2015 IEEE 15th International Conference on Data Mining Workshops

Fun in The Philippines: Automatic Identification and Sentiment Analysis of Tourism-related Tweets

Julia Camille L. Menchavez
Department of Computer Science
University of the Philippines Cebu
Cebu City, Philippines
jcmenchavez@gmail.com

Kurt Junshean P. Espinosa
Department of Computer Science
University of the Philippines Cebu
Cebu City, Philippines
kpespinosa@up.edu.ph

***Abstract*—With the growing use of social media in the Philippines, tourism-related user-generated content is readily available. As a growing hub of tourism and culture, this could be particularly useful to the country. However, a large amount of this data has gone unanalyzed. This study discusses and develops a way that could help bridge that gap using automated tourism-related tweet identification with Support Vector Machines and sentiment analysis with Naïve Bayes. F-scores of 0.943 and 0.81 were obtained by these components respectively, with the overall system obtaining an accuracy of 84%. Mapbox was used for visualization, with tweets plotted based on their geolocations and sentiments. This study can be used as a way of gathering tweets from the Philippines, identifying which could be relevant in terms of tourism information and presenting these in a way that could be useful and easy to understand and interpret.**

Keywords-Natural language processing; Twitter; tourism; topic detection; sentiment analysis

I. INTRODUCTION

A hub of culture, eco-adventure, sandy shores and food, the Philippines draws millions of tourists from around the globe each year [8]. When the Philippine Department of Tourism (DOT) launched its campaign in 2012, the campaign's official hashtag #ItsMoreFunInThePhilippines trended worldwide less than an hour after the launch. With this and other official hashtags, the DOT is able to gain information on tourism sites and activities in the country.

However, the tourism information retrieved is mostly limited to tweets and posts with DOT's official hashtags, while many social media users tweet about their travels in the Philippines and their views on tourism destinations without these hashtags. If tweets can be automatically identified as tourism or non-tourism related, then tourism-related tweets, even those without DOT's official hashtags, can be used to determine travel sentiment, and provide useful information regarding different Philippine tourist destinations.

A. Objectives

With the growing use of social media in the tourism industry, a substantial amount of user-generated content containing tourism information and sentiment is readily available. However, a large amount of this data goes unanalyzed. The research attempts to address this problem in the context of tweets made in the Philippines with the use of natural language processing and machine learning techniques. Specifically, the research aims to do the following:

- Automatically classify tweets from the Philippines as "tourism" or "nontourism"-related based on their contents.
- After identifying tourism-related tweets, determine the polarity of these tweets, indicating whether they are positive or negative.
- Present the data in a clear manner as to show tourism-related tweets and their polarities for different tourist destinations in the Philippines.

II. RELATED LITERATURE

A. Topic Detection and Classification

In their study [22], Young-Woo et al. explored hierarchical clustering and iterative clustering. In the study, the hierarchical methods used achieved an accuracy of around 80% while the performance of the iterative methods fluctuated from 50% to 80% due to their heavy dependence on the initialization of the cluster centers. Wartena et al. [4], as well as Tsur et al. [19] also dealt with clustering with the use of high-frequency words and K-means clustering.

A study by Sriram et al. [18] in 2010 dealt with classifying tweets into different categories using eight features c. After testing, 8F, the approach in the research, obtained a 32.1% improvement in terms of accuracy over the standard Bag-Of-Words approach.

B. Opinion Mining and Sentiment Analysis

Riloff et al. [15] and Jindal et al. [9] explored subjectiveness and comparative sentences. In 2003, Riloff et al. developed a bootstrapping process for distinguishing between objective and subjective sentences and for learning extraction patterns for subjective expressions using high-precision classifiers yielding a precision from 71% to 85%. Meanwhile, Jindal et al.'s study also used machine learning and class sequential rule (CSR) mining to identify comparative sentences. When both CSRs and manual rules were used with SVM and NB, a high F-score was obtained, with NB outperforming SVM.

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.184

IEEE computer society

In a more recent study [5], the Naive Bayes classifier and four others were used to extract small investor sentiment from stock message boards, with the majority vote among the five as the final sentiment. Other studies focused on applying these opinion mining techniques on tweets. Exploring Twitter as a corpus for opinion mining and sentiment analysis, Pak and Paroubek [13] developed a sentiment classification system, training two Naive Bayes sentiment classifiers, one using N-grams as features and the other using POS-tags. Barbosa et al. [2] attempted to clean biased and noisy Twitter data from several sources with subjectivity and polarity detection using two classifiers. Another study [7] enhanced Twitter sentiment learning further through the analysis of hashtags and smileys.

