Conference papers   School of Computer Science

2012

A Case-Based Approach to Cross Domain Sentiment Classification

Bruno Ohana
Technological University Dublin

Sarah Jane Delany
Technological University Dublin, sarahjane.delany@tudublin.ie

Brendan Tierney
Technological University Dublin

Follow this and additional works at: https://arrow.tudublin.ie/scschcomcon

A Case-Based Approach to Cross Domain Sentiment Classification

Bruno Ohana, Sarah Jane Delany, and Brendan Tierney
bruno.ohana@student.dit.ie, {sarahjane.delany,brendan.tierney}@dit.ie

Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland

Abstract. This paper considers the task of sentiment classification of subjective text across many domains, in particular on scenarios where no in-domain data is available. Motivated by the more general applicability of such methods, we propose an extensible approach to sentiment classification that leverages sentiment lexicons and out-of-domain data to build a case-based system where solutions to past cases are reused to predict the sentiment of new documents from an unknown domain. In our approach the case representation uses a set of features based on document statistics, while the case solution stores sentiment lexicons employed on past predictions allowing for later retrieval and reuse on similar documents. The case-based nature of our approach also allows for future improvements since new lexicons and classification methods can be added to the case base as they become available. On a cross domain experiment our method has shown robust results when compared to a baseline single-lexicon classifier where the lexicon has to be pre-selected for the domain in question.

Keywords: case-based reasoning, sentiment classification, sentiment lexicons

1 Introduction

Subjective text reflecting people's opinions is now widely available in online forums, product comparison sites, in social media websites and many other user-generated content outlets. Such repositories can provide a new valuable layer of sentiment information to business intelligence applications, and devising computational methods for their efficient use is the realm of sentiment analysis research. In particular, for a given piece of text, one can ask whether its sentiment can be considered generally positive or negative, favorable or unfavorable. Sentiment classification is the area of sentiment analysis concerned with predictive methods for determining the sentiment orientation of subjective text. It is often characterized as a binary classification problem where possible outcomes are positive or negative sentiment. Alternatively a sentiment classifier may attempt to place sentiment along a range of possible values, including neutral, as is the case for example on a star-rating system of film reviews.

Supervised learning methods for sentiment classification have been extensively studied in the past decade having achieved considerable success and have

been applied to various domains [1–3]. One downside of such methods however is the requirement for labelled in-domain data for training. Considering the wide range of potential areas sentiment classification can be applied to - product and content reviews on every possible industry for example - the cost of compiling data sets for each target domain becomes prohibitively expensive. Moreover, it is known that sentiment clues derived from supervised learning models are strongly associated with the domain used in training, and not easily reusable on a different domain [4]. This has encouraged research on cross domain approaches that minimize the requirement for in-domain training data.

Current research in cross domain sentiment classification methods have focused broadly on two approaches: one is to use out-of-domain data to build supervised learning models capable of performing well on other domains. The second class of methods essentially amounts to an unsupervised approach where documents are evaluated with the assistance of pre-existing knowledge, requiring no training data. The unsupervised sentiment classifier takes into account a document's linguistic clues, and often relies on a sentiment lexicon: a language resource that associates terms with sentiment information. For example, a lexicon would encode a priori knowledge of sentiment contained on words such as excellent, good or terrible. With a sentiment lexicon, an unsupervised classifier can make predictions by identifying opinion-bearing terms in a target document and making decisions according to the overall orientation of terms found.

Large numbers of sentiment lexicons are available in the literature, and various unsupervised techniques for using them in sentiment classification have been proposed [5–7]. However, the choice of lexicon and how it is going to be used is generally fixed in advance in a classification problem. We claim that in a multi domain setting there is no single fixed lexicon technique that will consistently generate good predictions for all documents. Instead, a more flexible approach would be to determine, out of all the available lexicons, which ones obtain good predictions for specific documents, and use those documents to build a case base of past examples for later reuse.

