Automatic Detection of Geo-tagged 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 culture of the chosen destination. We aspire to develop a system that, by utilizing geo-tagged videos, enhances understanding of global cuisines and enriches the enjoyment of food tourism. We propose a method to automatically detect food-related videos from YouTube. Our method 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. Experiments conducted to test the effectiveness of the proposed method show that the proposed method outperforms the method using ChatGPT-4. Using the proposed method, we constructed a system to map food-related videos detected from YouTube.

Keywords

Aspect-based sentiment analysis, Food tourism, YouTube

1. Introduction

A wide variety of foods exist in various parts of the world. Food tourism is the exploration of food as the purpose of tourism. Food culture encompasses many things, from how to choose ingredients, how to plan a menu, how to cook, how to choose tableware, with whom to eat, how to eat, manners and etiquette, and so on. Restaurant search sites are often used to find information on food. However, information obtained from such search sites alone is not sufficient in terms of enjoying food culture and learning how local foods are evaluated by travelers. On the other hand, travel blogs and SNS such as image-sharing sites can be another source of information on the food culture of a destination. However, there has been no systematic compilation of information on food culture.

We aim to build a system that will help people understand more about food around the world and enjoy food tourism by using geo-tagged videos. This system is expected to deepen people's understanding of food culture and generate economic benefits through food tourism. Furthermore, by mapping this information on a map, we will construct a system that allows easy access to food information from around the world.

We use aspect-based sentiment analysis technology to analyze video comments to select videos related to food. Aspect–based sentiment analysis is a process that identifies aspect terms in the given text and evaluates the sentiment associated with each of these aspects. By looking at the aspect terms extracted from the video comments, we can determine if the video is food-related or not.

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.

RecSys Workshop on Recommenders in Tourism (RecTour 2023), September 19th, 2023, co-located with the 17th ACM Conference on Recommender Systems, Singapore

*Corresponding author.

nanba@kc.chuo-u.ac.jp (H. Nanba); fukuda.satoshi.3238@kc.chuo-u.ac.jp (S. Fukuda);

0000-0001-7191-9856 (H. Nanba);

2. Related Work

2.1. Aspect-based Sentiment Analysis

Aspect-based sentiment analysis is a task that identifies aspects and determines the polarity of each aspect when text is entered. 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 "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 sentiments of each aspect.

Various data sets for aspect-based sentiment analysis have been created. In SemEval workshop, a sentiment analysis task was conducted, and the data sets were released [Pontiki 2014, Pontiki 2015, Pontiki 2016]. SemEval uses review texts for laptops, restaurants, hotels, mobile phones, digital cameras, and museums as input for aspect-based sentiment analysis. The review texts were written in eight languages: English, Arabic, Chinese, Dutch, French, Russian, Spanish, and Turkish. Aspect-based sentiment analysis for restaurant review texts is also relevant to our study. However, some of the food-related videos include not only restaurants, but also the local food itself, which was introduced and evaluated. Therefore, our research needs a system that can be used universally, rather than an aspect-based sentiment analysis specific to SemEval's task.

Yang et al. [Yang 2022] develop an easy-to-use aspect-based sentiment analysis framework for beginners. In addition, they publish models built using data from a variety of aspect-based opinion analysis tasks[*]. We use this model to extract aspect terms from video comments and determine whether they are food-related.

2.2. Food Tourism

Partarakis et al. [Partarakis 2021] proposed a tool that allows the representation and presentation of the tangible and intangible dimensions of culinary traditions as cultural heritage, including their socio-historical context. They analyzed culinary traditions and the trends of food culture therein.

Fujii et al. [Fujii 2016] propose a method for automatically categorizing travel blog entries into five categories: Watch, Experience, Buy, Dine, or Stay. Using this classifier, they classify travel blog entries from foreign visitors to Japan and aggregate them by region to analyze how travelers behave in each region of Japan. The classification of Dine travel blog entries is related to our research topic, food tourism. We classify geo-tagged videos instead of travel blogs.

3. Automatic Detection of Geo-tagged Food-related Videos

We collect metadata of geotagged food-related videos from YouTube. Section 3.1 describes how to collect candidate food-related videos. Section 3.2 describes aspect-based sentiment analysis to detect food-related videos among the candidates collected in Section 3.1. Section 3.3 describes how to create a list of food-related terms necessary to determine whether a video is food-related using the results of aspect-based sentiment analysis. Section 3.4 proposes a method for detecting food-related videos using the results of aspect-based sentiment analysis and the food-related term list.

