A Hybrid Approach to Sentiment Analysis with Benchmarking Results

Orestes Appel[1], Francisco Chiclana[1], Jenny Carter[1], and Hamido Fujita[2]

[1] De Montfort University, Centre for Computational Intelligence (CCI),
Leicester, United Kingdom
orestes.appel@email.dmu.ac.uk,{chiclana,jennyc}@dmu.ac.uk,
[2] Iwate Prefectural University (IPU), Takizawa, Iwate, Japan.
issam@iwate-pu.ac.jp

Abstract. The objective of this article is two-fold. Firstly, a *hybrid approach* to Sentiment Analysis encompassing the use of Semantic Rules, Fuzzy Sets and an enriched Sentiment Lexicon, improved with the support of SentiWordNet is described. Secondly, the proposed hybrid method is compared against two well established Supervised Learning techniques, Naïve Bayes and Maximum Entropy. Using the well known and publicly available *Movie Review Dataset*, the proposed hybrid system achieved higher *accuracy* and *precision* than Naïve Bayes (NB) and Maximum Entropy (ME).

Keywords: Sentiment Analysis, Fuzzy Sets, Semantic Rules, Natural Language Processing, Computational Linguistic, SentiWordNet

1 Introduction

In this section we will cover the basics of Sentiment Analysis (SA), or Opinion Mining (OM) as it is frequently called as well, and the motivation that led us to explore the solution and results we are discussing in this article.

1.1 Sentiment Analysis – Basics

Sentiment Analysis (SA) has been at the front of research efforts for the last few years. The data volumes generated through multiple channels and media are too bulky and complex for human digesting, hence the need for a computer-aided process capable of telling the end-user (a product consumer, a researcher, a teacher, a political analyst, etc.) whether a document, a sentence or a tweet are carrying an opinion or factual information. If it is the former, the users will be keen on telling positive opinions from negatives. Furthermore, there is even room for understanding the degree of positiveness or negativeness of a given piece of information. Typically, SA is performed at specific levels, such as feature/aspect level, sentence level, document level, etc. In this research, we are focusing at carrying SA at the sentence level. For a complete review of the evolution of the Sentiment Analysis field, please refer to the work of [1] and [2].

1.2   Motivation

Most of approaches to address the SA problem belong either to the category of Supervised or Unsupervised Machine Learning. However, it seems to the author that fuzzy sets, considering their mathematical properties and their ability to deal with vagueness and uncertainty, are well equipped as well to model sentiment-related problems. It can be hypothesised that a combination of different techniques could be more effective at succeeding at addressing the SA challenges than specific techniques used in isolation. In the next few paragraphs we will address our motivation to explore this realm of possibilities.

Dzogang et al. stated in [4] that usually authors refer mainly to psychological models when addressing the SA problem. However, other models may be successful as well. As per Dzogang et al. "...it must be underlined that some appraisal based approaches make use of graduality through fuzzy inference and fuzzy aggregation for processing affective mechanisms ambiguity and imprecision...". On the other hand, Bing Liu [10], one of the main world experts in SA, says that "...we probably relied too much on Machine Learning" when dealing with SA. Hence, the following arguments combined together have sparked the research here reported: (i) the concept of graduality expressed through fuzzy sets; (ii) the idea that other alternatives, besides Supervised Machine Learning, may be viable as well when extracting sentiment from text; (iii) the positive contribution that semantic rules and a solid opinion lexicon can have in identifying polarity; and (iv) the success brought in by the use of effective NLP techniques, like parsing and smart-tokenisation.

During the rest of this article we will cover: (I) the Research Methodology used; (II) the Proposed Hybrid Classification Method for SA; (III) a comparison of the experimental results obtained; and (IV) our Conclusions.

2   Research Methodology

The research methodology used is discussed from two different perspectives: the process to follow and the data to use for measuring the performance of the proposed SA solution.

2.1   The Process

In order to measure success, any proposed solution should perform *same or better* than today's most accepted solutions. In the specific case of the SA problem, the proposed hybrid solution is compared against two Supervised Learning methods that enjoy a high level of acceptance and credibility in the classification research community and that are relatively easy to implement: Naïve Bayes (NB) and Maximum Entropy (ME) [10, 14, 15, 13]. The comparison will focus on two aspects at the sentence level: (1) *Subjectivity Determination* (being able to tell an opinion from a fact), and (2) *Opinion polarity/graduality Classification* (assigning a value to an opinion inside a given range [positive, negative, neutral, etc.]).

