Procesamiento del lenguaje Natural, nº 41 (2008), pp. 201-208 recibido 7-05-2008; aceptado 16-06-2008

Determining the Semantic Orientation of Opinions on Products – a Comparative Analysis

Análisis comparativo de métodos para determinar la polaridad de opiniones sobre productos

Alexandra Balahur
Universidad de Alicante, DLSI
Apartado de Correos 99
E-03080, Alicante
abalahur@dlsi.ua.es

Andrés Montoyo
Universidad de Alicante, DLSI
Apartado de Correos 99
E-03080, Alicante
montoyo@dlsi.ua.es

Abstract: The high volume of user feedback on products under the form of reviews and forum or blog posts is helpful both to prospective buyers, as well as to producer companies. However, automatically determining the semantic orientation of the opinions expressed on different products and their features is a complex problem, requiring a series of steps: identifying the product features, extracting the opinion words present in a text and finally classifying them as positive or negative. This article concentrates on three approaches to solving the latter problem. One method employed determines polarity of the opinions expressed on the product features using on the one hand the sentiment bearing words in WordNet Affect (Strapparava and Valitutti, 2004). Two other methods explored involved determining the polarity of opinion holders (feature attributes) using Support Vector Machines Sequential Minimal Optimization (Platt, 1998) machine learning with the Normalized Google Distance (Cilibrasi and Vitanyi, 2006) and, respectively, with Latent Semantic Analysis (Deerwester et al., 1990) on a specialized versus a non-specialized corpus of user reviews. We comparatively analyze the methods, show the advantages and disadvantages resulted from using each of them and the results obtained by performing an evaluation on our opinion mining and summarization system.

Keywords: opinion mining, summarization, Support Vector Machines Sequential Minimal Optimization, Normalized Google Distance, Latent Semantic Analysis.

Resumen La gran cantidad de opiniones que los usuarios emiten sobre las características de los productos en blogs, foros y en documentos en internet, son de gran ayuda para los posibles compradores o para las compañías que los producen. Sin embargo, determinar de forma automática si un usuario tiene una opinión positiva o negativa de las características de un producto o del propio producto es un problema complejo que requiere de varios pasos para su resolución. Inicialmente hay que identificar las características del producto, extraer los términos que expresan la opinión del usuario y finalmente clasificar el producto de forma positiva o negativa. Este artículo describe un método para resumir los comentarios positivos o negativos sobre el producto a partir de las opiniones que los usuarios expresan a través de las características de los productos. Este problema se resuelve utilizando varias aproximaciones. Inicialmente se utilizan las palabras que aparecen en WordNet Affect (Strapparava and Valitutti, 2004) que expresan sentimiento. Finalmente se utiliza el método de aprendizaje automatico (Support Vector Machines Sequential Minimal Optimization (Platt, 1998)) aplicado a las medidas de similitud denominadas Normalized Google Distance (Cilibrasi and Vitanyi, 2006) y Latent Semantic Analysis (Deerwester et al., 1990). Los resultados obtenidos por estas medidas de similitud se comparan, para posteriormente ser analizados y presentar las ventajas y los inconvenientes cuando se aplican al sistema de minería y resúmenes de opiniones.

Palabras clave: minería de opiniones, resumir, Support Vector Machines Sequential Minimal Optimization, Normalized Google Distance, Latent Semantic Analysis.

1 Introduction

The multitude of products of any category presently available on the market offer the prospective buyer both the opportunity to best choose according to personal needs, as well as the difficulty of choice and the need for detailed information on product capabilities. On the other hand, recent years have brought about a large amount of public user feedback on products, in the form of reviews on e-commerce sites, forums or blogs. Nevertheless, the high volume of text containing this information makes it impossible for a potential customer to review all relevant data, while partially reading reviews can result in misinformation or biases.

