

SHELLFBK: An Information Retrieval-based System For Multi-Domain Sentiment Analysis

Mauro Dragoni

Fondazione Bruno Kessler

Via Sommarive 18

Povo, Trento

dragoni@fbk.eu

Abstract

This paper describes the SHELLFBK system that participated in SemEval 2015 Tasks 9, 10, and 11. Our system takes a supervised approach that builds on techniques from information retrieval. The algorithm populates an inverted index with pseudo-documents that encode dependency parse relationships extracted from the sentences in the training set. Each record stored in the index is annotated with the polarity and domain of the sentence it represents. When the polarity or domain of a new sentence has to be computed, the new sentence is converted to a query that is used to retrieve the most similar sentences from the training set. The retrieved instances are scored for relevance to the query. The most relevant training instant is used to assign a polarity and domain label to the new sentence. While the results on well-formed sentences are encouraging, the performance obtained on short texts like tweets demonstrate that more work is needed in this area.

1 Introduction

Sentiment analysis is a natural language processing task whose aim is to classify documents according to the opinion (polarity) they express on a given subject (Pang et al., 2002). Generally speaking, sentiment analysis aims at determining the attitude of a speaker or a writer with respect to a topic or the overall tonality of a document. This task has created a considerable interest due to its wide applications. In recent years, the exponential increase of the Web for exchanging public opinions about events, facts,

products, etc., has led to an extensive usage of sentiment analysis approaches, especially for marketing purposes.

By formalizing the sentiment analysis problem, a “sentiment” or “opinion” has been defined by (Liu and Zhang, 2012) as a quintuple:

$$\langle o_j, f_{jk}, so_{ijkl}, h_i, t_l \rangle, \quad (1)$$

where o_j is a target object, f_{jk} is a feature of the object o_j , so_{ijkl} is the sentiment value of the opinion of the opinion holder h_i on feature f_{jk} of object o_j at time t_l . The value of so_{ijkl} can be positive (by denoting a state of happiness, bliss, or satisfaction), negative (by denoting a state of sorrow, dejection, or disappointment), or neutral (it is not possible to denote any particular sentiment), or a more granular rating. The term h_i encodes the opinion holder, and t_l is the time when the opinion is expressed.

Such an analysis, may be *document-based*, where the positive, negative, or neutral sentiment is assigned to the entire document content; or it may be *sentence-based* where individual sentences are analyzed separately and classified according to the different polarity values. In the latter case, it is often desirable to find with a high precision the entity attributes towards which the detected sentiment is directed.

In the classic sentiment analysis problem, the polarity of each term within the document is computed independently of the domain which the document’s domain. However, conditioning term polarity by domain has been found to improve performance (Blitzer et al., 2007). We illustrate the intuition behind domain specific term polarity. Let us

consider the following example concerning the adjective “small”:

1. The sideboard is **small** and it is not able to contain a lot of stuff.
2. The **small** dimensions of this decoder allow to move it easily.

In the first sentence, we considered the Furnishings domain and, within it, the polarity of the adjective “small” is, for sure, “negative” because it highlights an issue of the described item. On the other hand, in the second sentence, where we considered the Electronics domain, the polarity of such an adjective may be considered “positive”.

Unlike the approaches already discussed in the literature (and presented in Section 2), we address the multi-domain sentiment analysis problem by applying Information Retrieval (IR) techniques for representing information about the linguistic structure of sentences and by taking into account both their polarity and the domain.

The rest of the work is structured as follows. Section 2 presents a survey on works about sentiment analysis. Section 3 provides a description of the SHELLFBK system by described how information are stored during the training phase and exploited during the test one. Section 4 reports the system evaluation performed on the Tasks 9, 10, and 11 proposed at SemEval 2015 and, finally, Section 5 concludes the paper.

2 Related Work

The topic of sentiment analysis has been studied extensively in the literature (Pang and Lee, 2008; Liu and Zhang, 2012), where several techniques have been proposed and validated.

