Context Incorporation Using Context - Aware Language Features

Aggeliki Vlachostergiou
School of Electrical and
Computer Engineering
National Technical University of Athens
aggelikivl@image.ntua.gr

George Marandianos
School of Electrical and
Computer Engineering
National Technical University of Athens
Gmarandianos@image.ntua.gr

Stefanos Kollias
School of Computer Science
University of Lincoln
Lincoln, UK
skollias@lincoln.ac.uk

***Abstract*—This paper investigates the problem of context incorporation into human language systems and particular in Sentiment Analysis (SA) systems. So far, the analysis of how different features, when incorporated into such systems, improve their performance, has been discussed in a number of studies. However, a complete picture of their effectiveness remains unexplored. With this work, we attempt to extend the pool of the context - aware language features at the sentence level and to provide the foundations for a concise analysis of the importance of the various types of contextual features, using data from two different in type and size datasets: the Movie Review Dataset (MR) and the Finegrained Sentiment Dataset (FSD).**

***Index Terms*—Human language technology, Sentiment Analysis, Context - aware language features, CRF, discourse RST**

I. Introduction

Incorporating context into human language technology and particular to SA systems has been successfully applied to various applications and domains. Among these, opinion expression tasks, opinion summarization, sentiment and opinion retrieval, analysis of market trends, business decisions [1], etc. are included. Context-aware Sentiment Analysis (SA) systems, as a more "novel" type of SA systems, are able to additional take context into consideration to further identify longer segments along with their aligned features that are indicative of the existence of sentiment.

For example, a person may choose to review a product positively or negatively based on his current attitude, opinion or emotion towards that certain product. Even in his most enthusiastic reviews, there is still a possibility of mentioning negative aspects of that particular product. Conversely, in very negative reviews there might still be mentions of several positive aspects of the product. Moreover, it is possible to retrieve different opinions, which can even be uttered in the same sentence. Considering for instance, the sentence: "Despite the pretty design I would never recommend it, because the sound quality is unacceptable" which expresses both assertive and negative opinions about a product. In this light, to determine both negative and positive sentences in product reviews, exploring the type of context-aware features should be examined in depth. Researchers believe that SA systems cannot stand alone without considering context, as such information w.r.t. the understanding of the sentiment of sentences assists to unfold the opinions collected online [2].

During the past decade, several context-aware SA systems have been proposed to increase the systems' performance. Figure 1 summarizes a number of the predefined parameters during the experimental setting which affect the systems' performance. In a more detailed way, it can be inferred that different approaches on how the subjectivity is extracted from natural language text [2], [3], how the sentiment is measured, which are the sentiment-carrying words in texts and how are we measuring them (e.g. word frequencies), what is the size and the source of the examined opinion repositories, at which level of analysis have they been examined (sentence-level, phrase-level, document-level) [2], [3] and finally how complicated is the sentence structure (e.g. when conjuction words and comparisons are included), may reduce or make more challenging the analysis of the performance.

At the same time, sentiment incorporation through context - aware features has reached some encouraging results. Specifically, different machine learning techniques investigate patterns in text's vector representations, while at the same time lexicon-aware methods [4] account for semantic orientation (i.e. positive, negative or neutral opinion) in individual words. However, the main strength of the lexicon-aware approaches is also their weakness. Considering that the lexicons' content is predefined, they cannot adapt to novel (domain) forms of expressions [3].

Additional attempts of accounting for structural aspects of opinionated text is the analysis of the documents' rhetorical structure to retrieve sentiments and opinions [4], [5]. To fulfill this requirement, the Rhetorical Structure Theory (RST) [6] is applied, to identify the rhetorical roles of text segments. This is of particular importance in SA, considering that sentiment words are expected to contribute differently to the overall sentiment depending on the examined text segment. For instance, within the sentence: "we saw a movie that was awful, but it was nice walking with you after the cinema", the subsentence (but it was nice walking with you after the cinema) could be counted as negative here, in order to account for the contrasting rhetorical relation between the two segments.