The present paper concentrates on methods to solve the issues involved in determining the semantic orientation of opinions in the task of automatically mining user reviews on products and presenting the potential buyers with summaries of positive and negative opinions expressed on the product and its features. This task is known in the literature under the name of "feature-based opinion mining and summarization" (Hu and Liu, 2004). For a given product, producing a feature-driven summary of the opinions expressed on its features is equivalent to producing an output in the form (feature, percentage of positive opinions, percentage of negative opinions). The process consists of three distinct steps. The first one involves discovering the potential features that will be commented on in the product reviews; the second step is identifying the opinions expressed in reviews on each of the features; the last step consists in summarizing the polarity of the opinions expressed on each feature as percentages of positive and negative opinions.

It is important to note the difference between this task and the classical definition of summarization (Ding, Liu and Yu, 2008), as this particular type of summarization only refers to summarizing the polarity of opinions expressed about features.

The present paper has the following structure: in section 2 we describe related work in feature-based summarization of customer reviews. Section 3 delimits the problem we intend to solve and our contribution, relating to our feature-driven opinion summarization system, whose extended description can be found in (Balahur and Montoyo, 2008). The contribution of the present work is described in the next two sections: in section 4 we present a comparative analysis of two methods for the classification of the opinion polarity using SVM SMO with the Normalized Google Distance and Latent Semantic Analysis, respectively. In section 5, we explore a method to extract feature polarity using subjective phrases constructed with the help of emotion words found in WordNet Affect (Strapparava and Valitutti, 2004). The next section shows the results obtained when evaluating our system employing the two methods in section 4 and the approach described in section 5. Finally, we conclude on our approach and sketch the directions for future work.

2 Related work

Recent years have brought about a growing interest in the field of opinion mining and sentiment analysis. The high number of applications, such as multi-perspective question answering, automatic market research or recommender systems, have determined extensive research - in classifying documents for polarity (Riloff and Wiebe, 2003; Dave, Lawrence and Penncock, 2003; Pang, Lee and Vaithyanan, 2002; Turney, 2002), sentences (Wilson, Wiebe and Hwa, 2004; Hatzivassiloglou and Wiebe, 2000).

The idea of feature-based summarization of opinions expressed in customer reviews was proposed in (Hu and Liu, 2004). The approach the authors describe is lexicon-based and consists in discovering frequent features using association mining and determining the semantic orientation of opinions as polarity of adjectives (as opinion holders) that features are described by. The classification of adjectives is done using an initial list of seeds which is completed using the WordNet synonymy and antonymy relations. Infrequent features are deduced using the opinion holders. However, the fact that there is no well-organized structure of features and sub-features of products leads to the fact that, for example, the summarization of opinions is done for 720 features for an mp3 player (Ding, Liu and Yu, 2008). The question that arises is: would a user in a real-life situation be interested on whether the edges of a camera are round or flat and what the previous buyers think about that, or would a potential buyer like to see if the design of the product is fine or not, according to the many criteria developed by buyers to assess this feature? The work does not approach implicit features and does not classify the orientation of adjectives depending on the context. A solution to the latter problem is presented in (Ding, Liu and Yu, 2008), where the authors take a holistic approach to classifying adjectives, that is, consider not only the local context in which they appear next to the feature they determine, but also other adjectives appearing with the feature and their polarity in different contexts. In (Popescu and Etzioni, 2005), a more complex approach is used for feature-based summarization of opinions, employing web PMI (Pointwise Mutual Information) statistics for the explicit feature extraction and a technique called *relaxation labeling* for the assignation of polarity to the opinions. In this approach, dependency parsing is used together with ten extraction rules that were developed intuitively.

Our approach differs from and improves on previous work in a series of aspects. Firstly, we employ *anaphora resolution* and *dependency parsing* to ensure that features extracted and the opinions expressed are on the product we are interested in, and the review does not comment on a related one, nor does a near opinion word express a positive or negative thought about other topics than the product. Secondly, we employ an *offline method to determine product features and sub-features,* which allows the system to *gather for summarization the opinions expressed on sub-features* (as in the case of "edge") *to the feature they correspond* ("design"). We also *determine product specific feature attributes with polarity,* as well as *compute the set of learning examples* of feature-specific opinion words. This is accomplished using a corpus of opinions over the same product class, which is structured in two sections: arguments in favor and against ("pros and cons-style reviews"). Thirdly, the *classification* of opinion words is *feature-dependent* and does not rely on the local context in which they appear, but to a larger semantic context. In the case of employing the Normalized Google Distance scores for the machine learning algorithm, the context is the World Wide Web; in the case of LSA scores, the context is given by the corpus from which the model is learnt. We show the manner in which all these factors influence the system performance and at what cost.

