Title	Improving customer decisions using product reviews: CROM - Car Review Opinion Miner
Author(s)	Dabrowski, Maciej; Acton, Thomas; Jarzebowski, Przemyslaw; O'Riain, Sean
Publication Date	2010-04-07
Publication Information	Dabrowski, M., P. Jarzebowski, et al. (2010). Improving customer decisions using product reviews: CROM - Car Review Opinion Miner. 6th International Conference on Web Information Systems and Technologies. Valencia, Spain, Springer.
Item record	http://hdl.handle.net/10379/1420

IMPROVING CUSTOMER DECISIONS USING PRODUCT REVIEWS
CROM – Car Review Opinion Miner

Maciej Dabrowski, Thomas Acton
J.E. Cairnes School of Business & Economics,
National University of Ireland Galway, Ireland
maciej.dabrowski@deri.org, thomas.acton@nuigalway.ie

Przemyslaw Jarzebowski, Sean O'Riain
Digital Enterprise Research Institute, National University of Ireland Galway, Ireland
przemyslaw.jarzebowski@deri.org, sean.oriain@deri.org

Abstract: Online shopping is a very goal-oriented activity. Consumers have a set of preferences for a product or service that is used as criteria for assessment of the available alternatives. However, crucial information about products is often available as text reviews. Finding a product with specific features is extremely time-consuming using the typical search functionality found in existing shopping sites. In this work we propose a method for the seamless integration of unstructured information from product reviews with structured product descriptions using opinion mining. We demonstrate our method through shopping for a used car based on 148240 car reviews. Evaluation results using a user study and simulations show that the technique enables customers to assess more product characteristics and potentially make better decisions.

1. INTRODUCTION

The increasing availability of product reviews enables ubiquitous use among customers shopping online or seeking additional or missing information about products and services. Gretzel and Yoo (2008) demonstrate that 97.7% of travel booking decisions are made after consulting other travellers' opinions, of which 77.9% involve the use of customer reviews as a source of information helping to make a better decision. In this paper we propose a method for the aggregation of information about products from online customer reviews. We deal with the contextual character of descriptive information using *cost-type* and *benefit-type* attributes (Yang, 2008). We show how this unstructured information can be used to complement structured product descriptions facilitating customer decisions. In particular, we discuss the impact of the method from a decision-making perspective.

2. RELATED WORK

Seamless integration of the information in product descriptions with customer reviews requires dealing with three tasks that have been investigated in the research literature so far: *extraction of feature terms*, *opinion mining* and *sentiment analysis*. Approaches for the extraction of feature terms proposed in OPINE (Popescu and Etzioni, 2005), in RedOpal (Scaffidi et al., 2007) and by Hu and Liu (2004) identify potential features using part-of-speech (POS) tagging for nouns and nouns phrases. Hu and Liu (2004) considered extracting neighbour opinion phrases using a window of a size k on the output of a noun phrase chunker. OPINE (Popescu and Etzioni, 2005) takes advantage of the syntactic dependencies computed by the MINIPAR parser. One group of approaches for opinion mining is based on using term dictionaries such as WordNet to identify opinion words in text reviews (Hu and Liu, 2004). The main disadvantage is the limited set of

terms available in their term dictionaries. Another group of approaches uses context-aware learned models of opinion words (Popescu and Etzioni, 2005). These models handle the limitation of the previous group but may not generalize well across product categories.

Our method differs from these approaches as follows. First, we use a hierarchy of product features based on domain knowledge, in contrast to flat lists. Second, we use local lists of trigger terms based on term occurrences for every element of a constructed feature hierarchy, and use explicit features to identify potential opinion phrases. Third, we compute opinion sentiments on three levels: word sentiment level, chunk sentiment level and context-dependent chunk sentiment.

3. METHOD

There are three generic tasks that an opinion mining system needs to fulfil: *identification of the product features*, *discovery of opinion phrases*, and *sentiment analysis* (see Popescu and Etzioni, 2005, Scaffidi et al., 2007). In our method, the first of these tasks is performed using domain knowledge and data from popular websites offering semi-structured car reviews. We identified and organized a list of all features of potential interest to customers. The list was evaluated in a study with 28 subjects who were interested in buying a car and organized into a hierarchy based on the domain knowledge available.

The second task was implemented using a modified version of the technique presented in Aciar et al. (2007). The hierarchy was extended with the list of "trigger terms" (phrases that symbolize features). As product features are typically nouns or noun phrases, we eliminated the infrequent and irrelevant phrases from the set and used association mining (performed with ARMiner software) to identify potential bigram features. The set of trigger phrases was later extended using bigram features, similarly to Aciar (2007), organized into a hierarchy and expanded to other parts of speech, for example: *driving (noun) -> drive (verb)*.