In this study we propose a case-based approach to document sentiment classification where out-of-domain labelled data is reused to make predictions on unseen documents as follows: a case base is built by evaluating labelled documents using various sentiment lexicons, and recording which lexicons yield correct predictions as the case solution. A case is represented by a set of features reflecting the structure and statistics which describe the document. A prediction is made by retrieving the most similar cases and reusing the lexicons found in the cases' solutions. This paper shows how a case-based approach can produce results comparable to single-lexicon unsupervised approaches while removing the need to determine the best lexicon in advance. In addition, the case-based approach is easily extensible, allowing for the addition of new cases where new sentiment lexicons and ways of applying them can be incorporated in the future as they become available.

In the remaining sections we discuss the research literature of sentiment classification, lexicons and the challenges of cross domain sentiment classification.

We describe the design of our case base, and present the results of a sentiment classification experiment using product reviews from different domains. We discuss the results obtained with our approach and propose avenues for future work.

2 Related Work

Supervised learning methods for sentiment classification using in domain training data have been extensively studied in the past decade: early work from Pang, Lee and Vaithyanathan [1] presents the results of different classifiers using features based on word n-grams on a data set of film reviews. In [8] this model is improved by eliminating objective sections from raw documents prior to training, while a similar approach seen in [9] builds multiple classifiers based on types of sentences found in a document. Extending the feature sets with document statistics and punctuation information is seen in [2].

The performance of supervised learning methods is strongly linked to the domain data used during training. Experiments seen in [10] and [4] illustrate how poor results can be obtained on combinations where domains used for training and evaluation have little in common. In the latter study more general methods to overcome this drawback suggest using out-of-domain data to build classifier ensembles, and extending training data with in-domain unlabelled documents. The use of small amounts of labelled and unlabelled in-domain data is also seen in [11, 12].

Sentiment lexicons are language resources that associate a vocabulary term with opinion polarity – positive, negative or neutral, often by means of a numeric score indicating certainty or opinion strength. Lexicons can be obtained via manual annotation of words, the *General Enquirer* being a well known example [13], however to overcome the limitations in size and cost of manual annotation, research has sought ways of creating lexicons by expanding a small set of seed terms using a pre-existing knowledge resource. Corpus based methods are first seen in the work of [14] where expansion is based on terms found near connectors such as "and", "or" and "but". Term proximity is also explored using larger corpora in [15] and more recently in a lexicon derived from web documents in [16]. Expansion via thesaurus relationships is explored in lexicons presented in [17]. The *SentiWordNet* lexicon [18] uses the WordNet database [19] as the source of information and is built first by exploring direct term relationships such as synonym and antonym information, and then performing a second step that uses a semi supervised method for detecting sentiment from term glosses. The *SentiFul* lexicon [20] finds new words via morphological rules that relate them to a word with known sentiment.

When considering their use in sentiment classification, sentiment lexicons appear as an additional source of information for engineering features on cross domain classifiers as seen in the use of SentiWordNet [21]. Alternatively sentiment lexicons are typically used in unsupervised approaches in conjunction with an algorithm for scanning a document and extracting a document sentiment score based on lexicon information and linguistic clues. Multi domain classifi-

cation has been explored using custom built lexicons [6, 5] and using multiple lexicons and a majority based scheme to obtain predictions [7]. Unsupervised methods based on lexicons have the advantage of requiring no training data. However, before applying them to a classification problem, the choice of lexicon and how they are to be applied needs to be determined and fixed. This may lead to sub optimal results where better suited combinations exist for a yet unseen target domain.

Case-based reasoning [22] is a problem solving approach aiming at the reuse of similar past examples to determine the outcome of a new unseen instance. To date, we found little evidence of this approach being used in sentiment classification. One example can be found in the literature in [10] where a document repository of labelled cases is indexed for keyword based searches. Predictions are made by first retrieving cases using a free-text search based on terms found in a target document. The sentiment of a target document is determined by the labels of similar documents and their rank in search results.