Machine learning techniques are the most common approaches used for addressing this problem, given that any existing supervised methods can be applied to sentiment classification. For instance, in (Pang et al., 2002) and (Pang and Lee, 2004), the authors compared the performance of Naive-Bayes, Maximum Entropy, and Support Vector Machines in sentiment analysis on different features like considering only unigrams, bigrams, combination of both, incorporating parts of speech and position information or by taking only adjectives. Moreover, beside

the use of standard machine learning method, researchers have also proposed several custom techniques specifically for sentiment classification, like the use of adapted score function based on the evaluation of positive or negative words in product reviews (Dave et al., 2003), as well as by defining weighting schemata for enhancing classification accuracy (Paltoglou and Thelwall, 2010).

An obstacle to research in this direction is the need of labeled training data, whose preparation is a time-consuming activity. Therefore, in order to reduce the labeling effort, opinion words have been used for training procedures. In (Tan et al., 2008) and (Qiu et al., 2009b), the authors used opinion words to label portions of informative examples for training the classifiers. Opinion words have been exploited also for improving the accuracy of sentiment classification, as presented in (Melville et al., 2009), where a framework incorporating lexical knowledge in supervised learning to enhance accuracy has been proposed. Opinion words have been used also for unsupervised learning approaches like the ones presented in (Taboada et al., 2011) and (Turney, 2002).

Another research direction concerns the exploitation of discourse-analysis techniques. (Somasundaran, 2010) and (Asher et al., 2008) discuss some discourse-based supervised and unsupervised approaches for opinion analysis; while in (Wang and Zhou, 2010), the authors present an approach to identify discourse relations.

The approaches presented above are applied at the document-level, i.e., the polarity value is assigned to the entire document content. However, for improving the accuracy of the sentiment classification, a more fine-grained analysis of the text, i.e., the sentiment classification of the single sentences, has to be performed. In the case of sentence-level sentiment classification, two different sub-tasks have to be addressed: (i) to determine if the sentence is subjective or objective, and (ii) in the case that the sentence is subjective, to determine if the opinion expressed in the sentence is positive, negative, or neutral. The task of classifying a sentence as subjective or objective, called “subjectivity classification”, has been widely discussed in the literature (Riloff et al., 2006; Wiebe et al., 2004; Wilson et al., 2004; Wilson et al., 2006; Yu and Hatzivassiloglou, 2003). Once subjective sentences are identified, the same

methods as for sentiment classification may be applied. For example, in (Hatzivassiloglou and Wiebe, 2000) the authors consider gradable adjectives for sentiment spotting; while in (Kim and Hovy, 2007) and (Kim et al., 2006) the authors built models to identify some specific types of opinions.

The growth of product reviews was the perfect floor for using sentiment analysis techniques in marketing activities. However, the issue of improving the ability of detecting the different opinions concerning the same product expressed in the same review became a challenging problem. Such a task has been faced by introducing “aspect” extraction approaches that were able to extract, from each sentence, which is the aspect the opinion refers to. In the literature, many approaches have been proposed: conditional random fields (CRF) (Jakob and Gurevych, 2010; Lafferty et al., 2001), hidden Markov models (HMM) (Freitag and McCallum, 2000; Jin and Ho, 2009; Jin et al., 2009), sequential rule mining (Liu et al., 2005), dependency tree kernels (Wu et al., 2009), and clustering (Su et al., 2008). In (Qiu et al., 2009a; Qiu et al., 2011), a method was proposed to extract both opinion words and aspects simultaneously by exploiting some syntactic relations of opinion words and aspects.

A particular attention should be given also to the application of sentiment analysis in social networks. More and more often, people use social networks for expressing their moods concerning their last purchase or, in general, about new products. Such a social network environment opened up new challenges due to the different ways people express their opinions, as described by (Barbosa and Feng, 2010) and (Bermingham and Smeaton, 2010), who mention “noisy data” as one of the biggest hurdles in analyzing social network texts.

