Automatic Detection of Geotagged Food-Related Videos Using Aspect-Based Sentiment Analysis

Hidetsugu Nanba[1] and Satoshi Fukuda[1]

[1] *Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551 JAPAN*

Abstract

Food tourism refers to travel focused on experiencing the distinctive culinary cultures at the chosen destinations. Using geotagged videos, we developed a system to enhance our understanding of global cuisines and enrich the enjoyment of food tourism. We propose a method to automatically detect food-related YouTube videos, which extracts aspect words from YouTube video comments using aspect-based sentiment analysis and detects food-related videos based on whether the aspect words are food-related or not. We conducted experiments to evaluate the effectiveness of the proposed method and found that the precision value of the proposed method outperformed a ChatGPT-4 method. Using the proposed method, we constructed a system to map food-related YouTube videos on a world map.

Keywords

Aspect-based sentiment analysis, Food tourism, YouTube

1. Introduction

There is a wide variety of food cultures worldwide; hence, food tourism is focused on the exploration of local cuisines. Food culture encompasses many elements of local cuisines, such as how to choose ingredients, how to plan a menu, how to cook, how to choose tableware, with whom to eat, how to eat, table manners and etiquette, and so on. On the one hand, tourists often use restaurant search sites to find information about their food offerings; however, the information obtained from these search sites alone is not sufficient in terms of enjoying local food cultures and understanding how travelers evaluate these local food offerings. On the other hand, travel blogs and social network services, such as image-sharing sites, can be another source of information about the food cultures at tourism destinations. Nevertheless, there has been no systematic compilation of food culture information.

We aimed to build a system using geotagged videos to help people understand more about food cultures around the world and enjoy food tourism. We expect that this system will deepen people's understanding of food cultures and generate economic benefits through food tourism. In addition, by mapping the geotagged videos on a world map, we build a system that allows easy access to food information from around the world.

We used aspect-based sentiment analysis (ABSA) technologies to analyze YouTube video comments to select food-related YouTube videos. ABSA is a process that identifies aspect terms in the given text and evaluates the sentiments associated with each of these aspects. As described below, we search for videos using food names as queries. Therefore, although the food name is always included in the video comments, it does not necessarily mean that the main focus of the video is that food. However, if sentiment information about the food is included in the comments, we can determine whether the video is food-related or not. Furthermore, the degree of food recommendation in a video can be determined by the sentiment information in the comments of each video.

The contributions of this paper are as follows:

- Proposal of a method for automatic detection of food-related videos using unsupervised learning.
- Development of a system that allows users to access food-related videos on a map.

2. Related Work

2.1. Aspect-Based Sentiment Analysis

AB SA identifies aspects and determines the polarity of each aspect when text is entered into the system. In the case of a restaurant review, the aspect corresponds to the food and service provided at the restaurant. For example, given a restaurant review sentence such as, "I liked the service and the staff, but not the food," {service, staff, food} are extracted as aspect terms, and {positive, positive, negative} are output as the sentiments for each aspect.

Various datasets for ABSA have been created. A sentiment analysis task was conducted during a SemEval workshop, and the datasets were released[6, 7, 8]. SemEval uses review texts for laptops, restaurants, hotels, smartphones, digital cameras, and museums as inputs for ABSAs. The review texts were written in eight languages (i.e., English, Arabic, Chinese, Dutch, French, Russian, Spanish, and Turkish). ABSAs for restaurant review texts are also relevant to our study; however, some of the food-related videos not only included restaurants, but also the local food cultures, which were then introduced and evaluated. Therefore, our research needs a system that can determine a wider range of aspects, rather than ABSAs specific to the task at the 2016 SemEval workshop.

Yang et al. [10] developed an easy-to-use ABSA framework for beginners. In addition, they published models built using data from a variety of aspect-based opinion analysis tasks.[1] We used this model to extract aspect terms from video comments and determined whether they were food-related.

2.2. Food Tourism

Food tourism research has made considerable academic and industry progress in the last two decades [5]. Major research areas include experiencing local food and beverages, culture and cuisine, gastronomy (science and cooking) [1], and the economic, socio-cultural, and global impacts of food tourism. This study is also positioned as one of these food tourism studies.

Partarakis et al. [6] proposed a tool that allows the representation and presentation of the tangible and intangible dimensions of culinary traditions as cultural heritage, including their sociohistorical context. Hence, they analyzed culinary traditions, and the food culture trends therein.