3.1 Opinion Mining

To extract opinion phrases and to select sentences containing potential opinions, we used term matching with terms in base morphological form of a given speech component, using WordNet to improve accuracy. Sentences containing potential opinions were annotated with POS tags using the GATE tagger, due to its accuracy on our corpus in comparison to other available taggers. We used the shallow-parsing method based on a set of rules that extract potential opinions as chunks of text. The rules are constructed to extract a consistent fragment of the sentence that contains a feature (opinion head) and the sentiment about the feature (opinion content), similar to other common approaches (see Aciar et al., 2007). The advantage of our method is that not only nouns and noun phrases are considered as features and not only adjectives are considered as opinions. Thus, our method allows dealing with context examples (e.g. "This car handles like a dream."). The method we used is similar to approaches based on term proximity windows (Hu and Liu, 2004), involving the computation of syntactic dependencies (Popescu and Etzioni, 2005). However, our approach accommodates language structure, in contrast to Hu and Liu's (2004) approach, and is more efficient than that proposed by Popescu (2005).

3.2 Sentiment Analysis

Our approach deals with sentiment analysis of three levels, word level, chunk level, and context dependant chunk level. To assess the sentiment of a given opinion we used a lexicon-based method first proposed by Kim and Hovy (2004). The initial list of sentiment words with known sentiment was enlarged with synonyms and antonyms based on WordNet. As proposed by Ding et al. (2008), lists adjectives, nouns, verbs and adverbs with positive and negative sentiment were created - word sentiment is based on the sum of all sentiment values. Using utility theory (Butler et al., 2001), and to avoid the negative effects of context, the features were divided into three classes: *Cost-type* – features with preference toward lower values (e.g. *fuel consumption*); *Benefit-type* – higher values are preferred (e.g *reliability*); *Neutral* – the character of a feature is context-dependent. Similarly, sentiment words were assigned to *cost* and *benefit* categories. Thus, occurrence of a *cost-type* sentiment word (e.g. "low") with a *cost-type* feature (e.g. "price") resulted in positive sentiment. Conversely, the same word occurring with a *benefit-type* feature resulted in negative sentiment (e.g. "low quality").

Sentiment of an opinion chunk describing a feature was computed based on all the sentiment words identified in the chunk. It is important to note here that our method dealt with negation by

changing the sentiment of a word to its opposite. To exploit full potential of our lexicon-based method we considered chunk context. In our method dependencies between chunks are assessed. If two chunks are combined with a sentiment changing word (e.g. *but, however, despite*), it is assumed that two chunks have the opposite sentiment polarity. If the chunks are connected using a word not from the list of the sentiment changing words, the same chunk sentiment is expected.

4. EVALUATION

To evaluate the method proposed here we gathered 148240 car reviews from popular websites. Of these, 12561 were pure text reviews in English available at *http://www.whatcar.com/* website, and 135679 were semi-structured reviews from other websites (eg. *Autocentrum.pl*).

4.1 Feature Extraction

The feature extraction approach proposed here was evaluated in a user study involving potential car buyers. First, we listed the features available for searching for a car at the most popular websites offering used cars and car reviews. In total, list of 27 features was composed that included both attributes from car sellers and car reviews. 32 participants were asked to perform a feature categorization task using a web application: 29 responded (91% response rate), with 28 valid cases. There was no time limit for task completion. Participants did not report any important car features missing from our list.

Table 1 Results of the categorization task.

Measure	VIF	FIF	NIF
Avg. # of features / cat.	10.52	12	5.48
Std. dev.	3.18	3.28	3.37
Score per feature	2	1	0

Subjects' responses were consistent, with standardized Cronbach $\alpha = 0{,}74$. The resulting categorization shows on average 12 *fairly important features (FIF)*, 10.52 *very important features (VIF)*, and 5.48 *not important features* (NIF) per subject. For convenience, we report a scoring system in which every category was awarded a score from 0 (least important) to 2 (very important) (see Table 1).

The experiments show high interest of customers in car features that are available in the reviews (C)

and are not available on typical shopping sites in structured product descriptions. We note that 60% of the TOP 10 highly ranked features was available only in customer reviews (see Table 2).

Table 2 Average score (Score) for TOP 10 features and number of votes for every category, *T* indicates source of the feature (S-seller, C-customer opinion).

