Vol.2014-DBS-159 No.24
Vol.2014-IFAT-115 No.24
2014/8/2

Cross-Domain Investigations of User Evaluations under the Multi-cultural Backgrounds

JIAWEN LE[†1] HAYATO YAMANA[†2 †3]

Twitter, as one of the most popular social network services, is widely used to query public opinions. In this research, a large corpus of Twitter data, along with online reviews, are used to apply sentimental and culture-based analysis, so as to figure out the cultural effects on user evaluations. Posts written in more than 30 languages from more than 30 countries are collected. In order to implement the cross-domain investigations, global restaurants and world attractions are taken as the research subjects, and a series of classifiers with high performances are trained and applied in the experiment steps. Then various analyzing methods are applied to obtain informative results and conclusions about the user evaluations for the targets. As the contributions, this research validates the capability and field transferability of the proposed methods for cross-lingual sentiment analysis, and arrives at the conclusions that the cultural effects on user evaluations for both restaurant domain and travel domain actually exist, and are obvious for some countries and cultural backgrounds.

1. Introduction

Twitter [a], one of the most popular social network services, owns a range of special characteristics, including the tremendous amount of posts, the great variety of tweet contents, and the rapid speed of information distribution. Actually, the huge volume of tweets can be used to survey public opinions. If many users post tweets that contain complimentary words of a restaurant, it is likely that this restaurant enjoys popularity among customers. Meanwhile, with the increase of the number of transnational enterprises and the development of transportation services, now people from all over the world can use the same product, savor the same meal, and appreciate the same scenery. However, it is quite common that people from different countries may have totally different feelings about these experiences, probably partly due to their diverse cultural backgrounds. In order to figure out the cultural effects on the evaluations of people with different cultural-backgrounds, tweets, as well as some website reviews, can serve as a good dataset to carry out the analysis.

However, there exist several challenges considering this issue. The problem of how to correctly figure out the sentiment of these short texts remains as the main task for many researchers. As for this task, a noted work from Liu [1] reviews the existing approaches and research in the field of sentiment analysis. The language barrier is another challenge. Most previous research in this field only focus on the English-written tweets, and posts in other languages are simply discarded. However, the fact that Twitter also enjoys great popularity in many non-English-speaking countries suggests that the strategy of ignoring non-English tweets will definitely lead to the biased and incorrect results in cross-culture analysis. Several works have studied cross-lingual sentiment analysis, but the target languages and text formats are very limited. OpinionIt [2] is an opinion mining system comparing the cross-lingual differences in opinions, and in the paper, the authors take the reviews written in English and Chinese as the main subject. One more challenge lies in the field transferability. Most research in the field of sentiment analysis only consider a single domain, and some of the most popular domains for this kind of study include the domain of films and political issues. For the reason that texts in different domains may actually have different vocabularies and stylistic features, it is quite questionable whether a sentiment analysis approach in one field applies to another.

Facing all these challenges, this research, which is based on our previous work [3], has made four main contributions: (1) Considering the multi-cultural background, tweets written in more than 30 languages from more than 30 countries are analyzed; (2) Cross-domain investigations are carried out that both tweets in restaurant and tourism domain are analyzed ; (3) The sequential three-step process of spam classification, subjectivity classification, and polarity classification been further modified to obtain better performances; (4) By carrying out a range of analyzing methods, an insight into people's attitudes towards the target restaurants or attractions is given, and informative conclusions considering the cultural effects are obtained.

2. Related Work

In the aspect of opinion mining, a noted work is presented by Pang and Lee [4], which gives a broad view of some existing approaches for sentiment analysis and opinion retrieval. Early research that tries to put forward new methods or improve existing approaches considering the particular study subject of tweets can be listed as followed. Go et al. [5] use the emoticons to query Twitter, and take the search results as the training set. They divide these tweets into negative ones and positive ones according to the sentiment of the query emoticons. They report

†1 Graduate School of Fundamental Science and Engineering, Waseda University
†2 Faculty of Science and Engineering, Waseda University
†3 National Institute of Informatics
a) http://twitter.com/.

that unigram feature model achieves the best performance, which cannot be gained by using bigrams and POS feature models. The work of Pak and Paroubek [6] characterizes in the method of the collection of objective training data. The source of this objective data includes several popular newspapers, whose sentences are usually considered as without special sentiment polarity. On the other side, the research of Barbosa and Feng [7] mainly focus on the syntax features, and combine them with the POS model. All these works only take English tweets into consideration, and have not touched upon the cross cultural backgrounds.

