end{cases} \tag{5}$$

G. Hybrid Schema 3

Hybrid classification schema 3 does not perform baseline filters, as shown in Figure 2(b).

All pre-processed messages are directly evaluated by CESA emphasizing positive message and, in case the current message is not positive, it is handed on to CBPI, which will classify the message as positive, neutral, or negative. Hybrid 3 is done using the three scores as given in (6).

$$P(m) = \begin{cases} positive & (S_D > 0) \vee (S_D \leq 0 \wedge S_M > 0); \\ negative & (S_D \leq 0 \wedge S_M < 0); \\ neutral & (\mathrm{S}_D \leq 0 \wedge S_M = 0). \end{cases} \quad (6)$$

H. Hybrid Schema 4

In a similar way to Hybrid 3, Hybrid 4 differs in emphasizing negative messages. All messages not classified at CESA as negative are treated by CBPI so as to be finally classified as positive, neutral or negative. Hybrid 4 is done using the three scores as given in (7).

$$P(m) = \begin{cases} positive & (S_D \geq 0 \wedge S_M > 0); \\ negative & (S_D < 0) \vee (S_D \geq 0 \wedge S_M < 0); \\ neutral & (\mathrm{S}_D \geq 0 \wedge S_M = 0). \end{cases} \quad (7)$$

VI. Analysis Results and Discussion

To evaluate the proposed hybrid approaches, we used confusion matrix, precision, accuracy and a ranking order scale. The division of true positive against both true positive and false positive defines precision, which is denoted as:

$$Precision_A = \frac{TP_A}{TP_A + FP_{BA} + FP_{CA}}$$

where TP_A represents the number of right predictions for class A while FP_{BA} are class B incorrectly classified as A, and FP_{CA} are class C incorrectly classified as A.

The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined which is denoted as:

$$Accuracy = \frac{Tpos + Tneg + Tneu}{Tpos + Fpos + Tneg + Fneg + Tneu + Fneu}$$

where Tpos, Tneg and Tneu are respectively true positives, true negatives and true neutrals while Fpos, Fneg and Fneu are false positive, false negative and false neutral respectively.

The rank is a comparative scale technique where 1 is assigned to the best accuracy obtained from an approach per dataset, and 6 is assigned to the worst accuracy. This rank is an ordinal scale that describes accuracies performance approaches, but does not revel distance between approaches. We use a final ranking that is calculated by performance average of each method over all datasets.

Our discussion of the results starts with the first question presented in Section I: “a) Which approach is more accurate?” Figure 3 presents a box plot of each approach along with their respective accuracies. The box plots show that the approaches Hybrid 1 and Hybrid 2 presented the highest average accuracy (75.00% and 77.00%) along with the highest values of quartile 1 (Q1) (65.50% and 65.75% accuracy).

In another view, the CESA approach presented the highest accuracy but in terms of stability it was not the best method because CESA’s box has a similar size to Hybrid 4 box that is clearly the most unstable. Although CESA presented a median value close to Hybrid 1 and Hybrid 2, it is very important to highlight that not only medians and maximum values of accuracies should be taken into consideration. If one of these methods should be selected, Hybrid 1 and Hybrid 2 are the most appropriate because of their stability presented in the box plot.

Thus, when maintaining stability is a matter of greater importance than obtaining the best accuracy in real scenario, we recommend Hybrid 1 or Hybrid 2. This is our conclusion regarding the first question once CESA presented either high and low results in terms of accuracy.

TABLE V. ACCURACIES AND RANKING ON THE DATASETS

Dataset	CESA	CBPI	Hybrid 1	Hybrid 2	Hybrid 3	Hybrid 4
BBC	83.13% (1)	50.39% (5)	76.74% (3)	77.34% (2)	48.64% (6)	67.67% (4)
Digg	84.15% (1)	43.91% (5)	74.63% (3)	78.47% (2)	41.59% (6)	63.42% (4)
Runners	58.95% (3)	40.76% (5)	68.24% (2)	69.71% (1)	46.47% (4)	32.65% (6)
Twitter	53.19% (3)	41.94% (4)	58.26% (1)	53.33% (2)	35.07% (5)	27.25% (6)
Youtube	70.66% (3)	42.12% (5)	76.99% (2)	77.78% (1)	58.11% (4)	41.30% (6)
Ranking	2.2	4.8	2.2	1.6	5.0	5.2

To answer the second question, “b) How do emoticons aid the classification?”, we performed experiments concerning the Hybrid approaches, and the results are shown in Figure 4(a). Note that Hybrid 1 and Hybrid 2 both take into consideration emoticons while Hybrid 3 and Hybrid 4 do not take advantage of that aspect. It is possible to see that, apart from which dataset the approach was performed on, considering emoticons always yielded higher results, as we can see by the fact that the bar plots show better results than the line plots for all datasets.