Several studies involved using sentiment analysis of tweets for prediction. Tumasjan et al. [20] predicted elections by profiling political sentiment in twelve dimensions, Asur et al. [1] dealt with movie box office revenues using linear regression and N-grams, and Bollen et al. [3], attempted to predict the stock market with fuzzy neural networks.

C. Tourism

A small number of studies have focused on applying machine learning techniques in the tourism sector. A recent study [11] aimed to create a system that would assist users in understanding tourism opinions on the web by finding and extracting subjective information from reviews in tourism websites. Aspect extraction was performed with the use of frequent nouns and the opinion was determined. Leaning closer to sentiment analysis, several researchers developed a system [14] that would assist tourists' decision-making by classifying comments in various travel websites with aspect classification and polarity identification, achieving an 85% accuracy. Another system [21] used DBPedia, GATE and SentiWordNet to automatically generate recommendations for trip planning.

An even smaller number of studies involving tourism have used Twitter data. In another study, Shimada et al. [17] found a way to measure tourism information likelihood on Twitter using SVM and ten features, including: bag of words, number of nouns and RT (retweet). Another earlier research [16] used seed words to determine the polarity of tweets with Naïve Bayes.

Another similar study by Claster et al. [5] used a keyword-based binary-choice algorithm with the Naïve Bayes algorithm to monitor sentiment of tweets over time regarding unrest in Thailand. They then extended the analysis of tweets to those pertaining to Colombo in Sri Langka and Cancun in Mexico, attaining an accuracy of 41.9%.

Although these studies have ventured into opinion mining with tourism-related tweets, none of them have focused on identifying whether a given tweet was tourism-related or not in the first place. One study [5] gathered tweets with keywords "Bangkok" and "Phukhet", and in another [16], "basic queries" containing names of facilities and events were used. In both cases, the tweets used as data were already assumed to be pertaining to tourism. On the other hand, the studies [18] that did delve into tweet classification focused on segregating tweets into other general categories. So far, there has not been a comprehensive study involving topic detection and opinion mining techniques in the identification and polarity detection of tourism-related tweets, particularly those from the Philippines.

III. METHODOLOGY

A. Dataset

For the identification of tourism and nontourism-related tweets, the dataset consisted entirely of tweets by users in the Philippines. The data used was retrieved from Twitter between January 16 and April 6, 2015. A random sample of these tweets was obtained and manually annotated by the researcher as being tourism and nontourism-related. For Experiment 2 onwards, 1,215 tweets were utilized, 513 of which were labeled tourism- related and 702 labeled nontourism-related.

For sentiment analysis, a dataset consisting of 3,920 tweets retrieved on February 2015 was used, evenly divided into positive and negative tweets. The positive tweets were obtained by querying Twitter for tweets containing a ":)", while the negative tweets were obtained by querying ":(".

B. Preprocessing

Each tweet was first cleaned. Tweets often contain links (in the form of http://t.co/iLb7Q0jTwr) and as these links are not utilized in the study, these were removed from the tweets. Several punctuation marks such as "!", "..." and "," were also removed. Tweets then underwent preprocessing in three steps: (1) tokenization, (2) conversion to lowercase, (3) lemmatization.

C. Identification of Tourism/Nontourism-related Tweets

1) Feature extraction

Several methods were experimented with for extracting features from the tweets to be used for training the classifiers.

a) Bag-of-words: A vector containing the set of unique tokens over all the tweets that comprised the training set was created. For each tweet, a dictionary was created, with the tweet's unique tokens or words as keys and their frequencies as the values. Words in tweets that are found in NLTK's Stopwords Corpus [23] for the English language were not added as unique tokens.

b) Top words: With tourism-related tweets, there are several words that are constantly repeated. Thus, another feature extraction method used involves only the top words obtained in the training set. In order to obtain the "top words" list, the top 15 frequently occurring words in the tourism dataset were obtained, filtering out stop words as in the bag-of-words method. Instead of a vector containing the set of unique tokens over all the training set tweets, the vector in the top words method contains only these top words.

top_words = ['travel', 'vacation', 'city', 'itsmorefuninthephilippines', 'travel', 'boracay', 'philippine', 'view', 'day', 'beach', 'morning', 'resort', 'good', 'cebu', 'island']

c) Top words with weights: In this case, all the words in the tweet were still taken into consideration, although the "tourism-related" words were given a larger weight. For each tweet, a dictionary was created, with the tweet's unique words as keys and their frequencies as the values. For words found in the "top words" list, extra weight was added.

