Extracting Service Aspects from Web Reviews

Jinmei Hao[1], Suke Li[2], and Zhong Chen[2]

[1] Beijing Union University, China
haomei99@yahoo.com.cn
[2] School of Electronics Engineering and Computer Science, Peking University, China
{lisuke,chen}@infosec.pku.edu.cn

Abstract. Web users have published huge amounts of opinions about services in blogs, Web forums and other review friendly social websites. Consumers form their judgements to service quality according to a variety of service aspects which may be mentioned in different Web reviews. The research challenge is how to extract service aspects from service related Web reviews for conducting automatic service quality evaluation. To address this problem, this paper proposes four different methods to extract service aspects. Two methods are unsupervised methods and the other two methods are supervised methods. In the first method, we use FP-tree to find frequent aspects. The second method is graph-based method. We employ state-of-the-art machine learning methods such as CRFs (Conditional Random Fields) and MLN (Markov Logic Network) to extract service aspects. Experimental results show graph-based method outperforms FP-tree method. We also find that MLN performs well compared to other three methods.

Keywords: service aspect extraction; opinion mining; web mining.

1 Introduction

Consumers are used to browsing online service related reviews and like to compare service quality of different service providers before making purchasing decisions in recent years. It is common for Web uses to publish Web reviews to describe a variety of service aspects of services due to convenient facilities of Web based service providers such as *tripadvisor*[1].

Service quality evaluation plays a very important role in traditional service management. Service quality has five dimensions [1] : reliability, responsiveness, assurance, empathy and tangibles. In this paper, we try to extract five dimensions related service aspects. Each dimension has different aspects. For example, in tangible dimension includes aspects: the appearance of physical facilities, equipment, and also other tangible evidence of the care and attention to detail that are exhibited by service providers. For instance, in the sentence "*The bathroom is great.*", *"bathroom"* is a service aspect.

The research challenge is how to extract service aspects from service reviews in order to conduct automatic service quality evaluation. To address this issue,

[1] http://www.tripadvisor.com

F.L. Wang et al. (Eds.): WISM 2010, LNCS 6318, pp. 320–327, 2010.

this paper proposes four different methods to extract service aspects. The first two methods are unsupervised methods and the last two methods are supervised methods. In the first method, we use FP-tree [2] to find frequent aspects. The second method is graph-based method. We employ CRFs (Conditional Random Fields) [3] and MLN (Markov Logic Network) [4] as the third and the fourth methods. Experimental results show graph-based method outperforms FP-tree [2] based method. We also found MLN performed well compared to other three methods.

There are several potential applications based on extracted service aspects. For instance, we can find what service aspects may affect purchasing behaviors of customers. And what service aspects make customers feel good or bad. We can also summarize reviews uses sentences containing the most important aspects. It is possible to rank the helpfulness of the reviews according to the extracted aspects. In service management research field, researchers usually design questionnaires or surveys to find the gap between customer expectations and perceptions. However, how to design proper questionnaires or surveys becomes a big challenge, because the designer needs to know what service aspects are important for most of customers. Efficient service aspect extraction methods are promising for survey design.

2 Related Work

Few publications focus on service aspect extraction. However, there are extensive research work focusing on product feature extraction. Hu and Liu's work [5] is early effort to summarize product opinions through association mining. Liu [6] proposed an opinion mining system to do feature-based opinion mining. In [5] and [6], nouns and noun phrases were used as product features. There are also some research focusing on implicit feature extraction, such as [7]. However, our work focuses on explicit service aspects only. Popescu et al. [8] introduced an unsupervised information extraction system, namely OPINE which parses the reviews and applies a simple pronoun-resolution module to the parsed data. It is not clear whether OPINE can been applied to service aspect extraction or other languages. Since our work only adopts shallow language processing techniques, the proposed methods are quite general and can be applied to other languages easily.

3 Service Aspect Extraction

Service aspect extraction is the first step toward further service opinion mining in our work. In this section, we show four methods to extract service aspects respectively. The first two methods are unsupervised methods, and the last two methods are supervised methods using CRFs and MLN to extract service aspects. In this work, service aspects are nouns and noun unit (consecutive nouns) that are related service quality evaluation. For example *hotel*, *room*, *breakfast*, *staff*, *location*, etc. are all service aspects. We judge whether nouns or noun units are service aspects by humans.