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

3.1. Automatic Collection of Candidate Videos about Food

To automatically collect terms related to food, we focused on the Wikipedia template, which contains entries for food name and place of origin†. We first extracted these items from Wikipedia. As a result, 2,404 items of information were obtained. For the birthplace, latitude and longitude information was manually added.

Next, videos were searched using the YouTube Data API. The name of the food, latitude and longitude information were entered in the search, and up to 30 videos were searched for each food within a 300 km‡ radius of the latitude and longitude information. For each video, we also collected the title, description, comments, and latitude and longitude information given to each video. As a result, we were able to collect 31,019 videos.

3.2. Aspect-Based Sentiment Analysis of Video Comments

Aspect-based sentiment analysis is performed on the comments collected using the method described in section 3.1, using PyABSA [Yang 2022]. Video comments are written in multiple languages, and since PyABSA supports multiple languages, the analysis itself is possible; however, considering the processing after the aspect-based opinion analysis, we used the M2M100 machine translation model§. Figure 2 shows an example of analysis by PyABSA. In the figure, "food" is extracted as an aspect, and the sentiment of "food" is positive.

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

Figure 1: Example of Analysis of Aspect-based Sentiment Analysis using PyABSA

Although some videos are searched because food terms happen to be included in the title or description, it is possible to determine whether a video is 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 in order of frequency from a YouTube video**. These aspect terms were analyzed using MyABSA. This video introduces food in the German city of Aachen. The numbers in parentheses indicate the frequency of each aspect term in the comments.

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 chips 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), ginger bread cookies (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 this video is about food. To automate this inference, we need to create a list of food-related terms. In the next section, we describe how to create the list.

† 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.

‡ The 300 km was determined based on the approximate size of a "state" or "province." However, since some foods are eaten over a wider or narrower area, there is room for further work.

§ https://huggingface.co/docs/transformers/model_doc/m2m_100

** https://www.youtube.com/watch?v=_AJeC4nVpYg

3.3. Creating a List of Food-related Terms

We use travel blog data collected from TravelBlog[††], which is one of the largest travel blog web sites, to automatically collect food-related terms. It hosted over 700,000 blog entries. In travel blogs, for example, there are 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, it is thought that terms related to food can be easily collected. In the case of the above sentence, using spaCy[‡‡], a Python library for natural language processing, we can extract "a hamburger" from the parsed result as shown in Figure 3.

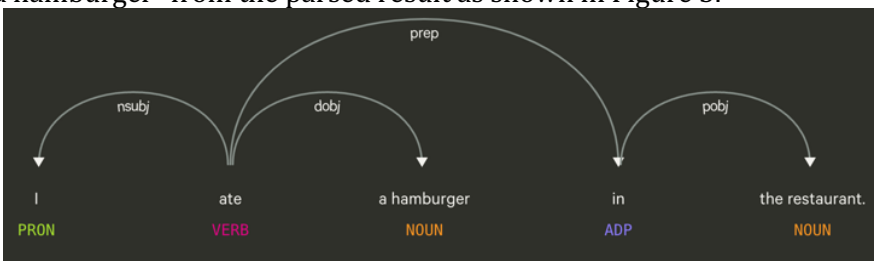


Figure 3: Syntactic Analysis Result of a Sentence "I ate a hamburger in the restaurant." using spaCy

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

As a result of syntactic analysis, there is a possibility that wrong terms are extracted as food-related terms. Therefore, from the extracted terms, terms that are considered inappropriate are automatically detected and excluded. For this detection, we used the language model BERT [Devlin 2019], whose usefulness has recently been confirmed in various natural language processing tasks. BERT takes a large amount of text data as input, masks some of the words, and learns to correctly estimate the terms. Using the BERT model, we collect food-related terms by the following procedure. To illustrate this procedure, we use the example shown in Figure 4.

(1) "I even ate sushi and watched MTV while there." (extract "sushi" as the object by spaCy)
(2) "I even ate [Mask] and watched MTV while there." (mask the object)
(3) Infer the masked word by BERT -> pizza, dinner, lunch, popcorn, breakfast

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

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

On the other hand, if spaCy did not correctly extract the object term, BERT would probably output a term not related to food. Therefore, we compare the word sets predicted by BERT for each sentence and exclude those that do not have any terms in common with the other sets.