Sample tweet with corresponding features and an added weight of 0.5:

Beach with the fambam! Love boracay and the view of the beach{'beach': 3, 'boracay': 1.5, 'fambam': 1, 'love': 1 'view': 1.5 }

d) Top words scraped from TripAdvisor with weights: In this method, words scraped from TripAdvisor [26] were used instead of the top 15 frequently occurring words in the training set. To obtain the words, Scrapy [25] was used to crawl TripAdvisor, specifically the top destinations and attractions of each of the top 20 places listed in "Things to do in Philippines". For each of the items on the list, the names of the destinations listed were obtained, along with the contents of the top comments. These destination names, as well as the top 100 frequently occurring words in the comments, were then used as the "top words", and words obtained from the tweets that were found on this list were given larger weight.

e) Words generated by LDA from TripAdvisor: For the last feature extraction method, the details scraped from TripAdvisor comments on different Philippine destinations and attractions were still used. A bag-of-words dictionary was generated using the content obtained from these comments. This dictionary was then passed into Gensim's [24] Latent Dirichlet Allocation model in order to obtain the topics in the content, along with the corresponding words for each of the topics. The number of topics for LDA to determine was set to 100. After undergoing preprocessing and cleaning, these words were then stored and used as the "top words" in feature extraction.

2) Classification

For the classification of tweets as "tourism" and "nontourism", three different supervised learning classifiers were utilized to obtain the optimum results. In training and assessing the different classifiers, each of the classifiers underwent 10-fold cross validation. In 10-fold cross validation, the training set is divided into 10 folds, each of which is used as the validation set at one point. The classifier configuration that obtains the best accuracy is selected. The accuracy is computed as the number of examples predicted correctly / the number of total examples. Other evaluation measures used to assess the models are precision, recall and f-score:

precision = #truepos / (#truepos + #falsepos) (1)

recall= #truepos / (#truepos + #falseneg) (2)

fscore = 2[(precision*recall)/(precision+recall)] (3)

The system experimented with three different classifiers: (1) Logistic Regression, (2) Naive Bayes, (3) Support Vector Machines.

a) Logistic Regression: With the logistic regression model, the probability of a tweet being classified positively (in this case, as "tourism"-related) is determined through the following hypothesis:

$$h_\theta x = g\left(\theta^T X\right) \text{ where } g(z) = 1/(1+e^{-z}) \quad (4)$$

and X is comprised of the tweet's features.

If a tweet obtains a result above 0.5, it is classified as "tourism"-related, whereas those that receive a lower result from the function are classified as "nontourism"-related.

b) Naive Bayes: For the study, the Nai ve Bayes classifier by NLTK [23] was used. With the Naive Bayes classifier, in order to choose a label ("tourism" or "nontourism") for a new example, a prior probability of each label is first generated by obtaining the frequency of each label in the training set. This probability is then combined with the contribution of the features in the example, forming the likelihood estimate for each of the labels. The label with the highest likelihood estimate is then the assigned label for the new example.

In order to predict the correct label of a tweet, the probability that the tweet will have a certain label given its set of features is calculated as such:

$$P(label \mid features) = \frac{P(label) * P(f1 \mid label) * ... * P(fn \mid label)}{P(features)} \quad (5)$$

c) Support Vector Machines: Unlike logistic regression, the output of SVM is an actual prediction of either 0, signifying nontourism and tourism respectively in this case. 0. In the process, the SVM classifier seeks to minimize a cost function similar to that of logistic regression with C initially set to 1.0 and a maximum iteration number of 1,000:

$$\min_\theta C\sum_{i=1}^{m}[y^i \text{cost}_1\left(\theta^T X^i\right) + \left(1-y^i\right)\text{cost}_0\left(\theta^T x^i\right)] + \frac{1}{2}\sum_{j=1}^{m}\theta_j^2 \quad (6)$$

D. Sentiment Analysis:

For sentiment analysis, tweets classified as being tourism-related were further classified as either expressing a positive or negative sentiment.

