Measuring public opinion on the import of US pork in Taiwan

Yu-Lun Huang
National Chengchi University, Taiwan
110753134@nccu.edu.tw

Yu-Yun Chang
National Chengchi University, Taiwan
yuyun@nccu.edu.tw

Jyi-Shane Liu
National Chengchi University, Taiwan
liujs@nccu.edu.tw

Chang-Yi Lin
National Taiwan Normal University, Taiwan
61015003E@gapps.ntnu.edu.tw

Abstract

As the import of pork with ractopamine from the United States (US) has stimulated a heated discussion on social media, the purpose of this exploratory study is to assess public opinion on the import of US pork by examining the interplay between the two sub-issues, namely food security (FS) and political economy (PE), at the word, sentence, and time series levels. First, we collect comments related to the US pork issue on PTT, the largest online discussion forum in Taiwan, throughout the year 2021. Second, at the word level, we apply Latent Dirichlet Allocation (LDA) to extract the main topics. Third, at the sentence level, we sample a subset of data for manual annotation. Taking the annotated subset as the gold standard, we use the pretrained BERT-base-Chinese model for the downstream topic classification task to predict the whole dataset. Fourth, at the time series level, we employ Detrended Fluctuation Analysis (DFA) to evaluate the predicted number of comments for FS and PE sub-issues. Two main results are obtained: (1) the theme of political economy interweaves with the concept of food safety at both word and sentence levels; (2) DFA shows the absence of a more salient perspective between the FS and PE sub-issues at the time series level.

1 Introduction

In Taiwan, social issues are typically divided into different discursive dimensions according to political ideology. Take the US pork issue as an example, the Democratic Progressive Party's (DPP) decision to lift the restriction on the import of pork with ractopamine was made to strengthen Taiwan's economic ties with the US, while the Kuomintang Party's (KMT) opposition to this policy heightens the potential risk to food safety and facilitates the holding of a national referendum banning the US pork importation. As the referendum is intended to represent the views of citizens for involvement in the policy-making process, the question then arises as to which perspective the political parties attempt to shape public opinion from. As social media mining has become a powerful tool to investigate the formation and change of public opinion (Anstead & O'Loughlin, 2015), this study evaluates the evolution of public views via the Professional Technology Temple (PTT), Taiwan's largest online bulletin board where the issue of US pork has been extensively discussed.

In their study of stance detection of comments posted on PTT, Chuang and Hsieh (2015) examine the cross-fertilization of ideas between netizens participating in online discussions and audiences reading the comments. Indirect participation in the conversations notwithstanding, the audiences still receive the messages conveyed by the interlocutors. Hence, discussion on PTT represents a dominant force shaping public attitudes toward a particular issue. However, studies to date on stance detection have frequently drawn on social media data at a document-level where each text was written by a single author (e.g., Faulkner, 2014; Walker et al., 2012). In fact, short comments posted below these articles are opinion-rich resources which influence the attitudes of netizens. This paper engages with the tendency of public opinion reflected in short comments provided by multiple authors.

Tu et al. (2021) analyze the issue of US pork from a media and communication perspective, exploring the behavioral patterns of netizens, such as posting time and posting frequency. However, they (2021) neglect the linguistic aspects in media discourse and mention without further elaboration that the US pork issue is recognized as both a food safety concern and a politico-economic issue. Building on Tu et al.'s (2021) study and taking it as a point of departure, the current research aims to examine public attitudes toward the US pork issue from a linguistic standpoint and measure the extent to which the two sub-issues, i.e., food security (FS) and political economy (PE), intersect and interact with each other through a quantitative lens.

In this study, topic modeling is conducted to retrieve linguistic information about the US pork issue at three different levels: word, sentence, and time series. First, at the word level, we adopt Latent Dirichlet Allocation (LDA), a topic model which has been extensively used to build clusters of documents, to extract the main topics. While LDA provides a useful lens for topic modeling, the topics extracted by LDA only model correlations among words without explicitly representing correlations among the topics (Li & McCallum, 2006). Therefore, at the sentence level, a semi-supervised transfer learning method based on the pretrained BERT model is applied to characterize the interaction between the FS and PE sub-issues by incorporating contextual information. Finally, at the time series level, we take the time axis into account by using Detrended Fluctuation Analysis (DFA) to evaluate the evolution of public views. The proposed method seeks to contribute to research on social media mining by assessing the interplay between sub-issues of a specific social issue at the word, sentence, and time series levels.

