Winter 12-4-2016

An Empirical Examination of Consumer Behavior for Search and Experience Goods in Sentiment Analysis

Jaehyeon Ju
KAIST, slamking@business.kaist.ac.kr

Dongyeon Kim
Korea Advanced Institute of Science and Technology, dykim88@business.kaist.ac.kr

Jae-Hyeon Ahn
Korea Advanced Institute of Science and Technology, jahn@business.kaist.ac.kr

Dong-Joo Lee
Hansung University, djlee@hansung.ac.kr

An Empirical Examination of Consumer Behavior for Search and Experience Goods in Sentiment Analysis

Jaehyeon Ju, KAIST, South Korea, slamking@business.kaist.ac.kr
Dongyeon Kim, KAIST, South Korea, dykim88@business.kaist.ac.kr
Jae-Hyeon Ahn, KAIST, South Korea, jahn@business.kaist.ac.kr
Dong-Joo Lee, Hansung University, South Korea, djlee@hansung.ac.kr

ABSTRACT

With the explosive increase of user-generated content such as product reviews and social media, sentiment analysis has emerged as an area of interest. Sentiment analysis is a useful method to analyze product reviews, and product feature extraction is an important task in sentiment analysis, during which one identifies features of products from reviews. Product features are categorized by product type, such as search goods or experience goods, and their characteristics are totally different. Thus, we examine whether the classification performance differs by product type. The findings show that the optimal threshold varies by product type, and simply decreasing the threshold to cover many features does not guarantee improvement of the classification performance.

Keywords: Sentiment analysis, Experience goods, Search goods, Feature extraction, Product features

INTRODUCTION

Widespread Internet access enables users to create user-generated content such as online product reviews. Reviews about products or services provide the representative medium through which consumers express their overall opinions or feelings. Reviews are increasingly important in attracting companies' attention, because companies want to identify what consumers think about their products or services in real time. Also, individual consumers want to know the opinions of the existing purchasers. Acquiring consumer opinions is an essential factor for decision makers and marketing staff. Sentiment analysis involves the process of analyzing opinions and emotions from consumers' reviews [10]. After collecting reviews, extracting the information contained in reviews presents an enormous task due to the huge amount of data involved. Individuals face difficulty extracting and summarizing opinions manually. Thus, automated sentiment analysis tools are important to streamline the process.

Sentiment analysis can be divided by three main classification levels: document level, sentence level, and feature level [4]. Document-level sentiment analysis assumes that the document presents an author's opinion on one subject. Given the product review, the analysis determines whether the review contains an overall positive or negative opinion about a product or service [3]. Sentiment analysis at the sentence level presumes that one document offers different opinions. That is, this analysis judges whether one sentence states a positive, neutral, or negative opinion [23]. Because document-level and sentence-level analyses do not identify what consumers like or hate exactly [8], feature-level analysis is usually conducted to analyze online consumer reviews. After purchasing products or services, consumers express their opinions (good vs. bad) about many attributes of products or services. Feature-based sentiment analysis is the method based on the fact that consumer opinion is comprised of the subject of an opinion and sentiment [8]. Identifying the subject of an opinion is an important element in understanding consumers' opinions. For example, the sentence "although the battery life of this mobile phone is not good, I like this mobile phone," exhibits a positive attitude. The sentence represents a negative tone about the feature, "battery," however, the main point of feature-based analysis is to find sentiments on attributes of a product or service.

There are three characteristics of product features according to Hu and colleagues [7]. First, product features consist of noun or noun phrases. Second, features are subjects in a sentence. Lastly, features are related to opinion words or phrases. A product (Smartphone) can have many features (CPU, memory, display size, storage capacity, and so on). Each feature can be expressed in different ways, even though the meaning is same (e.g., display quality: resolution, picture quality). Extracting similar features is not an easy task. Popescu and associates [17] suggested the use of association rule mining to extract noun phrases that appeared frequently as product features. They used the feature appraiser that could evaluate each candidate noun phrase by computing the PMI scores between the noun phrases and its related phrases.

As depicted above, it is important and necessary to recognize the attributes of products or services. Several studies have proposed methods to identify product features. Mejova and Srinivasan [12] shows that frequency-based selection, one of the methods applied, exhibits different performance depending on exclusion of the terms that appear fewer than c times in the dataset. Indeed, it is difficult to extract accurate product features, and the performance depends on how one extracts the product features.