1) Feature extraction

a) Parts-of-speech tags: As Pak and Paroubek [13] discovered, certain POS-tag distributions are more likely to be found in positive sentiments and others in negative sentiments. Thus, the study obtained the POS-tags of tweets using the NLTK POS-tagger [23] which were then used as features for the supervised learning models.

b) N-grams: Another feature set consisted of constructed N-grams out of consecutive words in the given tweets. As negations play a role in sentiment expression, an extra consideration for negations was made, patterned from a previous study [13]. For each negation term found in the tweets such as "no" or "not", the negation term was attached to the word that it followed or preceded.

2) Sentiment analysis

For predicting the sentiments of tweets, the study experimented with the use of NLTK'S[3] Naive Bayes classifier, with the tweets' POS-tags and the N-grams as features for prediction.

E. Visualization

For the visualization portion, Mapbox was used to plot the tweets on a map of the Philippines, based on their geolocations. Using Leaflet clustering, the tweets are automatically clustered and re-clustered upon zooming in to and out of the map. For each cluster, red and green circles are created. The sizes of these circles depend on the percentage of positive and negative tweets, as per their classified sentiments, in that cluster. The larger the green circle, the more positive tweets there are, and the larger the red circle, the more the negative tweets there are in the cluster. This allows for the easy identification of general sentiment in a particular area when it comes to tourism-related views. As new tweets classified as tourism-related arrive, the map is updated to reflect these additions.

IV. EXPERIMENTS AND RESULTS

A. Skewed Dataset

In the process of gathering and labeling tweets to be used as training data, there was initially a large disparity between the number of "tourism" and "nontourism" tweets. Given the skewed data set (number of tourism-related tweets = 65, number of nontourism-related tweets = 1,477), the bag-of-words used as features and no-cross-validation, the accuracies of the three classifiers in 3 runs (ordered from lowest to highest accuracy) are shown in Table 1.

TABLE I. ACCURACIES WITH A SKEWED DATASET

Runs / Classifiers	Logistic Regression	Naive Bayes	SVM
1	95.89%	60.47%	96.98%
2	96.33%	66.73%	97.13%
3	97.62%	69.33%	97.41%

During error analysis, it was observed that the abilities of logistic regression and SVM in detecting tourism-related tweets were actually poor, despite their high accuracies. Due to the dataset being skewed highly in favor of "nontourism"-related tweets, the logistic regression and SVM models merely predicted tweets as "nontourism", and because almost all the tweets in the test set were likely to be "nontourism", the two models achieved higher accuracy.

B. More tourism-related data with cross-validation

To improve the performance of the classifiers, more tourism-related tweets were added and several nontourism-related tweets were removed in order to make the dataset less skewed. For the following experiments, 1,215 tweets were utilized, 513 of which were tourism-related and 702 were nontourism-related. In addition to this, 10-fold cross validation was applied for better results.

After error analysis with the previous models, particularly Naïve Bayes, further pre-processing was also done to clean the tweets and remove links, certain punctuation marks, etc.

TABLE II. ACURRACY, PRECISION, F-SCORE, RECALL WITH EVEN DATASET

	Logistic Regression	Naive Bayes	SVM
Accuracy	0.897	0.823	0.872
Precision	0.967	0.755	1.0
Recall	0.714	1.0	0.762
F-score	0.822	0.86	0.865

As shown in Table 2, Naive Bayes was able to obtain the highest recall, signifying that it was able to correctly label all tweets that were in fact tourism-related. However, it labeled certain nontourism-related tweets such as "Early morning blessing." and "Eat your lunch na po" as tourism-related. Most of the tweets that were mislabeled contained words that can often be used in tourism-related contexts, such as the words "morning" and "eat" in the tweets above.