The remainder of this paper is organized as follows. It first reviews existing literature on topic modeling and time series data analysis (section 2). Then it presents data collection, the annotation scheme, and the use of LDA, the pretrained BERT model, and DFA (section 3). After that, it not only illustrates the word clouds generated by LDA and presents the predicted results obtained by the BERT model but also reveals the dynamics of public opinion on the US pork importation and demonstrates the DFA results (section 4). This paper concludes with a brief discussion of the significance of our results and implications for understanding the chronological development of public concern over a particular social issue (section 6).

2 Related Work

Prior studies pertaining to topic classification have typically utilized machine learning algorithms. One line of research has focused on unsupervised learning techniques in which words are clustered into a set of topics, such as LDA (e.g., Hong & Davison, 2010) and Latent Semantic Analysis (e.g., Valdez et al., 2018). Another line of research has concentrated on supervised learning algorithms, such as Naive Bayes, Support Vector Machine, Logistic Regression, and Decision Tree (e.g., Lee et al., 2011; Wang & Manning, 2012). In addition to traditional machine learning models, deep learning models, such as LSTM (Hochreiter & Schmidhuber, 1997) and BERT (Devlin et al., 2018), have received considerable attention in the field of natural language processing in recent years. However, due to high computation cost for training large models with numerous parameters, a transfer learning method using pre-trained models is widely employed to perform text classification, such as topic detection (e.g., Houlsby et al., 2019; Yin et al., 2019). In the current study, we present a complementary approach using both unsupervised and semi-supervised transfer learning for the investigation of public opinion on the import of US pork.

2.1 Latent Dirichlet Allocation

One promising approach to unsupervised topic modeling is the implementation of Latent Dirichlet Allocation (LDA) (Blei et al., 2003), a valuable tool for automated text summarization and visualization. On the basis of the Bag-of-Words paradigm, LDA is a generative probabilistic model that describes each document as a distribution of topics and each topic as a distribution of words. To generate a document, LDA first samples a per-document multinomial distribution over topics from a Dirichlet distribution. Then it repeatedly samples words in each document from the corresponding multinomial distribution; the procedure is repeated for all the words in each document.

2.2 Topic classification using BERT

As Devlin et al. (2018) elaborate, during the pre-training stage, the BERT-base-Chinese model is trained on two tasks, masked language modelling and next sentence prediction, to learn from the surrounding contextual information in the Chinese Wikipedia corpus. In the fine-tuning stage, the ⟨CLS⟩ token at the beginning of each sequence, whose embeddings are used as the representation of each input sequence, is fine-tuned for task-specific objectives. In this study, we annotate a subset of data and employ a transfer learning paradigm based on the pretrained BERT-base-Chinese model to solve the downstream topic classification task. After BERT makes predictions on the whole dataset, we obtain the predicted number of comments throughout the year 2021.

2.3 Detrended Fluctuation Analysis

Since the concept of timeline is integrated into the trend of public opinion, we apply the DFA to the time series analysis of the predicted number of comments in order to evaluate the changes in topics over time. More specifically, we quantify the scaling properties of time series movements in public opinion. The DFA is a useful technique for the detection of long-range correlations embedded in non-stationary time series (Peng et al., 1995; Kantelhardt, 2011). It is composed of the following steps: First, generate an integrated time series of N data points $\{x_1, x_2, \ldots, x_N\}$:

$$X(i) = \sum_{k=1}^{i} x_k - \langle x \rangle$$

where $\langle x \rangle$ is the mean of the above series.

Second, divide the integrated time series into $N_s = int(N/s)$ non-overlapping windows of equal length, each consisting of s number of samples.

Third, calculate the least square fit of preferred order to the data points in each window. This represents the local trend p_n. Next, subtract the local trend from the corresponding window to calculate the detrended series for windows of size s.

$$X_s(i) = X(i) - p_n(i)$$

where $p_n(i)$ is the fitting polynomial in window n, for each window $n = 1, 2, \ldots, N_s$.