We examined whether the classification performance differs by product type, experience goods and search goods. Extracting features of search goods appears to be a more straightforward than extracting features of experience goods, in which individual preferences tend to play an important role. On the other hand, consumer reviews of experience goods use much more different words to describe the same feature and description of product features and is more subjective than consumer reviews of search goods. Therefore, we propose a method that identifies product features and produces a dictionary depending on the product type. The approach offers better performance in feature-based sentiment analysis. We use an extensive dataset of online consumer reviews for products across various shopping sites and over a long period. Our findings show that the threshold of extracting features differs between search goods and experience goods. That is, the criteria of the threshold depends on the type of goods.

RELATED WORK

Product type

Most products available to sell online can be categorized into search goods and experience goods, according to the criteria established by Nelson [16]. Consumers find information about products before making an actual online purchase. The amount and quality of information collected through online search, such as product reviews, differ between two types. While search goods have product features that can be measured and compared more objectively, experience goods have product features which are difficult to evaluate in terms of product quality [14]. This is because consumers can evaluate the overall quality of search goods with relative ease, and they can purchase search goods online without difficulty [22]. For instance, a computer as the representative of search goods can be evaluated through memory capacity, CPU performance, and storage capacity, although experience goods such as a book do not have quantifiable features by which to evaluate the quality. Book reviews help consumers clarify information about books, but the evaluation of a book may require consumers' previous direct experience [15].

Sentiment Analysis

Sentiment analysis involves finding consumers' opinions, attitude towards a particular topic such as product reviews [20]. The analysis determines whether a consumer's opinion is positive, neutral, or negative about the topic. However, because product reviews contain objective information and subjective opinions, distinguishing both of them is important. Companies and consumers especially want to know the reasons the existing purchasers like/hate a product and finally identify consumers' overall opinions on the product. In this context, most previous studies are focused on product feature extraction [1, 19]. Still, the accuracy of feature extraction and performance of final classification is not satisfactory. To solve this problem, a new approach is required. We suggest performing sentiment analysis in terms of the two product types, experience goods and search goods, because the characteristics of product reviews depend on product type.

Product feature extraction

Consumers express positive and negative opinions through product features. The most important feature of sentiment analysis of product reviews is product feature extraction. For example, Ghani and associates [5] built a database of products through explicit and implicit features and linked the features by value. Putthividhya and Hu [18] proposed a named entity recognition (NER) system for extracting product features and combined NER with bootstrapping. Their system achieved 90.33% precision. These two approaches need manual labor for labeling process. Hu and Liu [6] utilized the association mining-based technique to extract product features in the online product reviews. This approach offers the advantage that the training process is not required. However, the drawback of this approach is that it extracted many redundant product features in spite of the feature cleansing [25]. Raju and colleagues suggested an approach which is unsupervised and domain independent. This method has 92% precision and 62% recall which is low.

DATA SET

Through crawlers, data were collected in NAVER shopping, which is one of the biggest online shopping website in Korea. This site provides consumer reviews from the other online shopping sites in Korea such as G-market, 11st, Interpark, and so on. The collected data contain product category, product name, price, review, etc. We chose two products, external battery charger for mobile phone and cream for women, as a representative of both search goods and experience goods. The total number of acquired reviews is 16,083,512. After gathering, we performed actions such as spelling check and grammar correction, not to correct all the errors but to eliminate as much of the noise of reviews as possible.

FRAMEWORK OF SENTIMENT ANALYSIS

The process of sentiment analysis includes feature extraction, opinion extraction, sentiment classification, and performance evaluation (Figure 1). When consumers write product reviews online, they may make errors such as inadvertent misspellings, which undermines the overall performance of the sentiment analysis. Part-of-speech (POS) tagging [11] can be greatly influenced by these errors. Thus, cleansing work is required prior to performing sentiment analysis.