C. Using top words as features

TABLE III. ACCURACY, PRECISION, RECALL, F-SCORE WITH TOP WORDS AS FEATURES

	Logistic Regression	Naive Bayes	SVM
Accuracy	0.872	0.811	0.786
Precision	1.0	1.0	1.0
Recall	0.725	0.682	0.725
F-score	0.841	0.811	0.841

When just the top 15 frequently occurring "tourism" words were used as features instead of the entire bag-of-words, all three models demonstrated a decrease in F-score. Although they all received high precisions, they obtained small recalls. After going through error analysis, the reason for this was that because the top “tourism” words were used as features, most likely those tweets containing those words would be predicted as “tourism”-related. Thus, when predicting a tweet as “tourism”-related, the classifiers could almost be certain that the prediction was correct. However, using this top tweets list means that tweets without these words are most likely labeled as “nontourism”. The issue with this is that there are many tourism-related tweets that in fact do not contain those words.

D. Using top words with weights as features

As shown in Table 4, the F-scores of the different classifiers increased and their recalls increased as well when top words with weights was used. Although the weights still had an effect in somehow making the classification process “stricter”, as it made tweets containing the top words more likely to be classified as “tourism”-related the effect was not as great as when only the top words were used as features.

TABLE IV. ACCURACY, PRECISION, RECALL, F-SCORE WITH TOP WORDS WITH WEIGHTS AS FEATURES

	Logistic Regression	Naive Bayes	SVM
Accuracy	0.848	0.812	0.889
Precision	0.971	0.764	0.971
Recall	0.825	1.0	0.791
F-score	0.892	0.866	0.872

Initially, an added weight of 0.5 was chosen because it was halfway between having no added weight and having an added weight of 1. As shown in Figure 1, the classifiers performed better when there was a lower added weight of 0.5 instead of 1 as an additional added weight, perhaps because of the same effects that using just the top words had on the classifiers.

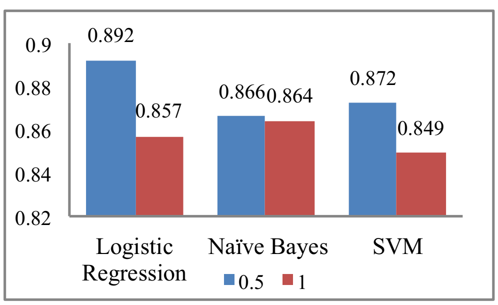

Figure 1. F-scores of different weights

E. Using top words from TripAdvisor with weights as features

In an attempt to broaden this scope with the use of more “tourism- related” words, especially those likely to be expressed by users, words were scraped from destinations and comments on these destinations on the tourism site, TripAdvisor [26].

TABLE V. ACCURACY, PRECISION, RECALL, F-SCORE WITH TOP WORDS FROM TRIPADVISOR WITH WEIGHTS AS FEATURES

	Logistic Regression	Naive Bayes	SVM
Accuracy	0.905	0.868	0.877
Precision	0.974	0.81	0.972
Recall	0.822	0.959	0.796
F-score	0.892	0.879	0.875

With more words being used to denote tourism-relatedness given larger weight than other words, more tourism-related tweets containing these words were correctly classified as tourism-related, as demonstrated by the increase in recall shown in the table above.

F. Using TripAdvisor topics determined by LDA with weights as features

When LDA was used to obtain 100 topics and their related words from the TripAdvisor comments on tourist attractions in the Philippines, there was a further increase in F-scores.

TABLE VI. ACCURACY, PRECISION, RECALL, F-SCORE WITH TOP WORDS FROM TRIPADVISOR GENERATED BY LDA WITH WEIGHTS AS FEATURES

	Logistic Regression	Naive Bayes	SVM
Accuracy	0.868	0.844	0.934
Precision	0.971	0.851	0.971
Recall	0.825	0.976	0.917
F-score	0.892	0.909	0.943

As shown in Table 6, this was particularly true for the Naïve Bayes and SVM classifiers. Adding more of these commonly-used tourism-related topics and words obtained from TripAdvisor allowed the classifiers to correctly classify even more tourism-related tweets in particular, as shown by their increase in recall.

Except for SVM which obtained a slightly lower value, the increased precision of the classifiers also showed that the tweets they were predicting as tourism-related were more likely to actually be tourism-related tweets. As a whole, SVM outperformed the other classifiers.