In this paper, we extract a number of context - aware language features at the sentence level and we further extend those features that are already used in SA systems. We also avoid at this point w.r.t. our experimental purposes to use

ISBN 978-0-9928626-7-1

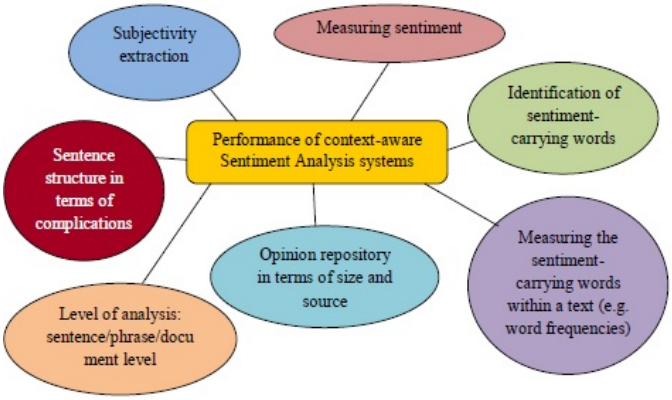

Fig. 1: Parameters which affect the performance of context-aware Sentiment Analysis systems.

existing affective lexica, as we would like to examine our system's performance without incorporating predefined lists of words that could not be adapted at a later point to other domains. So far, a wide gamut of them has been tested by the majority of research teams, mainly in constrained environment [7]–[10]. Table I summarizes the main characteristics based on which, the discussed studies have been performed in this area. With this in light, we assure that it is not yet confirmed the impact of every feature set as well as how the combination of some features behave with different in size and genre of information sources. Motivated by this, in this paper we explore the relative importance of a number of context - aware features incorporated into our SA system through the Conditional Random Fields (CRFs) and the RST methods and present our observations.

II. Related Work

An important research direction in context-aware SA is improving the robustness of SA systems after incorporating context. With this in mind, the recent advancements in SA have been revised in depth in [2], [3]. For a summarized survey on the SA field, the reader is referred to the works presented in [13]–[15]. Existing methods count on sentiment lexicons, which are enumerative lists of sentiment terms used to indicate the sentiment changes. Representative examples of such sentiment lexicons range from General Inquirer [16], Subjectivity Lexicon and Subjectivity Sense Annotations [3] to SentiWordNet [3]. Additionally, domain knowledge plays a key role, since a sentiment term's linguistic context often affects its sentiment charge. An initial attempt on SA was based on syntactic relations to identify new sentiment terms, that have been further considered as an early form of context exploitation [17]. Moreover, additional research works such as [18] and [19] started to realize that it difficult to identify context when processing sentences or paragraphs in isolation, underling the need of context incorporation to advance the human language technology systems w.r.t. SA. Particularly, in [18], the authors suggest that inferential language models outperform conventional models without context processing capabilities. Moreover, in [19], the impact of context linguistic features along with their combinations on the polarity of terms is examined. Their results suggest an improvement in accuracy.

On the whole, Sentiment Analysis research is organized into two main categories: the Machine Learning and the Lexicon-aware SA. According to Machine Learning SA, a typical text representation refers to a list of terms that appear in documents unordered. Particularly, a binary representation w.r.t. the presence or absence reached the accuracy percentage of 87.2% on Movie Review Dataset [9]. Moreover, adding extra phrases to further express sentiment as features in the binary form ended up into even higher results of 90.6% [20]. We should mention at this point that the highest result of 96.9% was achieved by [2], who used the term-frequency/inverse-document-frequency (tf-idf) weighting method. Furthermore, additional approaches include bigram feature selection mechanisms as the work presented in [21]. However, despite their high performance, these Machine Learning sentiment classifiers appear to adapt poorly. That could be explained if we consider that they often rely on particular features extracted from their domain's training data. Moreover, an additional deficiency of these methods is that it is time-consuming and costly as human-labeled data is required.

