Supervised Topic-Based Message Polarity Classification using Cognitive Computing

Daniele Stefano Ferru, Federico Ibba, and Diego Reforgiato Recupero

Department of Mathematics and Computer Science
University of Cagliari, Italy
d.s.ferru@outlook.it, federico.ibba@unica.it, diego.reforgiato@unica.it

Abstract. This paper describes a supervised approach we have designed for the topic-based message polarity classification. Given a message and a topic, we aim at (i) classifying the message on a two point scale, that is positive or negative sentiment toward that topic and (ii) classifying the message on a five-point scale, that is the message conveyed by that tweet toward the topic on a more fine-grained range. These two tasks have been proposed as subtasks of SemEval-2017 task 4. We have targeted them with the employment of IBM Watson that we leveraged to extract concepts and categories to enrich the vectorial space we have modeled to train our classifiers. We have used different classifiers for the two tasks on the provided training set and obtained good accuracy and F1-score values comparable to the SemEval 2017 competitors of those tasks.

Keywords: Sentiment Analysis, NLP, Polarity Detection, Cognitive Computation, Linear Regression, Decision Tree, Naive Bayes

1 Introduction

Social media platforms are commonly used to share opinions and thoughts about different subjects and topics in any domain. Their huge widespread and proliferation of content has created opportunities to analyze and study opinions, how and where emotions are generated, what the current feelings are on a certain topic and so on. It is straightforward therefore to understand that social media have more and more interest in identifying sentiment in document, messages or posts. The common task is to detect whether in a given text there are positive, negative, neutral opinions expressed, and whether these opinions are general or focused on a certain person, product, organization or event. A lot of research has been already performed to address this task and several variations and extensions of it [3, 13]. On the one hand, supervised and unsupervised approaches have been proposed based on Natural Language Processing (NLP) techniques, machine learning tools, statistics. On the other hand, semantics has already shown to provide benefits to supervised approaches for Sentiment Analysis [26, 10, 21] where extracted semantic features enrich the vectorial space to be fed to machine learning tools (classifiers) through augmentation, replacement and interpolation techniques leading to higher accuracy. Semantics has been leveraged in unsupervised approaches too for Sentiment Analysis: authors in [24, 14] have introduced Sentilo, a sentic computing approach to opinion mining that produces a formal representation (e.g. a RDF graph) of an opinion sentence that allows distinguishing its holders and topics with very high accuracy. They have also defined and extended an ontology for opinion sentences, created a new lexical resources enabling the evaluation of opinion expressed by means of events and situations and developed an algorithm to propagate the sentiment towards the targeted entities in a sentence.

Cognitive computation is a recent kind of technology that is specialized in the processing and analysis of large unstructured datasets by leveraging artificial intelligence, signal processing, reasoning, NLP, speech recognition and vision, human-computer interaction, dialog and narrative generation. Cognitive computing systems have earned a lot of attention for figuring out relevant insights from textual data such as classifying biomedical documents [5] and e-learning videos [4]. One of the most known systems is IBM Watson[1] which can understand concepts, entities, sentiments, keywords, etc. from unstructured text through its Natural Language Understanding[2] service.

In this paper we propose a supervised approach for topic-based message polarity classification formulated as follows: given a message and a topic, classify the message on a two-point scale (Task 1) and on a five-point scale (Task 2). These two tasks have been proposed within the task 4: Sentiment Analysis in Twitter of SemEval 2016 [19][3] and SemEval 2017 [25][4].

We used machine learning approaches to target the two tasks above and leveraged IBM Watson to extract concepts and categories from the input text and to augment the vectorial space using term frequency and TF-IDF. Training and test data consist of tweets and a given topic for each tweet. As for each topic we have several tweets, we created as many classifiers as the overall number of topics in the training set. During the prediction step for a given pair (tweet, topic), two possibilities might occur:

1. the topic was found within the training set and therefore we selected the classifier already trained on the tweets related to that topic;
2. The topic was not found in any tweets of the training set. To solve this case, we used the classifier on the closest topic to the one to predict. We leveraged the semantic features extracted by IBM Watson to find the closest topic in the training set to the one to predict.

The performance evaluation we have carried out indicates satisfying results for the Task 1 whereas for Task 2 they suffer from the low number of tweets per topic present within the training set with respect to the number of tweets in the test set.