Fourth, determine the variance of the detrended series in each window n by averaging over all data points:

$$F_s^2(n) = \langle X_s^2(i) \rangle = \frac{1}{s} \sum_{i=1}^{s} X_s^2\,[(n-1)s + i]$$

then, average over all segments and take the square root to compute the fluctuation function $F(s)$:

$$F(s) = \left[\frac{1}{N_s} \sum_{n=1}^{N_s} F_s^2\,(n) \right]^{1/2}$$

$F(s)$ increases with the growing value of s. If the data are long-range power-law correlated, then $F(s)$ will show in the form of a power-law:

$$F(s) \propto s^{\propto}$$

where $\propto$ is the scaling exponent of the time series. The case $0.5 < \propto < 1$ for all scales s confirms the presence of long-range correlations in a time series (Kantelhardt, 2011). The greater the value $\propto$ is, the higher the correlations in the signal are.

To the best of our knowledge, the literature on time-series measurement of public opinion is rather sparse, particularly Chinese textual data, in that previous relevant studies are predominantly from English-speaking countries (e.g., Brulle et al., 2012; Stimson, 1999). This study is expected to pave the way for a novel data-driven analytical framework for applying machine learning and time-series analysis to social media content.

3 Methodology

The methodology is made up of four phases. First, we collect comments related to the US pork controversy posted on PTT. Second, at the word level, we perform topic modeling using LDA and text visualization using word clouds. Third, at the sentence level, we annotate the dataset and then present a transfer learning method using the pretrained BERT model for the downstream topic classification task. Fourth, at the time series level, we take the results predicted by the BERT model and conduct the DFA to assess the public perceptions of the US pork importation.

3.1 Data collection

We collect a total of 74,040 comments posted in the year 2021 regarding the US pork importation from the Gossiping Board[1] and HatePolitics Board[2] on PTT according to Tu et al.'s (2021) finding that articles relevant to the import of US pork are constantly posted on these two boards to stimulate discussion. Each comment is required to satisfy two criteria in order to be included. First, the data are collected between January 1 and December 31, 2021. Moreover, to conduct time-series analysis of the evolution of public views, each comment is matched with its corresponding date. Second, on the basis of Tu et al.'s (2021) study, each comment must contain at least one of the six keywords listed below: 美豬 (US pork), 美國豬肉 (pork from the US), 萊豬 (pork containing ractopamine), 萊克多巴胺 (ractopamine), 萊劑 (ractopamine), and 瘦肉精 (leanness-enhancing feed additive). This data extraction method confirms all selected comments are relevant to the US pork issue.

Subsequent to data preprocessing, a proportion of the whole dataset is randomly sampled to construct a subset of 15,190 comments for manual annotation. Table 1 demonstrates the size of the corpus used in this research.

	Subset	All
# Comments	15,190	74,040
# Tokens	137,007	662,487
# Word types	14,806	47,666

Table 1. Corpus information.

3.2 Procedure

Firstly, at the word level, we set the number of clusters to be 2 based on a preliminary analysis of the top 200 keywords extracted by the TF-IDF method. Despite the problem of selecting the optimal number of topics, from the examination of the FS and PE sub-issues, we seek to investigate whether the topic keywords extracted by LDA are clustered together according to the relation between a food security standpoint and a politico-economic point of view.

Then, we exploit LDA to cluster comments on PTT into two topics and obtain 300 extracted keywords to represent each topic after removing stop words and expressions referring to the import of US pork into Taiwan (e.g., 美豬 US pork, 美國豬肉 pork from the US, 萊克多巴胺 ractopamine, 萊豬 pork with ractopamine, 台灣 Taiwan, and 進口 import). Through an analysis of the extracted topic keywords, we note that the FS and PE sub-issues are difficult to separate because the keywords representing the two sub-issues intersect with one another; thus, during manual labeling, we divide the comments into four categories by including two more categories (Both and Other).

