Construction of Enhanced Sentiment Sensitive Thesaurus for Cross Domain Sentiment Classification Using Wiktionary

P. Sanju and T. T. Mirnalinee

Abstract Sentiment classification is classification of reviews into positive or negative depends on the sentiment words expressed in reviews. Automatic sentiment classification is necessary in various applications such as market analysis, opinion mining, contextual advertisement and opinion summarization. However, sentiments are expressed differently in different domain and annotating label for every domain of interest is expensive and time consuming. In cross domain sentiment classification, a sentiment classifier trained in source domain is applied to classify reviews of target domain, always produce low performance due to the occurrence of features mismatch between source domain and target domain. The proposed method develops solution to feature mismatch problem in cross domain sentiment classification by creating enhanced sentiment sensitive thesaurus using wiktionary. The enhanced sentiment sensitive thesaurus aligns different words in expressing the same sentiment not only from different domains of reviews and from wiktionary to increase the classification performance in target domain. In this paper, the proposed method describes the method of construction of enhanced sentiment sensitive thesaurus which will be useful for cross domain sentiment classification.

Keywords Cross domain sentiment classification · Enhanced sentiment sensitive thesaurus · Domain adaptation · Data mining

P. Sanju (✉)
University College of Engineering, Anna University, Villupuram, Tamil Nadu, India
e-mail: Sanjupani_79@yahoo.co.in

T. T. Mirnalinee
SSN College of Engineering, Anna University, Chennai, Tamil Nadu, India
e-mail: mirna123@gmail.com

M. Pant et al. (eds.), *Proceedings of the Third International Conference on Soft Computing for Problem Solving*, Advances in Intelligent Systems and Computing 259,
DOI: 10.1007/978-81-322-1768-8_18,

1 Introduction

Nowadays purchasing of goods through online has been increased. In connection with online purchase, customers' opinions are shared to manufacturer through online. Such kind of opinions and sentiment information of the customers are overwhelming and growing exponentially which becomes a tedious work for the manufacture to classify the reviews manually. An automatic sentiment classifier is classification of reviews into positive or negative based on the sentiment words expressed in documents which is necessary to be developed for the manufacturer and the customer in order to analyze the reviews of the customers. The goal of sentiment classification is to discover customer opinion on a product. Sentiment classification has been applied in various tasks such as opinion mining [1], market analysis [2], opinion summarization [3] and contextual advertising [4].

Constructing general sentiment classifier is complex because in all domains the sentiment words have different connotation. Sentiment classification problem is more challenging because sentiments can be expressed in a more subtle manner. For example, the sentence, "How could anyone sit through this movie!" [5] contains not even a single sentiment words to express the sentiment but it is obviously negative. These kinds of sentences are available plenty while conveying the reviews of the product purchased. To develop sentiment classifier for these types of sentences is very complex.

Generally two types of approaches are being used for sentiment classification. The first approach is machine learning [5], in which classifier is trained using feature vectors and it produce more accurate results. The Second approach is semantic orientation [6, 7], it does not require prior training, instead it measures semantic orientation of sentences used in documents. Existing machine learning approaches [5, 8] rely on supervised learning models and performance of these models depends on manually labeled training data. However, such labeled data are not always available in practical applications and it is well known that sentiment classifier trained in one domain may not produce satisfactory results when it is used in another domain, since sentiments are expressed differently in different domain. Table 1 shows user review sentences from two domains books and electronics. In the books domain, the words "*well researched*" and "*interesting*"are used to express positive sentiments and "*disappointed*" is used to express negative sentiments. While in electronics domain "*fast*", *reliable, compact* and "*sharp*" are used to express the positive sentiment, "*blurry*" is negative sentiments. Due to the mismatch between domain specific words of source and target domain, classifier trained in one domain may not work well when directly applied to other domain.

