A Domain Independent Framework to Extract and Aggregate Analogous Features in Reviews

Archana Bhattarai, Nobal Niraula, Vasile Rus, King-Ip Lin

Departmet of Computer Science, The University of Memphis
209 Dunn Hall
Memphis, TN 38152, USA
{abhattar,nbniraula, vrus,davidlin}@memphis.edu

Abstract. Extracting and detecting user mentioned features in reviews without domain knowledge is a challenge. Moreover, people express their opinions on same feature of the product or service in various lexical forms. Thus it is also important to identify similar feature expressing terms and opinions in order to build fully automated, effective feature based opinion mining systems. In this paper, we present a novel framework to automatically detect, extract and aggregate semantically related features of reviewed products and services. Our approach involves implementation of the double propagation based algorithm [1] to detect candidate features in reviews, followed by a probabilistic generative model to aggregate ratable features in reviews. Our model uses the power of sentence level syntactic and lexical information to detect candidate feature words, and corpus level co-occurrence statistics to do grouping of semantically similar features to obtain high precision feature detection. The results of our model outperformed existing state of the art probabilistic models. Our model also shows a distinct advantage over double propagation by grouping like features together which in turns makes it easy and quick for users to perceive. We evaluate our model in two completely unrelated domains such as restaurant and camera-related textual reviews to verify its domain independence.

Keywords: Information Retrieval, Natural Language Processing, Text Analysis

1 Introduction

The informativity and reliability of reviews have embarked them as a standard to assess the quality of products and services. Various studies [2], [3] have shown that online reviews have real economic values on the products that the reviews target. However, the sheer volume of the reviews spread over many sites, makes it almost impossible to go through every review manually. There is a need to develop automated and intelligent tool that can process the reviews and turn them into useful information for users. One of the major challenges that are to be tackled is opinion detection and classification (e.g. positive, or negative). For example, "*I like this camera*" is a positive opinion and "*this camera did not produce good picture*" is a negative opinion. This is a much studied problem [4], [5], [6], [7], [8], [9]. Recently more interest has grown in fine grained feature based sentiment classification. A single review could have disparate opinions on different features of a product. For example, the sentence "*Although the picture quality of the camera is good, it is too bulky to carry*", expresses a positive opinion on the **picture quality** feature of the product and a negative opinion about the **size** feature of the product. Furthermore, users express their opinions about different features in variety of ways. For example, the opinion sentences "*I loved their sushi*" and "*the bread was really fresh*" although do not express exactly same thing, both the sentences seem to be positive about the food feature of some restaurant. Thus, an automated method for opinion detection and classification, like the one proposed in this paper, needs to be able to identify opinions expressed at different levels of granularity (e.g. one opinion per sentence or multiple opinions in one sentence) and in different lexical and structural forms as exemplified above.

In our approach presented here, we tackle the problem of identifying user rated product features and classifying sentences in reviews to one of the features. The application of this system is multifold. This becomes an inevitable part in feature based summarization of reviews. It can also be essential in extracting supporting textual cues for a rating of a product. Feature based sentiment analysis can also be more accurate and representative of existing data as it projects detailed information. Moreover, since sentiment and opinion are totally subjective and personal to individuals, it might be better to present overall impression rather than the polarity in terms of positivity and negativity as polarity does not provide details on why it is good or bad. A feature that is positive for a person might be negative for the other. Hence to better handle such issues, we first detect explicit features mentioned in each sentence of reviews and group them into major ratable features for overall impression.

The idea of topic models and other dimension reduction algorithms such as LSA [10], PLSA [11] and LDA [12] appear to be most suitable in unsupervised grouping of similar features in reviews and have been used in recent research works. However, these models have a very low precision and recall as it is completely uninformed of the domain, feature words and lexical/syntactic information at the sentence/document level. Another thread of research [1], [13] has exploited the lexical information and co-occurrence information at the sentence level. Both approaches have their own pros and cons. However, no work has been done to combine both the features to get the best of both. Hence, in this work, we tackle the problem exploiting both the local and global features preserving domain independency and least supervision. The only human intervention in our framework is the assignment of meaningful topics to LDA discovered topics. Our results show distinct advantage over both the standard methods combining both the advantages and mitigating the disadvantages.

1.1 Feature filtering at the sentence level

Some of the approaches that perform sentence level filtering are association rule based filtering [13] and the double propagation algorithm [1].With the lexical/syntactic information, product/service features can be extracted with very high precision and recall. However, these methods can only extract features that are explicitly identified by certain words. For example in the sentences, "*I've had a lot of cameras, but this one is ACTUALLY being used, and not stored in a desk: it's always in my shirt pocket*" is actually talking about the feature "**size**". However, with the lexical and association based features, this feature can never be inferred from the sentence. From our analysis in the camera dataset [1], x% of sentences have implicitly mentioned the features which seems very normal and usual in human language. Another important part is that people by nature use different words/phrases to refer to single feature, which makes it virtually impossible to only use this method to extract most prominent features from reviews. As an example, with the double propagation method, in x number of sentence, y number of features were extracted. As the number of reviews increases, the number of features increases exponentially too. As an example, in the restaurant dataset, we extracted yy number of features from xx number of sentences in zz reviews. Usually, reviews are in large number for most of the products and services and an effective review mining system should be able to provide the collective impression in shortest possible way. Thus, the real use of these extracted features is only seen when they are grouped to few important features.

