Tourist Behavior Analysis Using Instagram Hashtags

Daiju Kato
Nihon Knowledge Co. Ltd
Tokyo, Japan
d-kato@know-net.co.jp

Mitsuo Yoshida
Toyohashi University of Technology
Toyohashi, Japan
yoshida@cs.tut.ac.jp

Masaharu Hirota
Okayama University of Science
Okayama, Japan
hirota@mis.ous.ac.jp

Hiroshi Ishikawa
Tokyo Metropolitan University
Hino, Japan
ishikawa-hiroshi@tmu.ac.jp

Mitsuyoshi Nagao
Hokkaido Information University
Ebetsu, Japan
nagao@do-johodai.ac.jp

Abstract—**Many people use Instagram to submit their experiences related to travel destinations with many hashtag related keywords for their feelings and experiences. Those hashtags describe not only their experience,s but also some values related to the destination. Therefore, Instagram users search locations with those hashtags to obtain information about experiences, meals and sightseeing. Hashtags entail much value for tourists. We have started to find seasonal trends and a tendency of tourists' experiences from Instagram data. We regard it as effective for distributing information to tourists. This paper describes the accuracy and analytical method of tourists' behavioral analysis using hashtags.**

Keywords- e-tourism; hashtag; Instagram; social media; tourist behavior.

I. INTRODUCTION

Today, many tourists use and enjoy social media as Social Network Services (SNSs) before travel, during travel, and after travel. They search for sightseeing information reviews of travel destinations and post various information related to sightseeing spots visited by social media. The media provide accumulated reviews of sightseeing information and review blog sites such as TripAdvisor [1].

In Japan, social media of many kinds exist. Each medium has characteristic contents, as presented in Table I. Many users use several media to accommodate their feelings to post their experiences. Already, Instagram users [2] are more numerous than Facebook users [3] in Japan [4]. Instagram not only has many active users; it has many young female users. This social media site continues to expand.

Many researchers use social media data for tourist behavior or to provide specific travel information to achieve tourists' satisfaction and to attract interested tourists. We analyzed tourist behavior at sightseeing spots in Japan using social media and thought about using it for behavior analysis. This paper describes analyses of tourist behavior and trends using Instagram hashtags.

TABLE I. TYPES OF SOCIAL MEDIA IN JAPAN

	Social Media			
	Facebook	*Instagram*	*Twitter*	*Line*
Registered Users (August, 2018)	28 million	29 million	45 million	78 million
Types of contents	• Text • Carousel • Link • Image • Movie • Streaming • Story (24 hr limit)	• Image • Carousel • Movie • Story • Story by streaming	• Text (maximum of 280 characters) • Link • Image • Move • Streaming	• Text • Image • Link • Streaming (Line Live)
Feature	• Rich of contents • Formal attitude • Target accuracy	• Photographs and movies are main • Importance of a world view • Many active users • Hashtag	• Realtime property • Expectation of expandability • Hashtag	• Rich stamp • Having two of messages and timeline • Many mobile users with using talk or call

https://gaiax-socialmedialb.jp/post-29375/

II. PROPOSED ANALYTICAL METHOD FOR INSTAGRAM DATA

A. Target dataset

We analyze an example of tourists who visited Sapporo in Hokkaido, which, along with Kyoto, is famous as a Japan travel destination. We gather data about visiting Sapporo from Instagram. We search the tourist data in Instagram using the hashtag of '#札幌旅行 (Japanese, Sapporo Travel)'.

When browsing the data of the Sapporo travel hashtag, one can find posts for marketing use by restaurants, shops and so on in Sapporo city. These posts are not travelers. Therefore, we would like to raise the accuracy of tourist behavior analysis by suggesting a method of narrowing down travelers' submissions easily.

B. Proposed method specifically examining hashtag numbers

When extracting data with a hashtag of '#札幌旅行 (Japanese, Sapporo Travel)' from Instagram, we gather the following dataset.

- Period: 06/29/2013 – 12/19/2018
- Submission number: 8,229
- Number of unique submitting persons: 2,902

For these datasets, we ranked various factors and examined whether it was possible to conduct meaningful analyses. First, when ranking these datasets with users with many posts, as presented in Table II, the most frequent posts are those from owner id 4317813670, as shown in Figure 1. The post owner is a shop staff, posting information about handmade cookies to advertise them to Instagram users.