In literature, Pan et al. [9] proposed spectral feature alignment algorithms to solve feature mismatch problem by aligning domain specific words from different domains into unified cluster with the help of domain independent words and then unified cluster is used to train a classifier in target domain. Bollegale et al. [10] proposed cross domain sentiment classification by creating sentiment sensitive

Table 1 Cross domain sentiment classification examples: Reviews of books and electronics products. Bold faces are domain specific words and italic words are domain independent words

Label	Books source domain	Electronics target domain
+	This book is *excellent* and **well researched**	SanDisk products are *excellent*, **fast** and **reliable**
+	This is an **interesting** and *Good* story book	*Good* **Compac**t, easy to operate, looks **sharp**
−	When I read this book, I really **disappointed,** so *never buy* this book	It is **blurry** in dark settings, I would *never buy* this product

thesaurus which aligns different words that express the same sentiments. They expanded feature vector using created sentiment sensitive thesaurus while training a binary classifier. In existing approaches [10], they didn't give much more importance to the adjectives used in the reviews and also they collected more semantically similar sentiment elements from the reviews itself. In this paper, the proposed method creates an enhanced sentiment sensitive thesaurus (ESST) which aligns semantically similar words from various domains as well as sentiment elements from wiktionary. The idea behind this approach is, it is well known that adjectives are used to express sentiment in all domains, so, it is necessary to collect more adjectives or sentiment elements from any lexical knowledge base to solve the feature mismatch problem. The proposed method creates an enhanced sentiment sensitive thesaurus which collects more sentiment elements from wiktionary knowledge base because it has more semantic relatedness information.

To create an enhanced sentiment sensitive thesaurus, first, domain independent features and domain specific features are collected from the reviews. Second, co-occurrence matrix is computed between each domain independent features with domain specific features. Third, features are weighted using point wise mutual information. Finally, semantically similar words are collected for each domain specific word by computing similarity measure between two domain specific elements based on the PMI weighted ratio. At the same time semantically similar sentiment words for each adjective of the given reviews are collected from wiktionary with the help of java wiktionary library [11] and these sentiment words are added to the already created thesaurus.

In rest of the paper is organized as follows: Sect. 2 describes related work. Section 3 describes definitions and Methodology used in proposed work. Section 4 describes experimental setup and solution. Section 5 concludes the proposed work.

2 Related Works

Many machine learning techniques have been deployed for sentiment classification. The existing methods are helpful in classifying the sentiments at various levels i.e. Document level [9, 10], sentence level [12] and word level [13, 14]. The proposed work is based on document level that classifies the documents as positive or negative, based on overall sentiment words expressed in the documents. Two approaches have been used in existing sentiment classification studies; machine learning and semantic orientation. In machine learning approaches, documents are represented as feature vector and it is trained by various classification algorithms such as Naive Bayes, Maximum entropy and SVM [8]. Two types of semantic orientation approaches [15]: Corpus based and dictionary based techniques used in existing research work. Corpus based techniques aim to find co-occurrence patterns of the words to determine the sentiments. Dictionary based methods utilize synonyms, antonyms and hierarchies in wordnet or sentiwordnet to determine the sentiments.

Turney [6] used supervised learning techniques with mutual information to predict overall document sentiment by averaging out the semantic orientation of phrases in document. Turney [5] used Naive Bayes, Maximum Entropy and SVM for classifying movie reviews and received best results using SVM. Kennedy and Inkpen [16] used Contextual valence shifter to predict sentiment of the sentences. The Work on the papers [5, 6, 16] focuses only on classify the given reviews in particular domain. Lin et al. [17] proposed a novel probabilistic modeling framework called joint sentiment topic model which detects sentiments and topic simultaneously from text.

In sentiment classification, many researchers have been concentrated on online lexical resources such as sentiwordnet, wiktionary and Wikipedia which are publically available for research purpose. SentiWordNet [18] is a lexical resource, containing opinion information on terms extracted from the wordnet database where each term is associated with numerical scores indicating positive and negative information. Liu et al. [19] proposed movie rating and review summarization in mobile environment using Latent semantic analysis. Khan et al. [20] proposed sentiment classification by sentence level semantic orientation using sentiwordnet from online reviews and blog which classify the reviews into objective or subjective sentences. The semantic score of the subjective sentences are extracted from sentiwordnet to calculate their polarity as positive or negative.

