Extraction of Aspect Categories for Sentiment Analysis towards Technical and Vocational Education and Training (TVET) in Malaysia Using Topic Modelling Approach

Nurul Ashikin Abd Samad
Politeknik Mersing
nurulashikin@pmj.edu.my

Roliana Ibrahim
Universiti Teknologi Malaysia
roliana@utm.my

Abstract

This study highlights the neccesity of the analyses on public opinion through more transparent approach in order to observe the public perception towards Technical and Vocational Education And Training (TVET) in Malaysia. Aspect-based sentiment analysis (ABSA) enables the identification the aspect or features that reflect positive and negative sentiments in sentences. However, one of the challenges in ABSA is the extraction of aspect from the dataset. In this study, topic modelling is used to identify the aspect words from tweets related to TVET in Malaysia. One of the topic model that is proven useful in extracting topics from corpus of lesser known domain is Latent Dirichlet Allocation (LDA). LDA extracted aspects terms were compared with manually labeled aspect terms. The study shows means of accuracy at 0.96 while the means of precision is acceptable at 0.70 respectively after applying LDA in extracting the aspec categories. It is hope that by harvesting opinions from social media will enable the discovery of the most frequent aspects about TVET in Malaysia that often mentioned by the public. This research is hoping to contribute towards the improvisation of Malaysia's TVET in the future.

Keywords: aspect-based sentiment analysis, LDA, topic model, aspect extraction

1.0 Introduction

Poor perception towards technical and vocational education and training (TVET) was common among public in Malaysia (Lam & Hassan, 2018). In 2018, TVET and Industry Commission has been established to tackle the issue in order to make TVET as the favourable choice in the future, including to ensure the certifications received by TVET graduates are equal to other academic programmes, alongside with competitive salaries when they enter the workforce.

The expectations from technical and vocational education stream are not limited to the development of academic and technical knowledge among its students but also to help them acquire high employability skills. Public opinion and review towards TVET institutions are one of the excellent ways for improving the Malaysia TVET education in the future. By harvesting these opinions and reviews from social media; enable the discovery of relationship between TVET and its stakeholders such as parents and prospective students.

Public sentiments towards TVET in Malaysia were mentioned in the study by Ismail & Zainal Abidin (2014), Esa & Kannapiran (2014) & Esa, Razzaq, Masek & Selamat (2009). These studies involved traditional methods of data collection that targeted only certain groups of people and using surveys, questionnaires and interviews. Hence, there is a need to address the limitation of this research method.

At the widespread usage of social media, many sentiment analysis studies have been conducted in Malaysia utilizing social media data such as researches on employability, tourism, business, marketing, politic, health and public action (Shahid Shayaa et al., 2018; Ainin, Feizollah, Anuar & Abdullah 2020 & Drus & Khalid 2020). However, data from these platforms were not yet utilized to observe the sentiments towards TVET in Malaysia.

Due to this circumstance, the main research question was asked; *what are the sentiments towards TVET in Malaysia if an aspect-based sentiment analysis is performed on tweets related to it*?

Schouten & Frasincar (2016) in their extensive survey on aspect-based sentiment analysis has been highlighting three (3) important tasks that need to be addressed in performing sentiment analysis at aspect level; aspect identification, aspect sentiment classification, and sentiment aggregation. Among the three main tasks, the most crucial part is to identify and categorize the aspect especially the implicit aspects. Due to the nature of data collected from Twitters which consisted of unlabeled mixture of topics has led to the following research questions, can the topic modeling methods be used to extract and categorize the aspects that influence this sentiment and what is the suitable method for this study?.

2.0 Research background

Aspect level or aspect based sentiment analysis performs the finer-grained analysis. It is established on feature-based opinion mining that identifies the features values of the opinion and then summarizes the results of that opinion. Aspect-based sentiment analysis will extract aspects or features from the text and sentiment values are assigned to them (Shama & Dhage, 2018). There are three (3) important tasks that need to be addressed to perform sentiment analysis at aspect level; feature extraction, sentiment classification and sentiment aggregation (Schouten & Frasincar, 2016).