1.2 Feature Assembly at Corpus level

Topic models such as Latent Dirichlet Allocation (LDA), has been used to group and detect user mentioned features in reviews [14], [15]. Some form of filtering and localization has been introduced with sentence level LDA and choosing only nouns at the sentence level [14], [15]. However, still a lot of nouns that are not relevant to features reside in reviews which introduce impurities in the prediction. This reduces precision and recall in the system. Thus we came up with a strategy to filter the words in SLDA (sentence LDA) so only relevant words are left behind. These filtered sentences are then fed to LDA to get co-occurrence based grouping of features. With this approach employed in our framework, we have observed notable increase in precision and recall which we will present in subsequent sections.

The rest of the paper is organized as follows. Section 2 describes our problem in more detail and section 3 discusses related work. In section 4 we explain our methodology stepwise. Section 5 presents our experimentation evaluation and comparison to state of the art methods in different domain. In section 6 we conclude with final remarks and direction for future work.

2 Problem Statement

Given a collection of reviews related to a product/service, the first task is to collect domain dependent seed sentiment words in the domain with high precision. With these seed sentiment words, the next task is then to identify feature indicating words in each sentence of reviews. After identifying feature indicating words in sentences of reviews, the third task is to filter the reviews, just keeping these words and group them into few ratable features. After identifying the features of the product/service, the last step is to classify unseen sentences into corresponding feature topics so further processing can be done.

3 Related Work

Most works on review mining involves tasks of identifying related product features and classifying the sentiment of the review, either as a whole or by individual sentences/phrases. An example is the work by Hu and Liu [13]. They apply association rule mining to extract product features. To determine sentiment polarity, they use a seed set of adjectives which are expanded via WordNet [16]. Then they select a subset or rewrite some of the original sentences to capture main points as done in traditional text summarization.

Another line of research is using a supervised framework as done by [17]. However, these methods are only limited to certain domain and is highly dependent on the training data. Morinaga and others [18] use a search engine to collect statements regarding target products and then extract opinions using syntactic and linguistic rules derived by human experts. The result is then used to generate statistically meaningful information. (Dave et al., 2003) use information retrieval techniques to identify product attributes (feature extraction) and train a classifier using a corpus of self-tagged reviews available from web sites. After further refinement they apply it to sentences mined from broad web searches.

Works have also been done to make the whole process unsupervised. One example is OPINE [19]. It uses the concept of relaxation-labeling to determine the semantic orientation of potential opinion words in the context of the extracted product features and specific review sentences. Another work done by [20] proposes to generate rated aspect summary of short comments based on the overall rating provided. They propose to use unstructured and structured PLSA to identify group of words representing aspects.

One subtask of our problem is sentiment classification, which is mostly a binary classification of a text (document or sentence) as having positive or negative sentiment. Many such techniques depend on some external knowledge in the form of pre-classified word list or training set to predict the polarity [21], [7]. However, our method only depends on the overall rating of the review which is provided almost all of the time. This makes our method flexible enough for cross domain applications. Our work more closely follows the concept of unsupervised aspect classification as done by [14], [15].