Feature extraction

Part-of-Speech (POS) tagging

An online review consists of more than one sentence and includes opinions on aspects such as a shopping site, a product, or delivery. After splitting reviews into sentences, each sentence is parsed by a linguistic parser, which is widely used in natural language processing (NLP). The linguistic parser reads text and assigns parts of speech to each word, such as noun ('N') and verb ('V'), which is called part-of-speech (POS) tagging. For example, the POS tagger transforms a review "This cream product is good for oily skin" into "This (DT) cream (NN) product (NN) is (VBZ) good (JJ) for (IN) oily (JJ) skin (NN)". This task is consequential to the outcome of the sentiment analysis. As a proven tool for POS tagging, the Kokoma Korean morpheme analyzer (http://kkma.snu.ac.kr) was used.

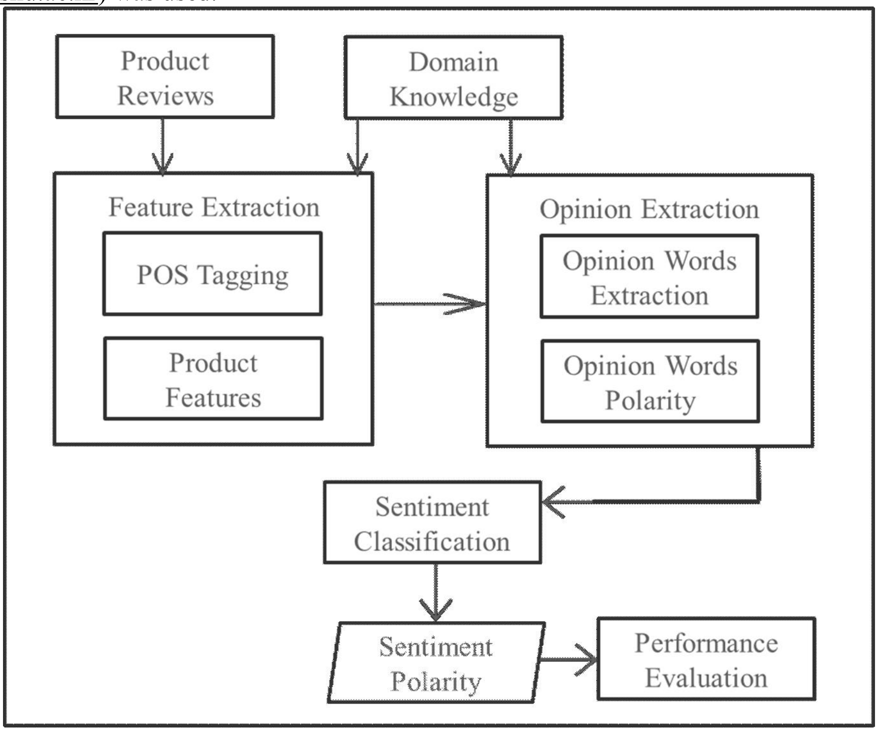

Figure 1:The Process of Sentiment Analysis

Product features

Product features exist in the form of nouns or noun phrases in online reviews [7, 19]. Proper selection of a product feature plays an important role for increasing classification accuracy [9]. The methods of feature extraction are divided into heuristic, statistical, clustering, and hybrid approaches [1, 7, 19, 24, 26]. The performance of some approaches depends on frequency thresholds [24], because features extraction faces the issue that primary features can be extracted, but it is difficult to extract minor features. These issues arise differently by product type. As mentioned in related works, the characteristics of product features that consumers mention are different. Thus, even though the same threshold is applied, the performance of sentiment analysis can be differentiated by product type. Table 1 shows the product feature list extracted from our data. Based on the naïve approach, if the threshold of frequency ratio (frequency/total number of sentence) is adjusted, the extracted product features varies with product type.

Table 1:Product Feature List based on Frequency

	Cream (Experience goods) # of reviews: 1,039,653			Battery (Search goods) # of reviews: 150,261	
1	Features	Frequency	1	Features	Frequency
2	moisture	74589	2	charge	16353
3	droughtiness	23910	3	design	4645
4	dry skin	20887	4	capacity	4567
5	trouble	11452	5	genuine	4055
6	humectant	11000	6	size	2697
7	absorption	10472	7	weight	2202
8	oily skin	9073	8	performance	2199
9	applying	6630	9	mobility	1545
10	smell	5312	10	color	926
11	nutrition	5109	11		
12	pimple	5058	12		
13	elasticity	3990	13		
14	wrinkle	3390	14		
15	sensitivity	2607	15		
16	texture	2565	16		
17	irritation	2164	17		
18	whitening	2150	18		
19	scent	1915	19		

