Social Media Mining on Taipei's Mass Rapid Transit Station Services based on Visual-Semantic Deep Learning

CHI-CHUNG TAO, YUE-LANG JONATHAN CHEUNG
Department of Transportation Management,
Tamkang University,
No.151, Yingzhuan Rd., Tamsui Dist., New Taipei City 25137
TAIWAN

Abstract: For public transport operators, passengers' comments towards their experience are valuable for promoting more friendly transportation services. This paper demonstrates that passenger-generated online comments can be used to assess railway transportation station services. The natural language processing and social media mining techniques that include establishing an opinion classification model through visual semantic fusion deep learning methods are applied to assess Taipei's Mass Rapid Transit (MRT) station services from the internet opinions. An opinion monitoring system includes: (1) opinion mining to build a social media comment dataset on the ontology of MRT stations.; (2) proposing intent-sentiment, image-text relationship, and content type categories to assist accessing of passengers' quality of experience; (3) constructing a classification model to classify the nature of opinions (4) proposing visualization to provide an intuitive information display dashboard to help Taipei's MRT operator sense the sentiment-intention trends of comments on each station and access the current service level as well as part of the quality management assessment is also proposed.

Key-Words:- social media analytics, opinion mining, visual semantic, deep learning, Taipei MRT station services, quality assessment

Received: July 20, 2021. Revised: January 24, 2022. Accepted: February 21, 2022. Published: March 31, 2022.

1 Introduction

In recent years, big data analytics with Artificial Intelligence (AI) becomes a trending topic to public transportation operators. Operators aim to improve their service quality and increase the number of rides. One of the common tactics is to understand the travelers' satisfaction, such as by questionnaire, dealing with customers' complaint, etc.

Meanwhile, social media now plays an important role in expressing opinions. According to the 2020 Taiwan Internet Report[1], people are using the internet during nighttime (18:00~23:59) for Instant Message (14.2%), Social Media (13.0%), Recreation (11.1%), News and Life Information (9.7%). According to the NDC Digital Opportunity Survey 2018[2], 46% of people have been posting on social media and the survey revealed that an increasing amount of people now tend to express an opinion online than through the official platform, such as leaving a comment on social media, some may even attach photos alongside. The rise of social media platforms has provided channels to the passenger to express their view towards the transport facilities or other topics they are concerning about.

When user-generated online comments accumulate day by day, valuable insights can be discovered by using social media analytics. During the pandemic period, face-to-face contact for questionnaire surveys will be difficult, a customized social media mining system can be a cost-effective alternative to serve as a 24-hour service center.

This study applies natural language processing and social media mining techniques from the comments collected about Taipei's Mass Rapid Transit (MRT) stations in order to assess MRT station services that include establishing an opinion classification model through visual semantic fusion deep learning methods. An opinion monitoring system that includes: (1) opinion mining to build a social media comment dataset on the ontology of railway transportation stations.; (2) proposing intent-sentiment, image-text relationship, and content type categories to assist accessing of travelers' quality of experience; (3) constructing a classification model to classify the nature of opinions; (4) proposing visualization to provide an intuitive information display dashboard to help Taipei's MRT operator sensing the sentiment-intention trends of comments on each station and accessing the current service level as well as part of the quality management assessment is also proposed.

2 Literature Review

Social media offers channels to update users status and photos concerning their recent activities. Self-presentation concept is becoming increasingly more popular in explaining the users' online participation[3]. Some studies have pointed out that information about users themselves is revealed to impress others[4,5].

Social media provide a platform for user-generated content (UGC). The motivations for people using social media is commonly classified as for social and functional purpose. In ref.[6], after investigating the properties and meanings of tweets, Twitter users can be divided as information sources, friends, information seekers, and Twitter users have the need for daily chatter, conversation, sharing, reporting. In ref.[7], function and sociality attributes can be found on ZoneTag, an online photo-sharing social media with place tagging and comment section, and ZoneTag users usually posted for social and self-use purposes.

Opinion mining is useful to analyze the sentiment and subjective ideas of people towards a specific topic. Some applications include subjectivity and polarity classification, opinion target identification, opinion source identification, opinion summarization[8].