As for the field of cross-lingual sentiment analysis, the noted opinion analysis system Oasys [8] allows the user to observe the change of intensity of opinion over countries and news sources. Guo et al. [9] construct a text mining system to detect the different sentiment in the Web texts written in different languages. The work of Cui et al. [10] uses emotion tokens to solve the problem of cross-lingual sentiment analysis. Gao et al. [11] research on Twitter and the Chinese version of Twitter---Sina Weibo, and make some simple statistical comparisons in several different aspects, such as the characteristics of user behaviors and the content of messages.

In comparison, we focus on the analysis of cross-lingual user evaluations in multi-field, which is based on the sentiment classification using the Twitter and review data. More than 30 languages and more than 30 countries are taken into account, so as to obtain more authentic and comprehensive results for culture-based analysis.

3. Methodology

3.1 Data

Restaurant domain. The data used in this research mainly comes from two sources, Twitter and restaurant review websites [b][c][d]. As for the Twitter dataset, 9,523,211 restaurant-related tweets were gathered from Sep. 2013 to Dec. 2013, by using Twitter Streaming API and Search API. All the data has been restricted by the names of target restaurants (i.e. *Burger King*, *McDonald's*, *KFC*, *Pizza Hut*, *Subway*, and *Starbucks*). Then, as an auxiliary dataset, 55,031 English-written reviews were collected from popular review websites.

Tourism domain. By using the same collecting method, 2,113,624 travel related tweets (from Sep. 2013 to Dec. 2013) and 42,769 travel related reviews are collected as Twitter dataset and the review dataset. The names of 12 world attractions (i.e. *Great Wall of China, Mountain Fuji, Matterhorn, Sydney Opera House, Statue of Liberty, Colosseum, Louvre Museum, Grand Canyon, Machu Picchu, Angkor Wat, Eiffel Tower, Taj Mahal*) are taken as the filtering condition.

As for the target languages, tweets originally labeled as written in the 34 languages (i.e. en, es, id, ja, fr, pt, tl, ru, tr, zh, ar, th, et, nl, it, de, ko, bg, sv, pl, vi, sk, da, ht, lt, lv, sl, fi, is, no, fa, hu, el, uk [e]) are taken into consideration for both domains. The target countries are listed in Table 1.

After basic preprocessing steps (i.e. Translation, spam filtering), two dictionaries are constructed for each domain. First, a total word dictionary (*tw_total_dict*) records words appeared in Twitter dataset more than 3 times. Then, an initiative polarity dictionary (*pol_dict_ini*) is constructed by combining the entries of several popular polarity dictionaries on the Internet, including SentiWordNet [f], MPQA [g], and the General Inquirer [h].

Table 1 Target Countries

Restaurant domain	United States (US), United Kingdom (GB), Australia (AU), Indonesia (ID), Malaysia (MY), Canada (CA), Philippines (PH), Singapore (SG), Brazil (BR), India (IN), South Africa (ZA), Japan (JP), Mexico (MX), France (FR), Netherlands (NL), Greece (GR), Thailand (TH), China (CN), Russia (RU), Spain (ES), Argentina (AR), Chile (CL), South Korea (KR), Germany (DE), Italy (IT), Ireland (IE), Venezuela (VE), Colombia (CO), Poland (PL), Egypt (EG), Ukraine (UA), New Zealand (NZ), Viet Nam (VN)
Tourism domain	United States (US), United Kingdom (GB), Australia (AU), Indonesia (ID), Malaysia (MY), Canada (CA), Philippines (PH), Singapore (SG), Brazil (BR), India (IN), South Africa (ZA), Japan (JP), Mexico (MX), France (FR), Netherlands (NL), Greece (GR), Thailand (TH), China (CN), Russia (RU), Spain (ES), Argentina (AR), Chile (CL), South Korea (KR), Germany (DE), Italy (IT), Ireland (IE), Venezuela (VE), Colombia (CO), Poland (PL), Egypt (EG), Viet Nam (VN), Salvador (SV), Slovenia (SI), Sweden (SE), Panama (PA), Norway (NO), Saudi Arabia (SA), Latvia (LV), Kazakhstan (KZ), Kuwait (KW), Cambodia (KH), Greenland (GL), Estonia (EE), Ecuador (EC), Denmark (DK), Czech (CZ), Switzerland (CH), Bulgaria (BG), Belgium(BE), Austria (AT)