3 The Case-Based Approach

Cases in our case base are derived from a training set of labelled out-of-domain opinionated documents, and have two essential components: the *case description* is a document signature used for later retrieval, while the *case solution* stores details about successful predictions obtained for the document during training. The case description attempts to broadly capture a document's characteristics into a set of features, leaving aside any potential domain specific aspects such as particular term presence. To that end we propose the use of an n-dimensional feature vector derived from document, sentence and term-level statistics, part of speech information, punctuation and other indicators of document complexity. These are described in Table 1.

Table 1. Features describing a case.

Category	Metrics
Document Statistics	Total words, tokens and sentences.
	Average sentence size.
	Part of speech frequencies.
Writing Style	Spacing Ratio.
	Stop words ratio.
	Average syllable count.
	Monosyllable ratio.
	Word to token ratio.
	Ratio of unique words.

The spacing ratio is the rate of empty spaces to characters, and the ratio of stop words is based on the SMART system stop word list [23]. There are

17 features in total, and all features are numeric and normalized with min-max normalization based upon values from the training set.

For the case solution we record all the sentiment lexicons that made a correct prediction during training on the document represented by the case. We use 5 different lexicons from the literature: The General Inquirer (GI) lexicon [13], a small manually annotated lexicon often used as a gold standard on sentiment analysis research; The Subjectivity Clues lexicon (Clues) [24] is also an annotated lexicon that includes words from the General Inquirer and other sources; SentiWordNet (SWN) [18] is an automatically built lexicon based on WordNet term relationships and gloss information. The Moby lexicon [7] is built from the Moby thesaurus by expanding from a set of seed terms. Finally the MSOL lexicon [17] is based on the expansion of the Macquaire thesaurus from a set of word pairs with opposing sentiment.

To make predictions we use an unsupervised classifier that takes a sentiment lexicon as input and computes a document sentiment *score* by querying the lexicon for sentiment information of terms found in the document. When a term is found in the lexicon, its sentiment orientation is retrieved. By convention, this is stored as a pair (pos, neg) of numeric values indicating positive and negative sentiment. We note that some lexicons such as SentiWordNet may record values for both positive and negative sentiment on a single term due to the process by which the scores are derived. The overall sentiment score of a document is also a pair of numeric values containing the accumulated scores of all individual terms obtained from the lexicon, and a prediction of document sentiment is based on the higher of the two values in the document score.

The unsupervised classifier works with the help of linguistic clues to improve its accuracy: a document part-of-speech tagger marks the grammatical role of each word in the document. To process documents in our experiment we use the *Stanford POS Tagger*[1]. Based on results from preliminary experiments, we select only terms tagged as adjectives and verbs during document scoring. In addition, certain lexicons are segmented by part of speech, thus a part-of-speech tagging pre-processing step also improves the accuracy of lexicon queries.

Negation detection is also another important element when detecting sentiment, for example, the sentences "I think this book is not good" and "I think this book is good" have opposing sentiment orientation, despite the presence of the largely positive term "good" on both. We apply a variation of the *NegEx* algorithm [25] to identify opinion in negated sentences. Our implementation works by scanning a document for known negating n-grams and inverting sentiment orientation of nearby terms in the same sentence.

Finally, when calculating document sentiment from the sum of individual term scores, we want to reduce the effect of terms frequently occurring in English from having a dominant effect on the predictions. To mitigate this effect we implement an adjustment factor based on relative term frequencies introduced by [7]. The adjustment is given by the formula:

---

[1] http://nlp.stanford.edu/software/tagger.shtml

$$S_{adj}(w) = S(w) * (1 - \sqrt{freq(w)}) \tag{1}$$

where $s(w)$ is the term score from the source lexicon and $freq(w)$ is the term frequency (valued between 0 and 1) of w relative to the most frequent term found in the lexicon. The relative frequencies are calculated for each lexicon according to term frequency information from a separate document corpus.

3.1 Populating the Case Base

The case base is built from a set of out-of-domain labelled documents. For each training document, we attempt to make predictions using each of the available sentiment lexicons and the unsupervised classifier described previously. If no lexicons can correctly predict a document, the document is discarded. Otherwise a new case is added where the case solution is the list of all lexicons that yielded a correct prediction. Algorithm 1 describes our method for populating the case base at training time.