4 System Framework

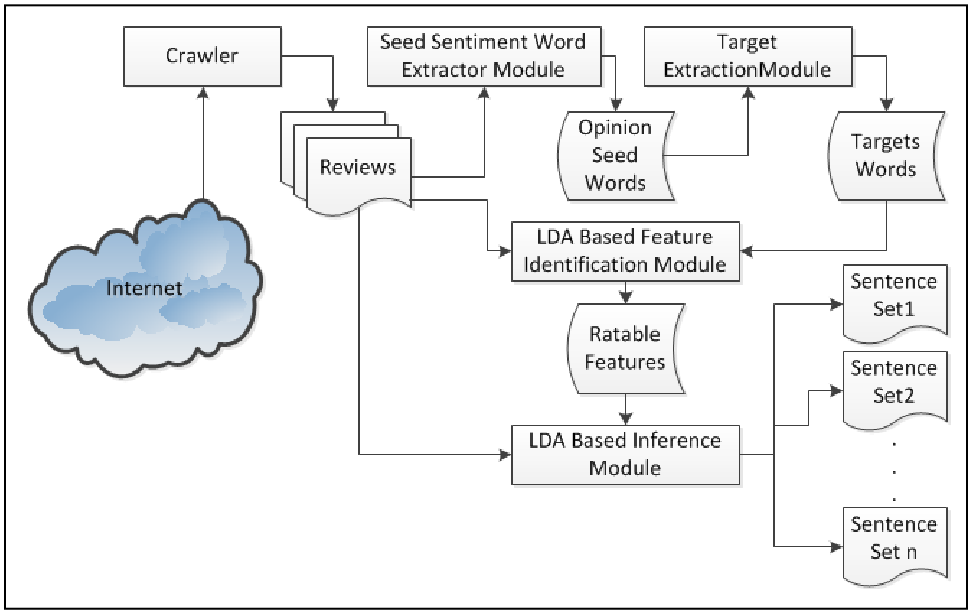


Figure 1. Overall System Framework

Figure 1 above depicts the overall framework and flow of our system. Customer reviews on a product are collected from review sites in the web. Most of the reviews have ratings (a number indicating overall sentiment of the review) with them. We use this number to extract seed sentiment words in that domain. These seed words are then fed in the target extraction module which essentially extracts candidate feature words and more sentiment words in a recursion. These candidate feature words are then grouped by the LDA based feature identification module to identify ratable features of the product/service. We then use LDA based inference module to identify ratable features in unseen sentences.

5 Methodology

The overall methodology can be seen as a combination of serial tasks such as domain specific seed sentiment words extraction, explicit feature candidate extraction based on sentence level features, assembly of target words to ratable features and classification of sentences to ratable features. In this section, we will explain each of the tasks in detail.

5.1 Domain specific seed sentiment words extraction

While the original double propagation method requires an opinion seed set to start with, our method does not require any external information. Since opinion words are completely domain dependent, it might not be so easy to obtain domain dependent opinion words for hundreds of products and services. We exploit user given ratings to collect domain dependent subjective opinions. Sentiment is expressed most of the time with adjectives. Thus, we only work with adjectives in a sentence. Now for each filtered word in a review, we count its frequency in reviews which have positive rating and negative rating. A review is considered positive if it has a 4 or 5 star and is considered negative if it has rating of 2 or fewer stars. Here, we use majority voting to detect the underlying polarity of these sentiment words. We define the semantic orientation of a sentiment word based on its minimum frequency in positive/negative reviews and the ratio of its occurrence in them. The semantic orientation is defined by the following equation.

$$SO\ (w) = \frac{|\mathrm{w_p}|}{(|\mathrm{w_p}|+|\mathrm{w_n}|)} \tag{1}$$

Here, w_p is the frequency of word w in positive reviews and w_n is the frequency of it in negative reviews.

If the semantic orientation (SO(w)) of a word w is greater than 60% in positive case and it appears at least in 2% of positive documents, the word is considered to have positive semantic orientation. Similarly, we also identify negative words. We changed the threshold values to obtain best precision and recall and hence found 60% and 2% is the optimal value. We also experimented in adverb phrase chunks and adjective phrase chunks with the same idea, but unigrams give the best results.

Negation handling

When considering a sentence as a bag of words, we lose the inherent structure of it and hence the semantic meaning of the sentence. A sentence may have a positive sentiment word; however it may have been modified by some negation word. For example, in a sentence, "*The **food** was **not** at all **good**",* the sentence talks about the aspect **food** and has a sentiment word "***good***". However, the overall meaning gets reversed because of the word "***not***" in it. To solve this problem, we define a window size of 5. If a negation word appears within a window size of 5 and if the sentiment word is the first to appear after the negation word, we consider the opinion to be reversed. We use a predefined set of negation words to detect the presence in sentences.

5.2 Explicit feature candidate extraction based on sentence level features

The feature extraction methodology is based on the syntactic relation between opinion words and the features. The process recursively discovers product/service features based on known opinion words and discovers more unknown opinion words based on known features.

