

# Multi-Language Sentiment Analysis for Hotel Reviews

Maleerat Sodani<sup>a</sup>

*Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand*

**Abstract.** Touristes and traveler use a variety of information sources (e.g. travel portals, blogs, or social networking sites like twitter) to help them decide for a hotel room. These sources all contain highly subjective text that expresses the opinions of many. We took a preliminary view on user generated hotel reviews from two travel portals in English and Thai. We developed a taxonomy of features and specifically investigated how accurately they can be predicted with three classification methods. The results indicate that support vector machines perform best for this specific domain.

## 1 Introduction

People clearly enjoy traveling. Travel and tourism is continuously on the rise over the past six decades and has become one of the biggest and fastest-growing economic sectors on the planet, regardless of any economic crisis. According to the World Tourism Organization UNWTO, international tourist arrivals have increased from 25m globally in 1950, to 278m in 1980, 527m in 1995, and have now reached an impressive 1133m in 2014. Likewise, tourist spending has risen from US\$ 2b in 1950 to US\$ 1245b in 2014, based on the 2015 edition of the World Tourism Highlights report [1]. The amount of information about hotels has increased with it and there was never a time where travellers had better and easier access to hotel reviews than today. This however also means that travellers have to base their decisions on a lot more detailed information about hotels and an even larger amount of generally very subjective information from previous guests about their quality. Decision making has therefore become more challenging and information systems use sentiment analysis to assist in the analysis of human opinion to help users make a sensible choice. In [4], for example, supervised learning with two types of information (frequency and TF-IDF) are used to realize polarity classification of reviews.

In this paper we took a preliminary view on user generated hotel reviews from two travel portals. We developed a taxonomy of features for these reviews and specifically investigated how accurately they can be predicted with three different classification methods.

We selected reviews for English, as the major spoken language in the travel domain, and Thai as a local comparison. After reviewing related work in the next section, we highlight our evaluation methodology and describe our data set in Section 3. Results are presented

and discussed in Section 4 before reaching our conclusions.

## 2 Related work

This section describes an overview of sentiment analysis related disciplines such as Text mining, Feature Extraction/Identification, Sentiment classification techniques and also performance evaluation. Text mining is conventional data mining done using text features. Text features are usually keywords which are extracted based on the language model. In this paper, the keyword means the word that can be used to classify reviews in terms of polarity. The accuracy of classification depends on the text features which are extracted from the reviews [2]. The goal of sentiment analysis is to detect opinion, identify the sentiments people express by classifying the contextual polarity of a given text. The polarity can be positive, negative or neutral. In general, there are three different levels in sentiment analysis: document, sentence, and feature/aspect level respectively. Document level in sentiment analysis is applied to detect the polarity of the whole product or service reviews [3], [8]. Sentence level is applied to define a given text usually a sentence or phrase while feature/aspect level determines the opinion expressed on different features or aspects (attribute or component) of entities [3]. The results of sentence level sentiment analysis are dependent on the definition of subjectively used when annotating texts [5]. The advantage of feature-based sentiment analysis is the possibility to extract about the object of interest and determine an opinion expressed on each feature/aspect is positive, negative or neutral [4].

In this paper, we applied three classification approaches, Naive Bayes (NB) [9], Decision Trees (DT) [10] and Support Vector Machine (SVM) [11], for the experimental comparison for Thai and English. These

<sup>a</sup> Corresponding author: msn@kmutnb.ac.th.

methods are selected as a common technique for this paper as they represent a very common and universally applicable approaches in an attempt to get initial results on our data.

### 3 Evaluation methodology

This section explained the main module that used to evaluate a reviews: One is reviews weight analysis, Second is classifier module which compared from three type of classification algorithms such as Naïve Bayes, Decision tree and Support Vector Machine (SVM).Third is testing module which the best classifier is applied.

#### 3.1 Proposed framework

There are three main parts of sentiment analysis of hotel reviews framework. Part 1 is a reviews analysis which an input is both Thai and English reviews and the output is a reviews separated into positive and negative. Weight calculation is applied in this part in order to assign weight for all properties (Room, Location, Service, Price and Facilities) Part 2 is a training module which reviews separated used as input and output is the best model of classification process. Part 3 is testing module which model is applied to classify a hotel review in to positive and negative as shown in Fig. 1.

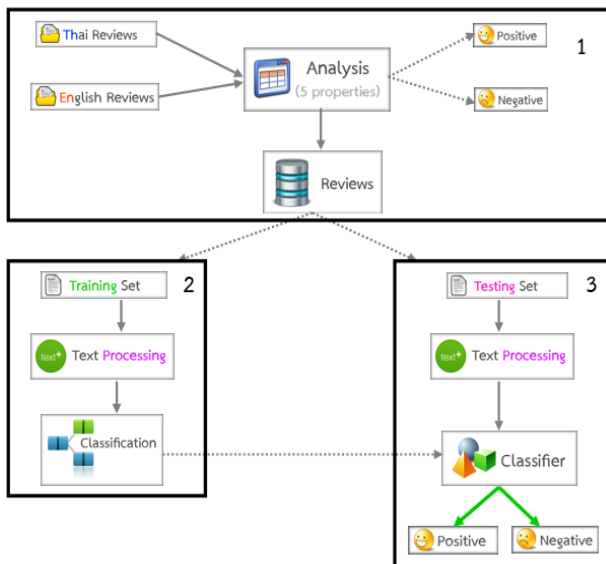


Figure 1. Proposed framework of sentiment analysis.