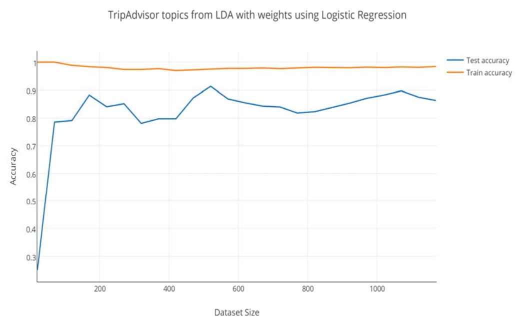

Figure 2. Logistic Regression training and testing accuracy

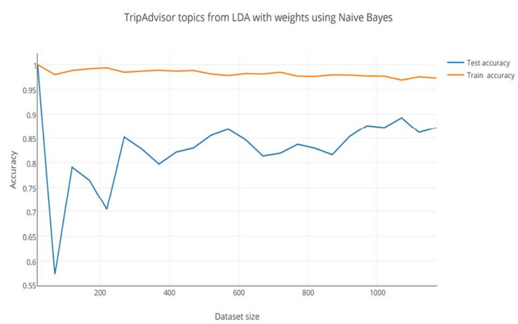

Figure 3. Naïve Bayes training and testing accuracy

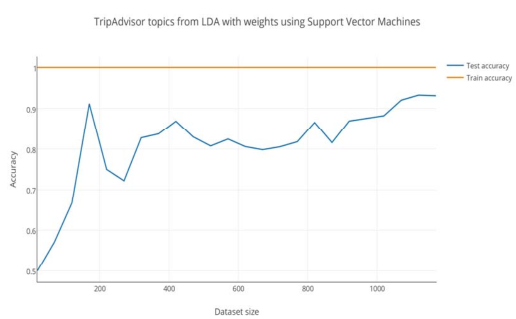

Figure 4. SVM training and testing accuracy

G. *Sentiment analysis using N-grams and Parts-of-speech tags*

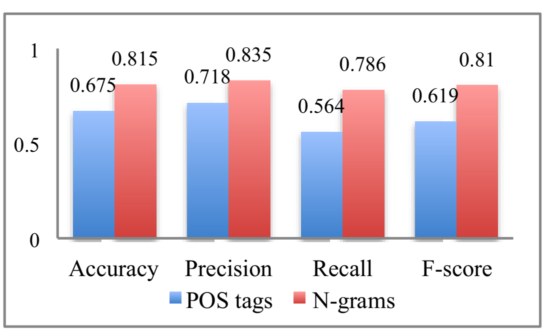

Figure 5. Accuracy, precision, recall with N-grams and POS tags as features

The first experiment using POS tags and their frequencies as features yielded a relatively low accuracy, precision, recall and F-score as shown in Figure 5 above. Meanwhile, using N-grams as features allowed the Naïve Bayes classifier to perform better. For N-grams, it was found that the use of bigrams in particular yielded the highest accuracy at 81.51%. The addition of the words “no” and “not” to their preceding and succeeding words were also shown to have positive contributions on the performance of Naïve Bayes, as the use of bigrams without these modifications yielded lower accuracies and F- scores.

H. *System Evaluation*

For the evaluation of the entire system, 50 random, manually labeled tourism-related tweets were selected, along with 50 random, manually labeled nontourism-related tweets. The tweets underwent the entire process, first undergoing preprocessing, followed by feature extraction and classification to identify whether they were tourism-related or not. The tweets that were labeled tourism-related then underwent further classification to determine whether these held positive or negative sentiments.

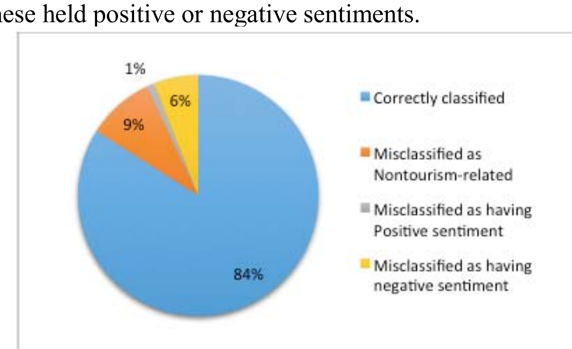

Figure 6. System evaluation results

After error analysis, it was shown that classifier had problems identifying short tweets or tweets with misspellings as being tourism-related. Some of the tweets used for the system evaluation, such as “Horseback riding” and “Ziiiiiiiipline!!!” fell into that category. However, there were also several tweets that did not fall into the category that the classifier had trouble with, such as "This calls for a quick dip baby! Beach life! #WhenInGuimaras #Alubihod #IslandHopping @ Raymen." In this case, although the word “beach” and most especially the hashtags conveyed tourism-related activity, dealing with hashtags comprising of more than one word and analyzing these words was not part of the study.