[1] https://www.ibm.com/watson/
[2] https://www.ibm.com/watson/services/natural-language-understanding/
[3] http://alt.qcri.org/semeval2016/task4/
[4] http://alt.qcri.org/semeval2017/task4/

The remainder of this paper is organized as follows. Section 2 describes background work on Sentiment Analysis techniques and how Semantics has been employed in that domain. Section 3 introduces the data we have used and how they are organized. Section 4 includes details on the method we have adopted to tackle the tasks and how Cognitive Computing has been leveraged. Section 5 shows results we have obtained and the evaluation we have carried out. Section 6 depicts concluding remarks.

2 Related Work

Several initiatives (challenges [22, 6, 23], workshop, conferences) within the Sentiment Analysis domain have been proposed. As mentioned in Section 1, the tasks we are targeting in this paper have been proposed by SemEval 2016 and SemEval 2017 task 4 where SemEval is an ongoing series of evaluations of computational semantic analysis systems, organized under the umbrella of SIGLEX, the Special Interest Group on the Lexicon of the Association for Computational Linguistics.

Authors in [28] investigated a method based on Conditional Random Fields to incorporate sentence structure (syntax and semantic) and context information to detect sentiments. They have also employed the Rethorical Structure Theory leveraging the discourse role of text segments and proved the effectiveness of the two features on the Movie Review Dataset and the Fine-grained Sentiment Dataset. Within the financial domain, authors in [9] proposed a fine-grained approach to predict real valued sentiment score by using feature sets consisting of lexical features, semantic features and their combination. Multi-domain sentiment analysis has been further targeted by authors in [7, 8] that suggested different general approaches using different features such as word embeddings. Semantic features can be extracted by several lexical and semantic resources and ontologies. Today, with the recent widespread of cognitive computing tools, we have one more tool we can leverage to refine our extraction. Cognitive computing systems [15, 16] are in fact emerging tools and represent the third era of computing. They have been used to improve not only the sentiment analysis [24], but also multi-class classification of e-learning videos [4], classification of complaints in the insurance industry [12] and within life sciences research [2]. These systems rely on deep learning algorithms and neural networks to elaborate information by learning from a training set of data. They are perfectly tailored to integrate and analyze the huge amount of data that is being released and available today. Two very well known cognitive computing systems are IBM Watson[5] and Microsoft Cognitive Services[6]. In this paper we have leveraged the former to extract categories and concepts out of an input tweet. Many others articles are presented every year within the Sentiment Analysis domain, and, therefore, several survey papers have been drafted to summarize the recent research trends and directions [27, 17, 20, 1, 11, 18].

[5] https://www.ibm.com/watson/
[6] https://azure.microsoft.com/en-us/services/cognitive-services/

3 The Used Dataset

The data have been obtained from SemEval[7]. The have been extracted from Twitter and annotated using CrowdFlower[8]. The datasets (training and test) for Task 1 included a tweet id, the topic, the tweet text and the tweet classification as positive, negative and neutral. The datasets for Task 2 (training and test) had the same structure except for the tweet classification that was an integer number ranging in [-2, +2]. Tables 1 and 2 show, respectively, five records of the dataset related to Task 1 and Task 2.

Table 1. Sample tweets for Task 1.

Tweet Id	**Topic**	**Tweet class**	**Tweet text**
522712800595300352	aaron rodgers	neutral	I just cut a 25 second audio clip of Aaron Rodgers talking about Jordy . Nelson's grandma's pies. Happy Thursday.
523065089977757696	aaron rodgers	negative	@Espngreeny I'm a Fins fan, it's Friday, and Aaron Rodgers is still giving me nightmares 5 days later. I wished it was a blowout.
522477110049644545	aaron rodgers	positive	Aaron Rodgers is really catching shit for the fake spike Sunday night.. Wtf. It worked like magic. People just wanna complain about the L.
522551832476790784	aaron rodgers	neutral	If you think the Browns should or will trade Manziel you're an idiot. Aaron Rodgers sat behind Favre for multiple years.
522887492333084674	aaron rodgers	neutral	Green Bay Packers: Five keys to defeating the Panthers in week seven: Aaron Rodgers On , Sunday ... http://t.co/anCHQjSLh9 #NFL #Packers

Moreover, Table 3 indicates the size of training sets and test sets for the two tasks whereas Table 4 and Table 5 show some statistics of the data.