Table 1. Taxonomy of categories and sub-categories.

Property	Sub-property	Property	Sub-property
Room	Clean	Location	Location
	Room size		Public services
	Room location		Environment
	Other satisfaction		
Service	Receptionist	Price	Room price
	Check-in, check-out		Accessories price
Facilities	Breakfast		
	Other facilities		
	Internet access		

#### 3.2 Data collection and process

We collected a total of 3,000 hotel reviews for English and Thai from the two well-known websites TripAdvisor<sup>a</sup> and Agoda<sup>b</sup> between January and March 2015. We had an equally sized set for each languages (1500 reviews each). The data was collected with the tool import.io<sup>c</sup>, and free web site scraping tool, to transform the review pages into text. A domain expert then assigned the text to the five categories Room, Location, Service, Price, and Facilities where each was labelled as a class of polarity (Positive and Negative). Then we manual separate and labelled polarity from each reviews into five categories: Room, Location, Service, Facilities, and Price. All reviews are labelled as a class of polarity (Positive and Negative) by experts. Data is spitted into three sets of training, validating and testing with 60%, 20% and 20% respectively. Then three classifier were tested based-on the same data in order to compare the performance in terms of accuracy. Fig. 2 and Fig. 3 show an example of Thai reviews and English reviews respectively.



Figure 2. Thai reviews from agoda.com/th-th.



Figure 3. English reviews from Tripadvisor.

#### 3.3 Weight calculation

In order to calculate weight property, five properties (Room, Location, Service, Price and Facilities) and fourteen sub-property need to be extracted from the reviews. The weight for all sub properties under the core property are sum into 1. The weight ratio of sub-property can be calculated as follows:

$$\% \text{ mentioned properties} = \frac{\text{No. of mentioned properties} * 100}{\text{Total of reviews}} \quad (1)$$

Then summarize the total weight for each properties as:

$$W_{\text{total}} = W_1 + W_2 + \dots + W_n \quad (2)$$

<sup>a</sup>www.tripadvisor.com

<sup>b</sup>www.agoda.com/th-th

<sup>c</sup>https://import.io

where  $W_{total}$  is a main property weight,  $W_i$  is sub-property weight where  $i=1,2,..,n$ , and  $n$  is a number of sub-property,  $W_i$  equal  $(+)W_i$  if the review is positive and  $W_i$  equal  $(-)W_i$  if the review is Negative. In our experiments, weight which calculated by equation 1 and 2 are illustrated in Table 2 and Table 3.

**Table 2.** Property weight for Thai reviews.

Property	Sub-property	%Reviews	Weight (+/-)
Room	Clean	38.33	0.3/ -0.3
	Room size	19.73	0.1/ -0.1
	Room location	12.13	0.1/ -0.1
	Other Satisfaction	50.47	0.5/ -0.5
Location	Location	18.60	0.1/ -0.1
	Public service	40.13	0.4 / -0.4
	Environment	22.73	0.2/ -0.2
Service	Receptionist	38.93	0.3/ -0.3
	Check-in, check-out	47.00	0.4 / -0.4
Facilities	Breakfast	22.8	0.2/ -0.2
	Otherfacilities	7.27	0.07/ -0.07
	Internet access	3.93	0.03/ -0.03
Price	Room price	16.13	0.1/ -0.1
	Accessoriesprice	6.47	0.06/ -0.06

**Table 3.** Property weight for English reviews.

Property	Sub-property	%Reviews	Weight (+/-)
Room	Clean	30.13	0.3/ -0.3
	Room size	27.93	0.2/ -0.2
	Room location	18.33	0.1/ -0.1
	Other Satisfaction	49.93	0.4/ -0.4
Location	Location	48.73	0.4/ -0.4
	Public service	46.20	0.4/ -0.4
	Environment	20.27	0.2/ -0.2
Service	Receptionist	67.13	0.6/ -0.6
	Check-in, check-out	43.00	0.4/ -0.4
Facilities	Breakfast	53.60	0.5/ -0.5
	Otherfacilities	41.07	0.4/ -0.4
	Internet access	11.60	0.1/ -0.1
Price	Room price	70.13	0.7/ -0.7
	Accessoriesprice	31.00	0.3/ -0.3

### 3.4 Evaluation measure

In order to evaluate classifier model, accuracy is used to measure how accurate of class that each classifier method performed. This paper compare the performance of the three classification models: Naïve Bayes, Decision Tree and Support Vector Machine. Accuracy is defined as:

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (3)$$

where

TP is number of correct items for positive polarity,  
 TN the number of correct items for negative polarity,  
 FP the number of incorrect items for positive polarity and  
 FN the number of incorrect for negative polarity.