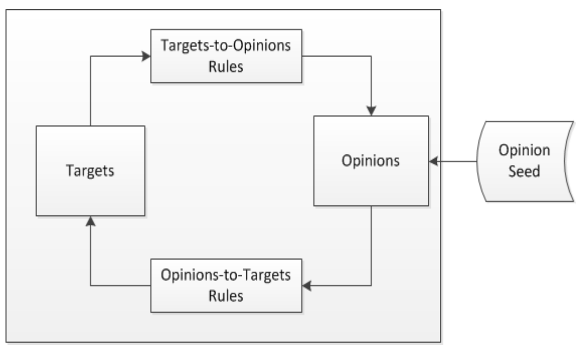


Figure 2. Conceptual framework for Target Extraction Module

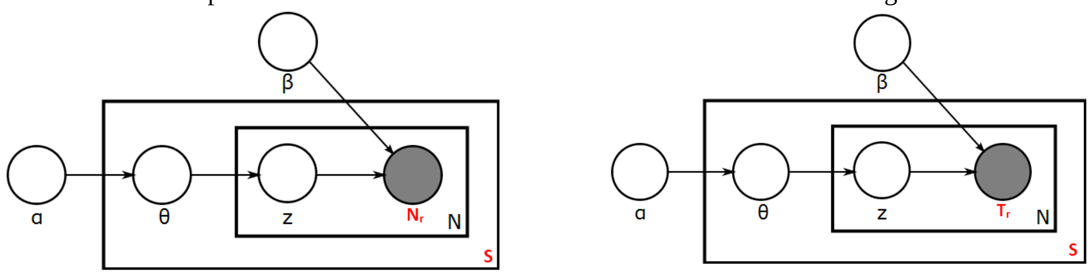
Figure 2 above shows the conceptual framework of target extraction module. Initially it requires a seed opinion word set, the extraction method of which has been described in the previous section. The process employs rule based strategy in dependency tree of a sentence. For example, If we know **great** is an opinion word and are given a rule like "*a noun on which an opinion word directly depends through mod is taken as the target*," we can easily extract **pictures** as the target. Similarly, if we know **picture** is a target, we could extract the adjective **great** as an opinion word using a similar rule. The paper [1] describes the extraction rules in detail.

5.3 Assembly of target words to ratable features

Opinions expressed in reviews are mostly associated for particular features of the entity. Each useful sentence should have a feature mentioned along with an opinion associated. So the next step is to identify representative features of the entity in terms of groups of words that represent the feature.

Latent Dirichlet Allocation (LDA) and its modifications [14], [15] have recently been applied to uncover the latent topics which are not directly observable in a document. We also followed the same idea of using bag of words in documents to identify the aspects in the reviews. LDA is a generative probabilistic model well suited for discrete distinct data such as text corpora. LDA can be seen as a three-level hierarchical Bayesian model which models each item of a collection in terms of finite mixture over latent set of aspects. Each aspect is then modeled as an infinite mixture of aspect probabilities. This allows it to capture significant intra-document statistical structure. Documents are represented as random mixtures over latent aspects, where each aspect is characterized by a distribution over words. Since we are interested in the fine grained features of the entity, we assume each sentence to be a single document [14]. Thus, the output of the model is a distribution over inferred aspects for each sentence in the data. We skip the details of LDA since it is out of scope in this work. Intuitively, features are mostly presented in noun forms. Thus, as a preprocessing step, Brody et al filtered noun form of words from the sentences in reviews and use them as candidate to generate hidden features of the entity.

We extend this filtering of words to another level in our work. Instead of using all nouns in the sentence, we filter the words to only keep candidate target words obtained from previous step to feed to LDA. LDA groups a set of representative words into pre-defined number of aspects. The following diagrams and the corresponding mathematical formulation represent sentence level noun filtered LDA and sentence level target filtered LDA.



(a) **(b)**

Figure 3. (a)Noun word based LDA and (b) Target word based LDA

We used the Gibbs sampling based LDA for estimating the parameters [21]. Let $\vec{N}_r$ and $\vec{Z}$ be the vectors of all nouns and their topic assignment in the collection. Then, the topic assignment for a particular noun in the review is computed as follows:

$$p(z_i = k \mid \vec{z}_{\neg i}, \vec{N}_r) = \frac{n_{k,\neg i}^{(t)} + \beta_t}{\left[\sum_{v=1}^{V} n_k^{(v)} + \beta_v\right] - 1} \; \frac{n_{m,\neg i}^{(k)} + \alpha_k}{\left[\sum_{j=1}^{K} n_m^{(j)} + \alpha_j\right] - 1}$$

(2)

Similarly, let $\vec{T}_r$ and $\vec{Z}$ be the vectors of all targets and their topic assignment in the collection. The topic assignment for a particular target in the review is computes as:

$$p(z_i = k \mid \vec{z}_{\neg i}, \vec{T}_r) = \frac{n_{k,\neg i}^{(t)} + \beta_t}{\left[\sum_{v=1}^{V} n_k^{(v)} + \beta_v\right] - 1} \; \frac{n_{m,\neg i}^{(k)} + \alpha_k}{\left[\sum_{j=1}^{K} n_m^{(j)} + \alpha_j\right] - 1}$$

(3)

Where $\vec{\alpha}$ and $\vec{\beta}$ are the Dirichlet parameters, V is the vocabulary size, $n_k^{(v)}$ is the number of times a topic k is assigned to the word v, $\neg i$ refers to the assignments excluding the current assignment.

5.4 Classification of sentences based on features

To each sentence in the data, the Target LDA model assigns a distribution {P(a)}a2A over the set A of inferred aspects.

6 Experimental Results

In this section, we will present our result in camera and restaurant domain for different tasks.