Zesch et al. [11] proposed extracting lexical semantic knowledge from Wikipedia. They proposed two application programming interfaces for Wikipedia and wiktionary which are designed for mining the lexical semantic information from knowledge bases. Chesley et al. [21] proposed automatic classification of blog posts using verbs, adjectives and information from wiktionary. They used wiktionary to determine polarity of the adjectives in the text. Chihli et al. [22] proposed using objective words in sentiwordnet to improve sentiment classification for word of mouth. Wiebe [23] proposed learning subjective adjectives from corpora. They identified subjectivity of sentences form corpora by clustering the words according to distributional lin's similarity measure, seeded by a small amount of detailed manual annotation. Hatzivassiloglou et al. [24] proposed predicting the semantic orientation of adjectives which describes conjunction between adjectives provides indirect information about semantic orientation of sentences. The proposed method is different from existing approaches [23, 24], more sentiment elements are extracted from the wiktionary with the help of seed adjectives of the given reviews which will be useful to predict the sentiment of target domain. Muller et al. [25] proposed Domain-Specific Information Retrieval using wiktionary and Wikipedia.

Many researchers have addressed the problem in cross domain sentiment classification [9, 10, 18, 26]. Blitzer et al. [26] addressed the problem in cross domain sentiment classification using structural correspondence learning algorithm where frequent words in both source and target domain were selected as candidate pivot features and linear predictors are trained to predict the occurrences of those pivot features. In structural corresponding learning-mutual information approach, the mutual information between a feature and the domain label is used to select pivot features instead of co-occurrence frequency. Pan et al. [9] proposed a spectral feature alignment algorithm to align domain specific words from different domain into unified clusters with the help of domain independent words as a bridge and bipartite graph is constructed between domain specific and domain independent features and then the domain specific clusters can be used to train a sentiment classifier in target domain. Yan et al. [27] used self growth algorithm to generate a cross domain sentiment word list which is used in sentiment classification of web news. Bollegale et al. [10] proposed cross domain sentiment classification by creating sentiment sensitive thesaurus which aligns different words that express the same sentiments. They expanded feature vector using created thesaurus while training a binary classifier.

All existing cross domain sentiment classification solution, identify the sentiment of unlabeled data of target domain with the help of labeled data in source domain by applying various methodology. The proposed solution for cross domain sentiment classification predict the unlabeled data of target domain by creating extended sentiment sensitive thesaurus using reviews and wiktionary which increases the classification performance in target domain.

3 Methodology

3.1 Definitions

In this section, some definitions are given to clarify basic terminology.

Domain-A domain D denotes a class of entities in the real world. For example, different types of product such as DVD, Kitchen appliances, books and electronics.

Sentiment-Given a specific domain D, sentiment data are the text documents which express the user opinion about the entities of the domain.

Generally, in sentiment classification tasks, single word (unigram), bigram, ngram are used as features. In this work, unigram and bigram are used as features.

Cross Domain Sentiment Classification-Given a set of labeled reviews $D_s = \{(x_i, y_i)\}$ from source domain where x_i represents features and y_i represent sentiment label y_i € {+1, −1}. To predict the label of unlabeled target domain $D_t = \{x_j\}$ where x_j represent features in target domain. Classifier is trained by labeled reviews source domain and it is applied to classify the unlabeled reviews of target domain.