Several of the errors also occurred during the sentiment analysis portion. Tweets with dual sentiments such as “View is great, but bad service. #summer #travel” were misclassified. The rest of the tweets were misclassified as having negative sentiments. Although some of the tweets clearly held positive sentiments, the other misclassified tweets merely stated where the users were or what they were doing, such as "Beach #subic #minibreak #summer #vacation #myboo @ Arizona Beach Resort, Subic Bay, Philippines."

Upon also looking through tourism-related tweets that were classified as having positive sentiments, while most of the tweets in fact did explicitly state their positivity, several of these also just stated activities, without necessarily showing whether the user was happy with what he or she was doing or not. Although it could be said that the positive sentiment is implied by the fact that the users tweeted about their experience, there is no explicit or objective classification for these tweets.

V. Conclusions and Recommendations

After experimenting with different features and classifiers, the system was able to automatically classify tweets as tourism-related or non-tourism related based on their contents. The use of topics and subsequent words generated by using Latent Dirichlet Allocation on TripAdvisor comments for tourism attractions and spots in the Philippines as features was shown to yield the best performance for all three classifiers. In particular, the SVM classifier had the best performance, achieving an accuracy of 93.42% and an F-score of 0.943. In this setting, the topics and words obtained by LDA had an added weight of 0.5, giving them extra importance in comparison to other words in the bag-of-words model.

However, although SVM obtained high accuracies and F-scores, after studying the tweets that were misclassified, several issues that most of the classifiers were unable to deal with were revealed. For example, certain tweets such as "Headin to Laguna latur", "Off to subic" and "heading to vigan" were constantly misclassified. A possible reason for this could be that the tweets are too short, thus the classifiers could not base on many present features in order to make correct classifications.

On the part of sentiment analysis, the system was able to determine the polarities of tweets, whether they contained positive or negatives sentiments, using the Naïve Bayes classifier. After several experiments, the use of N-grams, particularly bigrams, yielded better results in comparison to using POS tags as features, obtaining an F-score of 0.81. The addition of the words "no" and "not" to the words they preceded and succeeded during the construction of the bigrams was also shown to positively contribute to the performance of the classifier.

When the system as a whole was tested on 100 random tweets, it managed to attain an accuracy of 84%. Some of the errors occurred during tourism-relatedness identification, as the classifier was unable to correctly classify several short and misspelled tweets. Meanwhile, the sentiment analysis classifier struggled with tweets that did not explicitly state positive or negative sentiments, but rather neutral ones.

Although the performance of the tourism-relatedness classifier in particular was relatively good, it has not yet been tested on a larger dataset. The dataset that the classifier was trained and tested on was small, as it was difficult to obtain tourism-related tweets. In future studies, the classifier should be trained and tested on larger datasets to fully assess its performance. Also, in extracting the features and constructing the dictionary, some words were given more weight than others. While the values of certain words in the features dictionary were their frequencies, values of certain words, in the case of the best result, the top 100 topics and related words from TripAdvisor comments, were increased by more than 1 each time they occurred in the tweet. For the case of the experiments, the added weight used was 0.5, as it was shown to be better than utilizing no added weight at all, or an added weight of 1. In terms of this added weight, further work still has to be done to determine the optimum weight for these special or top words. In addition to this, the use of N-grams in tourism or nontourism-related classification should be studied as it may contribute to better determining whether a tweet is tourism-related or not. For example, during error analysis, some misclassified tweets contained tourism-related phrases such as "off to" and "best meal". These tweets may be correctly classified if a bigram model is used.

Several issues encountered by the classifiers as revealed during analysis also need to be addressed. For example, some of the tweets that were misclassified were the tweets that expressed tourism-related thoughts but were very short or the tourism-related thoughts were expressed in hashtags, such as #WhenInGuimaras and #IslandHopping. Separating the words in the hashtags and analyzing them could result in better classification performance. As important information beyond tourism-relatedness can be gathered from tweets such as these, these issues should be looked into and addressed. The same could also be said for tweets containing words that were misspelled or intentionally shortened or abbreviated. For example, "Ziiiiiiiipline!!!" should be normalized to the word "zipline". In sentiment analysis, objective tweets or tweets holding no or neutral sentiment should also be handled as these were shown to lower the performance of the sentiment analysis classifier. In addition to this, tweets containing both positive and negative sentiments should also be handled perhaps by treating these as two separate tweets with opposite sentiments.