Algorithm 1 Populating the case base

Input:

- **D**, set of labelled out-of-domain documents for training.
- **L**, set of all available sentiment lexicons.
- **f(L,d)**, unsupervised document sentiment classifier using lexicon L as input.

Output:

- **CB**, the populated case base.

$CB \leftarrow \{\}$
for all document d in D **do**
   $S \leftarrow \{\}$
   for all L_i in L **do**
      make prediction using $f(L_i, d)$
      if prediction is correct **then**
         $S \leftarrow S \cup L_i$
      end if
   end for
   if $S <> \{\}$ **then**
      compute case description $x(d)$
      $CB \leftarrow CB \cup (x(d), S)$
   end if
end for

We note that our approach allows for future expansion since the case base can grow iteratively as more out-of-domain data becomes available, while additional lexicons can be obtained and added to the algorithm in the future.

3.2 Case Retrieval and Reuse

Using Euclidean distance as the similarity measure, we retrieve the k cases nearest to an unseen instance. The solutions from the k-nearest cases are used as follows: where $k=1$ we can directly apply the obtained lexicons from the case solution to make a prediction. For larger values of k we establish a ranking of lexicons by counting their frequency of occurrence out of the k cases retrieved. By inspecting the ranking, we then obtain the most frequently found lexicons and use them on predictions, as illustrated in the example in Table 2.

Table 2. Example ranking of solutions using kNN and $k=3$.

Retrieved Solutions (k=3)	Ranking (Count)	Lexicons
case A: $\{L_1, L_2, L_5\}$	L_1 (3)	$\{L_1, L_2\}$
case B: $\{L_1, L_2\}$	L_2 (3)	
case C: $\{L_1, L_2, L_3\}$	L_3 (1)	
	L_5 (1)	

The outcome of retrieval and ranking may yield more than a single lexicon since a case solution may record multiple lexicons, and ties can occur. In this case we separately calculate the document sentiment scores using each lexicon and the unsupervised classifier, and aggregate all positive and negative scores. The accumulated scores are then used to make a prediction as before: the highest of the scores determines document sentiment.

4 Evaluation

We evaluate our proposed case-based approach on a cross domain sentiment classification experiment using 6 datasets of user generated product and film reviews in plain text: the IMDB dataset of film reviews [1], the hotel reviews dataset from [26], and product reviews for apparel, music, books and electronics from Amazon.com [27]. Each data set has an equal number of positive and negative documents and is detailed in Table 3.

Table 3. Customer review datasets.

Dataset	No. of Reviews	Avg. Size (tokens)	
		positive	negative
Film	2000	803.2	721.4
Hotels	2874	215.0	228.7
Electronics	2072	237.6	194.9
Books	2034	284.7	202.2
Apparel	566	137.1	110.5
Music	5902	246.4	195.6

4.1 Experiment Methodology

To assess our method for sentiment classification using out-of-domain data we created 6 distinct case bases by training on datasets of all but one of the domains. The case base is then used to classify documents on the hold out domain. The composition of each case base in terms of class distribution is presented in Table 4 along with the percentage of discarded documents. The case base names reflect the hold out dataset.

Table 4. Case base class distribution and discarded ratio.

Case base	Size	Pos. %	Neg. %	Discarded %
Books	9683	53.3	46.7	27.8
Electronics	9592	53.6	46.4	28.2
Film	9614	54.1	45.9	28.6
Music	6173	52.6	47.4	25.1
Hotels	11516	53.5	46.5	7.8
Apparel	11002	53.4	46.6	28.9

All of the case bases contain more documents with a positive rather than negative orientation, with negative documents being excluded more often during the case base population stage. This could be attributed to a relative difficulty in predicting negative sentiment compared to positive ones using lexicon based term scoring techniques - a behavior also noted on past research [6]. We also see a considerable number of discarded cases at training time, for which a correct classification could not be found using any of the available lexicons. The ratio of discarded entries stays around 25-28% on most case bases but shows a distinct lower ratio when the hotels dataset is left out, indicating this dataset makes a substantial contribution to the total of discarded cases.