20	absorbing ability	1695	20		
21	rash	1633	21		

Opinion extraction

Opinion words extraction

We limit our study to sentences containing product features from entire reviews, because we are interested in consumers' opinion on the product itself. To find opinion words, fundamental principles in natural language processing are used effectively. The first criterion is that opinion words are near product features [6]. Second, subjective sentences are related to the existence of adjectives positively [2]. In other words, adjectives and verbs which are close to product features act as opinion words.

Opinion words polarity

The Semantic Orientation from Point-wise Mutual Information (SO-PMI) algorithm, which estimates the semantic orientation by measuring the similarity of pairs of words, was used to identify the polarity of opinion words [21]. For example, when the reference words are "good" and "bad", the assumption of the algorithm is that a phrase is defined as a positive semantic orientation when a phrase is more strongly associated with good regard for the product. That is, determining proper reference words is critical when utilizing the SO-PMI algorithm. Although WordNet [13] is a good reference dictionary, it does not provide polarity information in Korean. Instead, OpenHangul (www.openhangul.com), the Korean sentiment dictionary based on collective intelligence, was used as a reference dictionary. Because OpenHangul is not a dictionary that judges based on domain knowledge, four people who have domain knowledge about e-commerce were hired to help reference the words chosen. With SO-PMI algorithm and the reference words, the polarity of adjectives and verbs is determined.

RESULTS

Table 2 and Table 3 show feature lists, which are extracted from the product feature in Table 1 based on the threshold. In the case of the battery product, all product features in Table 1 were extracted within the 0.5 percent threshold (Table 3). On the other hand, for the cream product, nine features out of twenty-one features in Table 1 were obtained (Table 2). While consumers need only a few features to express their opinion on a product in search goods, the lower threshold is required to utilize more feature words for experience goods. When choosing the threshold that is as low as possible to cover more features, irrelevant features can be included, that ultimately results in low performance of classification. We identify that applying different criteria by product type is necessary in spite of the use of the same algorithm.

Table 2:(Cream) Feature List based on Threshold

	Threshold	Frequency	Features								
Cream	2.5%	25991	moisture								
	2.0%	20793	moisture	dry							
	1.5%	15595	moisture	dry	dry skin						
	1.0%	10397	moisture	dry	dry skin	trouble	humectant				
	0.5%	5198	moisture	dry	dry skin	trouble	humectant	absorption	oily skin	applying	smell

Table 3:(Battery) Feature List based on Threshold

	Threshold	Frequency	Features								
Battery	2.5%	3757	charge	design	capacity	genuine					
	2.0%	3005	charge	design	capacity	genuine					
	1.5%	2254	charge	design	capacity	genuine	size				
	1.0%	1503	charge	design	capacity	genuine	size	weight	performance		
	0.5%	751	charge	design	capacity	genuine	size	weight	performance	mobility	color

To evaluate the classification results, we used 2,000 product reviews, with 1,000 reviews for each product type, as the test set. The performance metrics commonly used to evaluate the classification results are accuracy, precision, recall, and F1-score. Computation of these measures is as follows.

True positive (a) represents the number of messages correctly classified as positive, False positive (b) denoted the number of negative messages classified as positive, False negative (c) denotes the number of positive messages classified as negative, and True negative (d) represents the number of messages correctly classified as negative (Table 4).

Table 4:A Confusion Table

		Machine	
		Positive	Negative
Human	Positive	*a*	*b*
	Negative	*c*	*d*

$$Precision = a / (a + b) \quad (1)$$

$$Recall = a / (a + c) \quad (2)$$

$$Accuracy = (a + d) / (a + b + c + d) \quad (3)$$

$$F1\text{-}score = 2 \cdot (P \cdot R) / (P + R) \quad (4)$$

While decreasing the threshold to bring a significant improvement in the classification accuracy in experience goods from 95% to 98% (Figure 2), the accuracy for search goods declines, and the decrease in precision and recall is negligible (Table 5). This means that search goods include major features, even in the high threshold, which leads to high classification accuracy. Decreasing the threshold regardless of product type, it is not efficient in terms of performance. From these results, we concluded that the optimal threshold should differ by product type.