3.2 Features for Sentiment Classification

Syntax Features. The special syntax characteristics of tweets are quite informative in the task of sentiment analysis. In this research, 10 types of syntax characteristics (i.e. '!', '?', '#', '@', 'RT', upper-case words, capitalized words, URL links, emoticons, and slang words) are counted respectively, and this 10-dimension vector is regarded as the '*syn*' feature. Here, a manually built emoticon dictionary (300 entries) and slang dictionary (200 entries) are referred to during the counting process.

Modified Unigram. Compared to the standard unigram model with an extremely sparsity, an additional dimension-reduction is applied while processing the modified unigram features. First, for each word in *tw_total_dict*, the polarity score is set as 2, -2, and 0 if it is labeled as Positive, Negative, and Neutral in *pol_dict_ini* respectively. Then all the tweets are parsed to cal-

b) http://www.tripadvisor.com/.
c) http://www.yelp.com/.
d) http://www.zagat.com/.

e) http://en.wikipedia.org/wiki/ISO_639-1/.
f) http://sentiwordnet.isti.cnr.it/.
g) http://mpqa.cs.pitt.edu/.
h) http://www.wjh.harvard.edu/~inquirer/.

culate out the PMI (Pointwise Mutual Information) values of all the pairs of words in *tw_total_dict*. The PMI value of word w_1 and w_2 is given by

$$PMI(w_1, w_2) = \log\frac{p(w_1, w_2)}{p(w_1) \cdot p(w_2)}, \tag{1}$$

where, $p(w_1, w_2)$ is the co-occurrence probability of word $w_1$and w_2 in one tweet, and $p(w_1)$ and $p(w_2)$ are the occurrence probabilities of word $w_1$and w_2 in one tweet respectively. Then, for each word NOT appears in *pol_dict_ini*, sort its PMI values with the words in *pol_dict_ini*, and carry out majority voting among the top 10 sorted items. The 'positive inclined' word is then score as 1, the 'negative inclined' word is then scored as -1, and other words is then scored as 0. The output of this step is a new polarity dictionary (*pol_dict*) with the vocabulary of *total_dict*, and each word in it is mapped to a score of 5 scales (i.e. 2, 1, 0, -1, 2). Based on *pol_dict*, each tweet can be projected to a 5-dimension vector, and each dimension records the count of the unigram words in this category. This vector is named as the '5s' feature.

Review Dataset-based Average Score. The online reviews are always posted with a corresponding concrete score, which are quite informative considering the sentiment analysis. In our previously constructed review dataset, each entry has a tuple structure of $(text, score)$. In this step, all the text parts are first processed as a BoW model, and the total vocabulary of the review dataset is described as W_{rv}. For each word w_i in W_{rv}, the review dataset-based polarity score is calculated by

$$pol_{w_i} = \frac{\sum_{text_j \in TX_{w_i}} score_j}{|TX_{w_i}|}, \tag{2}$$

where, TX_{w_i} is the set of review texts, in which the word w_i occurs, $text_j$ is a review text in TX_{w_i}, and $score_j$ is the corresponding score of $text_j$. Then for each tweet tw_i in the Twitter dataset, the review dataset-based average score is given by

$$avg_{tw_i} = \frac{\sum_{w_j \in W_{tw_i}} pol_{w_j}}{|W_{tw_i}|}, \tag{3}$$

where, W_{tw_i} is the word set of tw_i, and pol_{w_j} is the polarity score of w_j given by formula (2). Here, length normalization is applied that the occurrence number of the word that has the highest frequency in a review or in a tweet is normalized into 1. The float average score calculated by formula (3) is named as the 'rv' feature.

Review Dataset-based CCA Score. Canonical correlation analysis (CCA) is a classical statistics method to figure out the latent relations among multiple variables. In the review dataset, each entry consists of a comment text and a 5-scale score, described as $(text, score)$. For the reason that there must be some consistency in the comment text and score from the same person, it can be safely concluded that there is some latent relationship between them. Thus, the CCA method can be used here to get the latent relationship between the users' sentiment and polarity words. Here, the first correlated variable is adopted as the measure criterion. The review dataset is taken as the condition set, and the first correlated variable parameters are decided by the CCA process. Then, for each tweet in the Twitter dataset, the first correlated variable is calculated and this float number is given to each tweet as the '*cca*' feature.