3.2 Enhanced Sentiment Sensitive Thesaurus (ESST) Using Wiktionary

In cross domain sentiment classification, sentiment classifier trained from one domain (source) is applied to another (target) domain. Obviously it produces poor performance because trained features are mismatched with target domain features. This feature mismatch problem in cross domain sentiment classification is solved by creating enhanced sentiment sensitive thesaurus using all domain reviews and knowledge from Wiktionary. Wiktionary is online dictionary which has glosses and synset for each word. Every word in the wiktionary is clearly defined with list of synonyms. Wiktionary also contains more information which is found in linguistic knowledge base (wordnet), like definitions, synonyms, and hyponyms, and also has additional types of information, e.g. abbreviations, compounds or contractions, which are usually not found in LKBs. Another difference to LKBs is that Wiktionary contains not only lexical knowledge for particular language, but also for other languages. Figure 1 shows the construction of enhanced sentiment sensitive Thesaurus. The Enhanced sentiment sensitive Thesaurus is constructed using the knowledge from the Wiktionary, labeled/unlabeled reviews from source domain and unlabeled reviews from target domain. To interact with wiktionary, JWKTL (java wiktionary library) is used which is released in Ubiquitous Knowledge Processing Lab [11].

Given a labeled or unlabeled review, first split the reviews into set of sentences, and then perform parts of speech (POS) tagging and Lemmatization is performed using RASP [28] system. Lemmatization is used to reduce the features by removing inflectional endings only and to return the base or dictionary form of a word, which is known as the *lemma.* By using a stop word removal algorithm based on the POS tagging to filter out unwanted words and retaining nouns, verbs, adjectives, adverbs. Table 2 shows how to extract unigrams, bigrams and sentiment elements from reviews. First, model the review as bag of words and then extract unigrams, bigrams from each sentence. Bigrams are necessary in sentiment classification, since semantic orientations of sentences are identified by bigrams. Generally unigrams, bigrams are called as lexical elements or sentiment elements. Next, from each source domain labeled reviews, sentiment elements are created by appending the label of the review to each lexical element. The notation *p to indicate positive sentiment elements and *N to indicate negative elements.

Domain independent features are extracted from all domains which occur frequently in all domains. Domain specific features from various domains are

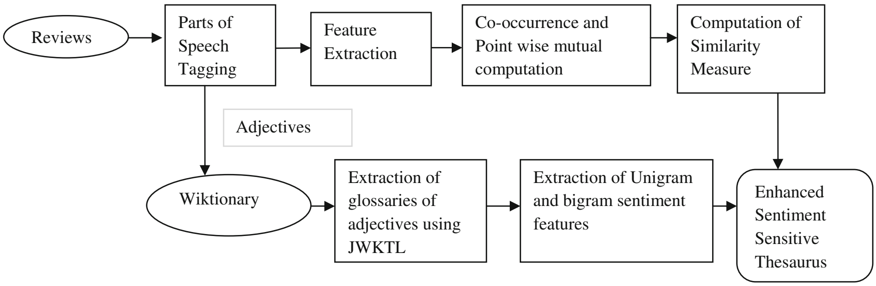

Fig. 1 Construction of enhanced sentiment sensitive thesaurus

Table 2 Example for extracting unigram, bigram and sentiment elements from positive review sentence

Sentence	An excellent workbook full of delicious recipes
POS tags	An_AT1 excellent_JJ workbook_NN1 full_JJ of_IO delicious_JJ recipes_NN2
Unigrams	Excellent, workbook, full, delicious, recipes
Bigrams	Excellent + workbook, workbook + full, full + delicious, delicious + recipes
Sentiment elements	Excellent * p, workbook * p, full * pl, delicious * p, recipes * p Excellent + workbook * p, workbook + full * p, full + delicious * p, delicious + recipes * p

aligned in ESST with the help of domain independent features. Features other than domain independent features are considered as domain specific features. Here, each domain independent feature is represented as feature vector based on the domain independent feature co-occurring with its distributional contexts or feature. Later on the co-occurrence matrix is computed between domain independent features with domain specific features. Semantic meanings of the words are found based on co-occurrences of the terms used in the documents. Co-occurrence [29] is used to measure the similarity between meanings of the two words based on words distributional contexts.

After the computation of co-occurrence between domain independent features and domain specific features, value of the domain specific features in the co-occurrence matrix are weighted using point wise mutual information [13]. Point wise Mutual information is an information theoretic measure of association between two lexical elements. Point wise mutual information is useful in numerous tasks such as similarity measurement [30], word classification [31] and word clustering [32]. To align domain specific features, transpose PMI weighted matrix and find similarity between domain specific words based on its distributional contexts.