4 The Proposed Method

In order to prepare the vectorial space, we have augmented the bag of words model resulting from the tweets of the training set with two kind of semantic features extracted using IBM Watson: categories and concepts. As an example, for the third tweet of Table 1 IBM Watson has extracted as categories *magic and illusion, football, podcasts* and as concepts *2009, singles*.

We have employed the augmentation method mentioned in [10] to create different vectorial spaces that we have adopted to evaluate the performances of

[7] http://alt.qcri.org/semeval2017/task4/index.php?id=data-and-tools
[8] https://www.crowdflower.com/

Table 2. Sample tweets for Task 2.

Tweet Id	**Topic**	**Tweet class**	**Tweet text**
681563394940473347	amy schumer	-1	@MargaretsBelly Amy Schumer is the stereotypical 1st world Laci Green feminazi. Plus she's unfunny
675847244747177984	amy schumer	-1	dani_pitter I mean I get the hype around JLaw. I may not like her but I get her hype. I just don't understand Amy Schumer and her hype
672827854279843840	amy schumer	-1	Amy Schumer at the #GQmenoftheyear2015 party in a dress we pretty much hate: https://t.co/j5HmmyM99j #GQMOTY2015 https://t.co/V8xzmPmPYX
662755012129529858	amy schumer	-2	Amy Schumer is on Sky Atlantic doing one of the worst stand up sets I have ever seen. And I've almost sat through 30 seconds of Millican.
679507103346601984	amy schumer	2	"in them to do it. Amy Schumer in EW, October amyschumer is a fucking rock star & I love her & Jesus F'ing Christ we need more like this" #NFL #Packers

Table 3. Sizes of the training and test sets for the two targeted tasks.

	Training Set	**Test Set**
#Task 1	16496	4908
#Task 2	23776	11811

Table 4. Statistics of the training and test sets for Task 1.

	# of Pos Tweets	**# of Neg Tweets**	**# of Neutral Tweets**
Training Set	9852	5649	995
Test Set	3780	914	214

Table 5. Distribution of the five classes for the training and test sets of Task 2.

	# Class -2	**# Class -1**	**# Class 0**	**# Class 1**	**# Class 2**
Training Set	210	2563	10216	10016	771
Test Set	172	3377	5871	2261	130

our methods. In particular we have employed the vectorial space consisting of: (i) tweets only (what we refer as baseline), (ii) tweets augmented with categories, (iii) tweets augmented with concepts, (iv) and tweets augmented with categories and concepts. We performed a set of cleaning steps to the resulting bag of words which included (i) lower casing the tokens of the input tweets, categories and concepts, (ii) removing of special characters and numbers, (iii) removing of stop words taken from StanfordNLP[9].

We employed machine learning classifiers and fed them with the produced vectorial spaces. In particular we used Linear Regression and Naive Bayes for the binary prediction of Task 1 where we have considered the positive/negative classes getting rid of the neutral class (as also suggested in the corresponding SemEval task). As far as the multi class classification of the Task 2 is concerned, we employed Decision Trees and Naive Bayes classifiers. To note that, because our data consisted of a set of tweets for each topic, we have trained a classifier for each topic in the training set feeding it with all the tweets with that topic. Both the tasks we targeted are topic-based and, therefore, given a tweet and a topic, we first had to find the most similar topic in the training set and then use the related classifier for the prediction step.

4.1 Associating Test Set and Training Set topics

Since the topics in the test set are completely different from those in the training set, we had to choose a strategy to associate the most similar topic of the training set (and therefore pick the related classifier) with each topic in the test set. To achieve this we used the categories obtained by IBM Watson. Every tweet in the training set has different related categories, thus a set with all the categories for each topic has been prepared. Similarly, for each topic in the test set, we prepared a set of all the categories extracted from each tweet related to that topic. Therefore, each topic in the training set and in the test set corresponded to a vector of categories. During the prediction of a given tweet with a certain topic t, we needed to use the classifier trained on the tweets having the most similar topic to t. To find the most similar topic in the training set to t, we counted how many categories the two lists (one corresponding to t and the other corresponding to each topic in the training set) had in common and took the one with the highest number.