Window Co-occurrence-based Average Score. Since that the neighboring relationship among words may contain useful information for sentiment analysis, the score based on the co-occurrence in a three-word window is calculated. Inspired by the previous research, a modified graph-based propagation algorithm is adopted here to obtain the polarity score of each word in *tw_total_dict* based on the three-word window neighboring relationship. First, a co-occurrence dictionary is constructed by parsing all the tweets in the Twitter dataset. The key of the item in this dictionary is the word pair $w_i\_w_j$, and the value of the item in this dictionary is the times $t(w_i, w_j)$ these two words appeared in the three-word window. Then, as an initial propagation graph, all the words in *tw_total_dict* are taken as the nodes of the graph. The value of each node is initiated as 1, and -1 for the words in the Positive category and Negative category of *pol_dict_ini* respectively. For other words, the initiated node value is set as 0. Then, for each iteration, the value of each node is updated by

$$v'_{n_i} = (1-\alpha) \cdot \frac{\sum_{n_j \in NEI_{n_i}} v_{n_j} \cdot \left(1 + \log\left(t(n_i, n_j)\right)\right)}{\sum_{n_j \in NEI_{n_i}} \left(1 + \log\left(t(n_i, n_j)\right)\right)} + \alpha \cdot v_{n_i} \tag{4}$$

where, NEI_{n_i} is the set of the nodes neighbored with node n_i, and $t(n_i, n_j)$ is the co-occurrence times of the words of node n_i and n_j, according to the previous built co-occurrence dictionary. α is a tuning parameter, which is set as 0.6 in this step. In the final graph where it converges, each node has a float value indicating the polarity of the word of this node. A polarity dictionary can be obtained by this final graph, and the average score of each tweet can be calculated based on the newly constructed polarity dictionary. This float score for each tweet is named as the 'win3' feature.

POS-based Feature. The POS (part-of-speech) information is usually used in the NLP analysis, and some POS pairs are especially sentiment expressive. Here, all the tweets are first processed by the Stanford Parser [i] to get the dependencies trees. Then 10 most common and sentiment expressive POS pairs (i.e. '*acomp*', '*advmod*', '*amod*', '*conj*', '*dobj*', '*neg*', '*nsubj*', '*purpcl*', '*rcmod*', and '*xcomp*'.) are chosen manually, and the sen-

i) http://nlp.stanford.edu/software/lex-parser.shtml/.

Vol.2014-DBS-159 No.24
Vol.2014-IFAT-115 No.24
2014/8/2

timent expressed in these pairs are decided according to some manually constructed rules (e.g. the sentiment expressed in a *'neg'* pair is the opposite of the sentiment of the polarity word in the pair). For each tweet in the Twitter dataset, each above-mentioned POS pair that appears in the tweet is given with a polarity label. To decide the polarity of the tweet, a simple majority voting method is applied, which means that the polarity label that has the biggest POS pair count passes its polarity to the tweet. This feature is called *'pos'* in the later analysis steps.

4. Experiment

In this section, the experiment is carried out over the restaurant domain and tourism domain, and the basic steps for these two domains are the same. Figure 1 shows the overview of the experiment.

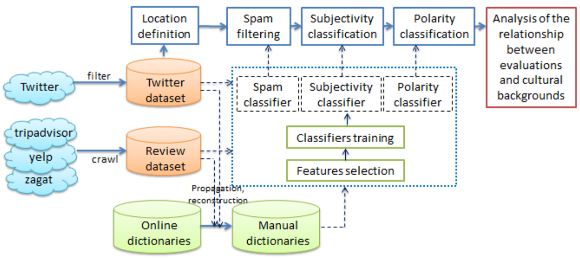

Figure 1　The Main Flow of the Experiment

As for the sentiment classification process, it is divided by two steps. The first step, subjectivity classification, is to classify the preprocessed dataset into the subjective dataset and the objective dataset. The second step, polarity classification, is to further classify the subjectivity dataset into the positive dataset and the negative dataset. In these two steps, pre-trained classifiers are applied to carry out the classification.