Our approach will consist of comparing our results with those attained by the Supervised Learning methods described above when identifying subjectivity

and estimating opinion polarity on the subjective content at the sentence level. The focus will be on *Accuracy* and *Precision*.

2.2 The Data

The main dataset used in our research is the *Movie Review Dataset*, published and utilised by Pang and Lee [14]. The results obtained by using the afore-mentioned dataset are addressed in [13, 15]. The fact that many articles in SA discuss this dataset and have used it to validate their own methods and approaches makes it an ideal candidate from the benchmarking angle. The dataset contains 5,331 positive snippets and an equal number of negative ones. Each line corresponds to a single snippet that could contain more than one sentence. As a results-validation dataset, we will use the data presented in *Sentiment140*, which is available at `http://help.sentiment140.com/for-students`.

2.3 Most commonly used measurements in the evaluation of SA

It has become customary to evaluate the performance of sentiment classification systems utilising the following four indexes, as defined in [18] (refer to the so-called *confusion matrix* given in Table 1):

- *Accuracy*: the portion of all true predicted instances against all predicted instances $\equiv \dfrac{(TP + TN)}{(TP + TN + FP + FN)}$
- *Precision*: the portion of true positive predicted instances against all positive predicted instances $\equiv \dfrac{TP}{(TP + FP)}$
- *Recall*: the portion of true positive predicted instances against all actual positive instances $\equiv \dfrac{TP}{(TP + FN)}$
- *F1-score*: a harmonic average of precision and recall $\equiv \dfrac{(2 \times Precision \times Recall)}{(Precision + Recall)}$

Table 1. Confusion Matrix

	Predicted Positives	**Predicted Negatives**
Actual Positive instances	# of True Positive in-stances (TP)	# of False Negative in-stances (FN)
Actual Negative instances	# of False Positive in-stances (FP)	# of True Negative in-stances (TN)

3 A Hybrid Approach to Sentiment Analysis - The Proposed Method

Let us discuss a little further what exactly we mean by utilising a 'Hybrid Approach', a concept that is key to our proposed solution. Our intention is to manage *hybrid concepts* at two different levels: (i) the methods employed by

the sentiment classifiers (Naïve Bayes, Maximum Entropy, Decision Tree, Fuzzy Sets/Logic, and others), and (ii) the techniques utilised to build key components of our approach, like the creation and population of the Sentiment/Opinion Lexicon. Our study will focus on addressing the SA problem at the *sentence level*. The following paragraphs will present the components and processes that encompass our proposed Hybrid Solution.

3.1   The Sentiment Lexicon

Dr. Liu compiled an Opinion Lexicon a few years ago, as mentioned at `http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon`: "it does include a list of positive and negative opinion words or sentiment words for English (around 6800 words) [...] compiled over many years starting from our first paper (Hu and Liu, KDD-2004)" [8]. it was decided to use Liu's Lexicon to re-use data resulting from a quality effort in words compilation. In generating our Opinion Lexicon, we have taken the following approach:

1. We have utilised the *opinion-conveying-words* that are part of the Opinion Lexicon used by Prof. Bing Liu et al. in [8] and other pieces of research work. They correspond to lists containing 'positive meaning words' and 'negative meaning words'. They include *only* nouns, verbs, adjectives and adverbs. These four elements of Part-of-Speech (PoS) have been proven to be capable of delivering opinions [6, 7, 9, 20].
2. We have used SentiWordNet [5] to extract *polarity or valence scores* for words carrying opinion sense.
3. We have combined both elements in (1) and (2) above. As such, we have substituted the words in the original Liu's opinion lexicons for their Synset-equivalent in SentiWordNet (at least partially). This way, we have added a *positive score* and a *negative score* to the existing words in Liu's lexicon, enriching the Lexicon. It is important to keep in mind as well that

$$0 \leq PositiveScore,\ NegativeScore,\ ObjectivityScore \leq 1$$

$$0 \leq (PositiveScore + NegativeScore) \leq 1$$

$$ObjectivityScore = 1 - (PositiveScore + NegativeScore)$$

   As such, when the sum of *PositiveScore* and *NegativeScore* is equal to 1 for a given word $Word_k$, then the term $Word_k$ is fully opinionated, as opposed to the case when the addition of these two scores is zero, in which case the term $Word_k$ is fully Neutral or Objective.
4. The above results in an *improved* Sentiment Lexicon containing lists of Positive and Negative words which have as attributes polarity/valence scores. Here is the description of the attributes of our proposed lexicon:
   Word: word in the lexicon (entries).
   PoS: part of speech (n=noun; v=verb; a=adjective; r=adverb; s=adjective satellite).
   PSC: Positive Score as taken from SentiWordNet [5].
   NSC: Negative Score as taken from SentiWordNet [5].