TABLE II.   RANKING OF POSTING USERS

Ranking	Owner Id	Submission number	Used hashtags number			
			total	*Min*	*Max*	*Avg*
1	4317813670	238	4,728	11	30	19.87
2	4027909537	198	5,902	23	31	29.81
3	2284144589	157	4,768	29	60	30.37
4	7583357588	138	2,918	8	29	21.14
5	5504659705	112	2,387	19	47	21.31
6	3839661803	99	1,544	8	25	15.60
7	4655960542	69	1,836	18	32	26.61
8	36184065	65	967	7	24	14.88
9	268071578	63	146	1	6	2.32
10	1958878171	58	1,282	14	30	22.10

TABLE III.   RANKING OF USERS WHO USE MANY HASHTAGS

Ranking	Owner Id	Submission number	Used hashtags number			
			total	*Min*	*Max*	*Avg*
1	2284144589	157	4,768	29	60	30.37
2	1981742792	16	482	30	31	30.13
3	6692218220	18	540	30	30	30.00
4	6231274057	11	330	30	30	30.00
5	229156278	5	150	30	30	30.00
6	290889411	5	150	30	30	30.00
7	280579655	10	299	29	31	29.90
8	1813842825	7	209	28	32	29.86
9	4027909537	198	5,902	23	31	29.81
10	1556307153	5	149	29	30	29.80

In addition, when ranking users with many hashtags used per post, as presented in Table III, one can find messages introducing stores, restaurants, and shops in the rankings. Owner Id 2284144589, which uses the greatest number of hashtags, is a lamb barbecue restaurant. The shop posts its time sales, as shown in Figure 2.

It is necessary to exclude users in marketing to improve the accuracy refinement of tourist-only datasets. However, because it is impossible to check and analyze large amounts of text data within the posts, we specifically examined the number of usage hashtags per post by one user and decided to see if it can choose whether to use marketing depending on the number of hashtags.

We have hypothesized that the similarity of sentences is high in postings by marketing users. We assume that persons who use Instagram for marketing use similar numbers of hashtags for their messages. To confirm that this hypothesis is correct, we choose to examine the similarity of sentences among posted users.

C. Investigation of sentence similarity

To obtain similarity of sentences by edit distance, the following are two methods for obtaining the edit distance.

- Levenshtein distance
- Jaro–Winkler Distance

The Levenshtein Distance is the distance represented by the minimum edit distance of a character string and another character string. Here we have a standardized Levenshtein distance. We calculate the similarity of sentences.

The Jaro–Winkler Distance calculates the distance from the number of characters that match in a character string different from a certain character string and the necessity of substitution. The Jaro–Winkler distance means that the possible values of the distance are 0 to 1. A larger distance value represents higher similarity between character strings.

The similarity between the previously submitted text and the next text was calculated on the posted sentence of the user who made two or more posts from the acquired dataset. We implemented these two editing distance algorithms and calculated the editing distance by posted users [5].

Using the edit distance, we graphed both the Levenshtein distance and the Jaro–Winker Distance similarity of the posted text. The texts submitted by users according to the number of posts never achieved higher similarity as the number of postings increased. Similarly, even if the number of uses of hashtags increases, the similarity of the submitted text will not increase, as shown in Figure 3. In other words, it turned out that the hypothesis that users who have many posts and hashtags described earlier are doing marketing activities is not applicable.

To find out if this hypothesis does not hold true, we examined whether it is specific to the dataset of '#札幌旅行 (Japanese, Sapporo Travel)', we calculated the same editing distance to the dataset of '#京都旅行 (Japanese, Kyoto Travel)'. The newly acquired dataset is the following.

- Period: 08/06/2011 – 01/08/2019
- Submits number: 303,013
- Unique number of submitted persons: 71,484

As in the case of Sapporo travel, the similarity of the posted text by post number and by hashtags was calculated, as shown in Figure 4.

For this dataset of Kyoto trips, even though the number of posts and hashtags usage increased, it was not apparent that the similarity of the text being posted would be high. Based on the number of hashtags of Instagram, it turned out that it can not readily exclude postings that are made as marketing tasks.