On the other hand, as far as the Lexicon-aware SA method concerns, a text representation or learning step prior to the analysis is not a presupposition step. From a general perspective, sentiment lexicons are either a) manually generated and in this case humans assign the sentiment polarity values usually of limited coverage, or b) automatically generated. Among the manually generated lexicons the General Inquirer [16] and the Opinion Lexicon, are included. W.r.t. the automatically generated lexicons, the corpus-based and the dictionary-based are the two most common methods, starting both with a small set of seed terms. To be more precise, having an assertive seed set such as 'good', 'kind' and 'excellent' and a negative seed set containing terms such as 'bad', 'evil' and 'wicked', the above presented methods investigate to unfold the connections between terms to expand these sets. The main difference between these two methods is that corpus-based uses collection of documents while the dictionary-based uses lexical resources (i.e. machine-readable dictionaries). Representative prior work in the field, w.r.t. the Lexicon-aware method is presented in [17]. At a later time, this work has been extended to include conjoining adjectives in a document collection based on the conjunctions 'and' and 'but', to further indicate similarity and contrasting polarities between the conjoining adjectives respectively.

III. Methodology and Experiments

A. CRF and RST context incorporation methods

Conditional Random Field (CRF): We set our experimental procedure as follows. We assume that the input is a set of m document:$\{d_1, d_2, ..., d_m\}$ along with the specified subject:$\{sub_1, sub_2, ...sub_m\}$. Each d_i contains n_i sentences S^i: $\{s^{i_1}, s^{i_2}, ..., s^i_{n_i}\}$. The output for all documents

ISBN 978-0-9928626-7-1 © EURASIP 2017

Work	Type of Data	Level of Analysis	Size	Classification type	Features used	Features that perform best
Pang et al. [7]	Movie Reviews	Docs.	1400	pos/neg	voc,pos,p	**voc**
Turney [8]	Reviews	Docs.	410	pos/neg	sp	**sp**
Pang and Lee [9]	Movie Reviews	Docs.	2000	pos/neg	voc,p	**voc**
Beineke et al. [10]	Movie Reviews	Sents.	2500	Summarization	voc,p	**voc+p**
Wiebe and Riloff [11]	Press articles	Sents.	9289	subj/obj	pos,sp	**pos**
Taboada et al. [4]	Reviews	Docs.	400	pos/neg	pos, sw, d	**pos + sw + d**
Heerschop et al. [5]	Movie Reviews	Docs.	1000	pos/neg	sw,d	**sw + d**
Katz et al. [12]	Hotel Reviews	Docs.	30.000	pos/neg	voc,pos,sp,p	–
Katz et al. [12]	Movie Reviews	Docs.	2000	pos/neg	voc,pos,sp,p	–

TABLE I: Main characteristics of publicly available datasets. The table reports the Type of Data, the Level of Analysis, the Size collection, the Classification type and the features considered. The feature set labeling is as follows: Vocabulary: unigrams and bigrams (voc), Part-Of-Speech (pos), Sentiment Words (sw), Syntactic Patterns (sp), Position (p) and Discourse (d).

is that for the j^{th} sentence in the i^{th} document s_j^i, it will assign a sentiment $o^{ij} \in \{P : positive, N : negative\}$ and a sentiment $o^{ij} \in \{S : subjective, O : objective\}$ respectively. Conditional Random Fields (CRFs) provide a probabilistic framework for calculating the probability of label sequences Y globally conditioned on sequences data X to be labeled. Parameters Θ=λ_k, μ_l are estimated by maximizing the conditional log-likelihood function $\mathcal{L}(\Theta)$ of the training data.

$$P(\mathrm{Y|X})=\frac{1}{Z_X} exp(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, X)+\sum_{i,l} \mu_l g_l(y_i, X)) \quad (1)$$

where Z_X is the normalization constant.

$$\mathcal{L}(\Theta)=\sum_{j=1....M} log(P(Y^j|X^j;\Theta)) - \sum_k \frac{\lambda_k^2}{2\sigma_k^2} - \sum_l \frac{\mu_l^2}{2\sigma_l^2} \quad (2)$$

An extensive explanation of the parameters' notation used in our experimental setting is explained in detail in [22].

Particularly, we use the CRF++ 0.58[1], to build our linear CRF chain, with a one-to-one correspondence between states and labels. Our aim is to capture the context information (e.g. the neighboring sentences within a document or the sentences connected by transition words within the sentences. In other words, the aim of our sequence labeling model is to give a label to each sentence corresponding to the sentence sequence.