COBJ: Calculated Objectivity Score [5].

VDX: Versioning index for identifying/managing synonyms (future use).

UPDC: Update Counter to keep track of every time a given entry in the lexicon is updated.

PL: Polarity Label (either **pos** for positive or **neg** for negative)

A typical occurrence of an item in the sentiment lexicon is represented as:
$$(\#(word_k \text{ PoS PSC NSC COBJ VDX UPDC}) \text{ PL}).$$

3.2 Semantic Rules (SR)

In this section we address those cases for which new rules need to be defined in order to model the problem of SA in a more accurate fashion. Indeed, a number of authors, among them [12, 19, 21], have pointed out the fact that negation and the use of specific part-of-speech particles, like 'but, despite, unless, ...' could affect the final outcome of a classification exercise. Thus, some rule strategies are needed to be put in place as the order of the different part-of-speech play a role in the semantic of a sentence. Researchers have been, through time, improving the quality of these semantic rules so that they are more encompassing of the possible cases that must be managed. These research efforts are summarised by Xie et al. in a very well organised, easy to read semantic rule tables in [21]. Despite the apparent completeness of existing Semantic Rules, *two new rules* are incorporated for managing particular part-of-speech particles: the particle **while** and the particle **however**, which were not included in the original set of rules provided in [21], resulting in the following set of Semantic Rules (Table 2).

3.3 Negation Handling

According to Dr. Christopher Potts from Stanford University, Linguistics Department `http://sentiment.christopherpotts.net/lingstruc.html`, "Sentiment words behave very differently when under the semantic scope of negation" [17]. The complex nature of *negation* suggests that it would be difficult to have a general *a priori* rule for how to handle negation. The technique that Dr. Potts favours for approximating the effects of negation is due to Das and Chen [3] and Pang, Lee, and Vaithyanathan [15]. This method utilises *Regular Expressions* and has been implemented in our research effort as an extension to the publicly available Tokenizer that we have used [16]. Notice that even long-distance negation effects can be managed using this technique.

3.4 Linguistic Variables and Fuzzy Sets

According to G.A. Miller [11], *7 plus or minus 2*, is the effective number of categories that a subject (individual or person) can maintain. In our case, we have chosen a conservative approach and have devised 5 labels (7 minus 2), symmetrically distributed in the interval $[0 \ldots 1]$. Our choice of trapezoidal function obeys to the fact that the latter generalises a triangular function and we have aimed for more generality and for more than one value at the top of every category. We have opted for linguistic modifiers that have the ability to change the level of granularity. In our research, the classification labels for the **intensity** of semantic orientation and/or polarity of a given sentence are identified as: (a) Poor,

Table 2. Semantic Rules for proposed Hybrid System

Rules	Semantic Rules	Example
R1	Polarity (not var_k) = -Polarity (var_k)	'not *bad*.'
R2	Polarity (NP_1 of NP_2) = *Compose* (NP_1, NP_2)	'*Lack* of crime in rural areas.'
R3	Polarity (NP_1 VP_1) = *Compose* (NP_1, VP_1)	'*Crime* has decreased.'
R4	Polarity (NP_1 be ADJ) = *Compose* (ADJ, NP_1)	'*Damage* is minimal.'
R5	Polarity (NP_1 of VP_1) = *Compose* (NP_1, VP_1)	'*Lack* of killing in rural areas.'
R6	Polarity (ADJ to VP_1) = *Compose* (ADJ, VP_1)	'*Unlikely* to destroy the planet.'
R7	Polarity (VP_1 NP_1) = *Compose* (VP_1, NP_1)	'*Destroyed* terrorism.'
R8	Polarity (VP_1 to VP_2) = *Compose* (VP_1, VP_2)	'*Refused* to deceive the man.'
R9	Polarity (ADJ as NP) = $1_{(Polarity(NP=0))}$ · Polarity(ADJ) + $1_{(Polarity(NP\neq0))}$ · Polarity(NP)	'As *ugly* as a rock.'
R10	Polarity (not as ADJ as NP) = -Polarity (ADJ)	'That wasn't as *bad* as the original.'
R11	If sentence contains "but", disregard all previous sentiment and only take the sentiment of the part after "but".	'And I've never liked that director, *but* I loved this movie.'
R12	If sentence contains "despite", only take the sentiment of the part before "despite".	'I love the movie, *despite* the fact that I hate that director.'
R13	If sentence contains "unless" followed by a negative clause, disregard the "unless" clause.	'Everyone likes the video *unless* he is a sociopath.'
R14 (New)	If sentence contains "while", disregard the sentence following the 'while' and take the sentiment only of the sentence that follows the one after the 'while'.	'*While* they did their best, the team played a horrible game.'
R15 (New)	If sentence contains "however", disregard the sentence *before* 'however' and take only the sentiment of sentence *after* 'however'.	'The film counted with good actors. *However*, the plot was very poor.'