We speculated that shops and restaurants that use Instagram for marketing are repeatedly posting similar texts. However, a guess can not be formulated even if the editing distance is obtained. At present, we can not exclude posts by restaurants and shops using marketing unless full-text analysis is used.

III. Tourist Behavior Analysis

According to Section II, it is not possible to exclude marketing activities using the number of hashtags for post messages under Instagram. We use the acquired data as is and analyze the tourist behavior. The analytical method is applied as follows and analysis is made as to whether there is periodicity.

1. Quarterly ranking of hashtags included in posting messages at 2018
2. Quarterly classification tendency of posted photographs at 2-18
3. Quarterly hashtags ranking of posts containing food in posted photographs at 2018
4. Trend analysis for the "Shime Parfait"

First, we created quarterly rankings of hashtags in 2018 included in posting text and investigated whether the posting tendency differs depending on the time, as presented in Table IV. Many hash tags attached with hashtags of Sapporo travel indicate that Sapporo City is a city in Hokkaido; most post users use hashtags for location of Hokkaido, Sapporo and someplace of Sapporo area. Additionally, because the hashtags of "#札幌グルメ(Japanese, Sapporo gourmet)" are raised to the top, one can see that it is possible to enjoy a meal on a Sapporo trip.

When the data are acquired using Instagram, the JSON element data 'accessibility_caption' are included for each post. These data show what is included in the posted pictures by Instagram's own classification result as follows.

"accessibility_caption": "Image may contain: food"

It is possible to classify the posted photographs using this parameter. Therefore, we created a quarterly ranking of 'accessibility_caption' at 2018, as presented in Table V. By this ranking, one can see that the images are always on top with the food and the self-portraits.

Instagram does not disclose what technology is used for 'accessibility_caption' or how it analyzes information related to photos. Therefore, detailed analysis can not be performed using this parameter. However, because it can be understood simply whether the photograph is food, person, indoor or outdoor, it is thought that analysis by hash tag can be explored and developed further.

According to the ranking of hashtags attached to pictures of which food is taken, filter by accessibility_caption

parameter, the top places of course have many hashtags of place names, but also include hashtags of specific foods such as sweets, Lamb meat barbecue called "Genghis Khan," and ice cream and gelato, as presented in Table VI. The rankings of sweets and cafés are nearly equal. Therefore, it is thought that they are eating sweets at a cafe.

Hashtag analysis can be used to check trends. For instance, magazines and web contents for Sapporo trips are written about "Sapporo Shime Parfait," a cold and sweet parfait that people can enjoy after a night of drinking. Since the Sapporo Parfait committee was established at September 1, 2015, it has advertised this item on its Web site [6].

Comparing "Shime Parfait" to the famous food of Sapporo, "Genghis Khan," tourists have gradually come to enjoy "Shime Parfait" as depicted in Figure 5 with a post and impressions. In addition, the extraction of the number of postings is done while considering synonyms.

By analyzing the use of hashtags, we can analyze the reasons for tourism given by visitors to Sapporo. Because these datasets are not limited to tourists, specific hashtags such as soup curry, lamb meat barbecue, and ice cream are assumed to be posted as a marketing strategy, so asking for seasonality by quarterly ranking can be regarded as difficult.

IV. Conclusion

The active use of Instagram is expanding. Many users enjoy posting pictures and videos to Instagram when they visit some place. Therefore, one can analyze the purpose of tourists visiting sightseeing spots by using the hashtags assigned to Instagram data. However, analysis of hashtags to analyze what meals tourists are enjoying in Sapporo, and what photographs tourists are taking when they are alone is impossible. Analyzing images of the posted photographs or analyzing the text body is necessary to conduct a detailed analysis.

Instagram can also post to Facebook and Twitter [7] simultaneously. Therefore, instead of conducting an analysis using Instagram data only, using methods such as analyzing the behavior of tourists in conjunction with analysis of tweets with location information can provide data for tourism behavior analysis without building a complicated system.

V. Future Work

As a method of finding similar posts, it is possible to calculate Jaccard coefficients by extracting hashtags from the contribution text in a regular expression. We would like to find a method of eliminating information posted by companies and shops to gather more accurate tourist data.

Because it is possible to post from Instagram to Twitter simultaneously, we conducted an analysis of places related to meals and places where seasonal variation occurs. By combining tweet data with positional information in Sapporo and hashtag analysis of Instagram, a more precise analysis of tourist behavior in Sapporo can be conducted.