3.1 Association Mining Based Aspect Extraction

Hu and Liu's work [5] is the early effort for feature-based product opinion mining using association mining algorithm Apriori [9]. Hu and Liu [5] believed when people comment on product features, the words that they use converge. Because most of service aspects are nouns or noun phrase, intuitively Hu and Liu's approach [5] can also be applied to our data set. In this paper, we take hotel industry as an example to extract service aspects. However, a sentence can have opinions about several service aspects and not every aspect is explicit. An aspect is explicit, if the service aspect words or phrases appear in the sentence. In this work, we only consider to extracting explicit service aspects.

3.2 Aspect Extraction Based on Conditional Random Fields

Conditional Random Fields (CRFs) [3] is used widely in sequence labeling. It is a framework for building probabilistic models to segment and label sequence data based on undirected graphical models. CRFs are also been applied in open information extraction [10] as well as POS tagging and phrase chunking. Suppose input data sequence is X and the label sequence is Y, then the joint distribution of Y given X is

$$p(y|x) \propto exp(\sum_{e \in E,k} \lambda_k f_k(e, y|_e, x) + \sum_{v \in V,k} \mu_k g_k(e, y|_v, x)), \quad (1)$$

where, x is a data sequence, y a label sequence. Function f_k and g_k are feature functions which can acquire from training data, and $y|_S$ is the set of components of y associated with the vertices in subgraph S. To obtain labeled data for CRFs, in this method, the beginning noun of a noun phrase is labeled as ***B-A***, other sequential nouns are labeled as ***B-I***. A training example for CRFs is as Fig. 1 shown.

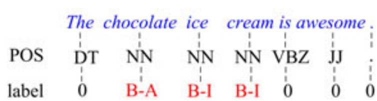


Fig. 1. A label sequence example for CRFs

3.3 Graph-Based Service Aspect Extraction

We use terms of sentences to construct directed graph. In order to simplify the process, we adopt a simple arc generation rule to generate graphs for service review. The rule is simple:

Arc Generation Rule. *If two terms in a sentence, for example term A and term B are consecutive, term A is term B's left neighbor, and if there is no arc from term A to term B, then an arc is generated from term A to term B.*

All service reviews in the data set will generate a directed graph, namely *term-tag* graph. Nodes in such a directed graph are terms which are extracted according to POS (Part-of-Speech) tags of sentences, and a tag is a attribute of a node. Each node is corresponding with one distinct term. It is obvious that even for the same term, it is possible that the term can have different POS tags due to the characteristics of natural languages. In this method, a node in a generated graph can have one tag attribute. If a term has more than one kind of tag, the tag with the most frequency will become the tag attribute of the node. For instance, the word *good* may have different POS tags in terms of different context. In the sentence *"He is a good man"*, the word good is labeled as ADJ. However, in the sentence *"I knew it was no good to say anything"*, the word *good* is a noun and should be labeled as NN. In our data set, because the frequency of tag ADJ is much bigger than the frequency of tag NN for the term *good*, the tag attribute of the term node of *good* is ADJ. In the opinion mining field, nouns and noun phrases are often used as candidate features or aspects. Intuitively, a major service aspect may be mentioned in different reviews and may have more in-links from other nodes or out-links to other nodes. Based on this observation, we use PageRank algorithm PageRank [11] on our constructed term-tag graph to rank all the terms. We then select noun or noun phrase terms as extracted service aspects according to their rankings. We consider the service aspect ranking problem as finding the top K most influential nodes in a term-tag graph. Let G be directed term-tag graph. The term ranking process is a random surfer process. Matrix A is the adjacency matrix or stochastic transition matrix of G which has n nodes. Let P is a n-dimensional column vector of aspect-opinion rank values:

$$P = (p_1, p_2, ..., p_n)^T \tag{2}$$

Matrix A is the adjacency matrix with

$$A_{ij} = \begin{cases} \frac{1}{O_i} & if(i,j) \in E, \\ 0 & otherwise \end{cases} \tag{3}$$

where O_i is the out-degree of node i. Node i can be a term. We can iterate to obtain rank values of all nodes them using

$$P = (1 - d)e + dA^T P \tag{4}$$

where $d \in (0, 1)$, d is a damping factor, and e is a column vector of all 1's. In our experiments, d is 0.85. For a random surfer in the graph, it has d probability to follow an out-link of the node and $(1 - d)$ probability to jump to a random node. Fig. 2 is a term-tag example graph.