The goal of aspect extraction task is the identification of all the discriminative aspect terms provided in each sentence. It is common wherein a sentence may consist of multiple aspects and all of this aspect needs to be extracted. There are two (2) types of aspects, explicit and implicit. Liu (2015), suggested four (4) main approaches in extracting explicit aspects, extraction established on frequent nouns and noun phrases (frequency-based), extraction by exploiting opinion and target relations (syntax-based), extraction using supervised learning and extraction using topic modeling (unsupervised learning). Extensive surveys on aspect extraction approach by Tubishat, Idris & Abushariah (2018) and Rana & Cheah (2016) have provided meaningful insights for this study.

Aspect-based sentiment analysis can be classified in different ways; one of them is based on the technique used in classifying the polarity of the sentiment. In most studies the classification can be achieved in three different approaches; the machine learning approach, lexicon-based approach and hybrid approach. Bayesian Network, Naïve-Bayes (NB), Support Vector Machine (SVM) and Maximum Entropy (ME) classifier are among probabilistic classifiers that can be used to calculate the degree of sentiment polarity of extracted aspect terms (Saberi & Saad, 2017).

The result of the analysis can provide a better understanding of the topic of interest to the public. Thus, sentiment aggregation is a vital task in delivering complete contextual information in regard to sentiment analysis. There are two suggested aggregation level, topic-wise and sentiment-wise.

According to Schouten & Frasincar (2016), the whole corpus is aggregated naturally when methods that applied topic models are used, thereby computing sentiment for each topic or aspect based on all the reviews. They are also suggesting the majority of approaches in unsupervised feature extraction are using topic modelling specifically LDA-based model.

Topic modelling approach was used in aspect extraction tasks in this study because it is one of the most commonly used extraction methods for unsupervised or semi-supervised aspect extraction techniques and has been extensively reviewed in many studies including by Alghamdi & Alfalqi (2015).

3.0 Research objectives

The objective of this study are:

a. to determine the number of topics that can be extracted using topic modeling from the collected corpus
b. to extract the aspect terms that appeared in each topics to represent the aspect categories
c. to analyse the result from the LDA approach in extracting the aspect categories from tweets related to TVET in Malaysia.

4.0 Methodology

To achieve the objectives, two (2) key activities have been performed. The first activity was topic extraction by using LDA and followed by aspect categories extraction. LDA's topic extraction has produced series of keywords which were used in aspect categories extraction. And finally, based on the categories; the aspect terms were identified for classification. The activities in this phase are shown in the dotted box in Figure 1.

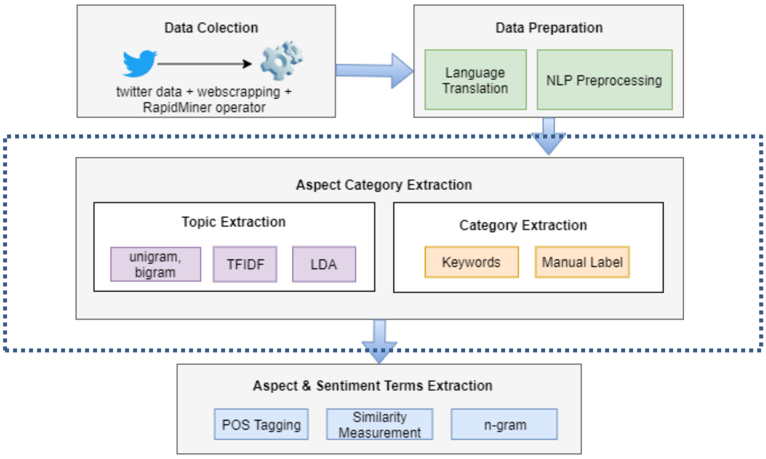


Figure 1: Activities in aspect extraction

4.1 Topic extraction

All the tweets crawled in the month of March 2019 were stored in a single comma delimited (csv) file and has gone through various pre-processing stage. The limited Malay resource for natural language processing tasks has been the motivation for this study to conduct translation process to any non-English tweets. This has proven beneficial in a study by Rahmat (2017) whereit showed significant contribution to the final result of the study. The processed dataset was used to generate document-term matrix (DTM) as an input to the LDA model. In this study, an established package; *doc2bow()* from Gensim library was used to create the DTM or the word vectors. Text feature (unigram and bigram) and term weighting were also applied for the DTM. The term weighting scheme, Term Frequency-Inverse Document Frequency (TF-IDF) determine the weight of a term by two factors:

a) how frequent the term j occurs in document i (the term frequency $tf_{i,j}$)
b) how frequent a term occurs in the whole document collection (the document frequency $df_{i,j}$).

As for the LDA model five (5) parameters were used; alpha and beta hyperparameter, data set, number of topics and number of iterations. The alpha and beta hyperparameters controlled the selection of topics and the discrete distribution respectively over possible topics. A high alpha value resulted in tweets being more similar in terms of what topic they contain, and similarly a high beta resulted in subjects being more similar in terms of what words they contain. The visual representation of LDA model is shown in Figure 2.

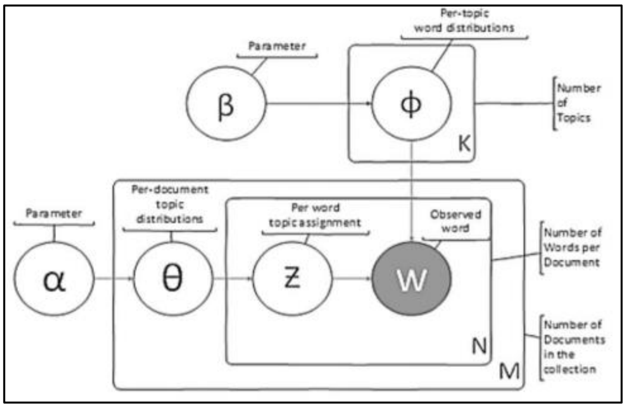


Figure 2: Latent dirichlet allocation (LDA) for topic modelling (Montenegro, Ligutom, Orio & Ramacho 2018)

In order to determine the number of topics to represent the aspect categories that can be produced from the dataset, several experiments have been performed. They were conducted to observe which of the models have generated the most coherence topic model. Each of the experiments were done on 5 runs using LDA package from Gensim library where the mean scores of the coherence values were recorded. The best number of topics were determined by using topic coherence scoring. The scores were summarized using Equation (1) where the pairwise score for frequently occurring words, wi…wn in each topic were computed and aggregated.

$$\text{Coherence} = \sum_{i<j} \text{score}(w_i, w_j) \quad (1)$$

The topic coherence score was measured using UMass coherent scoring and number of topics with the highest score is selected (Röder, Both, & Hinneburg, 2015). The UMass coherence score is computed by using the Equation (2) where $D(w_i, w_j)$ counts the number of documents containing words w_i and w_j and $D(w_i)$ counts the number of documents containing w_i.

$$\text{score}_{\text{UMass}}(w_i, w_j) = \log \frac{D(w_i, w_j) + 1}{D(w_i)} \quad (2)$$

Four (4) experiments were conducted to find out which of the textul features combined with LDA can produced the most coherence topic model. The textual features for the models were selected as shown in Table 1.

Table 1: Textual features selected for aspect category extraction using LDA

No	Features Name	Description
1	E1	unigram + LDA
2	E2	bigram + LDA
3	E3	unigram + TFIDF + LDA
4	E4	bigram + TFIDF + LDA

4.2 Aspect category extraction

A number of topics were generated in accordance with the respective keywords by the LDA model. The keywords were evaluated on the basis of the similarity and compared with manually labeled aspect categories to ensure the existing aspect categories were not disregarded

In this process, the keywords and manual label were converted into word vectors. Word vectors were obtained by using pre-trained GLoVE model provided by spaCy (en_vectors_web_lg) which consists of 300-dimensional word vectors (Pennington, Socher, & Manning, 2014). The similarity function in spaCy method enabled the comparison between one token with another and established its similarity.