Figure 1 shows, for each case base, the distribution of cases by the number of lexicons used in the solution. We note that across all case bases a total of 27-28% of cases were predicted correctly using just one or two of the available lexicons.

Cases with a single-lexicon solution reflect the situation where no other lexicon could make a correct prediction while building the case base. The average distribution of these cases over all case bases is given below to illustrate how a specific lexicon can sometimes be uniquely capable of making correct predictions on certain documents: GI: 0.9%; SWN: 2.4%; Clues: 0.7%; Moby: 7.7%; MSOL: 2.4%.

4.2 Results

In Table 5, we present accuracy results for the case-based method using the six hold out data sets and their corresponding case bases. For comparison we include baseline results from best single-lexicon accuracies using a similar scoring
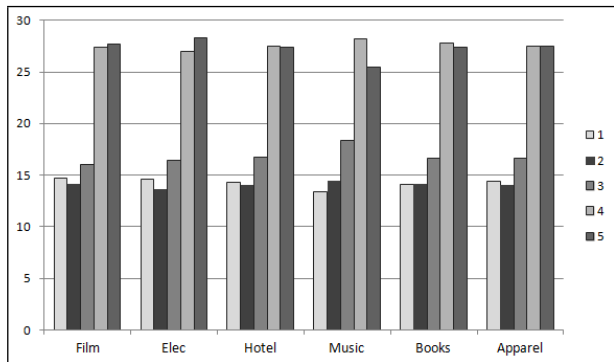
Fig. 1. Distribution of cases by number of lexicons in solution (% of case base size).

approach and the same data sets [7]. We marked in bold the experiment results that outperformed the single-lexicon baseline.

Table 5. Accuracy on the hold out data set for each case base.

				Baseline	
Case base	k=1	k=3	k=5	Accuracy	Lexicon
Film		67.88	66.73	66.78	68.18 Clues
Electronics	**68.06**	65.16	65.6	67.19 SWN
Hotel	**72.58**	71.43	70.7	71.67 SWN
Music		64.62	64.78	64.28	65.04 SWN
Books	**64.27**	62.54	62.0	63.73 Clues
Apparel	**66.96**	66.07	63.42	65.54 Clues

When compared to the best single-lexicon baseline, our results at $k=1$ gave improved accuracies in four of the six domains tested. Using the Wilcoxon signed-ranks test for comparing results over multiple data sets [28] at confidence $p=0.05$, we find the results for the case-based method and the baseline are not statistically significant. We note however that the baseline presents best accuracies from separate single-lexicon experiments run on each domain, and that the lexicon yielding best results is not necessarily the same on all experiments. Thus, our approach produces results similar to the baseline, but without the requirement for fixing a lexicon in advance.

4.3 Ranking Behavior

For different values of k, there is a slight reduction in performance as k grows and we begin applying the solution ranking algorithm. This trend can be seen in accuracies plotted on Figure 2 (note that the y axis is partial).
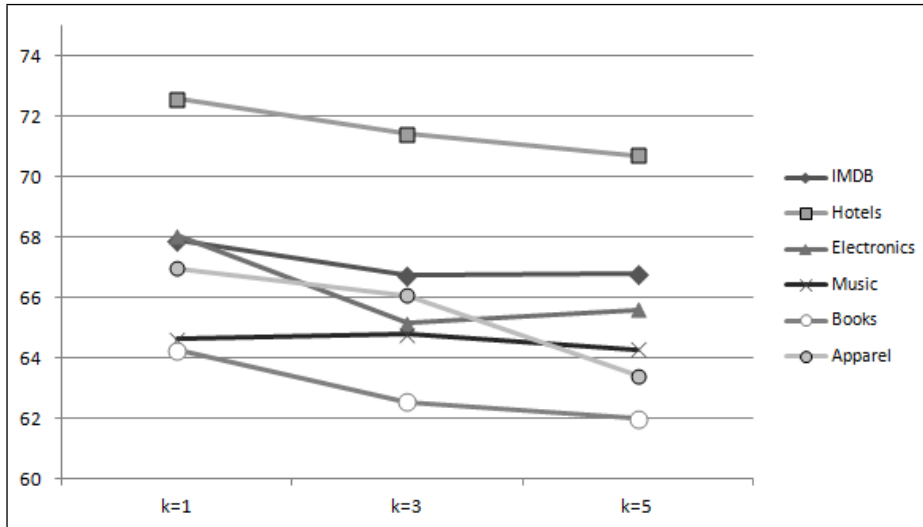
Fig. 2. Accuracies for varying k.