Secondly, at the sentence level, we manually annotate the subset by classifying each comment into four categories: FS, PE, Both, and Other. To reduce the burden of manual annotation, we present a semi-automatic corpus-based approach inspired by Liu (2012), in which TF-IDF, syntactic rules, and co-occurrence patterns are applied to produce the keyword lists. Nonetheless, due to the inevitability of errors caused by semi-automatic methods, a subsequent manual editing process is implemented by two well-trained annotators with linguistic background to ensure annotation quality.

Next, we adopt a transfer learning method in which the annotated subset is taken as the input to fine-tune the pretrained BERT-base-Chinese model (with 12-layer, 768-hidden, 12-heads, and 110M parameters) for the downstream topic classification task. During this phase, we divide the annotated subset into 80% training set and 20% test set. Then, our BERT model is trained for 20 epochs with a batch size of 64 and a learning rate of 5e-5. After the fine-tuning stage, our BERT model makes predictions on the whole dataset. Then, we count the predicted results to estimate the frequency distribution of the number of comments classified into the four categories per day in the year 2021, plotting the frequency distribution to show the evolution of public attitudes toward the US pork importation.

Finally, at the time series level, we apply DFA to the predicted frequency distribution so as to confirm the existence of long-range correlations in time series data.

[1] https://www.ptt.cc/bbs/Gossiping/index.html
[2] https://www.ptt.cc/bbs/HatePolitics/index.html

3.3 Annotation scheme

The annotation guidelines, listed in Table 2, are formulated to ensure accuracy and consistency of manual annotation. Chief among the annotation scheme is the target of the opinion as a mechanism to determine whether a comment is categorized as Food Safety or Political Economy; a comment is classified as FS if the target of the opinion is either the US pork or ractopamine, whereas a comment is classified as PE if the target of the opinion is the issue of US pork. However, there are a sizeable proportion of vague comments posted on PTT which express ambiguous meanings (Chuang & Hsieh, 2015) and are thus grouped into the Other category, including (1) vague expressions, such as the phrase 塔綠班 (tǎlyùban), a near homophonic pun on 塔利班 (Taliban); (2) truncated comments resulting from the word count limit of 27 Chinese characters; (3) vague messages which can only be interpreted as meaningful when processed together with the contextual information given by previous comments.

Category	Definition and example
Food Safety	The target of the opinion is either the US pork or ractopamine
	(e.g.) 有安全的豬肉為什麼要吃萊豬 'Why bother eating pork with ractopamine at all when we have safe pork to eat?'
Political Economy	The target of the opinion is the US pork issue
	(e.g.) 萊豬本來就是外交和國際政治的議題啊 'Pork with ractopamine has been a diplomatic and international political issue.'
Both	The comment satisfies the criteria of both Food Safety and Political Economy
	(e.g.) 進萊豬卻又不准賣場標示萊豬，有事嗎 '(The government allowed) the import of pork with ractopamine but banned the market from labeling it, is there a problem?'
Other	The topic of the comment is (1) related neither to Food Safety nor to Political Economy or (2) vague and ambiguous
	(e.g.) 把萊豬加個 i 就變潮啦 i 萊豬 'If we add an "i" before pork with ractopamine, it will become popular! iPork with ractopamine!'

Table 2. Annotation guidelines.

Besides the vagueness of language categorized into Other, we discuss the challenges annotators experience during the coding process. First, annotators have difficulties when deciding whether a comment that contains the word 吃 (eat) can be directly labeled as FS. In example (1), the phrase 吃虧 (being shortchanged) includes the word 吃 (eat), a keyword in the FS list. Considering this phrase along with 貿易組織 (trade organization), our semi-automated approach assign this comment with the Both pre-defined label; however, this comment does not convey the meaning of eating US pork, so the annotators made a correction by changing the label to PE. As for example (2), the phrase 進入校園 (introducing into the campus) implies that the author of the comment regards the US pork issue as a food security concern, because the author questions the safety of US pork with ractopamine. In example (3), "eating the US pork containing ractopamine" is interpreted as FS, whereas "breathing polluted air" is labeled as an issue of PE because of the author's ironic tone against the government's failure to address the problem of air pollution. Therefore, considering the whole sentence, example (3) is annotated as Both. Finally, example (4) is originally classified as FS through our semi-automated method due to the presence of the verb 吃 (eat), yet the agent of the verb is Taiwanese pigs rather than Taiwanese people. Since the food safety issue focuses on whether the importation of US pork has a harmful effect on the health of Taiwanese people rather than Taiwanese pigs, example (4) is neither related to the politico-economic dimension issue nor to the food security crisis and is thus categorized as Other.