Table 2 is showing an example of result obtained by using this function where the aspect category that was manually labelled was compared with the keywords generated by LDA to identify which of the keywords were corresponding to the aspect category. Precision, recall, and f-measure were used to evaluate the result of this comparison based on the similarity scores.

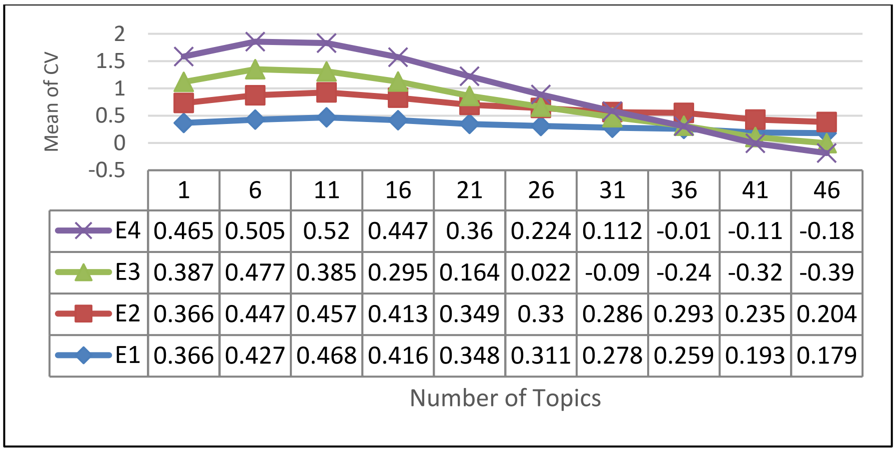
Table 2: Example of similarity score produced by spaCy

Manual Category	Keywords	Predicted Category	Similarity Score
student	vocational, graduate, skill, malaysia, student, study, professional, education, accredit, effort	student	1.00
course	vocational, skill, student, stream, malaysia, study, graduate, course, youth, field	course	1.00
admission	vocational, graduate, study, student, skill, higher, employment, malaysia	vocational	0.542104
skill competency	vocational, student, study, skill, malaysia, graduate, industry, education, course, want	skill	1.00
employability industry	industry, vocational, malaysia, skill, graduate, study, education, salary, student	industry	1.00

5.0 Analysis and discussion

The performance of four (4) topic models used in the experiments was visualized in Figure 3. The figure also shows that the highest means of topic coherence value was recorded from E4 where the means was 0.52 with eleven (11) topics found from the dataset. The highest coherence score for E1 was 0.47 and E2 (0.46) was also when the number of topics discovered were eleven (11) topics. However, in E3, the highest coherence was scored when the number of topics were 6.

The figure shows that higher the number of the topics, the lower coherence scores for the models. This mainly contributed by the limited number of documents used in this study thus allowed the overlapping of terms used in the topics and disabling the larger independence from one topic with another.



	1	6	11	16	21	26	31	36	41	46
E4	0.465	0.505	0.52	0.447	0.36	0.224	0.112	-0.01	-0.11	-0.18
E3	0.387	0.477	0.385	0.295	0.164	0.022	-0.09	-0.24	-0.32	-0.39
E2	0.366	0.447	0.457	0.413	0.349	0.33	0.286	0.293	0.235	0.204
E1	0.366	0.427	0.468	0.416	0.348	0.311	0.278	0.259	0.193	0.179

Figure 3: Means of coherence values per number of topics

Topics that were identified from the model were visualized using word clouds to show the most frequents keywords for each of the topics. The highest percentage of tweets were categorized under Topic 9 (15.34%) while lowest is Topic 6 (5.45%). Figure 4 is showing the word cloud for both topics. Distribution of tweets based on LDA generated topics is shown in Figure 5.