Rhetorical Structure Theory (RST): We use and extend the method presented in [5], [23] to take advantage of the discourse relations in the text to compute sentiment values. Specifically, these approaches use a parser that implements the Rhetorical Structure Theory (RST) [6] to identify the discourse elements in the text. As a result, important discourse parts, are given a higher weight, while parts that are less relevant a lower one. Particularly, when such discourse relations have been applied, the authors of [5] reported an F_1-score of 15%. To find the discourse relations in text, after completing the preprocessing of our data, we segment it into EDUs and finally we use the tool presented in [24] to parse the discourse (DPLP), to further create the RST trees for the individual sentences based on the suggested given feature set. The RST parser further generates the bracketing file for each document, which can be used for evaluation.

B. Context - aware language features

So far, a wide variety of context - aware features has been widely extracted from sentences for SA. In this work, with the term "context - aware language features" we consider the following set of features. For an even more extended description of the features we used, the reader is directed to [25].

- **Vocabulary (Uni., Uni. and Bigrams)**.These features are based on the existence of unigrams and bigrams within the sentence.
- **Length**.We include a number of positive and negative words respectively that appear in our sentences (1948 positive and 4550 negative words respectively) [26].
- **Sentiment words**. We include a list of 52 positive and 35 negative emoticons and a list of comparative adjectives, adverbs, superlative adjectives and adverbs or phrases ("compare to", "in contrast", etc.), as well as conjunction words and subordinating, coordinating and correlative conjunctions words.
- **Positional**. We include features that refer to the sentence position within the document. We define the beginning sentence as the one within the first 20% of the sentences, and the end sentence as the one within the last 20% respectively. Additionally, we consider the position of the positive and negative words respectively within the sentence.
- **Context-aware RST**. These binary features correspond to all types of the RST relationships.
- **All**. All features combined together.

C. Datasets

Table II shows the distribution of the datasets we are using: the Movie Review Dataset (MR) [9] and the Finegrained Sentiment Dataset (FSD) [27]. We use these datasets to evaluate the analysis performance, which we randomly split into a training and a test set of 75% and 25% respectively. Finally, the preprocessing we apply, includes tokenization and sentence splitting, part-of-speech (POS) tagging, lemmatization, NER, parsing, and coreference resolution based on the Stanford CoreNLP framework [28].

D. Experimental Results and Discussion

For our experimental purposes, we use the Support Vector Machines (SVMs) (linear classifiers). We select this classifier due to their capability to produce remarkable results than many other models [29]. After selecting the best classifier, we

[1]https://taku910.github.io/crfpp/

ISBN 978-0-9928626-7-1 © EURASIP 2017

	Test Collections			
Datasets	*Subj./Pos.*	*Obj./Neg.*	*Uni.*	*Bi.*
MR [9]	5000	5000	4948	9103
FSD [27]	923	1320	1275	1996

TABLE II: Test collections for investigating the two-class categorization analysis problem. The table includes the number of unigrams (Uni.) and bigrams (Bi.) after preprocessing.

optimize the classifiers by applying a 5-fold cross-validation on the training data. The aim of our evaluation is two-fold. First, we are interested in investigating the performance of our system compared to the baselines when context-aware language features are incorporated. Secondly, considering that the current work is an ongoing research work, which we would like to extend by applying our proposed method into additional social media platforms, we are interested in evaluating the contribution of each individual context-aware language feature of our system. For both collections, we further validate with the test set, the classifier that performed the best during the training time.

Table III, shows classification results on the MR and FSD datasets. Bold font corresponds to the best performance on a dataset and the significant difference from the baselines. It is shown that there is no differentiation between the subjective and objective sentences when length features are used. One possible explanation would be that they might not be more accurate that the baselines classifiers (unigrams and unigrams combined with bigrams). In the MR collection we do not have positional features. However, we suspect that they will work particularly well for detecting subjectivity content. Such a result is expected, if we consider that for example in the conclusion of a document it is easier to detect a subjective opinion. As far as the context-aware RST features concern, these features lead to decent improvements over the baselines. A potential explanation for that, would be that not much information is presented into the relations among the text's sentences compared to the baselines. According to our results w.r.t. the FSD collection, we observe that the length, positional and the context-aware RST features contribute to a limited extend to the overall analysis performance. Combining all features together performs best on the MR dataset compared to the FSD. Particularly, we achieved marginal improvement while the reverse was not reported on the FSD dataset. This can be attributed to the fact that context-aware features are more likely to appear in the MR than in FSD due to the lengthier documents.