REFERENCES

[1] https://www.tripadvisor.com

[2] https://www.instagram.com

[3] https://www.facebook.com

[4] https://blog.comnico.jp/we-love-social/sns-users, in japanese

[5] G. Dan, "Algorithms on strings, trees, and sequences." Cambridge University Press,1997, ISBN 0-521-58519-8.

[6] https://sapporo-parfait.com/en/

[7] https://www.twitter.com

[8] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from Social Big Data: Implications for Smart Tourism Destinations," Information Processing & Management, vol. 54, issue 5, 2018, pp. 847-860.

[9] K. Wakayama, "Travel routes extraction from Tweets", "Journal of School of Foreign Languages, Nagoya University of Foreign Studies," vol. 50, 2016, pp. 167-177, in Japanese.

Figure 1.　Submitted example by person ranked first (Contents are in Japanese).
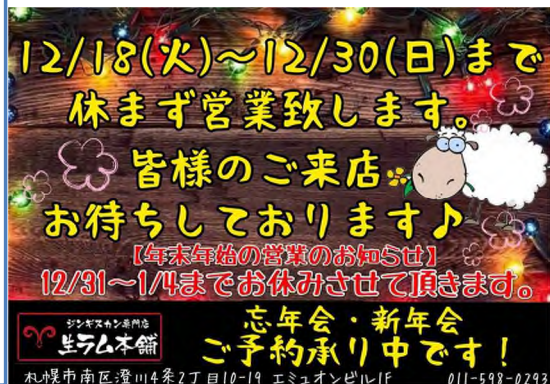

Figure 2.　Submitted example by person who uses the most hashtags (Contents are in Japanese).
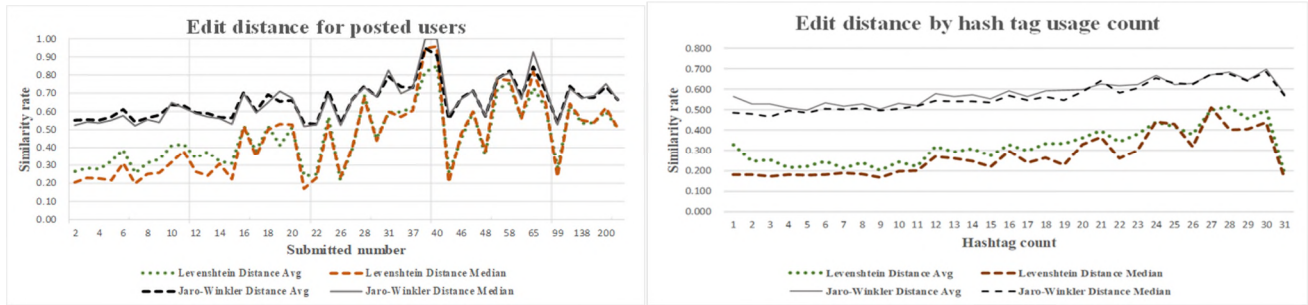
Figure 3.   Edit distance for posted users and by hash tag usage count.
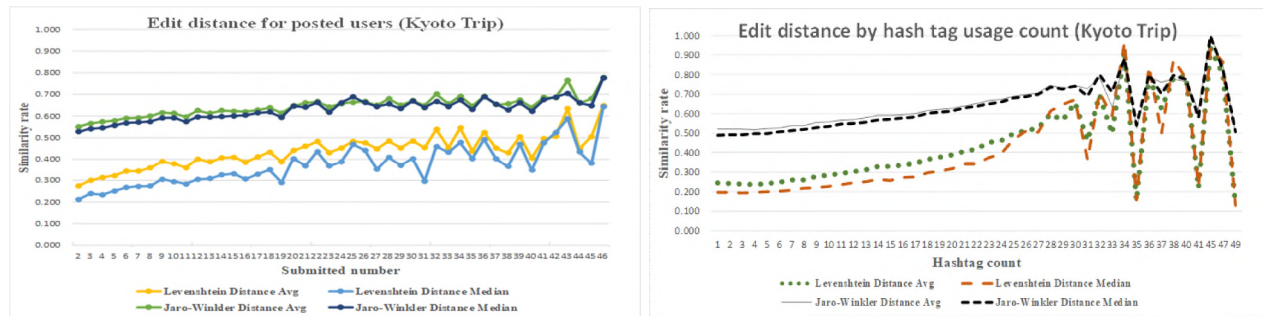

Figure 4.   Edit distance for posted users and by hash tag usage count (Kyoto trip).