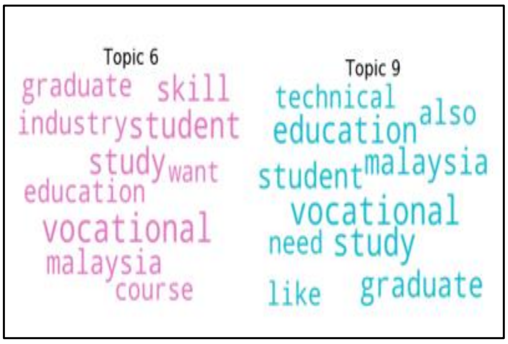


Figure 4: Word cloud for Topic 6 and 9

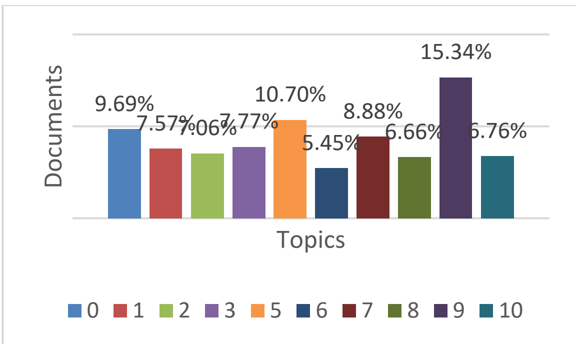


Figure 5: Topic-document distribution

Means of accuracy for all aspects were showing relatively high score with 96% and the means of precision is acceptable at 0.70 respectively. While the means for recall and F-measure are 0.43 and 0.49. There were four (4) aspect categories showing high precision at 100%; while most categories were indicating accuracy at more than 95% as shown in Table 3.

Table 3: Aspect category detection result from LDA generated keywords

No	Aspect Category	Precision	Recall	F-Measure	Accuracy
1	student	0.87	1.00	0.93	0.95
2	skills or competency	1.00	0.76	0.87	0.98
3	course offered	1.00	0.26	0.41	0.85
4	employability industry	1.00	0.37	0.54	0.93
5	accreditation	1.00	0.15	0.26	0.98
6	admission	0.00	0.00	0.00	0.95
7	alumni	0.83	0.93	0.88	0.99
8	financial	0.00	0.00	0.00	0.98
9	study fee	1.00	1.00	1.00	1.00
10	government policy	0.00	0.00	0.00	0.92
11	salary	1.00	0.31	0.47	0.99

It was worth to note that there are three (3) aspect categories that were recorded with zero precision value; admission, financial and government policy. We believed the reason for this result was due to the inability of LDA model to extract certain keywords related to the aspect categories from the given data. This limitation was also highlighted in a study by Ye (2017) where LDA tends to find the topic of different reviews rather than the aspects of these reviews. For this study, the tweets predicted with zero precision were not used further in aspect and sentiment terms extraction.

From the result of the experiment, we are suggesting the used of LDA as a topic modeling methods to extract and categorized the aspect in a lesser known domain. The eleven (11) topics that has been discovered by the used of LDA were supported with the usage of term weighting, TF-IDF and text-features unigram and bigram. The quality of the model was evaluated by using topic coherence score where the highest score obtained by our model was at 0.52 which were represented combination of LDA, TF-IDF and bigram. The results were evaluated by using UMass topic coherence measurement.

6.0 Conclusion

Topic modelling technique, LDA has been combined with bigram feature and weighted with TF-IDF was utilized to represent the aspect categories. The used of topic coherent measurement has ensured suitable number of topics were extracted from the dataset. The overlapping of keywords between each topic were expected with the generative nature of

LDA. The keywords from the topics then were used to predict the aspect categories in the dataset. The comparisons between the keywords with the manually labelled aspect categories were done and had shown a pretty good result where the highest score for accuracy was recorded at 96%. This shown that LDA topic models may help in categorizing aspects in the dataset especially in lesser know domain.