Table 5:Performance Evaluation

	Threshold	Accuracy	Precision	Recall	F1-score
Cream	2.5%	0.96	1.00	0.96	0.98
	2.0%	0.98	1.00	0.98	0.99
	1.5%	0.98	1.00	0.98	0.99
	1.0%	0.98	1.00	0.98	0.99
	0.5%	0.98	0.99	0.98	0.99
	Threshold	Accuracy	Precision	Recall	F1score
Battery	2.5%	0.94	0.96	0.98	0.97
	2.0%	0.94	0.96	0.98	0.97
	1.5%	0.94	0.96	0.96	0.96
	1.0%	0.93	0.95	0.97	0.96
	0.5%	0.93	0.95	0.97	0.96

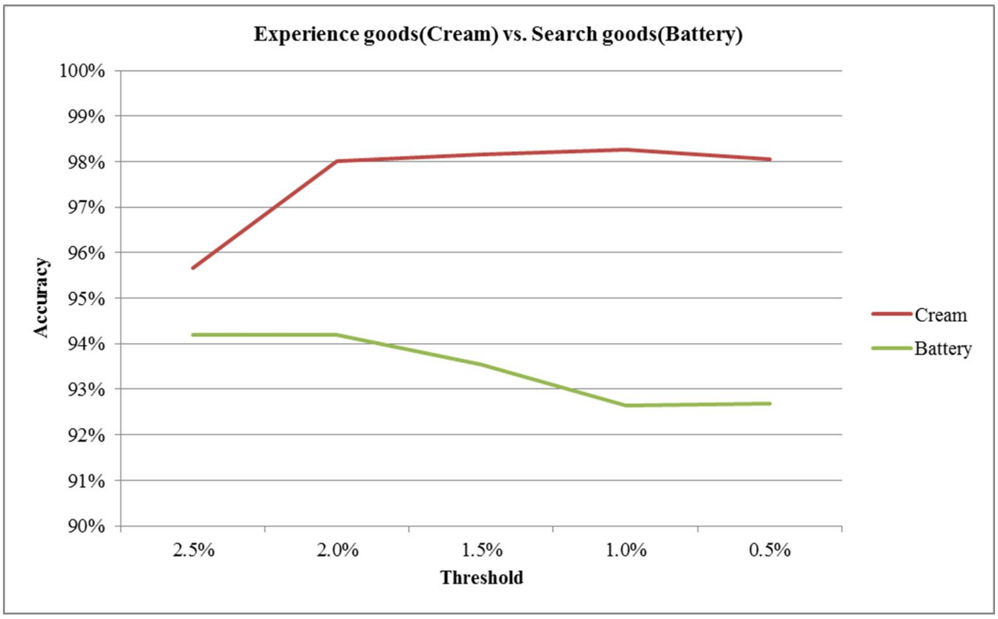

Figure 2:Accuracy Results

CONCLUSION AND FUTURE RESEARCH

With the explosively increasing use of user-generated contents such as product reviews and social media, sentiment analysis is becoming an important area of study. It is utilized in a wide variety of applications, such as product review analytics. Sentiment analysis is an especially important method in that through this method, companies can determine purchasers' opinions on a certain product. Feature extraction in sentiment analysis is a critical element, because it can impact classification performance. Previous studies examined various feature extraction techniques without considering product type. Our research confirmed that the results of classification performance differ by product type, experience goods and search goods. These findings mean that the optimal threshold according to product type may differ, and simply decreasing the threshold does not guarantee improvement of classification performance without product type.

We examined only one product by product type, so the results can be biased. To generalize our findings, we need to analyze at least two more products by product type. Our data are derived from the reviews collected from a Korean shopping site. Our findings must be scrutinized to determine whether they can be applied to reviews written in other languages, because the findings might be typical to the Korean context.

ACKNOWLEDGEMENT

This research was financially supported by Hansung University.

REFERENCES

[1] Asghar, M.Z., et al. (2014) 'A review of feature extraction in sentiment analysis', Journal of Basic and Applied Scientific Research, Vol. 4, No. 3, pp. 181-186.