5 Performance Evaluation

According to SemEval, the evaluation measure for Task 1 was the average recall that we refer as *AvgRec*:

$$AvgRec = \frac{1}{2} \cdot (R^P + R^N)$$

[9] https://bit.ly/1Nt4eMh

where R^P and R^N refer to the recall with respect to the positive and negative class. *AvgRec* ranges in [0,1] where a value of 1 is obtained only by a perfect classification and 0 is obtained in presence of a classifier that misclassifies all the items. The F1 score has further been used as secondary measure for Task 1. It is computed as:

$$F1 = 2 \cdot \frac{(P^P + P^N) \cdot (R^P + R^N)}{P^P + P^N + R^P + R^N}$$

As the task is topic-based we have computed each metric individually for each topic and then we computed the average value across all the topics to obtain the final score. Task 2 is a classification where we need to classify a tweet in exactly one class among those defined in C={highly negative, negative, neutral, positive, highly positive} represented in our data by {-2, -1, 0, 1, 2}. We used macro-average mean absolute error (MAE^M) defined as:

$$MAE^M(h, Te) = \frac{1}{|C|} \cdot \sum_{j=1}^{|C|} \frac{1}{|Te_j|} \cdot \sum_{x_i \in Te_j} |h(x_i) - y_i|$$

where y_i denotes the true label of item x_i, $h(x_i)$ is its prediction, Te_j represents the set of test documents having c_j as true class, $|h(x_i) - y_i|$ is the distance between classes $h(x_i)$ and y_i.

One benefit of the MAE^M measure is that it is able to recognize major misclassifications: for example misclassifying a highly negative tweet in highly positive is worse than misclassifying it as negative. We also used the standard mean absolute error MAE^μ, which is defined as:

$$MAE^\mu(h, Te) = \frac{1}{|Te|} \cdot \sum_{x_i \in Te} |h(x_i) - y_i|$$

The advantage of MAE^M with respect to MAE^μ is that it is robust to unbalanced class (as in our case) whereas the two measures are equivalent in presence of balanced datasets. Both MAE^M and MAE^μ have been computed for each topic and results averaged across all the topics to obtain one final score.

Tables 6 and 7 show the results we obtained for our proposed Task 1 whereas Tables 8 and 9 include results for Task 2. Results for both the tasks have been obtained by using the training and test sets of the data released from SemEval and also using a 10-cross validation by merging them. In the latter case, we did not consider the topic information during the learning step and trained one single classifier that used for the test.

5.1 Discussion of the results

In this section we discuss the obtained results for the two tasks we targeted in this paper. On the one hand, the employment of the semantic features had an impact for the classification within Task 1. As the Tables 6 and 7 show, adding the categories to the baseline improved the overall results. The addition

Table 6. Results of AvgRec and F1 values for Task 1 using the test set of SemEval.

	Baseline	**Tweets+Ctg**	**Tweets+Conc**	**Tweets+Ctg+Conc**
		AvgRec		
Linear Regression	0.4438	0.4942	0.4515	0.4982
Naive Bayes	0.4628	0.4946	0.4604	0.4969
		F1-value		
Linear Regression	0.5566	0.6339	0.5856	0.6316
Naive Bayes	0.5159	0.5200	0.5052	0.5104

Table 7. Results of AvgRec and F1 values for Task 1 using 10-cross validation on the union of training and test sets.

	Baseline	**Tweets+Ctg**	**Tweets+Conc**	**Tweets+Ctg+Conc**
		AvgRec		
Linear Regression	0.485	0.505	0.484	0.506
Naive Bayes	0.492	0.522	0.493	0.522
		F1-value		
Linear Regression	0.649	0.654	0.647	0.651
Naive Bayes	0.619	0.606	0.613	0.603

Table 8. Results of MAE^M and MAE^μ for Task 2 using the test set of SemEval.

	Baseline	**Tweets+Ctg**	**Tweets+Conc**	**Tweets+Ctg+Conc**
		MAE^M		
Decision Trees	3.628	4.207	3.745	4.242
Naive Bayes	9.548	12.02	9.882	12.34
		MAE^μ		
Decision Trees	0.472	0.552	0.488	0.559
Naive Bayes	1.219	1.556	1.256	1.601

Table 9. Results of MAE^M and MAE^μ for Task 2 using 10-cross validation on the union of training and test sets.