Point wise mutual information (PMI) is used to weight the features based on the co-occurrence value in the co-occurrence matrix. For each domain specific feature s, domain independent feature v that co-occurs with domain specific features s contributes feature vector **s**. The value of the feature v in vector **s** is denoted by M (**s,** v).

$$M(\boldsymbol{s}, v) = log\left(\frac{\frac{c(s,v)}{N}}{\frac{\sum_{i=1}^{n} C(i,v)}{N} X \frac{\sum_{j=1}^{m} C(s,j)}{N}}\right) \quad (1)$$

M(s, v) is the point wise mutual information between a domain independent features s and domain specific feature v. In Eq. 1, C(s, v) represents number of review sentences in which both domain independent feature s and domain specific feature v co-occur, C(s, j) represents number of times domain independent feature s occur in review sentences. C(i, v) represents number of times domain specific feature v occur in the review sentences. n and m denotes the total no of domain independent elements and domain specific features and N $= \sum_{i=1}^{n} \sum_{j=1}^{m} C(i, j)$.

After the computation of PMI values, the semantically similar features for each domain specific feature is computed by finding similarity score between domain specific features with every other domain specific features of PMI weighted matrix. For example, similarity score between two domain specific features s, t (both s and t have feature vectors **s**, **t**) is computed by following formula

$$\mathrm{T(s,t)} = \frac{\sum_{\mathrm{v} \in \{\mathrm{x}|\mathrm{M}(\mathbf{s},\mathrm{x}) > 0\}} \mathrm{M}(\mathbf{t}, \mathrm{v})}{\sum_{\mathrm{v} \in \{\mathrm{x}|\mathrm{M}(\mathbf{t},\mathrm{x}) > 0\}} \mathrm{M}(\mathbf{t}, \mathrm{v})} \quad (2)$$

The similarity score T(s, t) is the proportion of PMI weighted values of the domain specific feature t that are shared with domain specific feature s. The distributional hypothesis states that words that have similar distributions are semantically similar [30] i.e. Two words are semantically similar if two words occur with same distributional contexts.

Enhanced Sentiment sensitive thesaurus (ESST) aligns many domain specific features that are semantically similar from various domains by computing similarity measure Eq. (2) for every domain specific feature in PMI weighted matrix. For every domain specific feature s, ESST list up many domain specific features t based on descending order of the similarity score.

Moreover, Enhanced Sentiment Sensitive Thesaurus (ESST) collect more semantically similar features from wiktionary using java wiktionary library (JWKTL) which is very useful tool to extract more semantic relatedness from Wiktionary. To enhance sentiment sensitive thesaurus, first, seed adjectives are extracted from the given reviews. Second, XML dump file of wiktionary is downloaded from web. Third, using JWKTL, the XML file is parsed and data is extracted from wiktionary. Next, the extracted adjectives are given as input to the JWKTL and the corresponding glossaries of each adjective are extracted from the wiktionary. Finally, unigram and bigrams from the extracted glossaries are collected for each adjective and then these are appended to the ESST with similarity score of 0.9. This enhanced sentiment sensitive thesaurus is entirely different from normal thesaurus because it finds the similar meaning of the words based on the co-occurrence terms used in the sentences. The created ESST is very useful to classify the reviews of target domain in cross domain sentiment classification and also it will increase the classification performance in target domain.