TABLE IV.   QUARTERLY HASHTAGS RANKING AT 2018 (UPPER JAPANESE, LOWER TRANSLATION)

#	1Q		2Q		3Q		4Q	
	Submit No.	Hashtag	Submit No.	Hashtag	Submit No	Hashtag	Submit No.	Hashtag
1	782	#札幌 (#Sapporo)	1,076	#札幌 (#Sapporo)	1,528	#札幌 (#Sapporo)	2,154	#札幌 (#Sapporo)
2	484	#北海道 (#Hokkaido)	666	#rmet 北海道旅行 (#Hokkaido trip)	996	#北海道 (#Hokkaido)	1,676	#北海道 (#Hokkaido)
3	450	#sapporo	642	#北海道 (#hokkaido)	960	#北海道旅行 (#Hokkaido trip)	1,618	#北海道旅行 (#Hokkaido trip)
4	422	#北海道旅行 (#Hokkaido trip)	596	#札幌旅 (#Sapporo trip, abbreviation)	732	#札幌グルメ (#Sapporo gourmet)	1,108	#sapporo
5	388	#札幌旅 (#Sapporo trip, abbreviation)	578	#sapporo	696	#札幌旅 (#Sapporo trip, abbreviation)	1,050	#札幌観光 (#Sapporo sightseeing)
6	330	#札 (#Sappro abbreviation)	550	#札幌観光 (#Sapporo sightseeing)	680	#sapporo	942	#札幌グルメ (#Sappro gourmet)
7	316	#hokkaido	426	#札幌グルメ (#Sapporo Gourmet)	658	#札幌観光 (#Sapporo sightseeing)	706	#hokkaido
8	304	#北海 (#Hokkaido, abbreviation)	398	#札 (#Sapporo, abbreviation)	452	#札 (#Sapporo, abbreviation)	582	#北海道グルメ (#Hokkaido gourmet)
9	226	#札幌グルメ (#Sapporo gourmet)	344	#北海道旅 (#Hokkaido trip, abbreviation)	414	#北海道旅 (#Hokkaido trip, abbreviation)	538	#札幌旅 (#Sappro trip, abbreviation)
10	178	#北海道旅 (#sapporo trip, abbreviation)	320	#北海 (#Hokkaido, abbreviation)	384	#北海道グルメ (#Hokkaido gourmet)	490	#札 (#Sapporo, abbreviation)
11	154	#北海道グルメ	300	#札幌グル (#Sapporo gourmet,	368	#札幌スイーツ (#Sapporo sweets)	490	#札幌カフェ (#Sapporo café)

		(#hokkaido gourmet)		abbreviation)						
12	124	#北海道観光 (#Hokkaido sightseeing)	294	#北海道グル (#Hokkaido gurmet, abbreviation)	356	#札幌カフェ (#Sapporo café)	410	#すすきの (#Susukino)
13	122	#札幌グル (#Sapporo gourmet, abbreviation)	288	#すすきの (#Susukino)	338	#すすきの (#Susukino)	406	#旅行 (#Trip)
14	122	#trip	270	#hokkaido	338	#hokkaido	392	#北海 (#Hokkaido, abbreviation)
15	116	#japan	266	#札幌カフェ (#Sapporo café)	324	#北海 (#Hokkaido, abbreviation)	324	#trip