(1) 欸進萊豬是我們吃虧 怎不先讓我們進貿易組織
We are shortchanged on the import of US pork containing ractopamine, so why not first allow us to join the trade organization. *[Political Economy]*

(2) 那幹嘛不讓萊豬進入校園
Then why not introduce US pork with ractopamine into the campus? *[Food Safety]*

(3) 過的很好 讓我吃到萊豬跟渣米 吸到空污
I live a good life. Let me eat US pork containing ractopamine along with rice dregs and breathe polluted air. *[Both]*

(4) ㄚ反正把萊豬給台豬吃就好辣
Anyway, just feed US pork with ractopamine to Taiwanese pigs. *[Other]*

4 Results

This study not only provides the topic modeling results through the LDA-based word clouds and reports on the predicted results obtained by our fine-tuned BERT model but also demonstrates the evolution of public attitudes toward the US pork issue and analyzes the time series data via DFA.

First and foremost, the word clouds of FS and PE comments created by LDA are shown in Figures 1 and 2 to visualize the lexical semantic information and word frequency via font size. As listed in Table 3, we observe that the keywords extracted by LDA appear to be clustered based on the correlation between a food safety perspective and a politico-economic standpoint. However, we argue that there is no clear-cut between the two sub-issues. Taking a closer look at the overlapping keywords which appear in both word clouds, we observe the politico-economic standpoint seems to be interwoven with the food safety perspective on the US pork issue. More precisely, analyzing the top 25 keywords, we find that the overlapping keywords are mostly associated with PE and that only the phrase 核食 (radiated food) is related to FS; hence, we can derive that FS-related keywords are more representative features than PE-related keywords.

Figure 1. LDA topic model for Food Safety comments.

Figure 2. LDA topic model for Political Economy comments.

Category	Keyword
Food Safety	吃萊豬 eat the US pork containing ractopamine, 萊劑 ractopamine, 瘦肉精 a leanness-enhancing feed additive, 不吃 never eat, 好吃 delicious, 用萊劑 use ractopamine, 豬肉 pork, 標準 standard, 標示 label, 美國 USA
Political Economy	反萊豬 oppose the US pork with ractopamine, 反美 oppose the US, 公投 referendum, 支持 support, 同意 agree
Overlapping	反美豬 oppose the US pork, 政府 government, 民進黨 DPP, 國民黨 KMT, 開放 lift a ban, 疫苗 vaccine, 高端 MVC COVID-19 vaccine, 中國 China, 塔綠班 tǎlyùban, 核食 radiated food

Table 3. List of top 25 LDA topic keywords.

Secondly, Table 4 indicates that our fine-tuned BERT model achieves an overall performance of 0.96 in F1-score. A possible explanation for the success of our BERT model in topic classification is the repetitive nature of data. This inference is compatible with Tu et al.'s (2021) finding that key opinion leaders regularly post and repost articles regarding the US pork issue on PTT to stimulate discussion among netizens. Furthermore, Table 5 gives a comparison of the number of comments for each category between the annotated subset and the whole dataset. Since the comments predicted as FS and PE by our BERT model account for more than half of the total comments, we can confirm that the large majority of netizens look at the US pork issue either from a food safety perspective or from a politico-economic point of view.

Category	Precision	Recall	F1-score	Support
FS	0.95	0.97	0.96	824
PE	0.98	0.97	0.98	1058
Both	0.95	0.96	0.96	503
Other	0.95	0.94	0.95	654
Macro avg.	0.96	0.96	0.96	3039

Table 4. BERT classification results.

Category	# Comments	
	Subset	All
Food Safety	4,168	24,018
Political Economy	5,187	20,454
Both	2,584	11,941
Other	3,251	17,627
Total	15,190	74,040

Table 5. Dataset information.