3.3 Algorithm: Creation of Enhanced Sentiment Sensitive Thesaurus

Input: labeled source Domain data $D_{sr} = \{X_{sr}, Y_{sr}\}$ and unlabeled source domain $D_{sr} = \{X_i\}$ and unlabeled target Domain $D_{tr} = \{X_{tr}\}$ and Wiktionary dump file
Output: Enhanced Sentiment Sensitive Thesaurus

1. Extract Domain Independent features and domain specific features from the given Reviews.
2. Create co-occurrence matrix between domain independent features with domain specific features.
3. Compute Point Wise Mutual Information for each features using Eq. (1).
4. Compute similarity measure using Eq. (2) for every domain specific features with every other domain specific features.
5. Enhanced sentiment thesaurus aligns many domain specific features from various domains for every domain specific feature.
6. Extract seed adjectives from the given Reviews and these seed adjectives are given as input to the wiktionary parser.
7. Glossaries for each seed adjectives are extracted using java wiktionary library(JWKTL)
8. Unigram and bigrams are generated from the glossaries that are added with Enhanced Sentiment sensitive thesaurus.

4 Experiments

4.1 Data Sets

Amazon product reviews are bench mark data set which consist of four different product types: Books, DVDs, electronics and kitchen appliances are chosen for the proposed work. Each review is assigned with rating (0–5 stars). Review with rating >3 are labeled as positive and review with rating <3 are labeled as negative. The data sets structure is shown in Table 3. For each domain; there are 1,000 positive reviews with 1,000 negative reviews. Each domain also has some unlabeled reviews.

For experiments, among four domains, three domains are selected as source domain and another domain as target domain. To create enhanced sentiment sensitive thesaurus, 800 positive reviews and 800 negative reviews are selected from each source domain and some unlabeled reviews are selected from all domain. Wiktionary dump file is downloaded and stored as separate file. More sentiment features are extracted from wiktionary with the help seed adjectives of the given reviews. Here enhanced sentiment sensitive thesaurus is created automatically by considering each domain as target domain. So four different types of

Table 3 Amazon product data sets

Domain	Positive	Negative	Unlabeled
Kitchen	1,000	1,000	16,746
DVDs	1,000	1,000	34,377
Electronics	1,000	1,000	13,116
Books	1,000	1,000	5,947

thesaurus is created to solve the feature mismatch problem in cross domain sentiment classification

The enhanced sentiment sensitive thesaurus (ESST) is automatically created by giving reviews of source and target domain. To study the effect of multiple thesauruses, different enhanced sentiment sensitive thesaurus is created for one source domain/one target domain and two source domain and one target domain. From the analysis of multiple created thesauruses, enhanced sentiment sensitive thesaurus created from three source domains may be effective because more sentiment features are identified from the reviews as well as from the wiktionary. Moreover, this enhanced sentiment sensitive thesaurus highly effective to classify the reviews of unseen target domain features because it collect more semantically similar features from wiktionary. For example, ESST align 880 features for the adjective good from reviews as well as from wiktionary

5 Conclusions

Enhanced Sentiment Sensitive Thesaurus is created automatically based on the given source domain and target domain reviews. The proposed ESST which aligns more semantically similar sentiment features from source and target domain as well as additional sentiment features from wiktionary in order to increase the classification performance in target domain. This ESST is suitable to the applications when source domain has labeled data and target domain has unlabeled data. Here the gap between source and target domain are bridged by leveraging additional knowledge from wiktionary. In future, cross domain sentiment classification will be performed by feature vector augmentation using created Enhanced Sentiment Sensitive Thesaurus during training and testing time of sentiment classification.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)
2. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceeding of the 10th ACM SIG KDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. New York, USA (2004)