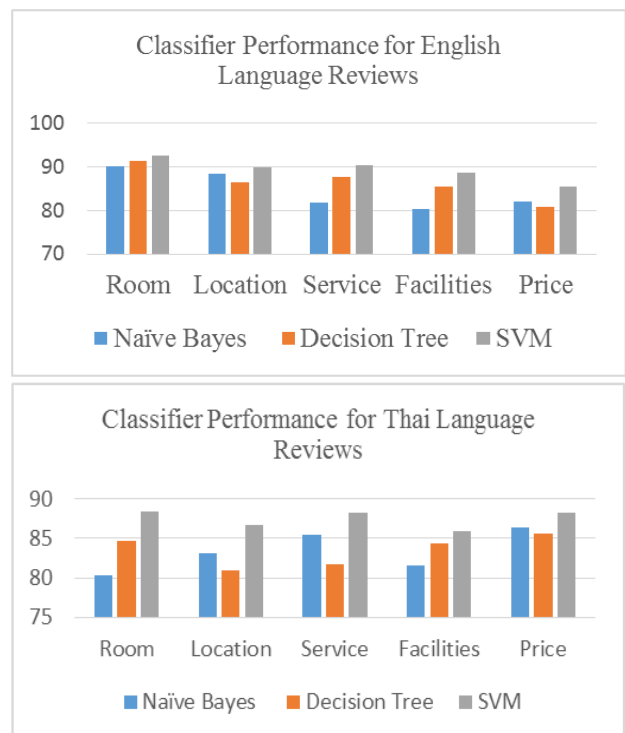
## 4 Results and discussion

The classification results are summarized in Table 4 and Fig 4. The result show the accuracies of the correct

classification as described in Table 1. We have separated them by properties and the base classifier and highlighted the language for each group in the last column. The highest accuracy for each partition and language is highlighted in bold. The performance in terms of accuracy depends on the text features used in part of text processing. The accuracy in Thai language is less than English in all properties because in Thai reviews, there is a little comment and not all sub-property are mentioned while in English language reviews, there are almost cover all sub-property which we used to calculate weight statistical.

**Table 4.** Performance comparison in terms of accuracy.

Properties	Naive Bayes	Decision Tree	SVM	
Room	<b>90.12</b>	<b>91.40</b>	<b>92.61</b>	🇺🇸
Location	88.52	86.56	89.91	🇺🇸
Service	81.70	87.70	90.39	🇺🇸
Facilities	80.37	85.47	88.75	🇺🇸
Price	82.02	80.91	85.50	🇺🇸
Room	80.36	<b>84.72</b>	88.34	🇹🇮
Location	83.20	80.93	<b>86.77</b>	🇹🇮
Service	<b>85.48</b>	81.72	88.19	🇹🇮
Facilities	81.54	84.40	85.92	🇹🇮
Price	86.37	85.63	88.25	🇹🇮



**Figure 4.** Performance comparison from three classifier of Thai and English language.

## 5 Conclusion

This paper proposed sentiment analysis based on statistic weight that is calculated from hotel reviews in both Thai and English language. We applied three classification methods — naïve Bayes, decision tree, and support vector machine have been used to classify the hotel reviews into positive and negative polarities. The

experiments were conducted using hotel reviews obtained from the TripAdvisor and the Agoda portals. The results revealed that support vector machines given the best accuracy, outperforming naïve Bayes and decision tree classification. However, properties weight which calculated using training data effected the performance of the system in terms of accuracy.

## References

1. WTO. UNWTO Tourism Highlights, 2015 Edition. WTO, (2015)
2. J. Yao, H. Wang, and P. Yin. Sentiment feature identification from Chinese online reviews. In H. Tan and M. Zhou, editors, *Advances in Information Technology and Education*, volume 201 of *Communications in Computer and Information Science*, pages 315–322. Springer Berlin Heidelberg, (2011)
3. M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, pages 168–177, New York, NY, USA, (2004). ACM
4. B. Liu, M. Hu, and J. Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web, WWW '05*, pages 342–351, New York, NY, USA, (2005). ACM
5. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, EMNLP '02*, pages 79–86, Stroudsburg, PA, USA, (2002). Association for Computational Linguistics
6. H.-X. Shi and X.-J. Li. A sentiment analysis model for hotel reviews based on supervised learning. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, volume 3, pages 950–954, (July 2011)
7. F. Su and K. Markert. From words to senses: A case study of subjectivity recognition. In *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1, COLING '08*, pages 825–832, Stroudsburg, PA, USA, (2008). Association for Computational Linguistics
8. P. D. Turney. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pages 417–424, Stroudsburg, PA, USA, (2002). Association for Computational Linguistics
9. Y. Huang and L. Li, "Naive Bayes classification algorithm based on small sample set," *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, Beijing, (2011), pp. 34-39
10. M. T. Cazzolato and M. X. Ribeiro, "A statistical decision tree algorithm for medical data stream mining," *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, Porto, (2013), pp. 389-392
11. R. Varghese and M. Jayasree, "Aspect based Sentiment Analysis using support vector machine classifier," *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, Mysore, (2013), pp. 1581-1586