We have arranged the terms thus collected in order of frequency, and the results are shown in Figure 5. The numbers in parentheses indicate the frequency of each word. As can be seen from Figure 5, most of the terms in the list are food-related, but some of them, 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.

[††] https://www.travelblog.org
[‡‡] https://spacy.io/
[§§] https://countwordsfree.com/stopwords

3132	lunch	274	lots	167	cake	99	supper
2434	dinner	244	sandwiches	162	chicken	95	eggs
1903	food	228	cream	156	rice	90	grass
1704	breakfast	225	heart	139	cheese	86	foods
717	meal	223	bread	139	chocolate	75	leaves
541	fish	195	buffet	134	pasta	74	sushi
503	pizza	192	sandwich	116	salad	74	snacks
473	lot	176	steak	108	gelato	74	bit
457	meat	174	fruit	107	seafood	73	noodles
299	meals	173	soup	100	restaurant	71	cookies

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 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 use the frequency of occurrence of the food-related term list as the confidence level of the term. We compute the score of video *m* using the following formula and judge *m* as a food-related video if the score is 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. The 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 Procedure 1. Of these, 274 were actual food-related videos.

Evaluation Measure
We evaluate using precision and recall.

Alternative Methods
To compare 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 by ChatGPT-4 using the following prompt instead of PyABSA 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, translate it into English.*

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 the videos were randomly judged as food-related, its precision value was 274/400 = 0.685, which is higher than that of Full-ChatGPT-4. In other words, we can conclude that Full-ChatGPT-4 performs worse than random.

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

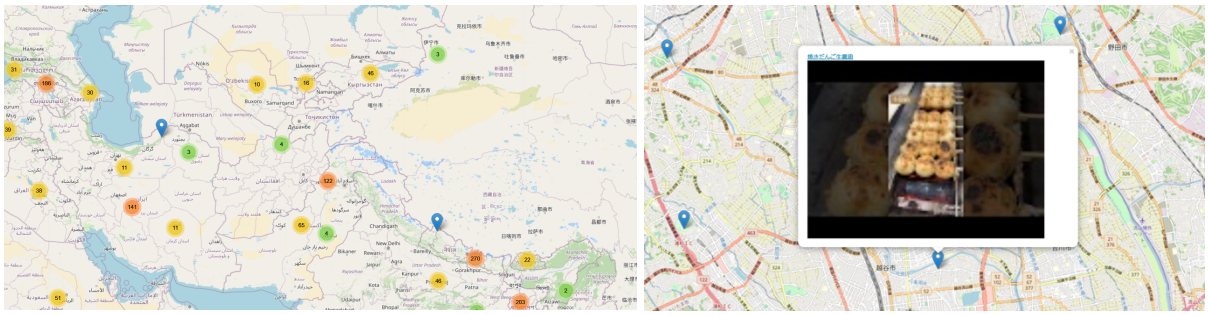
Methods	Precision	Recall
Our method	0.8476	0.3248
Full-ChatGPT-4	0.6703	0.6752
Aspect-ChatGPT-4	0.8182	0.2956

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

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

5. System Behavior

The data collected and analyzed using the method described in section 3 was mapped to YouTube videos on OpenStreetMap using the JavaScript Leaflet library[***]. Figure 6 (a) shows an example of how it works. Individual videos are displayed as pin icons on the map. On the other hand, 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)).



(a) Mapping geo-tagged videos with Leaflet (b) Video playback on a map

Figure 6: System Behavior

[***] https://leafletjs.com/

6. Conclusion

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

References

[1] [Devlin 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171–4186, 2019.

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

[3] [Partarakis 2021] Nikolaos Partarakis , Danae Kaplanidi , Paraskevi Doulgeraki , Effie Karuzaki , Argyro Petraki, Daniele Metilli, OrcIDValentina Bartalesi , Ilia Adami , Carlo Meghini and Xenophon Zabulis, Representation and Presentation of Culinary Tradition as Cultural Heritage, Heritage 2021, Vol. 4, No. 2, pp. 612-640, 2021

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

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

[6] [Pontiki 2016] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 19-30, 2016.

[7] [Yang 2022] Heng Yang and Ke Li, "PyABSA: Open Framework for Aspect-based Sentiment Analysis," arXiv:2208.01368 [cs.CL], 2022.