Fujii et al. [3] proposed a method for automatically categorizing travel blog entries into five categories: Watch, Experience, Buy, Dine, or Stay. Using these classifiers, they classified travel blog entries from foreign visitors to Japan and aggregated them by region to analyze how travelers behave in each region. Their classification of Dine travel blog entries is related to our research topic, namely, food tourism. Conversely, we classified geotagged YouTube videos instead of travel blogs.

Michail and Gavalas [4] implemented a food search application called Bucketfood. The majority of existing applications are based on searching for venues and focus primarily on the overall service (food, environment, service) rather than the food itself. In contrast, Bucketfood provides several means to search for food (by cuisine name, country, distance, category, popularity, etc.). This study also shares with Michail and Gavalas' study in that it focuses primarily on the food itself rather than on individual venues.

3. Automatic Detection of Geotagged Food-Related Videos

We collected the metadata from geotagged food-related YouTube videos. Section 3.1 describes how we collected candidate food-related YouTube videos, while Section 3.2 describes ABSA to detect food-related videos among the candidates collected in Section 3.1. Section 3.3 describes

[1] https://github.com/yangheng95/PyABSA

how to create a list of food-related terms necessary to determine whether a video is related to food using the results of ABSA. Section 3.4 proposes a method for detecting food-related videos using the results of ABSA in addition to the food-related term list.

3.1. Automatic Collection of Candidate Videos About Food

In this study, we use local food terms as queries and search videos on YouTube. In automatically collecting local food terms, we focused on the Wikipedia template, which contains entries for food names and places of origin,[2] and extracted these items from Wikipedia. As a result, 2,404 local food names were obtained. Considering the origin of these items, we manually added latitude and longitude information.

Next, we searched for YouTube videos using the YouTube data application programming interface (API). The names of the foods and their latitude and longitude information were entered into the search, and up to 30 videos were searched for each food within a 300 km radius of the latitude and longitude information. There were two reasons for searching using latitude and longitude information. One is to collect only videos with latitude and longitude information. This is used when finally mapping the search results on a world map. The other reason is that well-known local foods are also served in metropolitan areas around the world. The purpose of this study is to provide travelers with information on how to eat local food in the places where it is produced.

For each YouTube video, we also collected the title, description, comments, and latitude and longitude information tagged with each video. As a result, we were able to collect 31,019 videos.

3.2. Aspect-Based Sentiment Analysis of Video Comments

Using PyABSA, we performed ABSAs on the comments collected using the method described in Section 3.1 [10]. While YouTube video comments are written in multiple languages, PyABSA supports multiple languages; therefore, it is possible to perform the analysis. However, considering the processing after our ABSAs, we used the M2M100 machine translation model.[3] Figure 2 shows an example of analysis using PyABSA. In the figure, "food" is extracted as an aspect with a positive sentiment.

[input]
Very unique yet most traditional food experience I have ever seen.
↓
[output]
Aspect: Food
Sentiment: Positive

Figure 1: Example of aspect-based sentiment analysis using PyABSA

Although the system searched some YouTube videos because food terms happened to be included in the title or description, it was possible to determine whether a video was food-related or not by extracting all the aspect terms in the comment text set for a single video. Figure 2 shows the aspect terms from a YouTube video in their order of frequency.[4] These aspect terms were then analyzed using PyABSA. This YouTube video introduces food items from the German city of Aachen, where the numbers in parentheses indicate the frequency of each aspect term in the comments.

[2] Spaghetti is a popular food eaten around the world, but its place of origin is Italy. Assuming that no matter how famous a food may be, some people may want to eat the original in its place of origin, we included foods that are generally considered to have no regional characteristics.
[3] https://huggingface.co/docs/transformers/model_doc/m2m_100
[4] https://www.youtube.com/watch?v=_AJeC4nVpYg

food (10), cookies (5), *printen* (2), Oreos (2), watching (2), pineapple tart cookies (2), coffee (2), cup of tea (1), things (1), Christmas (1), versions (1), cookie (1), Berlin (1), chocolate chip cookies (1), sound (1), dom (1), gingerbread cookies (1), look (1), pastry (1), tour (1), city (1), area (1), *glühwein* (1), tea (1), glass (1), chocolate (1), eat (1), Frankfurt (1), Aachen (1), upload (1), cup (1), tip (1), *lebkuchen* (1), stock (1), chocolate chip (1), nut version (1)

Figure 2: Examples of aspect terms extracted from a YouTube video

Figure 2 allows us to infer that the target YouTube video is about food. This is because high-frequency aspect terms such as food and cookies are food-related terms. However, we must create a list of food-related terms to automate this inference. In the next section, we describe how to create this list.