	Baseline	**Tweets+Ctg**	**Tweets+Conc**	**Tweets+Ctg+Conc**
		MAE^M		
Decision Trees	1.292	1.317	1.299	1.320
Naive Bayes	1.930	2.196	1.984	2.250
		MAE^μ		
Decision Trees	0.586	0.603	0.586	0.605
Naive Bayes	1.058	1.196	1.085	1.221

of concepts only does not help the classification process as with the categories probably because the lower number of concepts ends up adding noise in the used classifiers (Naive Bayes and Linear Regression). Results are confirmed also with the 10-cross-validation.

On the other hand, Task 2 shows important differences between the baseline and the tweets with the semantic features as Task 1 but in the opposite direction. As Tables 8 and 9 show, adding semantic features never improves the classification results, indicating they act like noise. This might be justified given the unbalanced nature of the used dataset: typically, each topic contains more tweets for a few classes and much less for the others. This fact generate a lot of error in the classification task and produces poor results. Furthermore, one explanation of such a behaviour is that Task 1 only consisted of a binary classification whereas Task 2 consisted of the multiclass classification where the output class might be assigned to one of five different values. Predicting five values instead of two is much harder and, given the low number of tweets per topic, the classifiers could not be trained well enough on an appropriate dataset.

6 Conclusion

In this paper we have presented a supervised topic-based message polarity classification for two tasks proposed at SemEval. The first task aims at classifying a tweet on a two point scale (positive or negative) toward a given topic. The second task aims at classifying a tweet on a five-point scale. We have targeted the two tasks using a machine learning approach where the vectorial space has been created by augmenting the message (tweets) with semantic features (categories and concepts) extracted with IBM Watson, a well known cognitive computing tool. Moreover, categories and concepts have been used to calculate the distances between topics of the training set and test set in order to associate the latter to the former. Although the low number of tweets in the training set, for Task 1 we obtained good results whereas Task 2 suffered from the scarcity of training data. Obtained results showed that with few classes (Task 1), concepts and categories were important for the classification task. Conversely, given the strong unbalanced nature of the dataset, in Task 2 concepts and categories were not able to enrich the obtained vectorial space. To address this issue, and as next steps, we would like to further investigate the employment of semantic features extracted from other cognitive computing systems trying to combine and compare them with the results obtained using IBM Watson.

Acknowledgments

The authors gratefully acknowledge Sardinia Regional Government for the financial support (Convenzione triennale tra la Fondazione di Sardegna e gli Atenei Sardi Regione Sardegna - L.R. 7/2007 annualità 2016 - DGR 28/21 del 17.05.201, CUP: F72F16003030002).

References

1. E. Cambria, B. Schuller, Y. Xia, and C. Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, March 2013.
2. Ying Chen, JD Elenee Argentinis, and Griff Weber. Ibm watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clinical Therapeutics*, 38(4):688 – 701, 2016.
3. Keith Cortis, Andre Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis. Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. *Proceedings of the 11th International Workshop on Semantic Evaluation*, pages 517–533, 2017.
4. Danilo Dessì, Gianni Fenu, Mirko Marras, and Diego Reforgiato Recupero. Leveraging cognitive computing for multi-class classification of e-learning videos. In *The Semantic Web: ESWC 2017 Satellite Events - ESWC, Portorož, Slovenia, May 28 - June 1, 2017, Revised Selected Papers*, pages 21–25, 2017.
5. Danilo Dessì, Diego Reforgiato Recupero, Gianni Fenu, and Sergio Consoli. Exploiting cognitive computing and frame semantic features for biomedical document clustering. In *Proc. of the Workshop on Semantic Web Solutions for Large-scale Biomedical Data Analytics co-located with 14th Extended Semantic Web Conference, Portoroz, Slovenia, May 28, 2017.*, pages 20–34, 2017.
6. M. Dragoni and D. Reforgiato Recupero. Challenge on fine-grained sentiment analysis within eswc2016. *Communications in Computer and Information Science*, 641:79–94, 2016.
7. Mauro Dragoni and Giulio Petrucci. A neural word embeddings approach for multi-domain sentiment analysis. *IEEE Trans. Affective Computing 8(4)*, pages 457–470, 2017.
8. Mauro Dragoni and Giulio Petrucci. A fuzzy-based strategy for multi-domain sentiment analysis. *International Journal of Approximate Reasoning*, 93:59–73, 2018.
9. Amna Dridi, Mattia Atzeni, and Diego Reforgiato Recupero. Bearish-bullish sentiment analysis on financial microblogs. In *Proc. of EMSASW 2017 co-located with 14th ESWC 2017*, 2017.
10. Amna Dridi and Diego Reforgiato Recupero. Leveraging semantics for sentiment polarity detection in social media. *International Journal of Machine Learning and Cybernetics*, Sep 2017.
11. Ronen Feldman. Techniques and applications for sentiment analysis. *Commun. ACM*, 56(4):82–89, April 2013.
12. J Forster and B Entrup. A cognitive computing approach for classification of complaints in the insurance industry. *IOP Conference Series: Materials Science and Engineering*, 261(1):012016, 2017.
13. Thomas Gaillat, Manel Zarrouk, Andre Freitas, and Brian Davis. The ssix corpus: A trilingual gold standard corpus for sentiment analysis in financial microblogs. *11th edition of the Language Resources and Evaluation Conference*, 2018.
14. A. Gangemi, V. Presutti, and D. Reforgiato Recupero. Frame-based detection of opinion holders and topics: A model and a tool. *IEEE Computational Intelligence Magazine*, 9(1):20–30, Feb 2014.
15. J. O. Gutierrez-Garcia and E. López-Neri. Cognitive computing: A brief survey and open research challenges. In *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence*, pages 328–333, 2015.