These UGCs have found to be very useful for opinion mining. In ref.[9], the study collected Google Map review on several airports and summarize 25 latent topics matching the assessment of Airport Service Quality (ASQ). Compared to ASQ, a paid survey, Google Map review is an alternative to service quality survey for airports and has a good correlation between the ASQ rating and textual Google map review. In ref.[10], tweets are used to analyze the characteristics of celebrities by investigating the characterization and the popularity of the associated texts, using dataset of tweets from fifteen celebrities. In ref.[11], the study collects the online comments towards news article about comparison of healthcare systems across eight countries and have found some popular topics related to healthcare services mentioned and purposed a national healthcare systems ranking based on sentiment level.

Over past years, the use of smartphone encourages more online multimodal opinions that people comment not only by text but also by attaching images. The functions of image to the posting text can be summarized in 11 types of properties[12]. The logico-semantic relation can be classified by the subordinate relationship between text and image[13]. In ref. [14], the authors implemented a deep learning model to classify the image-text relation on the Weibo posts. Image-Text posts leads to advanced image-text opinion mining.

Deep learning is a computational model that consists of multiple layers to extract features(representation) to perform automatic classification, object recognition and many other domains[15]. Some famous model includes recurrent neural network that process well on sequential text material [16] and convolutional neural network that process well on images media[17].

Meanwhile, visual-semantic embedding (VSE) is the essential technique to input textual and visual subject to train a neural network. Studies have tried different method to fuse different modalities into common multimodal spaces as a form of VSE. In ref. [18], the authors use a pre-trained VSE to distinguish the commercials image-text at which degree of parallel and equivalent. In ref. [19], the authors classify the incentive of using Instagram and implement a deep convolutional neural network (DCNN) to determine the multimodal document in Instagram posts, in term of intent, semiotic and contextual correlations.

In the field of transportation, there are related sentiment analysis focusing on different mode of transport. There are opinion mining on the web forum to count the positive and negative comment statistics[20]. The sentiment analysis is also used to discuss the response of new traffic measures imposed, for example, opinion mining to inspect the effect on a toll company’s brand for the introduction of the new freeway electronic tolling collection scheme [21]. However, there are few studies on the social network opinion mining in the ontology of railway transportation stations.

3 Methodology

This paper proposes an opinion monitoring system to assess Taipei’s MRT station services. This includes opinion mining from the web, data preprocessing, classification model training, and finally comment trend visualization interface.

3.1 Data Source of Passenger’ Feelings Towards MRT Stations

This paper aims to collect passengers' opinions toward Taipei's MRT stations they have been to. Selecting a suitable station-based social media, the popular Google Map Review is finally chosen, where each Taipei's MRT station has its own place tag so that passengers can leave their opinion on that station.

The data is collected with crawler tool, collecting Google Map review on each the Taipei MRT, Taoyuan MRT (those in Taipei) and Tamsui LRT stations, a total of 135 stations. After removing meaningless comments and translating comments of foreign language into Chinese, we have collected 2179 opinion with image and text, ranging from 2017 to 2020.

According to literature review, we carried out data labeling on the dataset to build a MRT station online opinion dataset. This paper classifies each review into three categories: intent-sentiment, image-text relationship, and content type. The goal of these classes is to help the MRT operator to access their stations' quality of experience brought to the passengers. This is a multi-category classification prediction task and each category is independent.

"Intent-sentiment classification" is proposed to identify a passenger's sentiment polarity to help understand his or her feeling towards station services, for instance, if someone dislikes a station then the operator would have noticed through the sentiment polarity.

Intent-sentiment classification is divided into six categories: "very negative", "negative", "neutral (descriptive)", "near-neutral (informative)", "positive", "very positive". This is a six-level measurement of sentiment which previous sentiment studies follow similar taxonomy [20,21]. In addition, ref. [6,7] point out that the motivation of using social media contains the pattern of sociality and function, while neutral comment may contain subjective words implies neutral comment exist different level of neutrality[22].

Our dataset also matches the pattern that some opinions are not purely emotive, but give out descriptive stories or informative messages (with mild subjective info selection), such as "I commute this station everyday", "There is a big YouBike station outside the station", therefore the neutral(descriptive), near-neutral (informative) intent are created to classify these neutral opinions, respectively.