3.3. Creating a List of Food-Related Terms

We used travel blog data from TravelBlog[5] to collect food-related terms automatically. The food-related terms described here include not only the local food names described in section 3.1, but also more general food terms. This is because when food is mentioned in the comments of videos, generic names such as food or lunch are often used instead of local food names. TravelBlog is one of the largest travel blog websites with over 700,000 blog entries. Travel blogs may include, for example, descriptions of meals at travel destinations, such as "I ate a hamburger in the restaurant." By focusing on "eat" or its past tense, "ate," and using a syntactic analyzer to extract its object, food-related terms can be collected easily. In the case of the above sentence, using spaCy,[6] a Python library for natural language processing, we can extract "a hamburger" from the parsed result as shown in Figure 3.

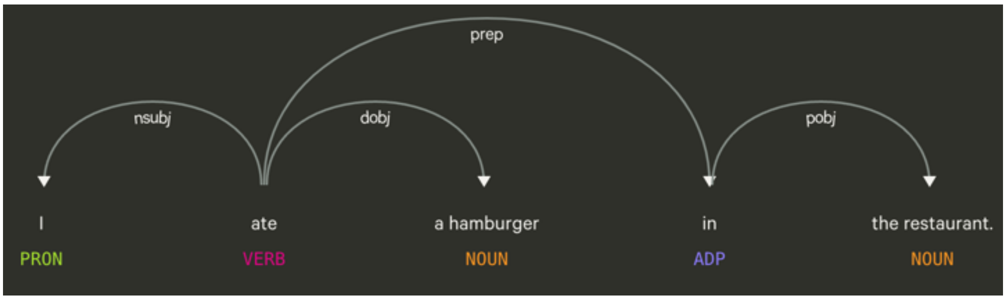


Figure 3: Syntactic analysis result for the sentence, "I ate a hamburger in the restaurant," using spaCy

Some of the terms obtained by the above method, such as "it" and "them," are inappropriate as food-related terms. Therefore, we used an English stop-word list[7] to exclude these terms in advance.

Next, food-related terms were automatically collected using the spaCy results. In Figure 3, if we mask the object of ate, a hamburger, and ask humans to guess the words that fit here, they would name a variety of food-related terms. The same task is performed using the BERT language model [2], which is one of the language models that has recently proven its usefulness in a variety of natural language processing tasks, and is capable of guessing what applies to the masked words in a given sentence. The example shown in Figure 4 reiterates the procedure for collecting food-related terms.

[5] https://www.travelblog.org
[6] https://spacy.io/
[7] https://countwordsfree.com/stopwords

(1) "I even ate sushi and watched MTV while there." (extract "sushi" as the object using spaCy) (2) "I even ate [mask] and watched MTV while there." (where the object is masked) (3) Infer the masked word using BERT, i.e., pizza, dinner, lunch, popcorn, or breakfast

Figure 4: Procedure for collecting food-related terms using spaCy and BERT

First, (1) the object of "eat" or "ate" was extracted from a sentence using spaCy. (2) The object was then masked and finally (3) inferred using BERT. If spaCy correctly extracted the "eat" or "ate" object, then BERT should guess a food-related term for the masked word.

However, if spaCy did not extract the object term correctly, BERT would probably output a term unrelated to food. Therefore, we compared the word sets predicted by BERT for each sentence and excluded those that did not have any terms in common with the other sets.