Last, but not least, many of the opinions on products are expressed in an *indirect manner*, that is, not relating the product or its features with polarity words, but expressing an emotion about them. We propose a set of *patterns to extract such indirectly expressed opinions* using the emotion lists from WordNet Affect.

3 Problem definition and contribution

In the task of feature-driven opinion mining and summarization, the aim is to identify in user reviews the opinions expressed on the product and its features, determine if they are positive or negative and summarize them as percentage.

In order to fulfill the first step, the features of the product that will be commented upon in the reviews must be determined. Methods proposed included association mining (Hu and Liu, 2004), the use of WordNet (Fellbaum, 1999) relations (Popescu and Etzioni, 2004) or WordNet and commonsense knowledge in ConceptNet (Liu and Singh, 2004) as shown in (Balahur and Montoyo, 2008). In the present approach, we decided to add to the extracted knowledge the structured one comprised on the same sites that contain product reviews under the chapter technical details.

In the second step, the extraction is done using patterns, in case of lexicon-based approaches or rules, in case of dependency parsing solutions.

The last step consists in classifying opinions as positive or negative. In the method described by (Hu and Liu, 2004), the words considered as opinion holders are adjectives and they are classified using a core of annotated adjectives and the synonymy and antonymy relations in WordNet. However, this approach has a serious problem – residing in the fact that the polarity of feature attributes (or opinion holders, as they are called by the authors) are feature dependent, i.e. for example, "large" in the context of an LCD screen is a positive attribute, whereas in the context of a mobile phone, for instance, it is a negative one. A remedy to this problem is sought in (Ding, Liu and Yu, 2008), where the larger context is taken into consideration in the classification of opinion holders – using conjunction rules and the polarity of the opinion holders the adjectives to be classified appear with. The shortcoming of this approach is that opinion expression in user review is mostly an enumeration of qualities and faults of the products. In (Popescu and Etzioni), a complex function is employed to classify the polarity of opinions, based on the polarity of the surrounding context, but this approach too has the shortcoming that user reviews tend to be short and the negative and positive aspects are mixed without any specific order. (Kim and Hovy, 2006), on the other hand, use the statistics given by the Pointwise Mutual Information score together with the number of search engine hits of target words and positive and, respectively, negative words.

In our approach in (Balahur and Montoyo, 2008), we use the classification of feature specific attributes on the one hand and of opinion words extracted from “pros and cons”-style reviews for each of the product categories we are interested in. For English, sites that contain such types of reviews are “newegg.com” and “eopinions.com”, which are American, or “shopping.com”, on which the regional site can be chosen also for European countries. For Spanish, sites containing reviews in the form of pros and cons (“a favor” and “en contra” or “ventaja” and “desventaja”) are “quesabesde.com” or “ciao.es”. If the opinion word is contained in a “pro” section, then the extracted words are classified as positive feature attributes. In the contrary case, the word is classified as negative. For example:

Pros: Beautiful pictures, ease of use, high quality, 52mm lens.

Cons: high price, a bit big and bulky.

Features encountered in text: picture, use, quality, lens, price.

Feature attributes extracted: (positive): beautiful (picture), easy (use), high (quality), 52mm (lens); (negative): high (price).

Feature attributes remaining: big, bulky, which are both negative and correspond, according to the feature categorization made in section 4, to the “size” feature.