After the fine-tuned BERT model predicts the whole dataset, the predicted results are given to illustrate the evolution of public opinion over time as can be seen in Figure 3. It is worthy of note that the growth in the number of comments is consistent with the corresponding social events and situations. Setting the number of 100 comments as the benchmark value, we observe four periods of time when a heated discussion among netizens took place, as marked in Figure 3. First, the US pork issue went viral between January and February as the administrative orders concerning the ban lift on US pork came into effect on January 1, 2021. Second, considering the appearance of the high-frequency keyword 疫苗 (vaccine), we maintain that since both the COVID-19 pandemic and the import of US pork are associated with public health, the US pork issue was frequently raised and discussed on PTT between May and June due to a surge in the number of COVID-19 cases and a COVID-19 vaccine shortage at that time. Third, given the high-frequency keyword 高端 (MVC COVID-19 vaccine), a sudden increase in late August may relate to the vaccination of the first dose of the MVC COVID-19 vaccine in Taiwan on August 23. Fourth, since the referendum against the US pork importation took place on December 18, 2021, netizens on PTT were engaged in a heated discussion over the US pork issue as the end of 2021 approached.

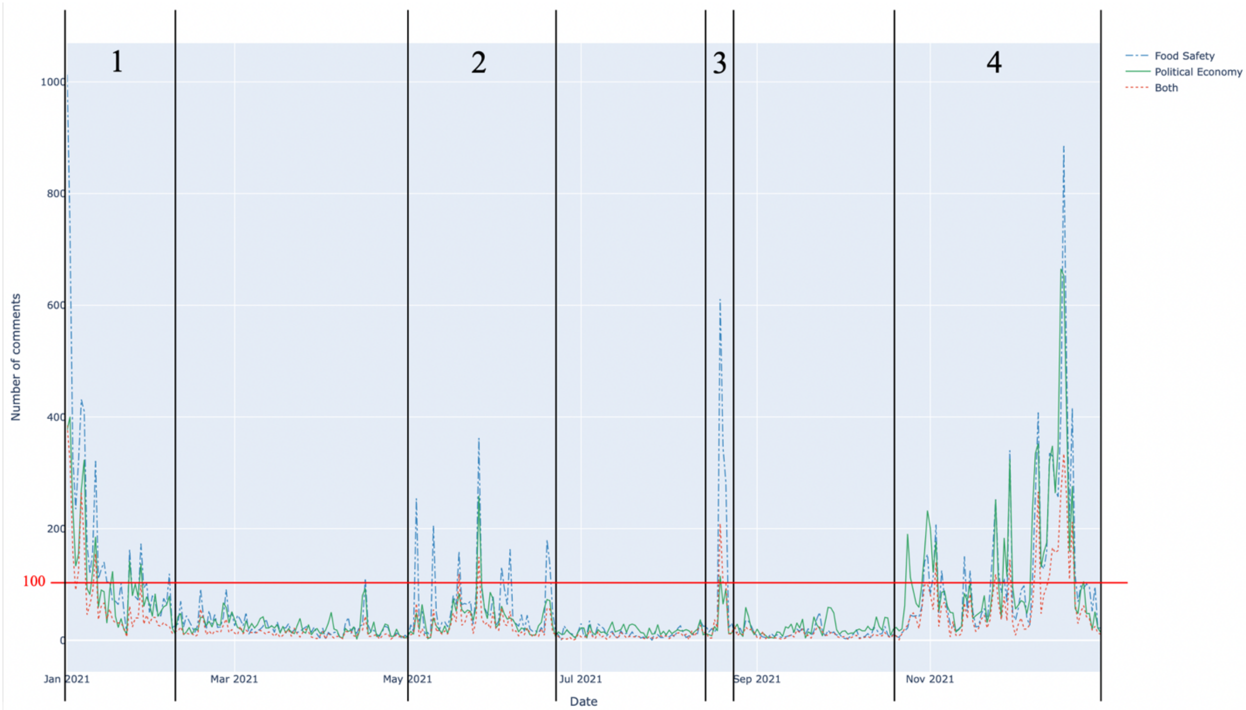

Figure 3. Evolution of public opinion.