Features selection. In previous sections, 6 groups of features (i.e. *'syn', '5s', 'rv', 'cca', 'win3'*, and *'pos'*) are introduced. All the combinations of these 6 groups of features are implemented in this experiment.

Training method. The SVM (Linear, RBF, and Polynomial) methods and the Naïve Bayes (Gaussian, Multinomial, and Bernoulli) methods are used in this experiment.

Training implementation. The total number of implementation variations is:

$$(2^6 - 1) \cdot 6 = 378$$

Validation method. The standard 10-fold cross-validation is applied here.

Training set. For the subjectivity classifier, 1000 tweets, half of whose subjectivity is objective and another half is subjective are selected from the manually labeled tweets (majority vote by 3 readers). For the polarity classifier, 1000 tweets, half of whose polarity is positive and another half is negative are selected from the manually labeled subjective tweets.

Test results. For the restaurant domain, the best-performed subjectivity classifier (with an accuracy of 78.4%) is obtained by the features combination of *'syn', 'rv', 'win3',* and *'pos'*, with SVM polynomial training method, while the best-performed polarity classifier (with an accuracy of 91.1%) is obtained by the combination of *'rv', 'win3', 'cca',* and *'pos'*, with the SVM linear training method. For the tourism domain, the best-performed subjectivity classifier (with an accuracy of 84.3%) is obtained by the features combination of *'syn', '5s',* and *'rv'*, with SVM RBF training method, while best-performed polarity classifier (with an accuracy of 96.4%) is obtained by the combination of *'rv', '5s', 'win3',* and *'cca'*, with the SVM polynomial training method.

5. Analysis

After applying the optimal subjectivity classifier and polarity classifier, the preprocessed Twitter dataset is divided into 3 polarity groups, i.e. positive, negative, and objective. While the positive, negative, and objective tweet is given a polarity score of 1, -1, and 0 respectively, the example sentiment maps for the target restaurants and attractions are depicted in Figure 2 and Figure 3 (Axis: green represents negative sentiment; red represents positive sentiment.)

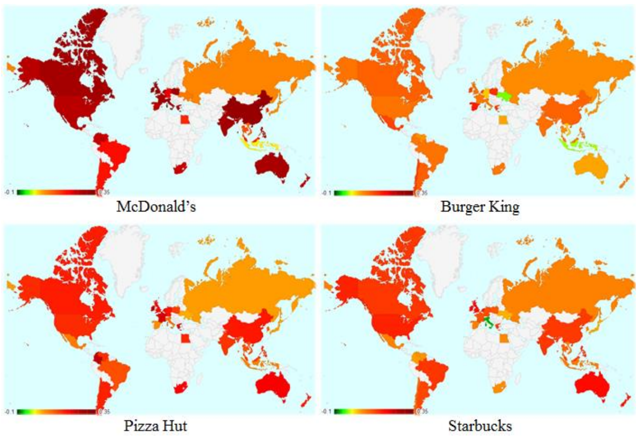

Figure 2　Example Sentiment Maps (Restaurant Domain)

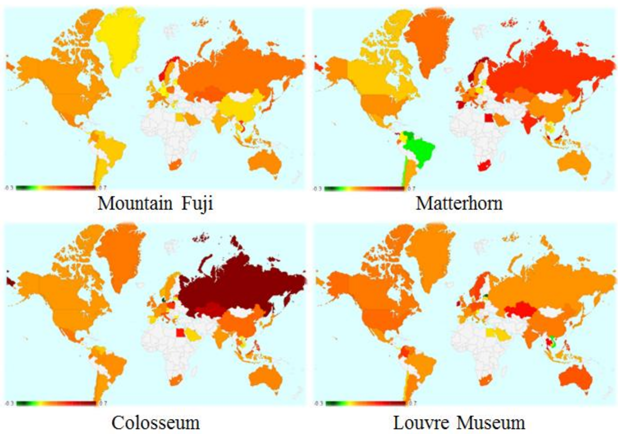

Figure 3　Example Sentiment Maps (Tourism Domain)

By representing the sentiment by gradient color, the above sentiment maps demonstrate the overall distributions of people's opinions for the example targets in the restaurant domain and tourism domain. However, as for the more specific reasons why people like or dislike a target, or the concrete characteristics of a target that shape people's attitudes, it still remains unclear and needs further exploration. To this end, the frequently occurred sentiment words, either positive or negative, are extracted with their frequencies for each target in the two domains, and the tag cloud is harnessed as a tool to describe these representative sentiment keywords. Figure 4 and 5 give example tag clouds for the two domains. White background indicates the positive sentiment, and black background indicates the negative sentiment. The size of the word denotes the occurrence frequency, and the multicolor of the word has no special significance.