Figure 4(b) illustrates our results regarding the question: “c) Is there an approach more suitable for a specific polarity?” This question is especially addressed in a case of discovering a specific polarity that is important and, therefore, making an analysis of the precision concerning one specific class would also be necessary.

In Figure 4(b), it is notable that the prediction of neutral opinions is not a problem for either Hybrid 1 and Hybrid 2. However, if it is necessary to achieve higher results about positive or negative, then Hybrid 1 must be selected to predict neutral and negative opinions (78.62%) while Hybrid 2 must be chosen to predict neutral or positive opinions (82.33%).

The last question in Section I: “d) Does the dataset influence the selection of an approach?” is illustrated by Table V. Each row in the table corresponds to accuracy, in percentage, and ranking of accuracies per dataset. The last row in the table represents the average ranking per approach. The average ranking shows that Hybrid 2 (1.6) is the best average method for evaluated datasets followed by Hybrid 1 and CESA (2.2). It is important to note that in datasets, such as BBC and Digg, that provide a more structured text than the Twitter scenario, the CESA approach performs with satisfactory results (e.g., 83.13% and 84.15%, respectively and ranking 1).

On the other hand, Hybrid 1 and 2 approaches achieve better results in datasets like Twitter (ranking 1 and 2 respectively) and YouTube (ranking 2 and 1 respectively) by taking emoticons into consideration. For example, the row for YouTube Dataset of Table V shows the accuracies on Hybrid

 ISBN: 978-1-61208-452-7

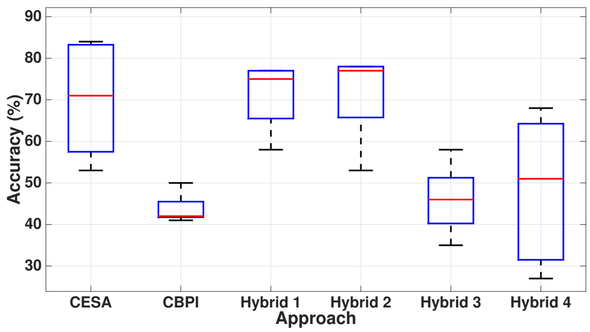
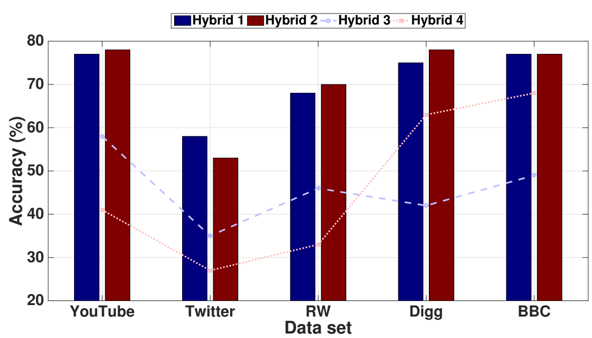
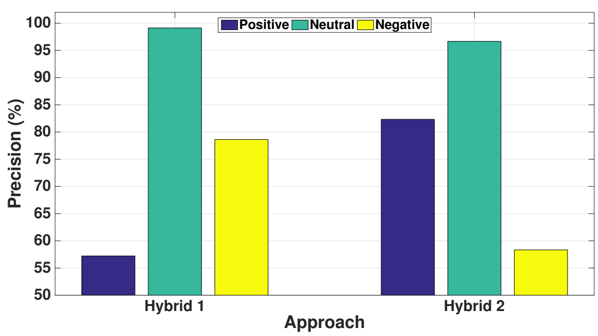


Figure 3. Box plot accuracies of each approach over all datasets

1 and Hybrid 2 achieved 76.99% and 77.78% (ranking 2 and 1 respectively) while the CESA approach achieved 70.66% (ranking 3).



(a) Hybrid approaches with the aid of emoticons



(b) Polarity Tendency on Hybrid 1 and Hybrid 2

Figure 4. Overview of the hybrid approaches Accuracy and Precision

Hybrid 3 and Hybrid 4 presented lower accuracy results than Hybrid 1 and Hybrid 2. As discussed in Figure 4(a), this results from the fact that they do not take advantage of emoticon analysis. Therefore, datasets providing a more informal use of language should indicate the selection of an approach concerning informal features like emoticons is a better choice.

The CBPI approach achieved lower results in terms of accuracy in the datasets used in our experiments. This can be justified by the fact that a better threshold from TF-IPF was not found and the words used as features on Naive Bayes might not have described the data properly. Still, CBPI is the most stable method in that it achieves similar results in different sets, as shown in Figure 3.

VII. Conclusion and Future Work

We provided in this paper a comparative study using reported datasets with distinct characteristics. Our results have shown that the content of datasets with an informal use of language is better classified by hybrid schemes reinforced with emoticon attribution and SWN.

Still, as to the emoticon attribution and SWN aid for the approaches Hybrid 1 and Hybrid 2, we noted that the precision is improved, reaching more than 95%, on single neutral classification independently of the dataset, whereas positive and negative single classification alternated good results (in precision terms) respectively.