As illustrated in Figures 4, 5, and 6, the notable finding based on DFA is a significant rising trend in the number of FS, PE, and Both comments over time in the year 2021 without any downward trend. The estimation of the scaling exponent $\boxed{\propto = 0.67}$, $\boxed{\propto = 0.65}$, and $\boxed{\propto = 0.67}$ for the FS, PE, and Both comments, respectively, demonstrates the long-range correlations in the number of comments per day. Though the overall number of FS comments per day surpasses that of PE comments, the proximity of the exponents obtained by the DFA method infers the absence of a more salient perspective between the FS and PE sub-issues. This finding is in accordance with the fact that the pork import referendum failed to pass but only by a narrow margin. While 51.21 percent of voters rejected the opposition against the import of US pork, there were still 48.79 percent of voters advocating a ban on the US pork importation. The referendum result implies that the potential disadvantage of food safety hazard does not outweigh the politico-economic benefits despite a larger volume of online discussion on the FS sub-issue throughout the year 2021.

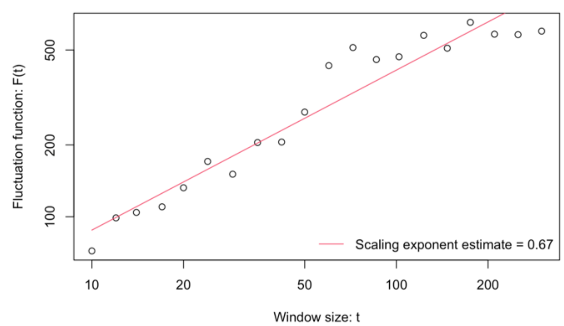

Figure 4. Detrended fluctuation analysis of Food Safety comments.

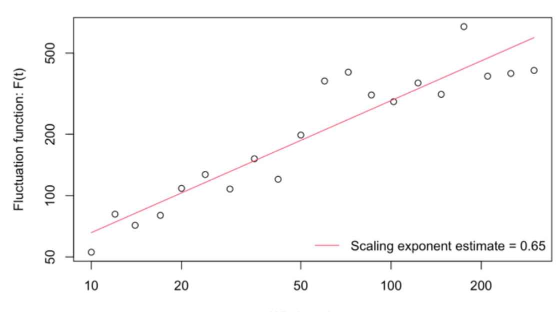

Figure 5. Detrended fluctuation analysis of Political Economy comments.

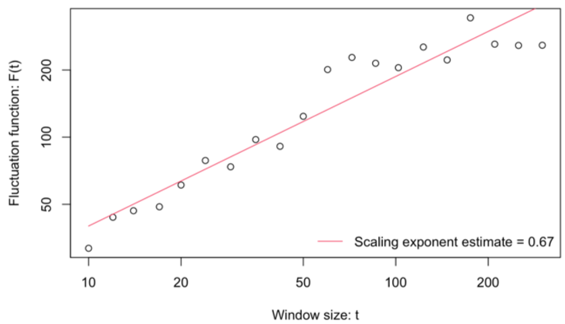

Figure 6. Detrended fluctuation analysis of Both comments.

5 Conclusion

This study investigates how netizens on PTT perceive the US pork importation by examining the interaction between the FS and PE two sub-issues through the combination of machine learning and quantitative analysis. In order to retrieve useful information from a large number of comments on PTT related to the import of US pork, we adopt a complementary approach using both unsupervised and semi-supervised transfer learning. First, at the word level, we incorporate LDA into word clouds to identify the main topics and assess whether the extracted keywords are clustered according to the correlation between the two sub-issues. Second, at the sentence level, we annotate the subset extracted from the whole dataset. Then, we treat the annotated subset as the gold standard to fine-tune a BERT model extended to make predictions on the whole dataset. Finally, at the time series level, we use the DFA to analyze the number of FS and PE comments per day.

Our main results are the following: (1) the finding that the FS sub-issue interweaves with the PE sub-issue in terms of the US pork importation at both word and sentence levels; (2) the absence of a more prominent perspective between the two sub-issues at the time series level; (3) the development of a semi-automated annotation approach and annotation guidelines for a domain-specific topic model; (4) the performance of 95% in F1-score for topic detection achieved by our fine-tuned BERT model; (5) a chronological view of public opinion predicted by the application of our fine-tuned BERT model to the downstream topic classification task; (6) the observation of a steady increase in the number of FS and PE comments over time in the year 2021 based on the DFA.