Our solution to the problem of feature attributes classification is using machine learning with two measures of similarity. On the one hand, we employ the Normalized Google Distance, which gives a measure of the strength of relationship between two considered words at the level of the entire WWW and on the other hand, we use the LSA, which gives the same measure of strength, but at a local corpus level. Classifying the feature attributes according to these scores and taking into consideration 6 anchor words that relate each word with the feature and known polarities, we show how the classification of feature attributes can be done in the feature context.

Last, but not least, in the reviews to be mined and summarized, however, other opinion words can be found and other manners of expressing opinion can be encountered, such as those describing emotional states related to the the product (for example, *“I love this camera”*) or to using it. Methods to solve these problems are discussed in section 5, where we show the list of patterns we used to extract from the reviews such phrases containing emotions to express opinions of the different product features using the words associated to different emotions from WordNet Affect. In the evaluation section, we show how the use of such patterns raised with 12% the recall of the system, while the precision of classification rose to the same degree.

4 Comparative experiments

In our previous approach, in order to assign polarity to each of the identified feature attributes of a product, we employed SMO SVM machine learning and the Normalized Google Distance (NGD). In this approach, we complete the solution with a classification employing Latent Semantic Analysis with Support Vector Machines classification.

For the NGD classification, we consider a set of anchors containing the terms *{featureName, happy, unsatisfied, nice, small, buy}*, that relate to all possible classes of products, as well as give an orientation to product feature attributes.

Further on, we build the classes of positive and negative examples for each of the feature attributes considered. From the list of classified feature attributes in the pros and cons reviews, we consider all positive and negative terms associated to the considered attribute features. We then complete the lists of positive and negative terms with their WordNet synonyms. Since the number of positive and negative examples must be equal, we will consider from each of the categories a number of elements equal to the size of the smallest set among the two, with a size of at least 10 and less or equal with 20. We give as example the classification of the feature attribute “tiny”', for the “size” feature. The set of positive feature attributes considered contains 15 terms such as “big”, “broad”, “bulky”, “massive”, “voluminous”, “large-scale” etc. and the set of negative feature attributes considered is composed as opposed examples, such as “small”, “petite”, “pocket-sized”, “little” etc. We use the anchor words to convert each of the 30 training words to 6-dimensional training vectors defined as v(j,i) = NGD(wi,aj), where aj with j ranging from 1 to 6 are the anchors and wi, with i from 1 to 30 are the words from the positive and negative categories. After obtaining the total 180 values for the vectors, we use SMO SVM to learn to distinguish the product specific nuances. For each of the new feature attributes we wish to classify, we calculate a new value of the vector vNew(j,word) = NGD(word, aj), with j ranging from 1 to 6 and classify it using the same anchors and trained SVM model.

In the example considered, we had the following results (by V1, V2, V3, V4, V5, and V6 we denote the six values corresponding to the NGD scores of the anchors and the words in the NGD classification scores, and the LSA scores, respectively; “pol” refers to the polarity of the feature attribute) :

FA	V1	V2	V3	V4	V5	V6	pol
small	1.52	1.87	0.82	1.75	1.92	1.93	pos
big	2.27	1.19	0.86	1.55	1.16	1.77	neg
bulky	1.33	1.17	0.92	1.13	1.12	1.16	neg
little	1.44	1.84	0.80	1.64	2.11	1.85	pos
tiny	1.51	1.41	0.82	1.32	1.60	1.36	pos

Table 1. Example NGD scores

The vector for the feature attribute “tiny” was classified by SVM as positive, using the training set specified above.

For the LSA classification, as in the case of NGD, we consider a set of anchors containing the terms *{featureName, happy, unsatisfied, nice, small, buy}*.