The purpose of "Image-text Relation" is to explore the relationship between the text and the image. The labels are "Image-text Related" and "Image-text Unrelated". This classification has been discussed in several visual semantic studies to understand the influence of visual semantic media [12,13,18,19,20].

The purpose of "Content Type" is to find out the depicted target of the opinion. Reviewing the dataset, opinion can be divided into "Station-related", "Scenery" and "Local" labels. In this category, both the text and image are used to judge the main focus of opinion.

These three proposed classifications require labelling to the dataset. Manual annotation is carried out to classify the intent-sentiment, image-text relationship, and content type of each image-text opinion. Annotators are required to follow a set of guidelines and look at both comment text and the attached image to label with consensus.

Dataset statistics are shown in Table 1 and examples are shown in Fig. 1. Half of the opinions are neutral tendencies, followed by the positive intent. This is because passengers may not want to leave strong emotive words on an open social media platform. The results show that opinion text and image have a consistent relationship and have a large proportion of station-related content.

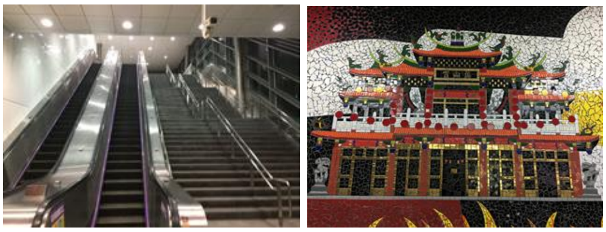

Fig. 1: Example of opinion [left]: So few people huh. (Negative/ Image-Text Related/ Station-related); Opinion [right]: Installation art…reflecting our culture…metro integration advances transport convenience! . (Very Positive/ Image-Text Related/ Station-related)

Table 1. Counts of different labels in each category

Intent-sentiment		Image-text Relation		Content Type	
Label	***Count***	***Label***	***Count***	***Label***	***Count***
Very negative	45	Related	1254	Station	1396
Negative	118	Unrelated	925	Scenery	572
Neutral(descriptive)	400			Local	211
Near-neutral (informative)	782				
Positive	648				
Very positive	186				

3.2 Classification Model to Sort Out Opinions Automatically

This neural network model identifies the intent-sentiment, image-text relationship, and content type of the online opinion towards MRT station services using a visual semantic deep-learning method. Rather than manual analysis of opinion, this model aims at providing a cost-efficient way to sort out online opinions.

Empirical studies are explained as followings:

3.2.1 Dataset Pre-Processing

This paper uses the above-mentioned dataset of 2179 samples as input to the model for training and testing. We take only the text and image within the opinion and do not use other metadata. Both text and image data are necessary to perform pre-processing,

The text on each opinion needs to be converted into word vector(features) beforehand. Due to the unique structure of Chinese wordings, text undergoes segmentation by CKIPtagger, a popular segmentation tool, to separate sentences into meaningful wordings. Since most opinions are focused on MRT stations, a transport-word supplement dictionary is fed into CKIPtagger for better segmentation, adding MRT station name and local transport slang, so that specific words like “Taipei Main Station” will not be wrongly cut into “Taipei/ Station” or “Zhongxiao Xinsheng Station” instead of “Zhongxiao/Xinsheng/Station)”. The segmented word list is then taken to remove common words for better model training, which is referred as stopwords consisting mostly of prepositional conjunctions. The text output is finally up to extract word vector.

This paper uses Word2Vec Skip-gram modal to extract text feature because the Skip-gram modal has better training result on rare words [23], making it suitable to this transport-terms filled dataset. The result of text segmentation is input into Word2Vec Skip-gram model, whose training dimension is set on 300, bring out the word vector needed for the computational model.

The image on each opinion needs to be converted into array beforehand. All image is compressed into RGB format with the size of 224x224 to save computational resources and then converted into a NumPy array of {image height, image width, RGB channel}, bring out the image representation needed for the computational model.