Moreover, the number of Sub./Obj. sentences and unigrams and bigrams respectively, which is about four times the size of the FSD, makes it more likely for MR to have a better lexical and length coverage, hence a better performance. Overall, our proposed method of merging the CRF and the RST methods to extend the pool of context-aware language features, performs better on the examined datasets, even when improvements are marginal. Finally, the improvement is more pronounced on the MR dataset.

On the whole, attempting to compare our overall results, we observe that even though some of our applied features are useful to detect subjective and objective comments, still they do not seem that able to detect positive and negative opinions (FSD collection), or at least more accurately compared to the baseline results.

IV. Conclusions

In this paper, we provide an extended set of context - aware language features for context - aware SA systems. We use the CRF and the RST methods to incorporate this feature set into two different in size and genre datasets. Moreover, we explore how each set of features behaves against these two datasets and we observe a number of interesting tendencies. Among these, the most interesting is that in both datasets the combination of the vocabulary feature set with the sentiment words feature set provides good performance. Also, the incorporation of the context - aware RST features provides slightly improvements compared to the baselines in the MR Dataset. Finally, for the future work, we intend to develop additional context - aware features that will be tested in an even wider range of datasets. Social media platforms that will differ from the other two in the size and their composition combined with the presented context incorporation methods we examined are expected to provide a more detailed comparative analysis of every context-aware language feature's role. Finally, we intend to compare our context - aware incorporation method with Deep Neural network approaches.

References

[1] D. Bal, M. Bal, A. Van Bunningen, A. Hogenboom, F. Hogenboom, and F. Frasincar, "Sentiment analysis with a multilingual pipeline," in *Web Information System Engineering–WISE*. Springer, 2011, pp. 129–142.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[3] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, no. 2, pp. 15–21, 2013.

[4] M. Taboada, K. Voll, and J. Brooke, "Extracting sentiment as a function of discourse structure and topicality," *Simon Fraser Univeristy School of Computing Science Technical Report*, 2008.

[5] B. Heerschop, F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong, "Polarity analysis of texts using discourse structure," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 1061–1070.

[6] W. C. Mann and S. A. Thompson, "Rhetorical structure theory: Toward a functional theory of text organization," *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 8, no. 3, pp. 243–281, 1988.

[7] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10. Association for Computational Linguistics, 2002, pp. 79–86.

[8] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on Association for Computational Linguistics*. ACL, 2002, pp. 417–424.

[9] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. ACL, 2004, p. 271.

[10] P. Beineke, T. Hastie, C. Manning, and S. Vaithyanathan, "Exploring sentiment summarization," in *Proceedings of the AAAI spring symposium on exploring attitude and affect in text: theories and applications*, vol. 39, 2004.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Features (MR)	Subjective			Objective			Microavg
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	F_1
Unigrams	0.8939	0.8910	0.8924	0.8916	**0.8944**	0.8930	0.8927
Length	0.8614	**0.8940**	0.8774	0.8901	0.8565	0.8730	0.8752
Positional	-	-	-	-	-	-	-
Sentiment-carrying words	0.8926	**0.8995**	0.8960	0.8989	0.8920	0.8954	0.8958
RST	0.8934	0.8910	0.8922	0.8915	**0.8939**	0.8927	0.8924
All	0.8876	**0.9005**	0.8940	0.8993	0.8862	0.8927	0.8934
Uni+Bigrams	0.9043	0.8942	0.8992	0.8956	**0.9055**	0.9005	0.8999
Length	0.8829	0.8811	0.8820	0.8816	**0.8834**	0.8825	0.8822
Positional	-	-	-	-	-	-	-
Sentiment-carrying words	**0.9016**	0.8964	0.899	0.8973	0.9024	0.8998	0.8994
RST	0.9054	0.8888	0.8970	0.8910	**0.9073**	0.8991	0.8980
All	0.8999	**0.9026**	0.9012	0.9025	0.8999	0.9012	0.9012
Features (FSD)	Positive			Negative			Microavg
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	F_1
Unigrams	0.6596	0.6175	0.6379	0.7302	**0.7647**	0.7471	0.7021
Length	0.6451	0.5195	0.5755	0.6897	**0.7889**	0.7360	0.6745
Positional	0.6720	0.6217	0.6459	0.7352	**0.7758**	0.7550	0.7104
Sentiment-carrying words	0.6936	0.6117	0.6825	0.7630	0.7808	**0.7718**	0.7345
RST	0.6690	0.6074	0.6367	0.7285	**0.7780**	0.7524	0.7055
All	0.6245	0.7348	0.6752	**0.7747**	0.6737	0.7207	0.6996
Uni+Bigrams	0.6801	0.5872	0.6302	0.7231	**0.7960**	0.7578	0.7073
Length	0.6618	0.4590	0.5421	0.6742	**0.8268**	0.7427	0.6705
Positional	0.6958	0.5855	0.6359	0.7260	**0.8109**	0.7661	0.7152
Sentiment-carrying words	0.7149	0.6578	0.6852	0.7614	0.8063	**0.7832**	0.7432
RST	0.6878	0.5734	0.6254	0.7194	**0.8078**	0.7610	0.7082
All	0.6297	0.7385	0.6798	**0.7786**	0.6793	0.7256	0.7045