Further on, we build the classes of positive and negative examples for each of the feature attributes considered. From the list of classified feature attributes in the pros and cons reviews, we consider all positive and negative terms associated to the considered attribute features. We then complete the lists of positive and negative terms with their WordNet synonyms. Since the number of positive and negative examples must be equal, we will consider from each of the categories a number of elements equal to the size of the smallest set among the two, with a size of at least 10 and less or equal with 20. We give as example the classification of the feature attribute `”tiny”, for the “size” feature. The set of positive feature attributes considered contains 15 terms such as “big”, “broad”, “bulky”, “massive”, “voluminous”, “large-scale” etc. and the set of negative feature attributes considered is composed as opposed examples, such as “small”, “petite”, “pocket-sized”, “little” etc. We use the anchor words to convert each of the 30 training words to 6-dimensional training vectors defined as v(j,i) = LSA(wi,aj), where aj with j ranging from 1 to 6 are the anchors and wi, with i from 1 to 30 are the words from the positive and negative categories. After obtaining the total 180 values for the vectors, we use SMO SVM to learn to distinguish the product specific nuances. For each of the new feature attributes we wish to classify, we calculate a new value of the vector vNew(j,word) = LSA(word, aj), with j ranging from 1 to 6 and classify it using the same anchors and trained SVM model.

We employed the classification on the corpus present for training in the Infomap software pack. The blank lines represent the words which were not found in the corpus; therefore a LSA score could not be computed.

FA	V1	V2	V3	V4	V5	V6	pol
Small	0.76	0.74	---	0.71	1	0.71	pos
Big	0.80	0.75	---	0.74	0.73	0.68	neg
Bulky	---	---	---	---	---	---	pos
Little	---	---	---	---	---	---	neg
Tiny	0.81	0.71	---	0.80	0.73	0.72	---

Table2. LSA scores on non-specialized corpus

On the other hand, we employed the classification on a corpus made up of reviews on different electronic products, gathered using the Google API and a site restriction on “amazon.com”.

In the table below, we show an example of the scores obtained with LSA on the features attributes classified for the feature “size”. The vector for the feature attribute “tiny” was classified by SVM as positive, using the training set specified above.

FA	V1	V2	V3	V4	V5	V6	pol
small	0.83	0.77	0.48	0.72	1	0.64	pos
big	0.79	0.68	0.74	0.73	0.77	0.71	neg
bulky	0.76	0.67	0.71	0.75	0.63	0.78	neg
little	0.82	0.76	0.52	0.71	0.83	0.63	pos
tiny	0.78	0.70	0.65	0.67	0.71	0.71	pos

Table 3. LSA scores on specialized corpus

In tables 4 and 5 below, we show the precision values in some example classifications we made with NGD and LSA for different product features for the examples of digital camera reviews and the mobile phones reviews, respectively and the kappa statistics values.

Feature (digital camera)	NGD		LSA			
			NonOpinion		Opinion	
	P	k	P	K	P	k
price	0.75	0.5	0.83	0.7	0.83	0.7
quality	0.78	0.45	0.84	0.7	0.84	0.7
design	0.75	0.45	--	--	0.85	0.65
size	0.80	0.6	--	--	0.85	0.7
resolution	0.83	0.5	0.84	0.6	0.85	0.7
zoom	0.8	0.6	--	--	0.86	0.6
display	0.78	0.6	0.78	0.5	0.84	0.7
software	0.8	0.5	--	--	0.82	0.5

Table 4. Results NGD versus LSA digital camera

Feature (mobile phone)	NGD		LSA			
			NonOpinion		Opinion	
	P	k	P	k	P	k
price	0.75	0.5	0.83	0.7	0.83	0.7
quality	0.75	0.5	0.84	0.7	0.84	0.5
design	0.78	0.45	--	--	0.85	0.65
size	0.8	0.5	--	--	0.85	0.7
display	0.8	0.45	0.7	0.4	0.83	0.5
memory	0.75	0.5	--	--	0.87	0.6
camera	0.75	0.45	--	--	0.84	0.7