[2] Bruce, R.F. & J.M. Wiebe. (1999) 'Recognizing subjectivity: a case study in manual tagging', Natural Language Engineering, Vol. 5, No. 02, pp. 187-205.

[3] Feldman, R. (2013) 'Techniques and applications for sentiment analysis', Communications of the ACM, Vol. 56, No. 4, pp. 82-89.

[4] Fersini, E., E. Messina, & F.A. Pozzi. (2014) 'Sentiment analysis: Bayesian ensemble learning', Decision support systems, Vol. 68, pp. 26-38.

[5] Ghani, R., et al. (2006) 'Text mining for product attribute extraction', ACM SIGKDD Explorations Newsletter, Vol. 8, No. 1, pp. 41-48.

[6] Hu, M. & B. Liu. (2004) 'Mining and summarizing customer reviews', Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.

[7] Hu, M. & B. Liu. (2004) 'Mining opinion features in customer reviews', Proceedings of the 19th national conference on Artifical intelligence.

[8] Joshi, N.S. & S.A. Itkat. (2014) 'A survey on feature level sentiment analysis', IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5, No. 4, pp. 5422-5425.

[9] Koncz, P. & J. Paralic. (2011) 'An approach to feature selection for sentiment analysis', 2011 15th IEEE International Conference on Intelligent Engineering Systems. IEEE.

[10] Li, N. & D.D. Wu. (2010) 'Using text mining and sentiment analysis for online forums hotspot detection and forecast',

Decision support systems, Vol. 48, No. 2, pp. 354-368.
[11] Manning, C.D. & H. Schütze. (1999) Foundations of statistical natural language processing. Vol. 999, MIT Press.
[12] Mejova, Y. & P. Srinivasan. (2011) 'Exploring Feature Definition and Selection for Sentiment Classifiers', Fifth International AAAI Conference on Weblogs and Social Media.
[13] Miller, G.A., et al. (1990) 'Introduction to WordNet: An on-line lexical database', International journal of lexicography, Vol. 3, No. 4, pp. 235-244.
[14] Mudambi, S.M. & D. Schuff. (2010) 'What makes a helpful review? A study of customer reviews on Amazon. com', MIS quarterly, Vol. 34, No. 1, pp. 185-200.
[15] Mudambi, S.M., D. Schuff, & Z. Zhang. (2014) 'Why aren't the stars aligned? An analysis of online review content and star ratings', 2014 47th Hawaii International Conference on System Sciences. IEEE.
[16] Nelson, P. (1970) 'Information and consumer behavior', Journal of political economy, Vol. 78, No. 2, pp. 311-329.
[17] Popescu, A.-M. & O. Etzioni. (2007) Extracting product features and opinions from reviews, in Natural language processing and text mining. Springer. pp. 9-28.
[18] Putthividhya, D.P. & J. Hu. (2011) 'Bootstrapped named entity recognition for product attribute extraction', Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.
[19] Somprasertsri, G. & P. Lalitrojwong. (2008) 'A maximum entropy model for product feature extraction in online customer reviews', 2008 IEEE Conference on Cybernetics and Intelligent Systems. IEEE.
[20] Tsytsarau, M. & T. Palpanas. (2012) 'Survey on mining subjective data on the web', Data Mining and Knowledge Discovery, Vol. 24, No. 3, pp. 478-514.
[21] Turney, P.D. (2002) 'Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews', Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics.
[22] Weathers, D., S. Sharma, & S.L. Wood. (2007) 'Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods', Journal of Retailing, Vol. 83, No. 4, pp. 393-401.
[23] Xu, K., et al. (2011) 'Mining comparative opinions from customer reviews for Competitive Intelligence', Decision support systems, Vol. 50, No. 4, pp. 743-754.
[24] Zhang, H., et al. (2011) 'Feature-level sentiment analysis for Chinese product reviews', Computer Research and Development (ICCRD), 2011 3rd International Conference on. IEEE.
[25] Zhou, F., R.J. Jiao, & J.S. Linsey. (2015) 'Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews', Journal of Mechanical Design, Vol. 137, No. 7, pp. 071401.
[26] Zhuang, L., F. Jing, & X.-Y. Zhu. (2006) 'Movie review mining and summarization', Proceedings of the 15th ACM international conference on Information and knowledge management. ACM.