As a result from this study there are six (6) aspect categories that have been detected from LDA generated keywords with precision value at 1.00. The aspect categories are skills or competency, course offered, employability and industry, accreditation, study fee and salary. These aspects are also mentioned in studies by Awang et.al (2011) and Ismail & Zainal Abidin (2014). It was also supported by Khazanah Research Institute (2018) where it mentioned that there are negative sentiment towards the employability of TVET graduates and academic progression in TVET institutions. The use of topic modeling techniques in extracting aspects category for this research has allowed the discovery of most popular topics among Twitter members about TVET in Malaysia. This will enable the TVET stakeholders to further increase the efforts in handling negative sentiments towards them.

Reference

Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah. (2020). Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, *34.*

Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal Of Advanced Computer Science And Applications*, *6*(1),147-153.

Awang, A. H., Sail, R. M., Alavi, K., & Ismail, I. A. (2011). Image and students loyalty towards technical and vocational education and training. *Journal of Technical Education And Training*, *3*(1), 13–28.

Drus, Z., & Khalid, H. (2019). Sciencedirect sciencedirect sentiment analysis in social media and its application : systematic sentiment analysis in social media and its application : systematic literature review literature review. *In The Fifth Information Systems International Conference 2019 Sentiment*. Surabaya, Indonesia.

Esa, A., & Kannapiran, S. (2014). Perception of male and female students towards higher education in technical and vocational study. *Journal of Education And Human Development*, *3*(2), 913–923.

Esa, A., Razzaq, A. R. A., Masek, A., & Selamat, D. A. (2009). The perception of students towards the community colleges' courses that offered in Malaysia. *Asian Social Science*, *5*(7),98-106.

Ismail, A., & Zainal Abidin, N. (2014). Issues and challenges of technical and vocational education and training in malaysia towards human capital development. *Middle-East Journal of Scientific Research , 19*, 7–11.

Khazanah Research Institute. (2018). *The school-to-work transition of young Malaysians.*

Lam, K. W., & Hassan, A. (2018). Instructional technology competencies perceived by technical and vocational education and training students in Malaysia. *International Journal Of Academic Research In Business And Social Sciences*, *8*(5), 343–366.

Liu, B. (2015). Se*ntiment analysis and opinion mining.* Cambridge University Press.

Montenegro, C., Ligutom, C., Orio, J. V., & Ramacho, D. A. M. (2018). Using latent dirichlet allocation for topic modeling and document clustering of dumaguete city twitter dataset. *In 2018 International Conference On Computing and Data Engineering (pp. 1-5).* Shanghai, China.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove : Global vectors for word representation. *In The 2014 Conference On Empirical Methods In Natural Language Processing (pp.1532–1543).* Doha, Qatar.

Rahmat, F. (2018). *Feature selection in sentiment analysis using naive bayes technique.* Bangi, Malaysia: Universiti Kebangsaan Malaysia.

Röder, M., Both, A., & Hinneburg, A. (2015). exploring the space of topic coherence measures. *In The 8th ACM International Conference on Web Search and Data Mining (pp. 399–408).* Shanghai, China.

Saberi, B., & Saad, S. (2017). Sentiment analysis or opinion mining: a review. *International Journal On Advanced Science, Engineering And Information Technology*, *7*(5), 1660–1666.

Schouten, K., & Frasincar, F. (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge And Data Engineering*, *28*(3), 813–830.

Shahid Shayaa, Wai, P. S., Chung, Y. W., Sulaima, A., Jaafar, N. I., & Zakaria, S. B. (2018). Social media sentiment analysis on employment in Malaysia. *In 8th Global Business and Finance Research Conference.* Taipei, Taiwan.

Shama, A., & Dhage, S. N. (2018). A meticulous critique on prevailing techniques of aspect-level sentiment analysis. *Proceedings of The 2018 International Conference On Current Trends Towards Converging Technologies, ICCTCT 2018 (pp.1–7).* Tamil Nadu, India.

Tubishat, M., Idris, N., & Abushariah, M. A. M. (2018). Implicit aspect extraction in sentiment analysis: review, taxonomy, oppportunities, and open challenges. *Information Processing and Management*, *54*(4), 545–563.

YE, M. (2017). *Aspect-based review extraction for e-commerce products*. Delft University of Technology, Netherlands.