6.1 Dataset

To demonstrate the proof of concept of our model and its domain independence, we chose two completely disparate domains, restaurant (service) and camera (product). For the restaurant dataset, we used the existing dataset from [22]. It consists of 652 reviews of restaurants in New York City, of which a subset is annotated with features and orientation, which we use extensively during evaluation. For the camera dataset, we collected reviews from amazon.com to develop the framework. For the evaluation purpose, we used the data from Bing et al. We manually annotated the sentences in the dataset to 9 ratable features. The dataset was annotated by 2 graduate students of computer science. The kappa factor for the annotation was YY. The low kappa factor indicates that some of the sentences are even confusing for human to identify the features as in significant number of sentences; the features are not explicitly mentioned in the sentence. Some useful statistics of the dataset is shown in the following table.

Dataset	Restaurant	Camera
# of Reviews	**52624**	
# of positive reviews	**38653**	
# of negative reviews	**9305**	
Avg # of sentences in positive reviews	**5.09**	
Avg # of sentences in negative reviews	**5.78**	
Average rating	**3.946**	

Table 2. Dataset Statistics

6.2 Seed Sentiment Words Extraction

Table 2 below shows some sample sentiment words that were detected as positive and negative based only on provided user rating using all the reviews in both the dataset. We evaluate our method using the manually annotated dataset described above. Each sentence in the dataset is positive, negative, neutral or has conflict. Among 3000, 2603 of them sentences are either positive or negative and we only use these for evaluation.

Polarity	Sample Words
Positive	absolute, luscious, golden, refreshingly, cozy, amazing, tastefully, phenomenal
Negative	unidentifiable, tasteless, unfriendly, unprepared, unaccommodating, discourteous

Table 2. Sample seed sentiment words

Accuracy	Precision	Recall
65	71	86
	100	

Table 3. Evaluation result on seed sentiment word extraction.

6.3 Candidate Target Extraction Based on Double Propagation

Since we did not have any conceptual contribution on the double propagation algorithm, we did not explicitly evaluate this task. However, we did apply the algorithm in a bigger and more realistic size dataset. In the following table, we present the statistics on number of targets generated by the algorithm on camera and restaurant domain. We also show sample target words extracted in both domains.

Domain	Sample Candidate Target Words	# of Targets	# of Sentences
Restaurant			
Camera			

Table 4. Target Extraction Statistics and Sample Target Words

6.4 Grouping of Targets to Ratable Features

Tables 5 and 6 show some sample words that were identified by the LDA model. For the evaluation purpose, we used the manually annotated dataset by Ganu et al. They have annotated around 3000 sentences from 652 restaurants for its sentiment polarity and aspect. For each sentence, they assign one or more aspects among six aspects; food, staff, ambience, anecdotes, miscellaneous and price. For the evaluation on camera domain, we used Bing data which we annotated for 9 features and we used amazon.com collected data for target extraction and grouping.

Topics	Noun Filtering Based grouping	Target Filtering Based Grouping
Price	Food, service, prices, price, quality, ambience, atmosphere, bit, ok, nothing, size, great	worth, prices, bit, price, quality, portions, service, cheap, reasonable, average, pricey,value
Ambience	Food, service, staff, atmosphere, Service, wait, décor, waiters, Very, attentive, bit, waitstaff	Bar, room, dining, music, décor, tables, area, space, atmosphere, cozy, nice, seating
Food	Menu, sushi, dishes, Food, fish, chef, everything, variety, dish, items, specials, choices, tasting	Pizza, steak, cheese, meat, side, burger, taste, fries, plate, chicken, burgers, bread, sauce
	Wine, drinks, meal, list, selection, appetizers, bar, glass, drink, course, bill, entrees, bottle, beer	Menu, wine, dishes, list, appetizers, selection, everything, entrees, glass, course, bottle, tasting
Anecdote	Dinner, night, time, friend, lunch, friends, day, brunch, birthday, party, evening, boyfriend, group	Place, recommend, love, fun, date, nice, friends, perfect, people, eat, enjoy, spot, big, anyone
Misc	Restaurant, experience, times, time, reviews, dining, years, place, meal, visit, couple, NYC	restaurants, experience, places, favorite, city, top, dining, years, neighborhood
Service	Table, waiter, minutes, time, order, people, reservation, hour, waitress, manager, hostess, wait,	Table, wait, minutes, order, reservation, bar, waiter, hour, waiting, reservations

Table 5. Sample words representing ratable features in restaurant domain

Topics	Noun Filtering Based grouping	Target Filtering Based Grouping
Shooting mode	Mode, auto, settings, focus, feature, photos, setting, iso, results, modes, control, not, thing	Shots, light, flash, shot, shooting, speed, shutter, night, indoor, lighting, grainy
Video/ memory	Video, hd, flip, videos, camcorder, hours, anything, minutes, sound, ultra, movies,	Video, videos, software, camcorder, download, record, sound, recording
Build/ look	Camera, reviews, lot, model, people, Kodak, things, review, research, others, problems	Settings, feature, auto, focus, photos, button, setting, macro, automatic, function
misc	Camera, reviews, lot, model, people, Kodak, things, review, research, others, problems	Camera, features, manual, user, learning, controls, functions, average, version
anecdote	Pictures, camera, pics, kids, lot, work, days, course, fun, expectations, detail, kind	Pictures, family, daughter, sharp, vacation, friends, friend, crisp, show, pix, wedding
size	Camera, size, use, pocket, something, everything, ease, purse, hand, weight	Camera, size, screen, compact, weight, bigger, portability, viewing, body
Price	Nikon, amazon, product, money, purchase, price, nothing, service, coolpix, deal, days	Good, price, deal, sale, reasons, buy, review, money, worth, purchase, return, pay, cost