Table 5. Results NGD versus LSA mobile phone

The conclusion that can be drawn from the results presented is that the main advantage in using the first method of polarity assignment is that NGD is language independent and offers a measure of semantic similarity taking into account the meaning given to words in all texts indexed by Google from the World Wide Web. On the other hand, using the whole Web corpus can also add significant noise. Therefore, we employ Latent Semantic Analysis at a local level, both on a non- specialized corpus, as well as on a corpus containing customer reviews. As we will show, the classification using LSA on a specialized corpus brings an average of 8% of improvement in the classification of polarity and a rise of 0.20 in the kappa measure, leading to an 8% overall improvement in the precision of the summarization system. However, these results were obtained using a specialized corpus of opinions, which was previously gathered from the WWW. To this respect, it is important to determine sources (web sites, blogs or forums) specific to each of the working languages, from which to gather the corpus on which the LSA model can be built. Using LSA on a non-specialized corpus improved the classification to the same degree as the classification on a specialized corpus in the cases where the specific pairs of words to be classified were found in the corpus. However, in 41% of the cases, the classification failed due to the fact that the words we tried to classify were not found in the corpus.

5 *Feature polarity extraction using subjective phrases*

As observed before, some opinions on the product or its features are expressed indirectly, with subjective phrases containing positive or negative emotions which are related to the product name, product brand or its features. In order to identify those phrases, we have constructed a set of rules for extraction, using the emotion lists from WordNet Affect. For the words present in the "joy" emotion list, we consider the phrases extracted as having a positive opinion on the product or the feature contained. For the words in the "anger", "sadness" and "disgust" emotion lists, we consider the phrases extracted as having a negative opinion on the product or the feature contained. Apart from the emotion words, we have considered a list of "positive words" (pos_list), containing adverbs such as "definitely", "totally", "very", "absolutely" and so on - as words positively stressing upon an idea – (Iftene and Balahur, 2007), that influence on the polarity of the emotion expressed and that are often found in user reviews. We present the extraction rules in table 6 (verb_emotion, noun_emotion and adj_emotion correspond to the verbs, nouns and adjectives, respectively, found in the emotion lists from WordNet Affect under the emotions "joy", "sadness", "anger" and "disgust"). In case of "surprise", as emotion expressed about a product and its features, it can have both a positive, as well as negative connotation. Therefore, we have chosen not to include the terms expressing this emotion in the extraction patterns.

Rule ID	**Rule pattern**
1	I [pos_list*] [verb_emotion] [this| the | my] [product_name| product_feature]
2	I ([am | 'm | was| feel | felt]) ([pos_list**]) [adj_emotion] [with| about| by] [product_name | product_feature]
3	I [feel | felt] [noun_emotion] [about |

	with] [product_name | product_brand]
4	I [pos_list*] [recommend] [this|the] [product_name | product_brand]
5	I ([don't]) [think | believe] [sentence**]
6	[It] ['s| is] [adj_emotion] [how| what] [product_name| | product_feature][product_action***]
7	[You | Everybody | Everyone | All | He | She | They] [will | would] [verb_emotion] [this | the] [product_name | product_brand | feature]

Table 6. Extraction patterns for subjective opinion phrases

6 *Evaluation and discussion*

We have performed a comparative analysis of the system employing the SMO SVM polarity classification using NGD and LSA on a specialized corpus, the subjective phrases and combined, with the corpus used in (Balahur and Montoyo, 2008) and also the corpus of 5 reviews from (Hu and Liu, 2004). Results obtained in table 7 are for the use of our own annotated corpus:

NGD		LSA		Rules		NGD+ Rules		LSA+ Rules	
P	R	P	R	P	R	P	R	P	R
0.80	0.79	0.88	0.87	0.32	0.6	0.89	0.85	0.93	0.93

Table 7. System results on own corpus

In the case of the (Hu and Liu, 2004) 5-reviews corpus, the observation that is important to make is that, as opposed to the annotation made in the corpus, we have first mapped the features identified to the general feature of the product (for example "fit" refers to "size" and "edges" refers to "design"), as we believe that in real life situations, a user benefits more from a summary on coarser-classes of product features. Also, a set of sentences that were not annotated in the corpus, such as *"You'll love this camera"*, which expresses a positive opinion on the product. The results are shown in table 8:

NGD		LSA		Rules		NGD+ Rules		LSA+ Rules	
P	R	P	R	P	R	P	R	P	R
0.81	0.80	0.85	0.88	0.28	0.5	0.89	0.85	0.93	0.93