At the same time, we see a distinct trend in the number of lexicons used in making a prediction as k grows: At $k = 1$ no solution ranking takes place and all lexicons obtained from the solution of the single nearest case are used, while for higher values of k ranking selects the most frequent lexicons out of all solutions retrieved. Figure 3 shows how the number of lexicons used in a prediction are distributed as k changes for the film review data set. For higher values of k ranking tends to favour predictions with fewer lexicons, and to rely less on aggregated predictions from many lexicons.

5 Conclusions and Future Work

In this study we present an case-based approach for performing sentiment classification. The case description is a feature vector based on document statistics, and the case solution contains all lexicons that made correct predictions during training. We evaluated our approach on user generated reviews from six different domains. When compared to a baseline using the best result from a single-lexicon classifier, our results remain competitive with no statistically significant performance difference while producing more robust results by eliminating the need to fix a lexicon prior to making predictions. This illustrates the potential of case-based methods as an important component in unsupervised sentiment classification.

We see the following areas as interesting paths to further improve the case-based sentiment classifier: in this study we have restricted the solution search space to five sentiment lexicons while populating the case base, and saw that a considerable number of cases were discarded as no correct predictions could be
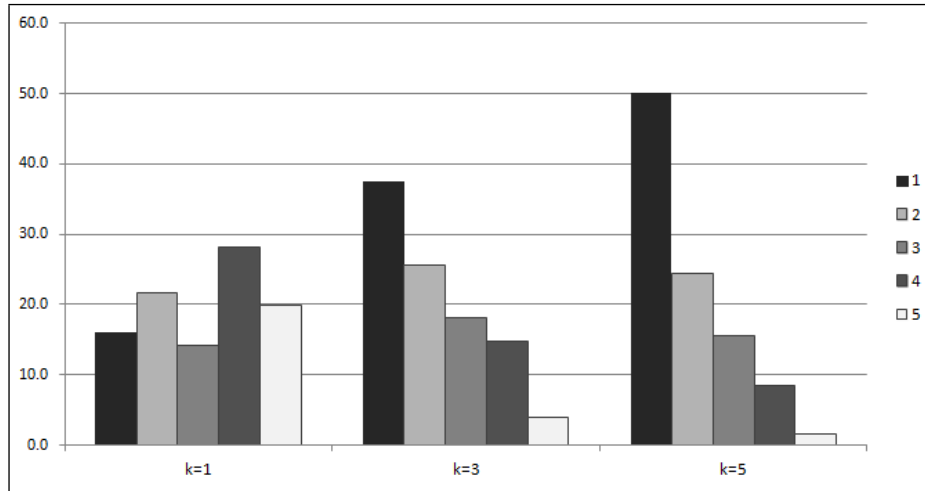
Fig. 3. Distribution of total number of lexicons used in prediction for varying k (film reviews case base).

found. Our approach can be easily be extended to consider more lexicons during the case base population stage, and this can help reducing the discard ratio by recording more cases for later reuse. Additional unsupervised algorithms can be added during training in a similar way.

Case reuse in our experiment relies on a ranking step to select the most frequently found lexicons from all solutions retrieved. The predictions of each selected lexicon are then aggregated to determine document sentiment. We have seen that, as the number of retrieved solutions grows ($k > 1$), ranking tends to favour single lexicon predictions causing a slight deterioration on performance. Experimenting with other ranking methods that benefit from aggregating predictions from many lexicons while still being able to produce scalable results for large search spaces is an interesting problem in developing better case-based approaches.