One of the first studies on sentiment analysis on micro-blogging websites has been discussed in (Go et al., 2009), where the authors present a distant supervision-based approach for sentiment classification.

At the same time, the social dimension of the Web opens up the opportunity to combine computer science and social sciences to better recognize, interpret, and process opinions and sentiments expressed over it. Such multi-disciplinary approach has been called *sentic computing* (Cambria and Hus-

sain, 2012b). Application domains where sentic computing has already shown its potential are the cognitive-inspired classification of images (Cambria and Hussain, 2012a), of texts in natural language, and of handwritten text (Wang et al., 2013).

Finally, an interesting recent research direction is domain adaptation, as it has been shown that sentiment classification is highly sensitive to the domain from which the training data is extracted. A classifier trained using opinionated documents from one domain often performs poorly when it is applied or tested on opinionated documents from another domain, as we demonstrated through the example presented in Section 1. The reason is that words and even language constructs used in different domains for expressing opinions can be quite different. To make matters worse, the same word in one domain may have positive connotations, but in another domain may have negative connotations; therefore, domain adaptation is needed. In the literature, different approaches related to the Multi-Domain sentiment analysis have been proposed. Briefly, two main categories may be identified: (i) the transfer of learned classifiers across different domains (Yang et al., 2006; Blitzer et al., 2007; Pan et al., 2010; Bollegala et al., 2013; Xia et al., 2013; Yoshida et al., 2011), and (ii) the use of propagation of labels through graph structures (Ponomareva and Thelwall, 2013; Tsai et al., 2013; Tai and Kao, 2013; Huang et al., 2014). Independently of the kind of approach, works using concepts rather than terms for representing different sentiments have been proposed.

3 The SHELLFBK System

The proposed system is based on the implementation of an IR approach for inferring both the polarity of a sentence and, if requested, the domain to which the sentence belongs to. The rationale behind the usage of such an approach is that by using indexes, the computation of the Retrieval Status Value (RSV) (da Costa Pereira et al., 2012) of a term or expression, automatically takes into account which are the elements that are more significant in each index with respect to the ones that, instead, are not important with respect to the index content. In this section, we present the steps we carried out to implement our IR based sentiment and theme classification system.

3.1 Indexes Construction

The proposed approach, with respect to a classic IR system, does not use a single index for containing all information, but a set of indexes are created in order to facilitate the identification of the correct polarity and domain, of a sentence during the validation phase. In particular, we built the following set of indexes:

- **Polarity Indexes:** from the training set, the positive, negative, and neutral sentences have been indexed separately.
- **Domain Indexes:** a different index has been built for each domain identified in the training set. This way, it is possible to store information about which terms, or expression, are relevant for each domain.
- **Mixed Indexes:** by considering the multi-domain nature of the system, this further set of indexes allows to have, for each domain, information about the correlation between the domain and the polarities. This way, we are able to know if the same term, or expression, has the same polarity in different domains or not.

For each sentence of the training set, we exploited the Stanford NLP Library for extracting the dependencies between the terms. Such dependencies are then used as input for the indexing procedure.

As example, let's consider the following sentence extracted from the training set of the Task 9:

"I came here to reflect my happiness by fishing."

This sentence has a positive polarity and belongs to the "outdoor_activity" domain. By applying the Stanford parser, the dependencies that are extracted are the following ones:

```
nsubj(came-2, I-1)
nsubj(reflect-5, I-1)
root(ROOT-0, came-2)
advmod(came-2, here-3)
aux(reflect-5, to-4)
xcomp(came-2, reflect-5)
poss(happiness-7, my-6)
dobj(reflect-5, happiness-7)
prep_by(reflect-5, fishing-9)
```

Each dependency is composed by three elements: the name of the "relation" (R), the "governor" (G) that is the first term of the dependency, and the "dependent" (D) that is the second one. We extract,

from each dependency, the structure "field - content" shown in Table 1 by using as example the dependency "dobj(reflect-5, happiness-7)". Such a structure is then given as input to the index.