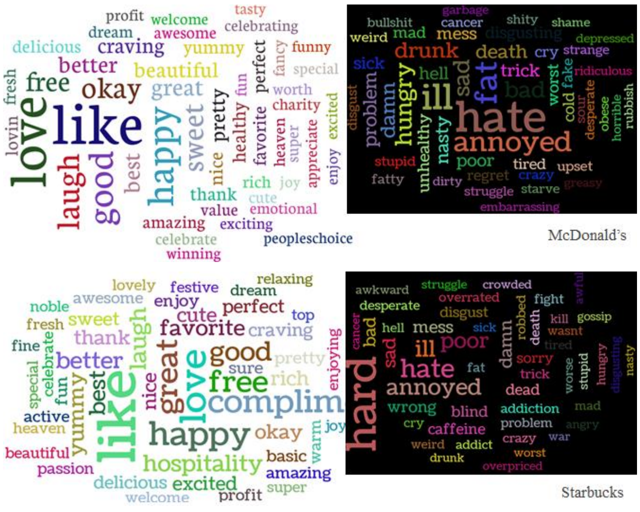

Figure 4　Example Tag Clouds (Restaurant Domain)

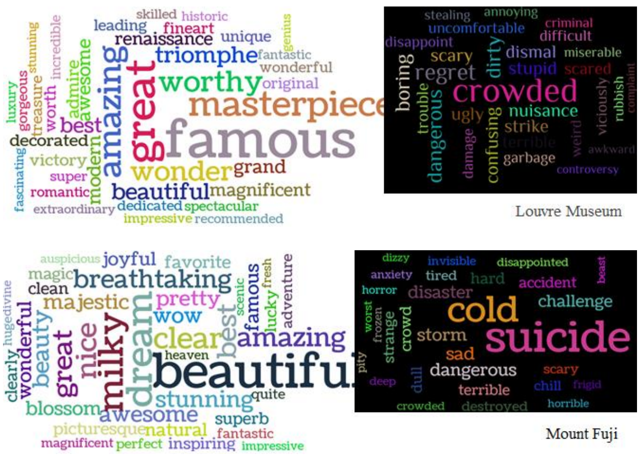

Figure 5　Example Tag Clouds (Tourism Domain)

Compared with the tag clouds of the restaurant domain, the tag clouds for world attractions seem to be more informative and meaningful. For instance, words like 'famous', 'masterpiece', 'renaissance', 'worthy', 'treasure', 'gorgeous' are particular or representative for the positive aspect of *Louvre Museum*, and words like 'crowded', 'dirty', 'boring', 'confusing' may relate to the negative aspect of *Louvre Museum*. Also, as for *Mount Fuji*, the positive features can be described by 'beautiful', 'clear', 'milky', 'blossom', 'picturesque', and 'fresh', while the negative descriptions include words like 'suicide', 'cold', 'dangerous', 'frozen', and 'invisible'. Based on these special keywords, we may easily obtain some important hints or underlying facts for the pros and cons of the target attractions. In contrast, as for the restaurants, comparatively little specific information can be acquired due to the big overlap of vocabulary among the different targets. Thus, to compensate for this deficiency, *(Attribute, Value)* pairs are used to describe the target restaurants.

Based on the Stanford dependency trees obtained in the sentiment classification step, we select out the sentiment expressive word pairs (explained in Section 3.2), each of which typically but not restrictedly consists of one noun (attribute) and one adjective (value), to construct the *(Attribute, Value)* list for each target. Table 2 and Table 3 give parts of the *(Attribute, Value)* lists of *McDonald's* and *Starbucks* as examples. Red color and green color represent positive and negative sentiment respectively. Numbers following the value words denote frequencies.