Picture quality	Quality, picture, image, pictures, photo, images, stabilization, lighting, situations, friend	Quality, picture, image, photo, stabilization, stabilizer, crystal, piece, versatility, blast
Battery life	Camera, battery, time, life, year, wife, daughter, Christmas, gift, great, son, problems	Battery, batteries, life, time, charge, portable, power, charger, stick, run, shirt

Table 6. Sample words representing ratable features in camera domain

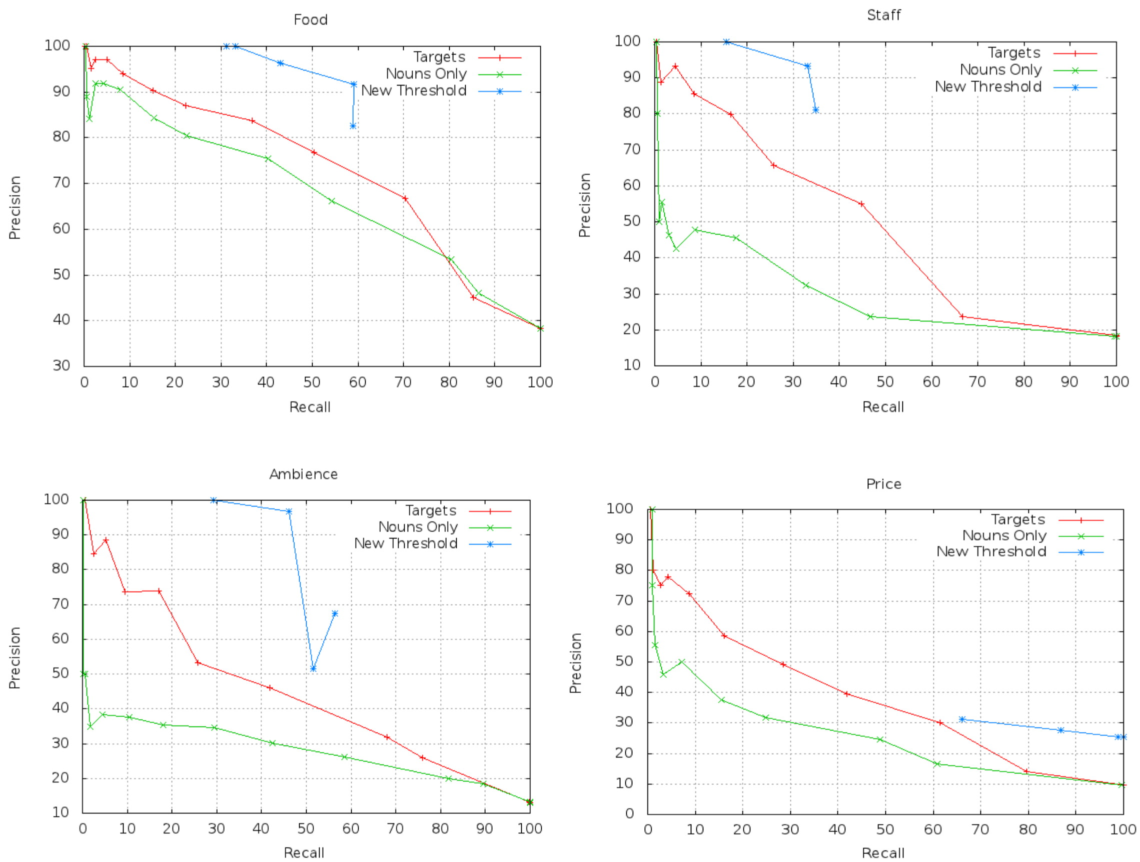


Figure 4. Precision/Recall curves for ratable aspects food, staff, Ambience and Price in restaurant domain with noun filtering and target filtering with different thresholds