Field Name	Content
RGD	"dobj-reflect-happiness"
RDG	"dobj-happiness-reflect"
GD	"reflect-happiness"
DG	"happiness-reflect"
G	"reflect"
D	"happiness"

Table 1: Field structure and corresponding content stored in the index.

The structure shown in Table 1 is created for each dependency extracted from the sentence and the aggregation of all structures are stored as final record in the index.

3.2 Polarity and Domain Computation

Once the indexes are built, both the polarity and the domain of each sentence that need to be evaluated, are computed by performing a set of queries on the indexes. In our approach, we implemented a variation of classic IR scoring formula for our purposes. In the classical TF-IDF IR model (van Rijsbergen, 1979), the *inverse document frequency* value is used for identifying which are the most significant documents with respect to a particular query. This value is useful when we want to identify the uniqueness of a document with respect to a term contained in a query, with respect to the other documents stored into the index. In our case, the scenario is different because if a term, or expression, occurs often in the index, this aspect has to be emphasized instead of being discriminated. Therefore, in our scoring formula we consider, as final score of a term or an expression, the document frequency (DF) value (i.e., the inverse of the IDF). This way, we are able to infer if a particular term or expression is significant or not for a given polarity value or domain.

The queries are built with the same procedure used for creating the records stored in the indexes. For each sentence to evaluate, a set of queries, one for each dependency extracted from the sentence is performed on the indexes and the results are aggre-

gated for inferring both the polarity and domain of the sentence.

As example of how the system works, let's consider the following sentence:

"I feel good and I feel healthy."

For simplicity, we only consider the following two extracted dependencies:

```
acompl(feel-2, good-3)
acompl(feel-6, healthy-7)
```

From these two dependencies, we generate the following two queries:

```
Q1: "RGD:"acompl-feel-good"
    OR RDG:"acompl-good-feel"
    OR GD:"feel-good" OR DG:"good-feel"
    OR G:"feel" OR D:"good"
Q2: "RGD:"acompl-feel-healthy"
    OR RDG:"acompl-healthy-feel"
    OR GD:"feel-healthy" OR DG:"healthy-feel"
    OR G:"feel" OR D:"healthy"
```

For computing the polarity of the sentence, the queries are performed on the three indexes containing polarized records: positive (*POS*), negative (*NEG*), and neutral (*NEU*). From the computed ranks, we extract only the *DF* associated to each field *F* contained in the query:

$$DF(F) = 1/IDF(F) \quad (2)$$

where *DF* is the value extracted.

As a direct consequence, for each index *I*, the value representing the *RSV* of a sentence is:

$$RSV(I) = DF(RGD_{Q1}) + DF(RDG_{Q1}) + DF(GD_{Q1}) + DF(DG_{Q1}) + DF(G_{Q1}) + DF(D_{Q1}) + DF(RGD_{Q2}) + DF(RDG_{Q2}) + DF(GD_{Q2}) + DF(DG_{Q2}) + DF(G_{Q2}) + DF(D_{Q2}) \quad (3)$$

Finally, the polarity of the sentence *S* is inferred by considering the maximum *RSV* computed over the three indexes:

$$Polarity(S) = \operatorname{argmax}_{P \in POS, NEU, NEG} RSV(S, P) \quad (4)$$

In case of domain assignment, given a set *D* of *k* domains, the domain is computed by:

$$Domain(S) = \operatorname{argmax}_{i \in 1 \dots k} RSV(S, D_i) \quad (5)$$

4 Results

The SHELLFBK system participated in three SemEval 2015 tasks: 9, 10, and 11. All three tasks were about the sentiment analysis topic with the following differences:

- Task 9 (Russo et al., 2015): this task is based on a dataset of events annotated as instantiations of pleasant and unpleasant events previously collected in psychological researches as the ones on which human judgments converge (Lewinsohn and Amenson, 1978), (MacPhillamy and Lewinsohn, 1982). Task 9 concerns classification of the events that are pleasant or unpleasant for a person writing in first person. This task was organized around two subtasks: (A) identification of the polarity value associated to an event instance, and (B) identification of both the event instantiations and the associated polarity values. The SHELLFBK system has been tested on both tasks.
- Task 10 (Rosenthal et al., 2015): this task aims to identify sentiment polarities in short text messages contained in the Twitter microblog. This task contains five subtasks: (A) expression-level, (B) message-level, (C) topic-related, (D) trend, and (E) a task on prior polarity of terms. The SHELLFBK has been tested only on the subtask (B).
- Task 11 (Ghosh et al., 2015): this task consists in the classification of tweets containing irony and metaphors. Given a set of tweets that are rich in metaphor and irony, the goal is to determine whether the user has expressed a positive, negative, or neutral sentiment in each, and the degree to which this sentiment has been communicated. With respect to the other tasks, here the polarity is expressed through a fine-grained scale in the interval [-5, 5].

In the following subsections, we will briefly report the performance obtained on each task.

4.1 Task 9

Table 2 reports the results obtained in Task 9. This task consisted in the identification of the polarity of a sentence written in first person (subtask A) and in

Task	Precision	Recall	F-Measure
Subtask A	0.555	0.384	0.454
Subtask B	0.261	0.155	0.197

Table 2: Results obtained by the SHELLFBK system on Task 9.

the identification of both the polarity and the domain of the sentence (subtask B). Precision, recall and F-Measure have been computed. As expected, the accuracy obtained on the sole prediction of the sentence polarity is higher with respect to the one obtained on the subtask combining the inference of both the domain and the polarity itself. Unfortunately, the recall values obtained on both subtasks are quite low, especially for the subtask B.

4.2 Task 10

Performance obtained by the SHELLFBK system on Task 10 have been reported in Table 3. For this task, the SHELLFBK system has been tested only on the message-level polarity subtask (B). By observing either the overall f-measure and the ones obtained on the different portions of the dataset, the performance of the system are too low for considering it a reliable solution for being used in contexts where short texts are taken into account.

4.3 Task 11

Results of the proposed system concerning Task 11 are shown in Table 4. In this task, due to the fine-grained nature of the polarity predictions, the cosine similarity and the mean square error with respect to the gold standard have been computed. In the first result-line, the values obtained on the four figurative categories are reported, while in the second one, the overall results. By observing the results, for the “Sarcasm” and “Irony” topics the obtained results are acceptable; while, for the “Metaphor” and for the “Other” category, both the cosine similarity and the MSE are significantly worse with respect to the first two. These results, either with the ones obtained on Task 10, confirm that the analysis of short texts is the first issue to address for improving the general quality of the system.

5 Conclusion

In this paper, we described the SHELLFBK system presented at SemEval 2015 that participated in SemEval 2015 Tasks 9, 10, and 11. Our system makes use of IR techniques to classify sentences by polarity, domain and the joint prediction of polarity and domain, effectively providing domain specific sentiment analysis. The results demonstrated that, while on well-formed sentences the system obtained good performance, the method performs less well on short texts like tweets. Therefore, future work will focus on the improvement of the system in this direction. In future work, we intend to explore the integration of sentiment knowledge bases (Dragoni et al., 2014) in order to move toward a more cognitive approach.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful suggestions and comments.