Table 3. Compose functions referenced in Table 2

Compose Functions	Algorithms
Compose1 (arg1, arg2)	1. Return **-**Polarity(*arg2*) if *arg1* is negation.
	2. Return Polarity(*arg1*) if (Polarity(*arg1*) = Polarity(*arg2*).
	3. Otherwise, return the majority term polarity in *arg1* and *arg2*.
Compose2 (arg1, arg2)	1. Return Polarity(*arg2*) if *arg1* is negative and *arg2* is not neutral.
	2. Return **-1** if *arg1* is negative and *arg2* is neutral.
	3. Return Polarity(arg2) if *arg1* is positive and *arg2* is not neutral.
	4. Return 2*****Polarity(arg1) if Polarity(*arg1*) = Polarity(*arg2*).
	5. Return Polarity(*arg1*) + Polarity(*arg2*) if *arg1* is positive and *arg2* is neutral.
	6. Return Polarity(*arg1*) + Polarity(*arg2*) if *arg2* is positive and *arg1* is neutral.
	7. Otherwise, return **0**.

(b) Slight, (c) Moderate, (d) Very and (e) Most. In essence, we either classify a sentence as *Objective* or *Subjective*. If it is considered as *Subjective*, then it could either be *Negative* or *Positive*, with an intensity as qualified by the modifiers belonging in the set $\{Poor, Slight, Moderate, Very, Most\}$. A generic trapezoidal membership function would take the following form as shown in Fig. 1 and described in Equation 1.
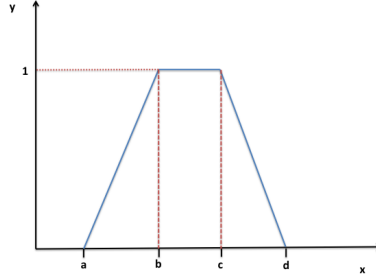

Fig. 1. General form of the trapezoidal family of curves

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{x-a}{b-a} & \text{if} \quad a \leq x \leq b; \\ 1 & \text{if} \quad b \leq x \leq c; \\ \frac{d-x}{d-c} & \text{if} \quad c \leq x \leq d; \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

A trapezoidal membership function (MF) can be represented as well using the following 4-tuple (a, b, c, d). Our MFs in 4-tuple format are as follows:

– MF (Poor): $(0, 0, 0.050, 0.150)$
– MF (Slightly): $(0.050, 0.150, 0.250, 0.350)$
– MF (Moderate): $(0.250, 0.350, 0.650, 0.750)$
– MF (Very): $(0.650, 0.750, 0.850, 0.950)$
– MF (Most): $0.850, 0.950, 1, 1)$

Let us refresh some fuzzy sets theory concepts.

– T-norm $(\min(a, b))$: it is a binary function $T : [0, 1] \times [0, 1] \to [0, 1]$ that is commutative, associative, monotonic, and $T(a, 1) = a$.
– T-conorm $(\max(a, b))$: it is a binary function $S : [0, 1] \times [0, 1] \to [0, 1]$, that is commutative, associative, monotonic, and $S(a, 0) = a$.
– Any t-conorm, S, can be generated by a t-norm, T, and viceversa: $S(a, b) = 1 - T(1 - a, 1 - b)$. When this is the case (T, S) is said to be a dual pair of t-norm and t-conorm, as it is the pair $(\min, \max)$.