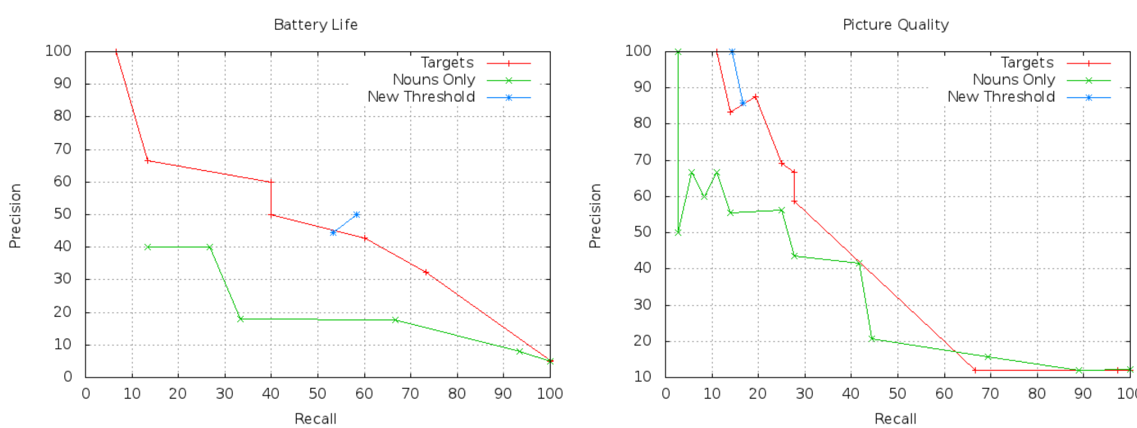


Figure 5. Precision/Recall curves for ratable aspects Battery life and picture quality with noun filtering and target filtering with different thresholds

Figures 4 and 5 above show the precision/recall curve for some of the ratable aspects in restaurant and camera domain. We experimented with two different types of threshold values to classify sentence to one of the inferred aspects. In the first type, a threshold ta for each aspect is defined and we assume a sentence as associated to aspect a if P(a) > ta. By varying the threshold ta we created precision-recall curves. This is similar to Brody et al. For a direct comparison to their method, we also implemented noun based filtering along with target based filtering and drew the curve that depicts significant improvement in precision/recall with the target based filtering method. With the other type, a threshold ta for each aspect is defined as the difference in probabilities, P(a)-P(b) of two features a and b with highest probabilities. With the result shown in the graph, this threshold seemed to be more useful in producing high precision results.

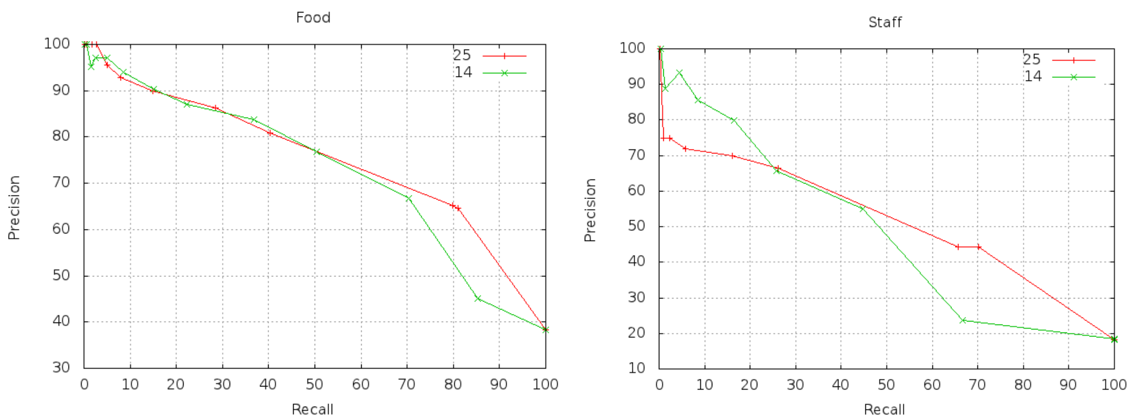


Figure 6. Precision/Recall curves with 14 and 25 topics for ratable aspects, food and staff

The problem of determining the model order (number of aspects) still persists in our method as in most unsupervised learning scenario. We thus performed a small experiment varying the number of features in precision/recall graph as shown in figure 6. As the result shows, although choosing right number of topics did seem to improve the result, the improvement was not significant. Hence we decided to use some practical number of features that users would think of while reviewing a product/feature. For the restaurant domain, to be able to evaluate our system we chose to use 14 topics with the mapping of 6 manually annotated features. For the camera domain, we tried to manually annotate the dataset from Bing liu dataset and found that there are nine practical features. So we used nine features for camera dataset.

7 Discussion and Future Work

We presented a framework to identify and aggregate ratable features with minimal supervision. Our method has shown significant improvement on the identification and grouping of features in reviews. We also introduce the idea of using the star rating as a way to classify sentiments without an external corpus. Since all the steps were performed in a domain-independent way, the system is flexible enough to be equally applicable to any other entity of any domain. Though, the recall of aspect identification system is not high, in real life scenario, most of the products/services have sizable amount of reviews and hence even a low recall result could be representative and helpful to customers.

As future work, we intend to extend the system to collect all the reviews for a particular entity from the web and produce succinct information. This work can be seen as a milestone to build a system that does not perform keyword based document retrieval but actually processes relevant documents to produce precise information.