References

- Nicholas Asher, Farah Benamara, and Yvette Yannick Mathieu. 2008. Distilling opinion in discourse: A preliminary study. In *COLING (Posters)*, pages 7–10.
- Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. In *COLING (Posters)*, pages 36–44.
- Adam Bermingham and Alan F. Smeaton. 2010. Classifying sentiment in microblogs: is brevity an advantage? In *CIKM*, pages 1833–1836.
- John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, pages 187–205.
- Danushka Bollegala, David J. Weir, and John A. Carroll. 2013. Cross-domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Trans. Knowl. Data Eng.*, 25(8):1719–1731.
- Erik Cambria and Amir Hussain. 2012a. Sentic album: Content-, concept-, and context-based online personal photo management system. *Cognitive Computation*, 4(4):477–496.
- Erik Cambria and Amir Hussain. 2012b. *Sentic Computing: Techniques, Tools, and Applications*, volume 2 of *SpringerBriefs in Cognitive Computation*. Springer, Dordrecht, Netherlands.
- Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery: opinion extraction

Task	Live Journal 2014	SMS 2013	Twitter 2013	Twitter 2014	Twitter 2014 Sarcasm
Progress Test	0.3406	0.2614	0.3214	0.3220	0.3558
Task	F-Measure	-	-	-	-
Overall	0.3245	-	-	-	-

Table 3: Results obtained by the SHELLFBK on Task 10.

	Mean Square Error				Cosine Similarity			
	Sarcasm	Irony	Metaphor	Other	Sarcasm	Irony	Metaphor	Other
Detailed Results	4.375	4.516	9.219	12.16	0.669	0.625	0.35	0.167
	MSE	Cosine	-	-	-	-	-	-
General Result	7.701	0.431	-	-	-	-	-	-

Table 4: Results obtained by the SHELLFBK on Task 11.

- and semantic classification of product reviews. In *WWW*, pages 519–528.
- Mauro Dragoni, Andrea G.B. Tettamanzi, and Célia da Costa Pereira. 2014. Propagating and aggregating fuzzy polarities for concept-level sentiment analysis. *Cognitive Computation*, pages 1–12.
- Dayne Freitag and Andrew McCallum. 2000. Information extraction with hmm structures learned by stochastic optimization. In *AAAI/IAAI*, pages 584–589.
- Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, Antonio Reyes, and John Barnden. 2015. SemEval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval '2015*, Denver, Colorado, June.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford University.
- Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *COLING*, pages 299–305.
- Sheng Huang, Zhendong Niu, and Chongyang Shi. 2014. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowl.-Based Syst.*, 56:191–200.
- Niklas Jakob and Iryna Gurevych. 2010. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *EMNLP*, pages 1035–1045.
- Wei Jin and Hung Hay Ho. 2009. A novel lexicalized HMM-based learning framework for web opinion mining. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, pages 465–472, New York, NY, USA. ACM.
- Wei Jin, Hung Hay Ho, and Rohini K. Srihari. 2009. Opinionminer: a novel machine learning system for web opinion mining and extraction. In *KDD*, pages 1195–1204.
- Soo-Min Kim and Eduard H. Hovy. 2007. Crystal: Analyzing predictive opinions on the web. In *EMNLP-CoNLL*, pages 1056–1064.
- Soo-Min Kim, Patrick Pantel, Timothy Chklovski, and Marco Pennacchiotti. 2006. Automatically assessing review helpfulness. In *EMNLP*, pages 423–430.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, pages 282–289.
- Peter M. Lewinsohn and Christopher S. Amenson. 1978. Some relations between pleasant and unpleasant events and depression. *Journal of Abnormal Psychology*, 87:644–654.
- Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In C. C. Aggarwal and C. X. Zhai, editors, *Mining Text Data*, pages 415–463. Springer.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *WWW*, pages 342–351.
- Douglas MacPhillamy, and Peter M. Lewinsohn. 1982. The pleasant event schedule: Studies on reliability, validity, and scale intercorrelation. *Journal of Counseling and Clinical Psychology*, 50:363–380.
- Prem Melville, Wojciech Gryc, and Richard D. Lawrence. 2009. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *KDD*, pages 1275–1284.
- Georgios Paltoglou and Mike Thelwall. 2010. A study of information retrieval weighting schemes for sentiment analysis. In *ACL*, pages 1386–1395.
- Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. 2010. Cross-domain senti-