Table 2　the (Attribute, Value) List of McDonald's

Attribute	Value
food	fast 1861, great 451, good 359, best 302, worst 281, healthy 244, favorite 216, new 151, leftover 133, unhealthy 118, delicious 104, bad 104, terrible 90, fat 82, better 75, nasty 67, indigestible 60, nice 60, mexican 58, greasy 46, normal 45, regular 35, asian 35, expensive 35, fresh 34, organic 34 ,lethargic 33, nutritious 31, awful 30, healthier 29, indian 27, filthy 27, healthiest 26, horrible 25
burger	delicious 606, double 590, cheese 307, better 243, best 184, free 151, mcbusted 107, big 101, good 92, large 45, fat 38, fish 33, nice 32, special 32, great 29, bad 28, disappointing 21, small 17, expensive 17, huge 16, nasty 16
chicken	real 153, good 143, large 115, fried 92, bad 61, best 60, cheese 58, grilled 52, big 37, french 31, fresh 29, better 22, crispy 21, small 19, garlic 19, hot 18, nasty 15, classic 14, delicious 14
meal	happy 1574, free 475, big 456, large 323, whole 277, full 136, extra 115, unhappy 110, happier 97, best 89, traditional 73, favorite 67, good 67, romantic 66, healthy 54, cheeseburger 38, nice 33, great 30, breakfast 28, bad 22, worst 16,despicable 15, regular 15, small 15, delicious 14, terrible 13
fries	large 627, fresh 369, french 314, good 184,cheese 136, best 94, small 71, hot 66, cold 58, great 39, greasy 32, big 29, nasty 28, favorite 27, yummy 21, delicious 16, famous 15

Table 3　the (Attribute, Value) List of Starbucks

Attribute	Value
coffee	good 1494, best 988, favorite 845, hot 639, , expensive 431, black 295, great 245, bad 231, breakfast 218, exploitative 201, delicious 200, nice 189, poor 151, iced 124, healthy 104, cold 99, fresh 93, instant 88, terrible 87, packaged 74, nasty 73, normal 68, yummy 65, daily 53, different 52, special 52, worst 49, horrible 39,classic 34, overpriced 33
tea	green 3398, hot 344, bubble 297, black 217, sweet 181, good 108, best 85, iced 56, great 42, favorite 35, nice 27, breakfast 25, herbal 18, nonfat 18, red 17, poor 17, bad 17, chamomile 16, classic 16, daily 16, refresh 16
barista	favorite 163, cute 152, best 74, temporary 62, friendly 38, good 26, happy 22, attractive 16, beautiful 15,rude 15, certified 15
latte	delicious 104, french 76, good 65, hot 58, brûl ée 57, yummy 46, chocolate 45, best 44, favorite 35, breakfast 28, great 27, nonfat 18, fat 18, iced 18, nice 16
cake	chocolate 70, cheese 59, marble 45, good 31, best 28, lemon 25, classic 24, new 24, bad 23, complimentary 23, fetid 19, sweet 19, birthday 18, crumble 18, delicious 17, fat 16, favorite 16, festive 16, great 15, healthy 15, nice 15, obnoxious 15, truffle 15

Finally, in both domains, we have tried to use k-means method to cluster the target countries according to their average sentiment scores for the target restaurants or attractions. Figure 6 shows the clustering results while k is set as 7, 8, 9, and 10 for the restaurant domain.

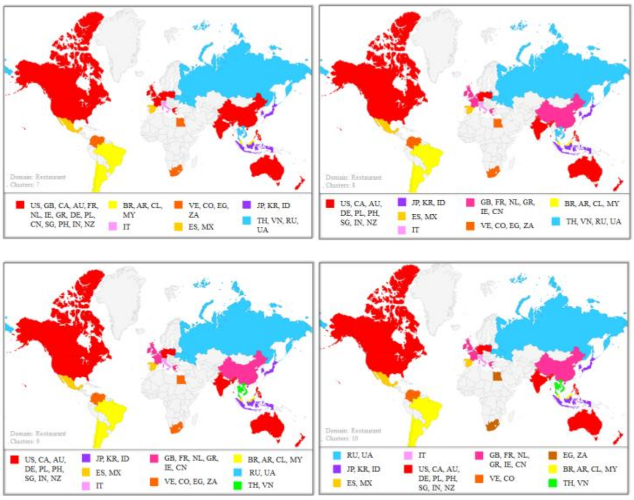

Figure 6　Example Clustering Results for Restaurant Domain

Upon observing the changing process of the clustering result, we have following information: a) While k is set as 8, a few European countries form a separate cluster, which suggests they share more similar attitudes towards the target restaurants, compared to North American countries and English speaking countries in other areas; b) While k is set as 9, RU and UA, TH and VN become two separate clusters, which reflects the location-based cultural effects; c) While k is set as 10, CO and VE, EG and ZA become two separate clusters, which may also demonstrate the location-based cultural effects.