16. John E. Kelly and Steve Hamm. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia University Press, New York, NY, USA, 2013.
17. Bing Liu. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
18. Andrés Montoyo, Patricio Martínez-Barco, and Alexandra Balahur. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, 53(4):675 – 679, 2012.
19. Preslav Nakov, Alan Ritter, Sara Rosenthal, Veselin Stoyanov, and Fabrizio Sebastiani. SemEval-2016 task 4: Sentiment analysis in Twitter. In *Proc. of SemEval '16*, San Diego, California, June 2016. ACL.
20. Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.
21. Diego Reforgiato Recupero, Sergio Consoli, Aldo Gangemi, Andrea Giovanni Nuzzolese, and Daria Spampinato. A semantic web based core engine to efficiently perform sentiment analysis. In Valentina Presutti, Eva Blomqvist, Raphael Troncy, Harald Sack, Ioannis Papadakis, and Anna Tordai, editors, *The Semantic Web: ESWC 2014 Satellite Events*, pages 245–248, Cham, 2014. Springer International Publishing.
22. D.R. Recupero, M. Dragoni, and V. Presutti. Eswc 15 challenge on concept-level sentiment analysis. *Communications in Computer and Information Science*, 548:211–222, 2015.
23. D. Reforgiato Recupero, E. Cambria, and E. Di Rosa. Semantic sentiment analysis challenge at eswc2017. *Communications in Computer and Information Science*, 769:109–123, 2017.
24. Diego Reforgiato Recupero, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Giovanni Nuzzolese. Sentilo: Frame-based sentiment analysis. *Cognitive Computation*, 7(2):211–225, Apr 2015.
25. Sara Rosenthal, Noura Farra, and Preslav Nakov. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proc. of the 11th International Workshop on Semantic Evaluation*, SemEval '17, Vancouver, Canada, August 2017. ACL.
26. Hassan Saif, Yulan He, and Harith Alani. Semantic sentiment analysis of twitter. In Philippe Cudré-Mauroux, Jeff Heflin, Evren Sirin, Tania Tudorache, Jérôme Euzenat, Manfred Hauswirth, Josiane Xavier Parreira, Jim Hendler, Guus Schreiber, Abraham Bernstein, and Eva Blomqvist, editors, *The Semantic Web – ISWC 2012*, pages 508–524, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
27. Mikalai Tsytsarau and Themis Palpanas. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3):478–514, May 2012.
28. Aggeliki Vlachostergiou, George Marandianos, and Stefanos Kollias. From conditional random field (crf) to rhetorical structure theory(rst): Incorporating context information in sentiment analysis. In Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Ławrynowicz, Fabio Ciravegna, and Olaf Hartig, editors, *The Semantic Web: ESWC 2017 Satellite Events*, pages 283–295, Cham, 2017. Springer International Publishing.