We arranged the terms thus collected in their order of frequency (Figure 5), where the numbers in parentheses indicate the word frequency. As can be seen from Figure 5, most of the terms in the list are food-related, but some words, such as "lots," are inappropriate as food terms. Therefore, in the next section, we propose a method to detect food-related videos with some robustness even if the list contains inappropriate terms.

lunch (3,132)	lots (274)	cake (167)	supper (99)
dinner (2,434)	sandwiches (244)	chicken (162)	eggs (95)
food (1,903)	cream (228)	rice (156)	grass (90)
breakfast (1,704)	heart (225)	cheese (139)	foods (86)
meal (717)	bread (223)	chocolate139	leaves (75)
fish (541)	buffet (195)	pasta (134)	sushi (74)
pizza (503)	sandwich (192)	salad (116)	snacks (74)
lot (473)	steak (176)	gelato (108)	bit (74)
meat (457)	fruit (174)	seafood (107)	noodles (73)
meals (299)	soup (173)	restaurant (100)	cookies (71)

Figure 5: Part of the list of food-related terms

3.4. Detection of Food-related Videos

If most of the aspect terms extracted from the YouTube video comments are included in the food-related term list in Figure 5, the video is food-related. Here, as mentioned in Section 3.3, the food-related term list contains inappropriate terms. Therefore, we used the frequency of occurrence of the food-related term list as the confidence level for the term. We computed the score of video *m* using the following formula and judged *m* as a food-related video if the score was above a threshold value:

$$Score(m) = \sum_{t \in Asp} P(t) \cdot freq(t)$$

where *Asp* is the set of aspect terms extracted from video *m*, *P(t)* is the probability of occurrence of aspect term *t* in video *m*, and *freq(t)* is the frequency of *t* in the food term list. These thresholds were determined using 99 videos prepared separately from the experimental data described in the next section.

4. Experiments

We performed some experiments to confirm the effectiveness of our method.

4.1. Experimental Conditions

Data

We used 400 arbitrarily selected YouTube videos collected using the procedure described in Section 3.1 and identified 274 food-related videos.

Evaluation Measure

We evaluated the videos using precision and recall.

Alternative Methods

To compare our findings with the method proposed in Section 3, we also experimented with the following two baseline methods.

- **Full-ChatGPT (Baseline Method 1):** Detecting food-related videos using ChatGPT-4
  The following sentence was added before the comment extracted from each video as a prompt and judged using ChatGPT-4.
  [prompt] *The following text is a comment on a YouTube video. Is this video about food or a restaurant or meal? Answer with "yes" or "no."*
- **Aspect-ChatGPT (Baseline Method 2):** Aspect extraction using ChatGPT-4 and detection of food-related videos using aspect words.
  After extracting aspect terms using ChatGPT-4 with the following prompt instead of PyABSA, as described in Section 3.2, we used the methods described in Sections 3.3 and 3.4 to detect food-related videos.
  [prompt] *Perform an aspect-based sentiment analysis of the following sentence and extract the aspect. Note that if the aspect word is not an English word, please translate it into English.*

The reason for using ChatGPT-4 for aspect-based sentiment analysis as Baseline method 2 is that we could not find a universally available ABSA tool other than PyABSA. The reason for using ChatGPT-4 as Baseline method 1 is that the data used in this experiment is 400 videos, which is too small for supervised learning[8], so zero-shot learning was necessary. ChatGPT-4 performs best in zero-shot learning on a variety of natural language processing tasks.

4.2. Results and Discussion

Table 1 shows the experimental results. Among the three methods, Full-ChatGPT-4 had the highest recall value, while its precision value was the lowest. When all the videos were judged as food-related, the precision value was 274/400 = 0.685, which was higher than that of Full-ChatGPT-4. In other words, we can conclude that the Full-ChatGPT-4 method performs worse than the random method.

Table 1
Evaluation results for the proposed and baseline methods for detecting food-related videos

Methods	Precision	Recall
Our method	0.8476	0.3248
Full-ChatGPT-4	0.6703	0.6752
Aspect-ChatGPT-4	0.8182	0.2956

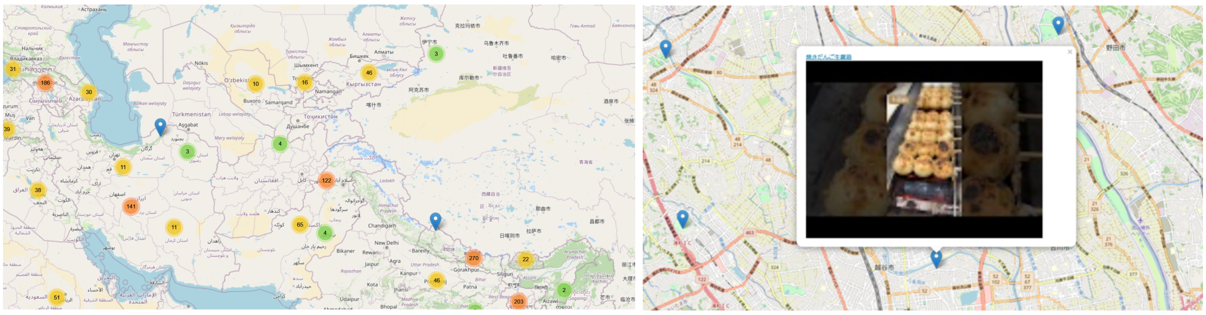
[8] The reason for the small number of correct data is that there are many comments for one video, or in some cases, the comments are written in multiple languages, which makes the judgment process time-consuming.