3.5 Calculating Polarity Scores for Sentences

In order to illustrate the mechanism used, let us calculate the polarity of a given sentence S_k. As an example, let us assume that out of the 11 words included in sentence S_k, 4 are in our lexicon and will be used to determine the polarity of the sentence. *Sentence (S_k)*: "Many good actors. However, the film was simplistic, silly, unrealistic and tedious". The *sentiment-conveying words* in S_k, are:

– Good: (#(good s 0.75 0.0 0.25 0 0) pos)
– Simplistic: (#(simplistic s 0.0 0.0 1.0 0 0) neg)
– Tedious: (#(tedious no-info no-info no-info no-info 0 0) neg)
– Unrealistic: not in our Opinion Lexicon
– Silly: (#(silly n 0.125 0.0 0.875 0 0) neg)

Our approach to determining the qualified semantic orientation of the sentences belonging in our test dataset, involves a two-step approach, as described below. Please notice that we will present a summarised version of the processes behind the proposed Hybrid Method.

Step I: Hybrid Standard Classification (HSC): This process is performed by using a number of techniques, as follows. Every intermediate step has as an outcome a list of features that is passed to the next sub-step in the process chain.

1. Takes as input the output from the Tokeniser, PoS tagging and Smart Parsing (negation is partially handled here)
2. Applies the Semantic Rules presented in Table 2.
3. Extracts essential particles that convey sentiment/opinion (adjective, nouns, verbs and adverbs). It generates a list with those key PoS particles
4. Searches the Sentiment Lexicon and substitutes the words in the list-expression with the associated matches in the lexicon
5. Generates an exception list for those key-words that are not in the Opinion Lexicon
6. Calculates the semantic orientation of each sentence by taking into consideration: (a) the word-label semantic orientation present in the Lexicon (POS, NEG or OBJ), and (b) the Positive and/or Negative Scores of the words in the sentence that appear in the lexicon. The calculations are performed as per the Semantic Rules and a classification label belonging in the set {Positive, Negative, Objective, No-Semantic-Orientation} is assigned to each sentence.
7. Re-scans the resulting list in search for words still labelled as OBJ that could be now converted into either POS or NEG labels. This situation is resolved by using the services of a dictionary previously generated for all sentences being processed. The structure of the Dictionary is [Word, Sentence-numbers-where-Word-appears, Frequency-of-Appearances].
8. Generates a new list with the classification POS/NEG of all sentences in the test dataset (particles initially marked as OBJ, are substituted with their final POS/NEG labels as per resolved in the previous step).

Step II: Hybrid Advanced Classification (HAC): This approach enhances the standard classification process by incorporating:

1. Determination of the degree with which a given sentence leans towards being *positive* or *negative* (a fuzzy approach rather than a crisp method). Let us assume that IP corresponds to the Intensity of Polarity of a given word, $Word_i$

$$\min\Big(IP(Word_1) \ \ldots \ IP(Word_n)\Big) = \delta,$$

where $Word_i$ is the i^{sm} sentiment-carrying word in a given sentence, and $i = 1 \dots numwords$, where $numwords$ corresponds to the total number of sentiment-conveying words, n, found in the Opinion Lexicon for a given sentence S_k; $\delta \in [0 \dots 1]$.

$$\max \Big(\mu_{poor}(\delta), \ \mu_{slight}(\delta), \ \mu_{moderate}(\delta), \ \mu_{very}(\delta), \ \mu_{most}(\delta) \Big) = \beta,$$

where, μ_j, with $j \in \{poor, slightly, moderate, very, most\}$, corresponds to the evaluation of the membership function μ_j for the sub-index j; $\beta \in [0 \dots 1]$. In essence, we calculate the T-conorm represented by the value β.

2. Diagnosing when a given sentence could be considered rather *objective/neutral* as opposed to either *positive* or *negative*. Not all sentences have been created equal, and even in the test dataset that has been carefully chosen, there are some sentences that one could argue are rather neutral (not leaning towards negative or positive). With the Advance Classification system we will ponder which sentences could be borderline neutral/objective.

4   Experimental Results

In this section we will look at the experimental results, starting with the outcome of using two ML supervised learning methods (Naïve Bayes and Maximum Entropy), and then we will show the results obtained applying our hybrid method.