Feature	Score	VIF	VIF	FIF	T
Overall Reliability	1.89	25	3	0	**C**
Major problems	1.82	24	3	1	**C**
Price	1.75	22	5	1	S
Mileage	1.61	18	9	1	S
Engine	1.61	19	7	2	**C**
Maintenance cost	1.54	16	11	1	**C**
Overall value	1.50	15	12	1	**C**
Year	1.43	14	12	2	S
Mechanical quality	1.39	15	9	4	**C**
Make	1.39	14	11	3	S

4.2 Opinion Mining

To evaluate the framework we designed a simulation using a subset of product reviews we gathered from *whatcar.com*. Due to limited resources we annotated a corpus of 203 reviews (1233 sentences) of Ford Focus cars, the most popular model in the dataset based on the number of reviews and number of online adverts available (2692 adverts, 4,73% of all cars for sale at *carzone.ie*). For every sentence in the set an annotation was provided by a group of human annotators. Every annotation consisted of a list of featured mentioned explicitly and implicitly in a sentence together with the expressed sentiment for the feature using a 5 step scale: -2 (very negative), to 2 (very positive). Annotators negotiated inconsistencies to avoid a potential negative impact of subjective opinion on polarity and strength of the sentiment. We evaluated the performance of our method with precision and recall metrics for our test dataset and accuracy of the sentiment analysis technique (see Table 3).

Table 3 Opinion Mining evaluation results

Opinion sentence extraction and classification		Sentiment Analysis
Precision	Recall	Accuracy
76,3%	77,5%	82,6%

4.3 Decision Making Impact

Consumers often face a task to select from a large set of alternatives, such as choosing a car to buy.

Consumer websites often provide functionality to search for alternatives, usually by asking a user to provide his criteria for a desired product. Although prevalent, both users and retailers can find such functionality unsatisfying (Hagen et al., 2000). One of the major reasons users are rarely provided with the information they need is that they are often not able to transform their preferences into requirements (Viappiani et al., 2008), or simply the information they are looking for is not available in the appropriate format.

Our method addresses such drawbacks of existing websites by extracting opinions about products and features from product reviews. Our evaluation shows that reviews provide consumers with information about products that is valuable to them, and which is not available on standard shopping websites. Further, the extracted features are available together with the existing product attributes so that no additional action is required from consumers, decision-making effort is lower and less time is required to make a decision (Scaffidi et al., 2007). Moreover the customers can avoid time-consuming analyses of product reviews. The direct implication of such approach is the lower decision-making effort, as less time and information processing is required. Todd and Benbasat (2000) point out that decision makers tend to trade off decision quality for minimization of decision-making effort: a reduction of the decision-making effort from using our method can therefore increase decision quality

5. CONCLUSION

We described an opinion mining system that extracts and integrates opinions about products and features from very informal, noisy text data (product reviews) using a hierarchy of features from a number of websites and domain knowledge. The major contribution of this paper provides a decision making perspective on integration of consumer reviews in customer product selection and evaluation of customer information needs in the used car market.

Our method is of value not only to consumer-based web providers and potential customers but also to product manufacturers. Without additional effort, the approach enables consumers to consider further features of products only available in customer reviews. Our approach can be of value in various domains to both customers and product sellers.

REFERENCES

Aciar, S., Zhang, D., Simoff, S. & Debenham, J. (2007) Informed recommender: Basing recommendations on consumer product reviews. *Ieee Intelligent Systems,* 22, 39-47.

Butler, J., Morrice, D. J. & Mullarkey, P. W. (2001) A multiple attribute utility theory approach to ranking and selection. *Management Science,* 47, 800-816.

Ding, X., Liu, B. & Yu, P. S. (2008) A holistic lexicon-based approach to opinion mining. *WSDM '08: Proceedings of the international conference on Web search and web data mining.*

Gretzel, U. & Yoo, K. H. (2008) Use and Impact of Online Travel Reviews *Information and Communication Technologies in Tourism.* Innsbruck, Austria.

Hagen, P. R., Manning, H. & Paul, Y. (2000) Must Search Stink? *The Forrester Report*

Hu, M. & Liu, B. (2004) Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.* Seattle, WA, USA, ACM.

Kim, S.-M. & Hovy, E. (2004) Determining the sentiment of opinions. *COLING '04: Proceedings of the 20th international conference on Computational Linguistics.*

Popescu, A.-M. & Etzioni, O. (2005) Extracting product features and opinions from reviews. *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.* Vancouver, British Columbia, Canada, Association for Computational Linguistics.

Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., Jin, C. & Acm (2007) Red Opal: Product-Feature Scoring from Reviews. *8th Annual Conference on Electronic Commerce (EC 2007).* San Diego, CA, Assoc Computing Machinery.

Todd, P. & Benbasat, I. (2000) Inducing compensatory information processing through decision aids that facilitate effort reduction: An experimental assessment. *Journal of Behavioral Decision Making,* 13, 91-106.

Viappiani, P., Pu, P. & Faltings, B. (2008) Preference-based search with adaptive recommendations. *AI Commun.,* 21, 155-175.

Yang, I. T. (2008) Utility-based decision support system for schedule optimization. *Decision Support Systems,* 44, 595-605.