In the application portion, as suggested by a representative from DOT, it may be better to focus on a specific city or province rather than the entire country; and identify and retrieve tourism-related aspects through the tweets about the city or province that may be helpful to the local government.

References

[1] Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492–499). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5616710

[2] Barbosa, L., & Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (pp. 36–44). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1944571

[3] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8.

[4] Christian Wartena, & Rogier Brussee. (2008). Topic Detection by Clustering Keywords.

[5] Claster, W. B., Cooper, M., & Sallis, P. (2010). Thailand -- Tourism and Conflict: Modeling Sentiment from Twitter Tweets Using Naive Bayes and Unsupervised Artificial Neural Nets(pp. 89–94). IEEE. doi:10.1109/CIMSiM.2010.98

[6] Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science, 53(9), 1375–1388. doi:10.1287/mnsc.1070.0704

[7] Davidov, D., Tsur, O., & Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (pp. 241–249). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1944594

[8] Department of Tourism(2013). Philippines Travel and Tourism Statistics 2013 [Brochure]. Makati, Philippines.

[9] Jindal, N., & Liu, B. (2006). Identifying comparative sentences in text documents. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 244–251). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1148215

[10] Kim, S.-B., & Kim, D.-Y. (2011). Travel information search behavior and social networking sites: The case of US college students. Retrieved from http://scholarworks.umass.edu/gradconf_hospitality/2011/Presentation/80/

[11] Marrese-Taylor, E., Velasquez, J. D., & Bravo-Marquez, F. (2013). Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web (pp. 261–264). IEEE. doi:10.1109/WI-IAT.2013.193

[12] Milano, R., Baggio, R., & Piattelli, R. (2011). The effects of online social media on tourism websites. In ENTER (pp. 471–483). Retrieved from http://www.iby.it/turismo/papers/baggio_socialmedia.pdf

[13] Pak, A., & Paroubek, P. (2010). Pak and Paroubek (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining.pdf. Retrieved May 22, 2014, from http://incc-tps.googlecode.com/svn/trunk/TPFinal/bibliografia/Pak%20and%20Paroubek%20(2010).%20Twitter%20as%20a%20Corpus%20for%20Sentiment%20Analysis%20and%20Opinion%20Mining.pdf

[14] Palakvangsa-Na-Ayudhya, S., Sriarunrungreung, V., Thongprasan, P., & Porcharoen, S. (2011). Nebular: A sentiment classification system for the tourism business. In Computer Science and Software Engineering (JCSSE), 2011 Eighth International Joint Conference on (pp. 293–298). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5930137

[15] Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In Proceedings of the 2003 conference on Empirical methods in natural language processing (pp. 105–112). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1119369

[16] Shimada, K., Inoue, S., Maeda, H., & Endo, T. (2011). Analyzing Tourism Information on Twitter for a Local City (pp. 61–66). IEEE. doi:10.1109/SSNE.2011.27

[17] Shimada, K., Inoue, S., & Endo, T. (2012). On-site Likelihood Identification of Tweets for Tourism Information Analysis (pp. 117–122). IEEE. doi:10.1109/IIAI-AAI.2012.32

[18] Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., & Demirbas, M. (2010). Short text classification in twitter to improve information filtering. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 841–842). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1835643

[19] Tsur, O., Davidov, D., & Rappoport, A. (2010). ICWSM-A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews. In ICWSM.

[20] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM, 10, 178–185.

[21] Varga, B., & Groza, A. (2011). Integrating DBpedia and SentiWordNet for a tourism recommender system. In Intelligent Computer Communication and Processing (ICCP), 2011 IEEE International Conference on (pp. 133–136). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6047856

[22] Young-Woo, S., & Sycara, K. (2004). Text Clustering for Topic Detection.

[23] NLTK Project. (2015). Natural Language Toolkit. http://www.nltk.org

[24] Radim. (2015). Gensim. http://radimrehurek.com/gensim/

[25] Scrapy. (2015). Scrapy. http://scrapy.org/

[26] TripAdvisor LLC. (2015). TripAdvisor Philippines. http://www.tripadvisor.com.ph