3.2.2 Model Build Up

This paper builds up a visual-semantic neural network by using Python, Tensorflow and Keras as shown in Fig.2. First, the pre-processed image and text are input to the visual input layer and the textual input layer respectively to do encoding, then the output embedding from visual and textual modalities are concatenated in the fusion layer and pass through several fully connected layers, until the output layer give out the three classification results, which are intent-sentiment, image-text relationship, and content type.

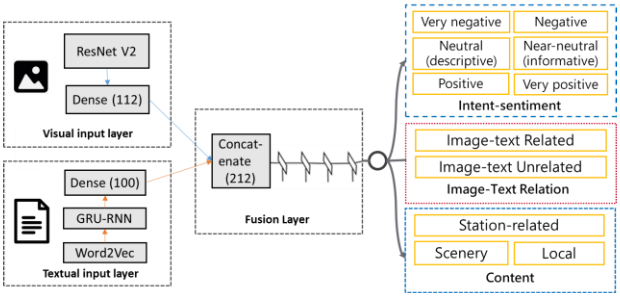

Fig. 2: The architecture of the model

The textual input layer takes the word embedding from the Word2Vec and undergoes encoding through Gated Recurrent Unit(GRU) and Recurrent Neural Network(RNN)[16]. RNN is a neural network which has a recurrent structure that hold memory state, feeding the output features of previous steps into the current steps. GRU consists of update gate and reset gate, holding the existing features while adding new content into current steps. [24] The GRU-RNN network is used to extract the features of comments’ word embedding and then pass to dense layer of dimension 100.

Table 2. Structure of the pre-trained RESNET V2 Model

Layer Name	Filter size	Resnet V2 101-layer
input	224×224	
conv1	112×112	7×7×3,64，stride 2
maxpool 1	56×56	3×3×3 maxpool，stride 2
conv2×x	56×56	$\begin{bmatrix} 1\times1\times3,64 \\ 3\times3\times3,64 \\ 1\times1\times3,256 \end{bmatrix}\times3$
conv3×x	28×28	$\begin{bmatrix} 1\times1\times3,128 \\ 3\times3\times3,128 \\ 1\times1\times3,512 \end{bmatrix}\times4$
conv4×x	14×14	$\begin{bmatrix} 1\times1\times3,256 \\ 3\times3\times3,256 \\ 1\times1\times3,1024 \end{bmatrix}\times23$
conv5×x	7×7	$\begin{bmatrix} 1\times1\times3,512 \\ 3\times3\times3,512 \\ 1\times1\times3,2048 \end{bmatrix}\times3$

maxpool 2	1×1	Average pool

The visual input layer takes the preprocessed image representation and undergoes through pre-trained ResNet101 V2 model[17], an advanced convolutional neural network(CNN) that consists of convolution layers, pooling layers, Batch Normalization, activation function layers, full connected layers as well as shortcut connections and pre-activation design, shown in Table 2, and then pass to a dense layer of dimension 112.

The outputs features of both the textual input layer and the visual input layer are then passed to the fusion layer that concatenate the features outputs into a common multimodal embedding space of dimensions 212. The fused embedding is further converged in the fully connected layer of dimensions 212, 128, and 64. At last, the output layer consists of three different layers that give out the comments' intent-sentiment, image-text relationship, and content type respectively.

3.2.3 Training

80% of the data(1743 samples) is split into training set, and 20% of data is split into verification set (436 samples). Since the data is an imbalanced dataset, some categories account for a large number in the total dataset, therefore the data is split while maintaining the same stratification ratio to ensure a fair performance assessment of the model.
We performed a stratified 5-fold cross validation during the training and takes one of the best-performed split to final evaluation. We trained with the Adam optimizer(learning rate of 0.001) with the epochs and batch size of 200 and 7, respectively.
For evaluation, we reported the classification accuracy and also F1-score as shown in Table 3. Both the accuracy and F1-score indicators show that the visual-semantic model has a predictive ability. The predictive effect is in an order of Image-Text Relation (72.7%), Intent-Sentiment (73.9%), Content Type (61.2%). The performance is similar to that of some multimodal model like intent classifying [19].