3.4 Aspect Extraction Base on MLN (Markov Logic Network)

Markov Logic Network (MLN) [4] combines first-order logic and probabilistic graphical models. A MLN can be viewed as a template for constructing Markov Random Fields,

$$P(X = x) = \frac{1}{Z} exp\langle \sum_i w_i n_i(x) \rangle = \frac{1}{Z} \prod_i \phi_i(x_i)^{n_i(x)}, \tag{5}$$

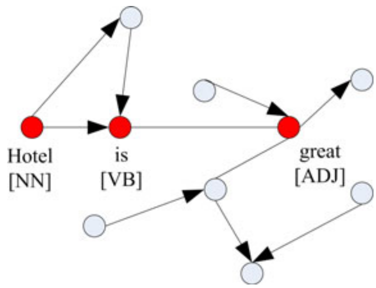


Fig. 2. A term-tag example

where $n_i(x)$ is the number of true groundings of F_i in X, x_i is the state (true values) of atoms appearing in F_i. In this method, a term is classified into two classes: aspect class and non-aspect class using MLN.

Table 1 contains all predicates that are used by MLN. For instance, $isSeqNVA(x, y)$ means term x and term y are in a subsequence of a tag sequence of a sentences. The tag sequence is $< NP, VP, ADJP >$. After we construct MLN, we can answer probabilistic queries such as $isAspect(x)$ which means the probability of x is a service aspect. In this work, if we get $p(isAspect(x)) > 0.5$, then we deem that x is a service aspect.

Table 1. Predicates and Their Descriptions

Rule	Description
isAdj(x)	X is an adjective.
isNon(x)	X is a noun.
isModifier(x,y)	X is a modifier of y.
isSeqNVA(x,y)	X and y are in the sequence of $< NP, VP, ADJP >$.
isInNPPhrase(x)	X is in NP phrase.

MLN needs to employ some logic formulas to work. These formulas are shown as follows.

$$\forall x(isNon(x) \wedge isAdj(y) \wedge isModifier(x, y)) \rightarrow isAspect(x)$$
$$\forall x(isNon(x) \wedge isAdj(y) \wedge isSeqNBA(x, y)) \rightarrow isApsect(x)$$
$$\forall x(isNon(x) \wedge \neg isInNPPhrase(x) \rightarrow \neg isAspect(x)$$
$$\forall x(isNon(x)) \rightarrow \neg isAdj(x)$$
$$\forall x(isAdj(x)) \rightarrow \neg isAspect(x)$$

4 Experiments

4.1 Experiments with Aspect Extraction

Our data set contains 500 reviews that are uniformly randomly sampled from the global data set without replacement. The global data set contain about 25

Table 2. Data Set Description

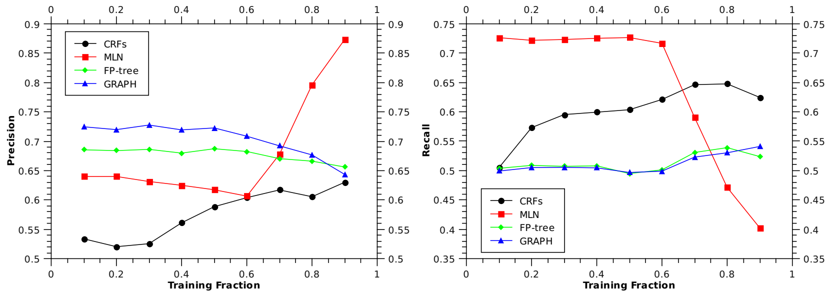
Method	Precision	Recall	F-Score
Pure Noun	0.4174	0.9024	0.5708
NP Noun	0.6172	0.4059	0.4898

thousands reviews crawled from *Triadvisor*[2]. The labeled data set contains 500 reviews and it has been segmented into 4545 sentences. The number of distinct labeled aspects is 1706 and they are distributed in 3421 sentences. These aspects are judged by humans. The work of labeling service aspect is labor intensive and time consuming. We also use OpenNLP[3] to get all the Part-of-Speech tags for the sentences. Noun and noun phrase can be used as features [5]. Table 2 shows some results of using *Pure Noun* and *NP Noun* as service aspects. *Pure Noun* means we only consider nouns only. *NP Noun* means we use nouns in noun phrases only. We can see if we use pure noun as service aspects, the precision is only 0.4174. NP Noun has higher precision but lower recall.