Hybrid approaches 1 and 2 resulted in more stable results in terms of accuracy among different datasets, which constitutes the relevant contribution of this paper. For future research, we suggest focusing on the computational cost and specifically on a deeper research into sarcasm, which we believe to be intrinsically connected to OM as a great challenge.

References

[1] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining Text Data. Springer, 2012, pp. 415–463.

[2] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining." in LREC, 2010.

[3] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011, pp. 30–38.

[4] F. H. Khan, S. Bashir, and U. Qamar, "Tom: Twitter opinion mining framework using hybrid classification scheme," Decision Support Systems, vol. 57, 2014, pp. 245–257.

[5] "Cyberemotions consortium," [accessed October-2015]. [Online]. Available: http://sentistrength.wlv.ac.uk/

[6] F. L. Cruz, J. A. Troyano, F. Enríquez, F. J. Ortega, and C. G. Vallejo, "Long autonomy or long delay? the importance of domain in opinion mining," Expert Systems with Applications, vol. 40, no. 8, 2013, pp. 3174–3184.

[7] G. Petz, M. Karpowicz, H. Fürschuß, A. Auinger, V. Stříteskỳ, and A. Holzinger, "Computational approaches for mining user's opinions on the web 2.0," Information Processing & Management, vol. 50, no. 6, 2014, pp. 899–908.

[8] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, and K. Filippova, "Multi-lingual opinion mining on youtube," Information Processing & Management, 2015.

[9] E. S. Erdem and M. Sert, "Efficient recognition of human emotional states from audio signals," in Multimedia (ISM), 2014 IEEE International Symposium on. IEEE, 2014, pp. 139–142.

[10] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of german emotional speech." in Interspeech, vol. 5, 2005, pp. 1517–1520.

[11] J. G. Ellis, W. S. Lin, C.-Y. Lin, and S.-F. Chang, "Predicting evoked emotions in video," in Multimedia (ISM), 2014 IEEE International Symposium on. IEEE, 2014, pp. 287–294.

[12] "Opinion finder," [accessed October-2015]. [Online]. Available: http://mpqa.cs.pitt.edu/opinionfinder/

[13] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, vol. 311, 2015, pp. 18–38.

[14] C. Potts, "Sentiment symposium tutorial," in Sentiment Symposium Tutorial. Acknowledgments, 2011.

[15] S. Park, W. Lee, and I.-C. Moon, "Efficient extraction of domain specific sentiment lexicon with active learning," Pattern Recognition Letters, vol. 56, 2015, pp. 38–44.

[16] A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, and F. de Jong, “Multi-lingual support for lexicon-based sentiment analysis guided by semantics,” Decision support systems, vol. 62, 2014, pp. 43–53.

[17] Y. Yu and X. Wang, “World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans tweets,” Computers in Human Behavior, vol. 48, 2015, pp. 392–400.

[18] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, “News impact on stock price return via sentiment analysis,” Knowledge-Based Systems, vol. 69, 2014, pp. 14–23.

[19] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, “Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis,” Knowledge-Based Systems, vol. 69, 2014, pp. 24–33.

[20] A. Montejo-Ráez, M. Díaz-Galiano, and L. Ureña-López, “Crowd explicit sentiment analysis,” Knowledge-Based Systems, 2014.

[21] M. Zimmermann, E. Ntoutsi, and M. Spiliopoulou, “Discovering and monitoring product features and the opinions on them with opinstream,” Neurocomputing, vol. 150, 2015, pp. 318–330.

[22] M. R. Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, and L. Ureña-López, “Experiments with svm to classify opinions in different domains,” Expert Systems with Applications, vol. 38, no. 12, 2011, pp. 14 799–14 804.

[23] J. M. Chenlo and D. E. Losada, “An empirical study of sentence features for subjectivity and polarity classification,” Information Sciences, vol. 280, 2014, pp. 275–288.

[24] E. D’Avanzo and G. Pilato, “Mining social network users opinions’ to aid buyers shopping decisions,” Computers in Human Behavior, 2014.

[25] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment strength detection for the social web,” Journal of the American Society for Information Science and Technology, vol. 63, no. 1, 2012, pp. 163–173.

[26] H. Saif, M. Fernández, Y. He, and H. Alani, “Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold,” 2013.

[27] J. Read, “Using emoticons to reduce dependency in machine learning techniques for sentiment classification,” in Proceedings of the ACL Student Research Workshop. Association for Computational Linguistics, 2005, pp. 43–48.

[28] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining.” in LREC, vol. 10, 2010, pp. 2200–2204.

[29] E. Gabrilovich and S. Markovitch, “Computing semantic relatedness using wikipedia-based explicit semantic analysis.” in IJCAI, vol. 7, 2007, pp. 1606–1611.

[30] S. D. Kamvar and J. Harris, “We feel fine and searching the emotional web,” in Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011, pp. 117–126.