4.1   Naïve Bayes Classifier (NB)

We trained the NB classifier using some of the recommendations presented by Perkins in [16]. The classifier uses the concept of 'bag of words'. Using this representation model of a sentence, the classifier creates 'feature vectors' exhibiting the main traits of such a sentence. The NB classifier is a binary classifier. It will classify a sentence either as 'negative' or 'positive', and these categories are exclusive. The classifier returns a probability value that represents the *probability* that the sentence belongs with a specific label (negative or positive). The probability value has to be *0.5* or higher for a sentence to belong in a specific category. The results of applying the NB classification algorithm to the data set created by Pang and Lee in [13] are summarised in Table 4.

4.2   Maximum Entropy (ME)

We trained the ME classifier using Perkins's recommendations as presented in [16]. We used the Generalized Iterative Scaling (GIS) learning method to train the ME classifier. As in the case of the Naïve Bayes classifier, the ME classifier returns a probability value that represents the *probability* that the sentence belong with a specific label (negative or positive). The probability value has to be of *0.5* or higher for a sentence to belong in a specific category. Once we applied the trained classifier to our test dataset we obtained the results presented in Table 5.

4.3 Our Hybrid Method: HSC/HAC

Our Hybrid Method utilises a number of different components, coming from different disciplines, to achieve the results presented in Table 6, which are very encouraging when compared to the outputs obtained by applying the NB and ME techniques. In addition, when we incorporate the fuzzy set approach (HAC) we can provide a much better granularity level in the classification process. See Table 7 and Table 8 for details.

Table 4. NB Results

Accuracy	0.67
Precision	0.63
Recall	0.85
F1-score	0.72

Table 5. ME Results

Accuracy	0.68
Precision	0.63
Recall	0.86
F1-score	0.73

Table 6. HSC/HAC Results

Accuracy	0.76
Precision	0.73
Recall	0.83
F1-score	0.77

Table 7. HAC Classifier Results for **POS** Dataset - *Increased Granularity*

False Negatives	929
No Semantic Orientation	35
True Positives	4,402
Poor	577
Slight	1,106
Moderate	1,041
Very	1,365
Most	313
Number of Snippets	5,331

Table 8. HAC Classifier Results for **NEG** Dataset - *Increased Granularity*

False Positives	1,646
No Semantic Orientation	76
True Negatives	3,685
Poor	770
Slight	1,089
Moderate	789
Very	864
Most	173
Number of Snippets	5,331

As an example, the intensity of sentence 'The movie was simplistic, silly and tedious' (classified as *Most Negative*) is certainly stronger than the one exhibited by the phrase 'An alternately frustrating and rewarding experience' (labelled by HAC as *Slightly Negative*).

4.4 Comparison of Results

In this subsection we will take a closer look at the results that so far we have obtained. We present the metrics compiled for two other methods and our two hybrid approaches. The results are very encouraging, especially with respect to Accuracy, Precision and F1-score. However, the Recall indicator -known as well as Sensitivity- is better for the results shown by the NB/ME methods. Experiments show that the NB/ME method does slightly better than HSC/HAC for identifying positive sentences, but it does much worse than HSC/HAC when classifying negative snippets. As Recall represents the portion of true positive predicted instances against all actual positive instances, it seems reasonable that NB/ME carries a better Recall than HSC/HAC. In general, our method HSC/HAC shows good improvements when compared to the results achieved by the NB/ME method. The main difference between the results achieved with

HSC and HAC, is that the latter adds a fuzzy approach that provides not only a classification result (POS or NEG), but in addition, supplies important data in establishing the intensity or strength with which a given sentence is positive or negative. This outcome could be used in the future to determine additional properties and characteristics, like sentences that are borderline between being subjective or objective.

5   Conclusions

In general, our proposed hybrid system works very well with a high level of *accuracy* and *precision*. Indeed, the fact that our hybrid system improved the results obtained when we applied Naïve Bayes (NB) and Maximum Entropy (ME) to the same dataset *satisfies* one of our initial hypotheses, that a hybrid method using natural language processing techniques, semantic rules and fuzzy sets should be able to perform well. Additionally, by the utilisation of fuzzy sets we can determine when a given sentence has a stronger/weaker intensity in terms of polarity. In closing, there are some lessons learned and observations that we would like to share: (a) using efficient NLP techniques (like tokenising, parsing, negation handling, etc.), contribute positively to the application of our Hybrid Method; (b) the creation of an *improved* Sentiment Lexicon was decisive in obtaining good experimental results; (c) SentiWordNet became an important component of our proposed solution and certainly *enriched* dramatically the quality of our Lexicon. Our expectation is that the quality of the content of SentiWordNet would continue improving with time, reflecting positively in the performance of our Hybrid Method; (d) work should continue on improving the completeness and quality of Semantic Rules. In essence, hybrid techniques can play an important role in the advancement of the SA discipline by combining together the elements we described in our research contribution.