The precision scores for both our method and Aspect-ChatGPT-4 were much higher than the value of 0.685 for the random method, suggesting that the food-related video detection method using the results from aspect-based sentiment analysis was effective. Comparing our method with the Aspect-ChatGPT-4 method, the precision and recall scores for our method outperformed the Aspect-ChatGPT-4 scores, which is due to the superiority of PyABSA trained on various ABSA datasets compared to the zero-shot trained ChatGPT-4.

Although the recall score of our method was low, this is not a significant problem because, as mentioned in Section 3.1, the maximum number of videos retrieved using the YouTube data API is 30 for a single food, but the problem of low recall score can be solved by increasing this upper limit beyond 30.

5. System Behavior

The data collected and analyzed using the method described in Section 3 were mapped on a world map using OpenStreetMap with the Leaflet JavaScript library.[9] Figure 6(a) shows an example of how the system works. Individual videos are displayed as pin icons on the map. The locations where videos are clustered together are grouped into clusters and displayed as circle icons. The number of videos in a cluster is also displayed within the icon. In Figure 6(a), when the pin icon is clicked, the title and thumbnail of the video at that location are displayed in a pop-up window and the video is played when the title is clicked (Figure 6(b)). The system is currently available online[10].



(a) Mapping geotagged videos using Leaflet (b) Video playback on the map

Figure 6: System behavior

6. Conclusion

In this study, we proposed a method to automatically detect food-related YouTube videos. The proposed method first extracts aspect terms from video comments using the PyABSA ABSA tool and then compares the results with food-related terms to determine whether the video is food-related. The experimental results conducted to confirm the effectiveness of our proposed method confirmed that the use of ABSA for detecting food-related videos is effective and that the proposed method is superior to the ChatGPT-4 method.

References

[1] C. Barzallo-Neira and J. I. Pulido-Fernández: Identification of the Main Lines of Research in Gastronomic Tourism: A Review of the Literature. Sustainability, Vol. 15, No. 7, 5971, 2023.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North

[9] https://leafletjs.com/

[10] http://nlp.indsys.chuo-u.ac.jp/tools/travel/leaflet/food/

American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186, 2019.

[3] K. Fujii, H. Nanba, T. Takezawa, A. Ishino, M. Okumura, and Y. Kurata: Travellers' Behaviour Analysis Based on Automatically Identified Attributes from Travel Blog Entries, Proceedings of Workshop of Artificial Intelligence for Tourism, PRICAI 2016, 2016.

[4] A.-M. Michail and D. Gavalas: Bucketfood: A Crowdsourcing Platform for Promoting Gastronomic Tourism, Proceedings of 2nd International Workshop on Context-Awareness for Multi-Device Pervasive and Mobile Computing, pp. 9-14, 2019.

[5] B. Okumus: Food Tourism Research: a Perspective Article, Tourism Review, Vol. 76, No. 1, pp. 38-42, 2020.

[6] N. Partarakis, D. Kaplanidi, P. Doulgeraki, E. Karuzaki, A. Petraki, D. Metilli, V. Bartalesi, I. Adami, C. Meghini, and X. Zabulis: Representation and Presentation of Culinary Tradition as Cultural Heritage, Heritage 2021, Vol. 4, No. 2, pp. 612-640, 2021.

[7] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar: SemEval-2014 Task 4: Aspect Based Sentiment Analysis, in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 27–35, 2014.

[8] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos: SemEval-2015 Task 12: Aspect Based Sentiment Analysis, in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 486–495, 2015.

[9] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Q., O. D. Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit: SemEval-2016 Task 5: Aspect Based Sentiment Analysis, Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 19-30, 2016.

[10] H. Yang and K. Li, PyABSA: Open Framework for Aspect-based Sentiment Analysis, arXiv:2208.01368 [cs.CL], 2022.