While only focusing on the k=10 situation, the following conclusions can be drawn: a) The location-based cultural effects are quite obvious, e.g. the cluster of BR, CL, AR, the cluster of TH, VN, and the cluster of most of the Western European countries; b) Some of the English-speaking Asian countries are clustered into the same group with North American countries, which suggests that the language-based cultural background may have some effect; c) Comparing to most of the European countries, some countries, such as ES and IT, seem to have quite different opinions for these restaurants, which may suggest that they have special attitudes considering the food culture.

As for the tourism domain, the clustering result when k is set as 8 is shown in Figure 7. Based on this result map, we can arrive at these conclusions: a) People from different countries have quite different opinions for the same attraction. Also, the discrepancy exists among the overall sentiment distributions for each attraction; b) The location-based cultural effects on the user evaluations for world attractions are obvious for some countries, such as the neighboring countries in Europe, North America, and Southeast Asia; c) The language-based cultural effects also exist that most typical English-speaking countries are in the same cluster, including US, GB, AU, and CA; d) While considering the opinions towards world attractions, the boundary between North America and South America is blurring, especially compared to the result in restaurant domain.

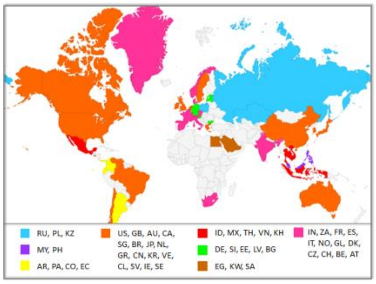

Figure 7　One Example Clustering Result for Tourism Domain

6.　Conclusion

In this research, various methods of sentiment analysis were carried out in restaurant and tourism domain. We used Twitter data of more than 30 languages from more than 30 countries as the dataset, and explored the cultural effects on user evaluations. The proposed approaches were testified to be capable of cross-cultural investigations and transferable among fields.

References

1) Liu, B.: Sentiment Analysis and Subjectivity. J. Handbook of natural language processing, 2, 627-666 (2010)
2) Guo, H., Zhu, H., Guo, Z., Zhang, X., Su, Z.: OpinionIt: a Text Mining System for Cross-lingual Opinion Analysis. In: Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1199-1208. ACM (2010)
3) Le, J., and Yamana, H.: A Comparative Study of User Evaluations for Global Restaurants under the Multi-cultural Background. In: Proceedings of WebDB Forum (2013).
4) Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. J. Foundations and trends in information retrieval, 2(1-2), 1-135 (2008)
5) Go, A., Bhayani, R., Huang, L.: Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report, Stanford, 1-12 (2009)
6) Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: Proceedings of International Conference on Language Resources and Evaluation (2010)
7) Barbosa, L., Feng, J.: Robust Sentiment Detection on Twitter from Biased and Noisy Data. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 36-44 (2011)
8) Cesarano, C., Picariello, A., Recupero, D. R., Subrahmanian, V. S.: The OASYS 2.0 opinion analysis system. In: Proceedings of the AAAI Conference on Weblogs and Social Media (2007)
9) Guo, H., Zhu, H., Guo, Z., Zhang, X., Su, Z.: OpinionIt: a Text Mining System for Cross-lingual Opinion Analysis. In: Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1199-1208. ACM (2010)
10)　Cui, A., Zhang, M., Liu, Y., Ma, S.: Emotion Tokens: Bridging the Gap among Multilingual Twitter Sentiment Analysis. In: Information Retrieval Technology, pp. 238-249. Springer Berlin Heidelberg (2011)
11)　Gao, Q., Abel, F., Houben, G. J., and Yu, Y.: A Comparative Study of Users' Microblogging Behavior on Sina Weibo and Twitter. In: User Modeling, Adaptation, and Personalization, pp. 88-101. Springer Berlin Heidelberg (2012)