- ment classification via spectral feature alignment. In *WWW*, pages 751–760.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL*, pages 271–278.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP 2002*, pages 79–86, Philadelphia, July.
- Natalia Ponomareva and Mike Thelwall. 2013. Semi-supervised vs. cross-domain graphs for sentiment analysis. In *RANLP*, pages 571–578.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009a. Expanding domain sentiment lexicon through double propagation. In *IJCAI*, pages 1199–1204.
- Likun Qiu, Weishi Zhang, Changjian Hu, and Kai Zhao. 2009b. Selc: a self-supervised model for sentiment classification. In *CIKM*, pages 929–936.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27.
- Ellen Riloff, Siddharth Patwardhan, and Janyce Wiebe. 2006. Feature subsumption for opinion analysis. In *EMNLP*, pages 440–448.
- Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. SemEval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval '2015*, Denver, Colorado, June.
- Irene Russo, Tommaso Caselli, and Carlo Strapparava. 2015. SemEval-2015 Task 9: CLIPeVal Implicit Polarity of Events. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval '2015*, Denver, Colorado, June.
- Swapna Somasundaran. 2010. *Discourse-level relations for Opinion Analysis*. Ph.D. thesis, University of Pittsburgh.
- Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen, and Zhong Su. 2008. Hidden sentiment association in chinese web opinion mining. In *WWW*, pages 959–968.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly D. Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.
- Yen-Jen Tai and Hung-Yu Kao. 2013. Automatic domain-specific sentiment lexicon generation with label propagation. In *iiWAS*, pages 53:53–53:62. ACM.
- Songbo Tan, Yuefen Wang, and Xueqi Cheng. 2008. Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples. In *SIGIR*, pages 743–744.
- Angela Charng-Rung Tsai, Chi-En Wu, Richard Tzong-Han Tsai, and Jane Yung jen Hsu. 2013. Building a concept-level sentiment dictionary based on common-sense knowledge. *IEEE Int. Systems*, 28(2):22–30.
- Célia da Costa Pereira, Mauro Dragoni, and Gabriella Pasi. 2012. Multidimensional relevance: Prioritized aggregation in a personalized Information Retrieval setting. *Inf. Process. Manage.*, 48(2):340–357.
- Peter D. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *ACL*, pages 417–424.
- Cornelis Joost van Rijsbergen. 1979. *Information Retrieval*. Butterworth.
- Hongling Wang and Guodong Zhou. 2010. Topic-driven multi-document summarization. In *IJALP*, pages 195–198.
- Qiu Feng Wang, Erik Cambria, Cheng Lin Liu, and Amir Hussain. 2013. Common sense knowledge for handwritten chinese recognition. *Cognitive Computation*, 5(2):234–242.
- Janyce Wiebe, Theresa Wilson, Rebecca F. Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30(3):277–308.
- Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2004. Just how mad are you? finding strong and weak opinion clauses. In *AAAI*, pages 761–769.
- Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2):73–99.
- Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2009. Phrase dependency parsing for opinion mining. In *EMNLP*, pages 1533–1541.
- Rui Xia, Chengqing Zong, Xuelei Hu, and Erik Cambria. 2013. Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IEEE Int. Systems*, 28(3):10–18.
- Hui Yang, Jamie Callan, and Luo Si. 2006. Knowledge transfer and opinion detection in the TREC 2006 blog track. In *TREC*.
- Yasuhisa Yoshida, Tsutomu Hirao, Tomoharu Iwata, Masaaki Nagata, and Yuji Matsumoto. 2011. Transfer learning for multiple-domain sentiment analysis—identifying domain dependent/independent word polarity. In *AAAI*, pages 1286–1291.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP 2003*, pages 129–136, Stroudsburg, PA, USA.