Table 3. Classification accuracy and F1-score

Classification \ Score	Accuracy	F1-score (Weighted)
Intent-Sentiment	72.7%	0.72
Image-Text Relation	73.9%	0.74
Content Type	61.2%	0.58

Class-wise performances are shown as Tables 4-6. The performance of intent-sentiment classification reflects that the model can classify some characteristics of different intent, which the neutral(descriptive) type performed best, followed by the positive intent. The number of intent is imbalanced, i.e. different types of intent class are unevenly trained, leading to different performance in each type. As a result, a small number of negative comments leads to poor classification performance. For the Image-Text Relationship classification, the model can distinguish the visual-semantic relation that both classes have an even result.

Table 4. Confusion matrix of intent-sentiment

		Predicted					
		Very Neg	***Neg***	***Neu***	***Near Neu.***	***Pos***	***Very Pos***
True label	Very Neg	4 (0.47)					
	Neg		4 (0.20)				
	Neu			140 (0.9)			
	Near Neu				50 (0.63)		
	Pos					105 (0.74)	
	Very Pos						14 (0.44)

For the content-type classification, the model can only classify little characteristics of different content, however the accuracy of station-related content classification reaches 0.75, but the other two classifications lower its overall accuracy. This classification is summarization task, so the model has to deal with abstract content that more training and more balanced data are needed to improve the content-type performance.

Table 5. Confusion matrix of image-text-relation

		Predicted	
		Related	***Unrelated***
True label	**Related**	135 (0.7)	
	Unrelated		187 (0.77)

Table 6. Confusion matrix of content type

		Predicted		
		Station	***Scenery***	***Local***
True label	**Station**	233 (0.75)		
	Scenery		30 (0.31)	
	Local			4

		Predicted		
		Station	***Scenery***	***Local***
				(0.13)

Integrating with the opinion mining classification model, it can be utilized to overview any emergent comment and display the results to the MRT operator for decision making.

3.3 Visualization as a Means of Monitoring

This paper proposes to visualize the results of social media findings with geographic and time factors.

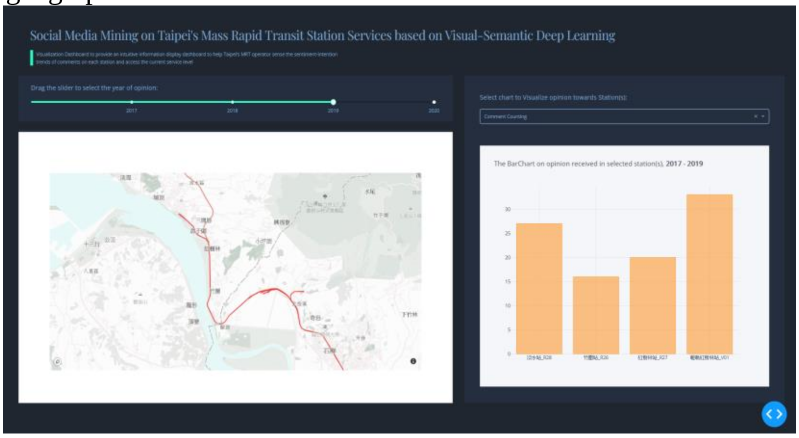

Fig. 3: The proposed GIS dashboard of opinion monitoring for MRT stations

The stations (place tag) are taken as the geographic reference of the opinions, i.e. a set of coordinate. These coordinates are extracted from the National Land Surveying and Mapping Center, and transferred to CRS WGS84 format.

As shown in Fig. 3, we propose a GIS dashboard for all comments of Taipei's MRT stations by using Python, Plotly. The MRT operator can analyze the distribution of intent-sentiment, content type, image-text relationship by selecting single or multiple station(s) on the map. The results are then shown on the right of the interface.

Take the MRT stations in the Tamsui District, the terminus section of the MRT Red Line, as an example. The intent-sentiment in this area is shown in Fig. 4. The majority were neutral comments, followed by positive comments. This matches the general characteristics of the dataset. On the contrary, the terminus LRT station Hongshulin station has a relatively large number of negative comments due to the long queue time on the new opening of LRT line. The new opening of LRT station has also increased the number of neutral (descriptive) comments that mainly depict the public art installations in the Hongshulin station.

social media study, forming a social media comment dataset on the ontology of MRT stations.