References

1. K. S. M. Anbananthen and A. M. H. Elyasir. Evolution of opinion mining. *Australian Journal of Basic and Applied Sciences*, 7(6):359–370, 2013.
2. O. Appel, F. Chiclana, and J. Carter. Main Concepts, State of the Art and Future research Questions in Sentiment Analysis. *Acta Polytechnica Hungarica - Journal of Applied Sciences*, 12(3):87–108, May/June 2015.
3. S. R. Das, M. Y. Chen, T. V. Agarwal, C. Brooks, Y. shee Chan, D. Gibson, D. Leinweber, A. Martinez-Jerez, P. Raghubir, S. Rajagopalan, A. Ranade, M. Rubinstein, and P. Tufano. Yahoo! for amazon: Sentiment extraction from small talk on the web. In *8th Asia Pacific Finance Association Annual Conference*, 2001.
4. F. Dzogang, M.-J. Lesot, M. Rifqi, and B. Bouchon-Meunier. Expressions of Graduality for Sentiments Analysis – A Survey. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–7, 2010.
5. A. Esuli and F. Sebastiani. SentiWordNet – A Publicly Available Lexical Resource for Opinion Mining. *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*, pages 417–422, 2006.

6. V. Hatzivassiloglou and K. McKeown. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In L. K. Schubert, editor, *ACL: 31st Annual Meeting of the Association for Computational Linguistics, 22-26 June 1993, Ohio State University, Columbus, Ohio, USA, Proceedings*, pages 172–182. ACL, 1993.

7. V. Hatzivassiloglou and K. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL. New Brunswick, NJ, USA: ACL*, pages 174–181, 1997.

8. M. Hu and B. Liu. Mining and summarizing customer reviews. *Proceedings – ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004 full paper), Seattle, Washington, USA, Aug. 22–25*, 2004.

9. J. KAMPS, M. Marx, R. J. Mokken, and M. de Rijke. Using WordNet to measure semantic orientations of adjectives. In *Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation*, volume IV of *LREC '04*, pages 1115–1118, 2004.

10. B. Liu. *Sentiment Analysis and Opinion Mining.* Morgan and Claypool Publishers: Synthesis Lectures on Human Language Technologies, 1st edition, 2012.

11. G. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Rev.*, 63:81–97, 1956.

12. S. Nadali, M. Murad, and R. Kadir. Sentiment classification of customer reviews based on fuzzy logic. *In Information Technology (ITSim), 2010 International Symposium in (Volume:2), Kuala Lumpur, Malaysia*, 2:1037–1040, June 2010.

13. B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL '05)*, ACL'05, pages 115–124, 2005.

14. B. Pang and L. Lee. *Opinion mining and sentiment analysis.* NOW: the essence of knowledge, Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1–2, pp. 1–135, 2008.

15. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 10:79–86, 2002.

16. J. Perkins. *Python Text Processing with NLTK 2.0 Cookbook.* Packt Publishing, 2010.

17. C. Potts. Sentiment Symposium Tutorial: Linguistic structure (part of the Sentiment Analysis Symposium held at San Francisco, november 8-9, 2011). Stanford Department of Linguistics, Stanford University. Accessed date: December 2014, November 2011.

18. M. Sadegh and R. I. Z. A. Othman. Combining lexicon-based and learning-based methods for twitter sentiment analysis. *International Journal of Computers & Technology*, 2(3):171–178, June 2012.

19. P. Subasic and A. Huettner. Affect Analysis of Text Using Fuzzy Semantic Typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496, August 2001.

20. J. Wiebe. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740. AAAI Press, 2000.

21. Y. Xie, Z. Chen, K. Zhang, Y. Cheng, D. K. Honbo, A. Agrawal, and A. N. Choudhary. MuSES: a multilingual sentiment elicitation system for Social Media Data. *Intelligent Systems, IEEE*, 29(4):34–42, July 2014.