Finally, easily adding new cases from additional training data is a beneficial aspect of case-based methods. Investigating how this can be best achieved by using case base maintenance policies for addition and deletion is also an interesting path for future research, in particular when considering large volumes of out-of-domain training data with uneven label and domain distribution.

References

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of 2002 EMNLP, ACL (2002) 79–86
2. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th WWW (2003), ACM (2003) 528

3. Abbasi, A., Chen, H., Salem, A.: Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Transactions on Information Systems (TOIS) **26**(3) (2008) 12

4. Aue, A., Gamon, M.: Customizing sentiment classifiers to new domains: a case study. In: Proceedings of Recent Advances in Natural Language Processing (RANLP). (2005)

5. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence **22**(2) (2006) 110–125

6. Taboada, M., Voll, K., Brooke, J.: Extracting sentiment as a function of discourse structure and topicality. In: Technical Report TR 2008-20, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada (2008)

7. Ohana, B., Tierney, B., Delany, S.J.: Domain Independent Sentiment Classification with Many Lexicons. In: Mining and the Web Workshop - AINA 2011, IEEE (2011) 632–637

8. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting of the ACL, ACL (2004) 271

9. Li, S., Huang, C.R., Zhou, G., Lee, S.Y.M.: Employing personal/impersonal views in supervised and semi-supervised sentiment classification. In: 48th Annual Meeting of the ACL, ACL (2010) 414–423

10. Sood, S., Owsley, S., Hammond, K., Birnbaum, L.: Reasoning through search: A novel approach to sentiment classification. In: Technical Report NWU-EECS-07-05, Northwestern University (2007)

11. Tan, S., Wu, G., Tang, H., Cheng, X.: A novel scheme for domain-transfer problem in the context of sentiment analysis. In: Proceedings of the sixteenth CIKM, ACM (2007) 979–982

12. Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Annual Meeting-ACL. Volume 45. (2007) 440

13. Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M., Others: The general inquirer: A computer approach to content analysis. MIT Press, Cambridge, MA (1966)

14. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: 35th Annual Meeting of the ACL. ACL '98, Association for Computational Linguistics (1997) 174–181

15. Turney, P.D., Littman, M.L.: Unsupervised learning of semantic orientation from a hundred-billion-word corpus. In: Technical Report ERB-1094, National Research Council of Canada (2002)

16. Velikovich, L., Blair-Goldensohn, S., Hannan, K., McDonald, R.: The viability of web-derived polarity lexicons. In: HLT'10: Annual Conference of the North American Chapter of the ACL, ACL (2010) 777–785

17. Mohammad, S., Dunne, C., Dorr, B.: Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: Proceedings of EMNLP 2009, ACL (2009) 599–608

18. Esuli, A., Sebastiani, F.: SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In: LREC 2006. (2006) 417–422

19. Miller, G.A.: WordNet: a lexical database for English. Communications of the ACM **38**(11) (1995) 41

20. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: SentiFul: A Lexicon for Sentiment Analysis. IEEE Transactions on Affective Computing (99) (2011) 1

21. Denecke, K.: Are SentiWordNet scores suited for multi-domain sentiment classification? In: ICDIM 2009., IEEE (2009) 1–6

22. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI communications **7**(1) (1994) 39–59

23. Salton, G., Lesk, M.: The smart automatic document retrieval systems an illustration. Communications of the ACM **8**(6) (1965) 391–398

24. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT/EMNLP'05, ACL (2005) 354

25. Chapman, W., Bridewell, W., Hanbury, P., Cooper, G., Buchanan, B.: A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of biomedical informatics **34**(5) (2001) 301–310

26. Baccianella, S., Esuli, A., Sebastiani, F.: Multi-facet rating of product reviews. Advances in Information Retrieval (2009) 461–472

27. Jindal, N., Liu, B.: Opinion spam and analysis. In: Proceedings of the conference on Web search and Web data mining (WSDM'08), ACM (2008) 219–230

28. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research **7** (2006) 1–30