This paper classifies each comment into three categories: intent-sentiment, image-text relationship, and content type, to better understand the nature of the comments. "Intent-sentiment" consists of "very negative", "negative", "neutral (descriptive)", "near-neutral (informative)", "positive", "very positive" labels. "Image-text Relation" consists of "Image-text Related" and "Image-text Unrelated" labels. "Content Type" can be divided into "Station-related", "Scenery" and "Local" labels.

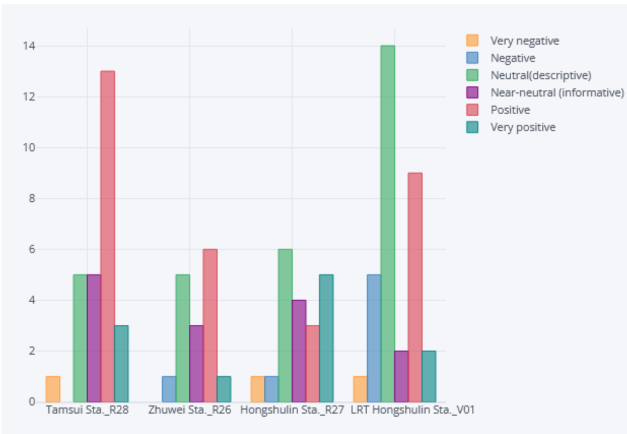

Fig. 4: Bar chart of the intent-sentiment in the Tamsui District area from 2017 to 2020

4 Conclusion

This paper gathered online comments from the Google Map of each tagged MRT station in the Taipei metropolitan area and demonstrated how these online sources fulfill the characteristics of

To monitor the comments of each MRT station, this paper has proposed a classification to sort out those comments from the scope of quality management, and built up a visual-semantic deep learning approach to foster the process.

The empirical study proved a satisfying predictive ability of the visual semantic classification model. The accuracy of results can be shown as follows: intent-sentiment (72.7%), image-text relationship (73.9%), and content type (61.2%). Among the content type, the station-related content has the accuracy of 75%, but its overall accuracy is lowered by the other two categories.

This paper has also proposed to visualize the online opinions, providing an intuitive information display dashboard for the MRT operator. These sorted comments, containing valuable information of the sentiment, content and quantity of passenger has voiced, would be useful to evaluate a particular station's overall reputation and screen out those under-score stations.

4.1 Contributions

This paper has contributed in demonstrating using a new source of information from social media. The Google Map review has rich sentiment material that passenger express their feelings toward the place tag on the Google map. This help analyzes the opinion mining on the ontology of railway transportation station since the review come from particular place tag that greatly reduce the process to identify the place subject.
This paper has also contributed in demonstrating the use of visual semantic deep learning model. The unsorted reviews contain both image and review text. The model establishes the visual and textual input layer, fusion layer that concatenate the features from two input layer, and finally output layer that sort out reviews into three categories.

This paper has also contributed in proposing an alternative opinion monitoring system for the metro operators. During the pandemic, face-to-face questionnaire is not encouraged and so raises the difficulty of accessing passenger attitude toward the metro service. Since this paper proposes an online opinion mining, this is a contactless passenger investigation and more cost-friendly solution to perform a service quality survey under these circumstance. Metro operators are able to monitor the status of each station from the passengers' reviews and look for possible causes of stations with low number of positive rating, with visualization tools.

4.2 Limitations and Future Study

There are some limitations with this study. The data source is limited to Google Map that restrict the number of opinion can be crawled. The Google Map review has many neutral comments resulting of a skewed opinion dataset. This may restrict the scope of the passengers' review received. For future studies, the data sources should expand to several social media platforms. There are many platforms with place tag, such as Facebook, Twitter, Instagram. It would be valuable to collect more comprehensive information.
The performance of the classification model can be improved. The accuracy is skewed as mentioned, since the number and the distribution of the dataset is unbalanced. With wider source of social media platforms, the dataset can be enriched and more balanced.
This paper focuses on the metro system in Taipei. For future studies, Taiwan Railway Authority(TRA), another important railway system for long-haul commuters, can be included in the assessment of the city-wide railway transportation system. This may bring more interesting results of the sentiment level between different type of railway stations.
For future studies, the interactions between online comments and the real-world situation can also be further discussed, such as how the negative comments on specific service attributes (e.g. temperature, tidiness) can correlate to the station improvement control measures.