Table 8. System results (Hu and Liu, 2004) corpus

The results shown are compared against the baseline of 0.20 precision and 0.41 recall, which was obtained using only the features determined as in (Balahur and Montoyo, 2008) and the feature attributes whose polarity was computed from the "pros and cons" –style reviews. As it can be seen, the best results are obtained when using the combination of LSA with the rules for subjective phrases extraction.
However, gathering the corpus for the LSA model can be a costly process, whereas NGD scores are straightforward to be obtained and classifying is less costly as time and resources used.
What is interesting to study is the impact of employing LSA for gradual learning and correction of a system that uses NGD for classifying the polarity of feature attributes. In such a self-learning scheme, the "online" classification would be that of NGD. However, the classification of the new feature attributes can be later improved "offline" using the classification given by LSA, which can then be used as better training for learning the polarity of new feature attributes by the "online" NGD classification.

7 *Conclusions and future work*

In this paper, we presented a method to assign polarity to feature attributes in a feature-dependent manner, using the scores showing relational strength between two words, given by the Normalized Google Distance and Latent Semantic Analysis and the classification using SMO SVM machine learning. The main advantage in using polarity assignment depending on NGD scores is that this is language independent and offers a measure of semantic similarity taking into account the meaning given to words in all texts indexed by Google from the World Wide Web. The main advantage in using LSA on a specialized corpus, on the other hand, is that it eliminates the noise given by the multiple senses of words.

We completed the opinion extraction on different product features with rules using the words present in WordNet Affect, as indicative of indirectly expressed opinions on products.

We showed how all the employed methods led to significant growth in the precision and recall of our opinion mining and summarization system.

Future work includes a more thorough and systematic organization of product categories and features with their corresponding attributes and the fuzzy analysis of text for the detection of misspellings and grammar errors.

References

Balahur, A., Montoyo, A. 2008. Building a Recommender System Using Community Level Social Filtering. To appear in the Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2008), ICEIS 2008.

Cilibrasi, D., Vitanyi, P. 2006. Automatic Meaning Discovery Using Google. IEEE Journal of Transactions on Knowledge and Data Engineering .

Dave, K., Lawrence, S., Pennock, D. 2003. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In: Proceedings of WWW-03.

Deerwester, S. Dumais, S., Furnas, G. W., Landauer, T. K, Harshman, R. (1990). Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 41 (6): 391-407.

Ding, X., Liu, B., Yu, P. 2008. A Holistic Lexicon-Based Approach to Opinion Mining. In Proceedings of ACM WSDM, 2008.

Fellbaum, C. (ed.). 1999. WordNet: An Electronic Lexical Database, MIT Press, Cambridge, Massachusetts.

Hu., M. Liu, B.: Mining and summarizing customer reviews. In Proceedings of KDD (2004)

Iftene, A., Balahur-Dobrescu, A. 2007. Hypothesis Transformation and Semantic Variability Rules Used in Recognizing Textual Entailment. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing.

Kim, S.M., Hovy, E.H. 2006. Identifying and Analyzing Judgement Opinions. In Proceedings of HLT-NAACL 2006, ACL, pp. 200-207

Liu, H., Singh, P. .2004. ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal, To Appear. Volume 22, forthcoming issue. Kluwer Academic Publishers.

Pang, B., Lee, L. and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. EMNLP 2002.

Platt,J.. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Microsoft Research Technical Report MSR-TR-98-14, (1998)

Popescu, A.-M., Etzioni, O. 2005. Extracting product features and opinions from reviews. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (EMNLP 2008).

Riloff, E. and Wiebe, J.. 2003. Learning extraction patterns for subjective expressions. EMNLP 2003.

Strapparava, C. Valitutti, A. 2004. WordNet-Affect: an affective extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, pp. 1083-1086.

Turney, P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. ACL'02.

Wilson, T., Wiebe, J. and Hwa, R.2004. Just how mad are you? Finding strong and weak opinion clauses. AAAI'04, 2004.