One potential limitation of this study is the inevitability of prediction error when applying the predicted results obtained by the fine-tuned BERT model to the DFA procedure. Accordingly, we manually label as much data as possible and make our BERT model as accurate as possible during the fine-tuning process in order to minimize errors. Despite possible errors predicted by the model, our proposed approach still offers a data-driven method for quantifying public perceptions of a particular issue and provides new insights into public opinion on the US pork issue based on the integration of machine learning and quantitative analysis.

This paper has contributed to the growing body of studies on social media mining by probing the dynamics of public opinion from a linguistic perspective and examining the interplay between sub-issues of a specific social issue at the word, sentence, and time series levels. It would be useful to conduct more research on the evolution of public attitudes toward different social issues to add further depth, richness, and insights to the findings of this study and provide a further development of topic classification and time series analysis of textual data on social media.

Acknowledgments

We are grateful to four anonymous reviewers for their constructive and insightful comments, which have been taken into account in the revised version to improve the quality of this study. We would also like to thank Dr. Hen-Hsen Huang for his valuable feedback. Finally, this project is supported by a research grant awarded by the National Science and Technology Council, Taiwan (NSTC 111-2410-H-004-098).

References

Anstead, N., & O'Loughlin, B. 2015. Social media analysis and public opinion: The 2010 UK general election. Journal of computer-mediated communication, 20(2), 204–220.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent dirichlet allocation. Journal of machine Learning research, 3, 993–1022.

Brulle, R. J., Carmichael, J., & Jenkins, J. C. 2012. Shifting public opinion on climate change: an empirical assessment of factors influencing concern over climate change in the US, 2002–2010. Climatic change, 114(2), 169–188.

Chuang, J. H., & Hsieh, S. K. 2015. An Arguing Lexicon for Stance Classification on Short Text Comments in Chinese. In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation: Posters (pp. 27–36).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Faulkner, A. 2014. Automated classification of stance in student essays: An approach using stance target information and the Wikipedia link-based measure. In Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference, 174–179.

Hochreiter, S., & Schmidhuber, J. 1997. Long short-term memory. Neural computation, 9(8), 1735–1780.

Hong, L., & Davison, B. D. 2010. Empirical study of topic modeling in twitter. In Proceedings of the first workshop on social media analytics (pp. 80–88).

Houlsby, N., Giurgiu, A., Jastrzębski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. 2019. Parameter-efficient transfer learning for NLP. In International Conference on Machine Learning (pp. 2790–2799). PMLR.

Kantelhardt, J. W. 2011. Fractal and multifractal time series. In Meyers, R. A. (Ed.), Mathematics of complexity and dynamical systems (pp. 463–487). Springer.

Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. 2011. Twitter trending topic classification. In 2011 IEEE 11th International Conference on Data Mining Workshops (pp. 251–258). IEEE.

Li, W., & McCallum, A. 2006. Pachinko allocation: DAG-structured mixture models of topic correlations. In Proceedings of the 23rd international conference on Machine learning (pp. 577–584).

Liu, B. 2012. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1–167.

Peng, C. K., Halvin, S., Stanley, H. E., & GoldbergerA, L. 1995. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos: An Interdisciplinary Journal of Nonlinear Science, 5(1), 82–87.

Stimson, J. A. 1999. Public opinion in America; moods, cycles and swings, CO: Westview Press.

Tu, S. T., Wu, C. Y., Lu, L. Y., & Hsieh, C. H. 2021. Opinion Leaders in Internet Public Opinion: A Case Study on the Controversy of US Pork Containing Ractopamine. In 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW) (pp. 1–2). IEEE.

Valdez, D., Pickett, A. C., & Goodson, P. 2018. Topic modeling: latent semantic analysis for the social sciences. Social Science Quarterly, 99(5), 1665–1679.

Walker, M., Anand, P., Abbott, R., & Grant, R. 2012. Stance classification using dialogic properties of persuasion. In Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 592–596).

Wang, S. I., & Manning, C. D. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 90–94).

Yin, W., Hay, J., & Roth, D. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLPIJCNLP 2019, Hong Kong, China (pp. 3912–3921). Association for Computational Linguistics.