References:

[1] Taiwan Network Information Center, "2020 Taiwan Internet Report", 2020
[2] National Development Council, "2018 Individual/Household Digital Opportunity Survey in Taiwan", 2018
[3] B. Hogan. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online." Bulletin of Science, Technology & Society 30: 377 – 386, 2010.
[4] M. Lucie and S. Josef, "Goffman's Theory as a Framework for Analysis of Self Presentation on Online Social Networks", MUJLT2019-2-5, 2019
[5] K. Trammell and A. Keshelashvili, "Examining the New Influencers: A Self-Presentation Study of A-List Blogs", Journalism & Mass Communication 82(4):968-982, 2005
[6] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter," Proceedings of the 9th

WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis -, 2007.
[7] M. Ames and M. Naaman, “Why we tag: Motivations for annotation in mobile and online media,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2007.
[8] K. Khan, B. Baharudin, A. Ullah, “Mining opinion components from unstructured reviews: A review”, Journal of King Saud University-Computer and Information Sciences, 26(3), 258-275,2014
[9] K. Lee and C. Yu, “Assessment of Airport Service Quality: A complementary approach to measure perceived service quality based on Google Reviews,” Journal of Air Transport Management, vol. 71, pp. 28–44, 2018.
[10] C. Pethe and S. Skiena, “The Trumpiest Trump? Identifying a Subject’s Most Characteristic Tweets”, arXiv [cs.CV], 2019.
[11] A. Ruelens, “Analyzing user-generated content using natural language processing: a case study of public satisfaction with healthcare systems.”, Journal of Computer Social Science, 2021.
[12] E. E. Marsh and M. Domas White, “A taxonomy of relationships between images and text,” J. Doc., vol. 59, no. 6, pp. 647–672, 2003.
[13] R. Martinec, “A system for image-text relations in new (and old) media,” Vis. commun., vol. 4, no. 3, pp. 337–371, 2005.
[14] T. Chen, D. Lu, M.-Y. Kan, and P. Cui, “Understanding and classifying image tweets,” in Proceedings of the 21st ACM international conference on Multimedia - MM ’13, 2013.
[15] Y. LeCun, Y. Bengio & G. Hinton, “Deep learning”, Nature 521, 436–444, 2015
[16] K. Cho et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” arXiv [cs.CL], 2014.
[17] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778
[18] M. Zhang, R. Hwa, and A. Kovashka, “Equal but not the same: Understanding the implicit relationship between persuasive images and text,” arXiv [cs.CV], 2018.
[19] J. Kruk, J. Lubin, K. Sikka, X. Lin, D. Jurafsky, and A. Divakaran, “Integrating text and image: Determining multimodal document intent in Instagram posts,” arXiv [cs.CV], 2019.
[20] T. Chen, “Sentiment Analysis of Internet Public Opinions After Introducing Distance-based Electronic Toll Collection on Taiwan"s Freeway” , 2015. Accessed on 2021. [online]. Available: https://hdl.handle.net/11296/9p2xrb
[21] Y. Tsai, “Internet Public Opinion Sentiment Analysis on Topic of Taiwan Freeway’s Distance-based Toll Collection Using Three-way Decisions Theory”, 2016. Accessed on 2021. [online]. Available: https://hdl.handle.net/11296/6k9272
[22] T. H. Park, J. Li, H. Zhao, and M. Chau, “Analyzing writing styles of bloggers with different opinions,” 2009.
[23] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” arXiv [cs.CL], 2013.
[24] J. Chung, C. Caglar, et. Al. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.”, arXiv [cs.CL], 2014

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
All authors have contributed equally to creation of this article.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
This paper is a part of the research project “MOST108-2410-H032-043”. The authors are grateful for the fund provided by the Ministry of Science and Technology of Taiwan.