3. Lu, Y., Zhai, C., Sundaresan, N.: Rated aspect summarization of short comments. In: The Proceedings of the 18th Conference on World Wide Web, pp. 131–140. ACM, New York (2009)
4. Fan, T.-K., Chang, C.H.: Sentiment-oriented contextual advertising. Knowl. Inf. Syst. **23**(3), 321–344 (2010)
5. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79–86. Philadelphia, July 2002
6. Turney, P.D.: Humbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: ACL'02 Proceedings of the 40th Annual Meeting on association for Computational Linguistics, pp. 417–424 (2002)
7. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: ACL 1997, pp. 174–181 (1997)
8. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: Proceedings 10th European Conference on Machine Learning (ECML), pp. 137–142. Springer Verlag, Berlin (1998)
9. Pan, S.J., Ni, X., Sun, J.T., Yan, Q., Chen, Z.: Cross-domain Sentiment classification via spectral feature alignment. In: WWW 2010, Raleigh, North Carolina, USA.ACM 978-1-60558-799-8/10/04, 26–30 April 2010
10. Bollegala, D., Weir, D., carroll, J.: Cross domain sentiment classification using a sentiment sensitive thesaurus. IEEE Trans. Knowl. Data Eng. **25**(8) (2013)
11. Zesch, T., Muller, C., Gurevych, I.: Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary (2008)
12. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: EMNLP 2003, pp. 129–136 (2003)
13. Pantel, P., Ravichandran, D.: Automatically labeling semantic classes. In: NAACL-HLT'04, pp. 321–328 (2004)
14. Breck, E., Choi, Y., Cardie, C.: Identifying expressions of opinion in context. In: IJCAI 2007 (2007)
15. Turney, P.: Similarity of semantic relations. Comput. Linguist. **32**(3), 379–416 (2006)
16. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Comput. Intell. **22**, 110–125 (2006)
17. Lin, C., He, Y., Everson, R., Ruger, S.: Weakly supervised joint sentiment- topic detection from text, J. Latex class files, 24 (2011)
18. Esuli, A., Sebastiani, F.: Sentiwordnet: a publicly available lexical resource for opinion mining. In: LREC 2006. pp. 417–422 (2006)
19. Liu, C.L., Hsaio, W.H., Lee, C.H., Lu, G.C., jou, E.: Movie Rating and review summarization in mobile environment. In: IEEE 2012, vol. 42, pp. 397–407 (2012)
20. Khan, A., Baharudin, B.: Sentiment classification by sentence level semantic orientation using sentiwordnet from online reviews and Blogs. Int. J. Comput. Sci. Emerging Tech. 2(4) (2011)
21. Chesley, P., Bruce Vincent, B., Xu, L., Srihari, R.K.: Using verbs and adjectives to automatically classify blog sentiment. In: American Association for Artificial intelligence (2005)
22. Hung, C., Lin, H.: Using objective words in sentiwordnet to improve the sentiment classification word of mouth. In: IEEE Intelligent Systems (2013)
23. Wiebe, J.: Learning subjective adjectives from corpora. In: American Association for AI (2001)
24. Hatzivassiloglou, V., MckeownK.R.: Predicting the Semantic Orientation of Adjectives .pp. 176–182 (1995)
25. Muller, C., Gurevych, I.: Using Wikipedia and Wiktionary in Domain-Specific Information Retrieval (2008)

26. Blitzer, J., McDonald, R., Pereira, F.: Domain adaptation with structural correspondence learning. In: EMNLP 2006 (2006)
27. Yan, L., Zhang, Y.: News sentiment Analysis based on cross domain sentiment word lists and content classifiers. ADMA 2012.LNAI 7713, pp. 577–588. Springer-verlag, Berlin (2012)
28. Briscoe, T., Carroll, J., Watson, R.: The second release of the RASP system. In: COLING/ACL 2006 Interactive Presentation Sessions (2006)
29. Weeds, J., Weir, D.: Co-occurrence retrieval: a flexible framework for lexical distributional similarity. Comput. Linguist. **31**(4), 439–475 (2006)
30. Harris, Z.: Distributional structure. Word **10**, 146–162 (1954)
31. Turney, P.: Similarity of semantic relations. Comput. Linguist. **32**(3), 379–416 (2006)
32. Lin, D.: Automatic retrieval and clustering of similar words. In: ACL, pp. 768–774 (1998)
33. Hu, K., Zhang, Y., Hu, X.: A multi-domain adaptation for sentiment classification algorithm based on class distribution. In: IEEE International Conference on Granular Computing (2012)
34. Remus, R.: Domain adaptation using domain similarity- and domain complexity-based instance selection for cross-domain sentiment analysis. In: IEEE 12th International Conference on Data Mining Workshops (2012)