Current work ignored the requirement of identifying opinion holder of the opinion. However, if the same concept needs to be applied in more general contexts such as blog to generate a summarized form of information for an entity such as a person or a company or location, opinion holder identification cannot be ignored. Moreover, in such a scenario, we would not even have explicit sentiment polarity clue such as rating. Thus an entity based faceted opinion summarization in such case would be more challenging. We intend to explore in that direction down the line.

References

1. Opinion Word Expansion and Target Extraction through Double Propagation. Qiu, Guang, et al. 1, 2011, Computational Linguistics, Vol. 37, pp. 9-27.
2. Dhar, Vasant, Chang, Elaine and Stern, Leonard N. Does Chatter Matter? The Impact of User-Generated Content on Music. Does Chatter Matter? The Impact of User-Generated Content on Music. 2007.
3. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. Ipeirotis, Panagiotis G. 2010, IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 99.
4. Bansal, Mohit, Cardie, Claire and Lee, Lillian. The power of negative thinking: Exploiting. The power of negative thinking: Exploiting.
5. Sentiment Retrieval using Generative Models. Eguchi, Koji and Lavrenko, Victor. [ed.] Dan Jurafsky and Éric Gaussier. s.l. : ACL, 2006. pp. 345-354. ISBN: 1-932432-73-6.
6. Determining the semantic orientation of terms through gloss classification. Esuli, Andrea and Sebastiani, Fabrizio. 2005. pp. 617-624.
7. Thumbs up?: sentiment classification using machine learning techniques. Pang, Bo, Lee, Lillian and Vaithyanathan, Shivakumar. Stroudsburg, PA, USA : Association for Computational Linguistics, 2002. pp. 79-86. DOI: http://dx.doi.org/10.3115/1118693.1118704.
8. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Pang, Bo and Lee, Lillian. 2004. pp. 271-278.
9. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Turney, Peter D. Stroudsburg, PA, USA : Association for Computational Linguistics, 2002. pp. 417-424. DOI: http://dx.doi.org/10.3115/1073083.1073153.
10. Using Latent Semantic Analysis To Improve Access To Textual Information. Dumais, Susan T., et al. s.l. : ACM, 1988. pp. 281-285.
11. Probabilistic latent semantic indexing. Hofmann, Thomas. New York, NY, USA : ACM, 1999. pp. 50-57. ISBN: 1-58113-096-1 DOI: http://doi.acm.org/10.1145/312624.312649.
12. Latent dirichlet allocation. Blei, David M., et al. 2003, Journal of Machine Learning Research, Vol. 3, p. 2003.
13. Mining opinion features in customer reviews. Hu, Minqing and Liu, Bing. s.l. : AAAI Press, 2004. pp. 755-760. ISBN: 0-262-51183-5.
14. An unsupervised aspect-sentiment model for online reviews. Brody, Samuel and Elhadad, Noemie. Stroudsburg, PA, USA : Association for Computational Linguistics, 2010. pp. 804-812. ISBN: 1-932432-65-5.
15. Modeling online reviews with multi-grain topic models. Titov, Ivan and McDonald, Ryan. New York, NY, USA : ACM, 2008. pp. 111-120. ISBN: 978-1-60558-085-2 DOI: http://doi.acm.org/10.1145/1367497.1367513.
16. WordNet: A Lexical Database for English. Miller, George A. 1995, Communications of the ACM, Vol. 38, pp. 39-41.
17. Movie review mining and summarization. Zhuang, Li, Jing, Feng and Zhu, Xiao-Yan. New York, NY, USA : ACM, 2006. pp. 43-50. ISBN: 1-59593-433-2 DOI: http://doi.acm.org/10.1145/1183614.1183625.
18. Mining product reputations on the Web. Morinaga, Satoshi, et al. New York, NY, USA : ACM, 2002. pp. 341-349. ISBN: 1-58113-567-X DOI: http://doi.acm.org/10.1145/775047.775098.
19. Extracting product features and opinions from reviews. Popescu, Ana-Maria and Etzioni, Oren. Stroudsburg, PA, USA : Association for Computational Linguistics, 2005. pp. 339-346. DOI: http://dx.doi.org/10.3115/1220575.1220618.
20. Rated aspect summarization of short comments. Lu, Yue, Zhai, ChengXiang and Sundaresan, Neel. New York, NY, USA : ACM, 2009. pp. 131-140. ISBN: 978-1-60558-487-4 DOI: http://doi.acm.org/10.1145/1526709.1526728.
21. Learning Subjective Adjectives from Corpora. Wiebe, Janyce M. 2000. pp. 735-740.
22. Ganu, Gayatree, Elhadad, Noemie and Marian, Amelie. Beyond the stars: improving rating predictions using review text content. Proceedings of the Twelfth International Workshop on the Web and Databases. 2009.