TABLE III: 2-class categorization problem for the Movie Review (MR) [9] and the Finegrained Sentiment (FSD) [27] datasets.

[11] J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," in *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 2005, pp. 486–497.

[12] G. Katz, N. Ofek, and B. Shapira, "Consent: Context-based sentiment analysis," *Knowledge-Based Systems*, vol. 84, pp. 162–178, 2015.

[13] O. Appel, F. Chiclana, and J. Carter, "Main concepts, state of the art and future research questions in sentiment analysis," *Acta Polytechnica Hungarica*, vol. 12, no. 3, pp. 87–108, 2015.

[14] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis," *Knowledge-Based Systems*, vol. 89, no. C, pp. 14–46, 2015.

[15] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[16] P. J. Stone, D. C. Dunphy, and M. S. Smith, "The general inquirer: A computer approach to content analysis." 1966.

[17] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th annual meeting of the association for computational linguistics and 8th conference of the European Chapter of the Association for Computational Linguistics*. ACL, 1997, pp. 174–181.

[18] R. Y. Lau, C. Lai, and Y. Li, "Leveraging the web context for context-sensitive opinion mining," in *2nd International Conference on Computer Science and Information Technology (ICCSIT)*. IEEE, 2009, pp. 467–471.

[19] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Computational linguistics*, vol. 35, no. 3, pp. 399–433, 2009.

[20] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proceedings of the 14th ACM CIKM*. ACM, 2005, pp. 625–631.

[21] R. Mukras, N. Wiratunga, and R. Lothian, "Selecting bi-tags for sentiment analysis of text," in *Research and Development in Intelligent Systems XXIV: Proceedings of AI-2007, the Twenty-seventh SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, 2008, pp. 181–194.

[22] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the 18th International Conference on Machine Learning, ICML*, vol. 1, 2001, pp. 282–289.

[23] C. Zirn, M. Niepert, H. Stuckenschmidt, and M. Strube, "Fine-grained sentiment analysis with structural features," in *IJCNLP*, 2011, pp. 336–344.

[24] Y. Ji and J. Eisenstein, "Representation learning for text-level discourse parsing." in *ACL (1)*, 2014, pp. 13–24.

[25] A. Vlachostergiou, G. Marandianos, and S. Kollias, "From conditional random field (crf)to rhetorical structure theory (rst): incorporating context information in sentiment analysis," in *Semantic Sentiment Analysis Workshop, in conjuction with the 14th Extended Semantic Web Conference, ESWC2017*, 2017.

[26] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language resources and evaluation*, vol. 39, no. 2-3, pp. 165–210, 2005.

[27] O. Täckström and R. McDonald, "Discovering fine-grained sentiment with latent variable structured prediction models," in *Proceedings of the 33rd European Conference on Advances in Information Retrieval*, ser. ECIR'11. Springer, 2011, pp. 368–374.

[28] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *ACL (System Demonstrations)*, 2014, pp. 55–60.

[29] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. ACL, 2012, pp. 90–94.

ISBN 978-0-9928626-7-1 © EURASIP 2017