Because the graph-based method and FP-tree method are unsupervised methods, and the CRFs method and MLN method are supervised methods, in order to make comparison, graph-based method and FP-tree method only use test data set to extract service aspects in our experiments. The same as [5], this work uses support value 0.01 to extract product features. For instance, in Fig. 3a, when training fraction is 0.2, graph-based method only extracts aspects from the test data set with fraction of 0.8. In our experiments, we adopt 0.01 as our support value for FP-tree. For the graph-based method, we first use FP-tree on test data set with support 0.01 to get the number of extracted aspect, namely K, than we select K ranked aspects generated from graph-based method. However, CRFs method and MLN method are supervised machine learning methods and need training data set to work. We adopt CRF++[4] for labeling sequential data.

Fig. 3a shows the precision distributions of different methods. Graph-based method performs best. However, CRFs method performs worst. With the increase of training fraction, the precision also increases except for MLN method. Fig. 3b illustrates distributions of recall of four methods. We can see CRFs based method perform better than MLN method. MLN method has good performance when the training fraction less than 0.6. Fig. 4 shows distributions of f-score of our methods. In this case, FP-tree has the worst performance when training fraction greater than 0.4. F-score of CRFs based method increases with the increase of training fraction. When the training data fraction is less than 0.7, MLN method has the best performance.

[2] http://www.tripadvisor.com
[3] http://www.opennlp.org
[4] http://crfpp.sourceforge.net/



(a) Precision distributions (b) Recall distributions

Fig. 3. Precision and recall distributions

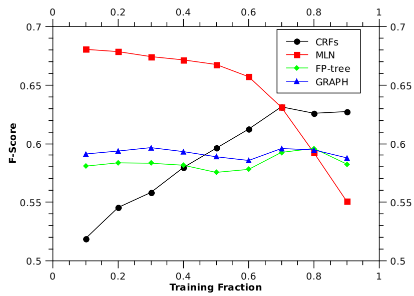


Fig. 4. F-Score distributions

5 Conclusions and Future Work

This work focuses on service aspect extraction. We propose four methods to conduct service aspect extraction. In the first method, we use FP-tree to find frequent aspects. The second method is graph-based method. We employ CRFs (Conditional Random Fields) and MLN (Markov Logic Network) as the third and the fourth methods. For measuring extraction precision, experimental results show graph-based method outperforms FP-tree based and other two methods in almost all cases. MLN method performs well in measuring extraction recall. In measuring F-score of service aspect extraction, We also find MLN outperforms other three methods when the fraction of training data set is less than 0.7. In the future, we will continue our research work toward automatic service quality evaluation based on extracted service aspects.

References

1. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: Servqual: A multiple-item scale for measuring consumer perceptions. Journal of Retailing 64 (1988)
2. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, SIGMOD 2000, pp. 1–12. ACM, New York (2000)

3. Lafferty, J.M.A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. the 18th International Conference on Machine Learning, pp. 282–289. Morgan Kaufmann, San Francisco (2001)
4. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning 62, 107–133 (2006)
5. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004, pp. 168–177 (2004)
6. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th international conference on World Wide Web, WWW 2005, pp. 342–351 (2005)
7. Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Swen, B., Su, Z.: Hidden sentiment association in chinese web opinion mining. In: Proceeding of the 17th international conference on World Wide Web, WWW 2008, pp. 959–968. ACM, New York (2008)
8. Popescu, A.M., Nguyen, B., Etzioni, O.: Opine: extracting product features and opinions from reviews. In: Proceedings of HLT/EMNLP on Interactive Demonstrations, pp. 32–33. ACL, Morristown (2005)
9. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB 1994, pp. 487–499 (1994)
10. Etzioni, O., Banko, M., Soderland, S., Weld, D.S.: Open information extraction from the web. Commun. ACM 51(12), 68–74 (2008)
11. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. 30(1-7), 107–117 (1998)