TABLE V.   QUARTERLY RANKING USING BY ACCESSIBILITY_CAPTION PARAMETER AT 2018

#	1Q		2Q		3Q		4Q		Total	
	Submit No.	*Value*	*Submit No.*	*Value*	*Submit No.*	*Value*	*Submit No.*	*Value*	*Submit No.*	*Value*
1	344	food	576	food	672	food	1,056	food	2,648	food
2	50	outdoor	64	food and indoor	82	sky and outdoor	136	sky and outdoor	314	sky and outdoor
3	38	indoor	62	sky and outdoor	76	drink	114	outdoor	278	outdoor
4	34	sky and outdoor	62	drink	76	1 person	102	food and indoor	268	food and indoor
5	28	food and indoor	54	indoor	74	food and indoor	90	indoor	244	indoor
6	26	text	54	outdoor	62	indoor	68	text	222	drink
7	24	food and text	46	drink and indoor	60	outdoor	66	night and outdoor	196	1 person
8	22	drink	42	1 person	50	drink and indoor	64	dessert and food	164	text
9	20	night and outdoor	36	dessert and food	38	dessert and food	64	night, sky and outdoor	146	dessert and food
10	20	night, sky and outdoor	32	text	38	text	62	drink	136	drink and indoor
11	18	1 person	28	1 person, food	34	1 person, food	60	1 person	122	night, sky and outdoor
12	16	drink and indoor	28	sky, tree and outdoor	28	sky, cloud and outdoor	56	sky, cloud and outdoor	116	1 person, food
13	16	1 person, outdoor	26	2 people	24	food and text	48	sky, tree and outdoor	112	sky, cloud and outdoor
14	16	sky, cloud and outdoor	24	drink and food	22	one or more people	42	1 person, food	110	night and outdoor
15	12	1 person, night and outdoor	24	one or more people	22	sky, tree and outdoor	38	1 person, outdoor	106	sky, tree and outdoor
16	12	1 person, food	20	people sitting, table and indoor	22	night, sky and outdoor	34	one or more people	92	one or more people
17	12	2 people	20	food and text	20	2 people	28	tree, sky, plant, outdoor and nature	84	2 people
18	12	text and food	16	night, sky and outdoor	14	people sitting, table and indoor	28	tree, plant, sky, outdoor and nature	74	food and text
19	12	1 person, sky and outdoor	14	one or more people and indoor	14	people sitting and food	26	2 people	68	1 person, outdoor
20	12	one or more people	14	night and outdoor	14	sky, cloud, tree and outdoor	26	table and indoor	68	drink and food

TABLE VI.   RANKING OF HASHTAGS CLASSIFIED BY ACCESSIBILITY_CAPTION PARAMETER AT 2018 (UPPER JAPANESE, LOWER TRANSLATION)

#	1Q		2Q		3Q		4Q	
	Submit No.	*Hashtag*	*Submit No.*	*Hashtag*	*Submit No*	*Hashtag*	*Submit No.*	*Hashtag*
1	58	#ジンギスカ (#Genghis Khan, abbreviation)	62	#プチギフト (#Smaill gift)	62	#グルメ (#Gurmet)	82	#スープカレー (#Soup curry)
2	54	#ラム	60	#ギフト	62	#ジェラート	76	#旅行

|		(#Lamb)		(#Gift)		(#Gelato)		(#trip)
3	50	#ラ (#Lamb, abbreviation)	58	#candle	60	#アイスクリーム (#Icecream)	62	#寿 (#Bridal)
4	50	#道産 (#Hokkaido made)	58	#キャンドル教室 (#Candle class)	60	#Shonpy	62	#アイスクリーム (#Icecream)
5	50	#成吉思汗 (#Genghis Khan, displayed by Kanji-word)	58	#キャンドルトマト (#candle tomato)	60	#ハンドパフェ (#Hand parfait)	62	#ラーメン (#Ramen)
6	48	#ラムタ (#Ramta)	56	#candles	60	#修学旅行 (#School excursion)	60	#Shonpy
7	48	#生ラ (#Raw lamb, abbreviation)	56	#ハンドメイド (#handmade)	58	#八景島シーパラダイス (#Hakkeijima Sea Paradise)	60	#修学旅行 (#School excursion)
8	48	#生ラム本舗澄川 (#Raw lamb restrant Sumikawa)	52	#海鮮 (#Seafood)	58	#アイシングクッキー (#Icing cookie)	60	#大通 (#Oodori)
9	48	#ラムしゃ (#Ram Shabushabu, abbreviation)	46	#ブライダル (#Bridal)	52	#ジンギスカ (#Genghis Khan, abbreviation)	60	#ジェラート (#Gelato)
10	48	#芋焼 (#Roasting poteto)	44	#個室居酒屋 (#Private room tavern)	50	#icingcookies	60	#アイシングクッキー (#Icing cookie)
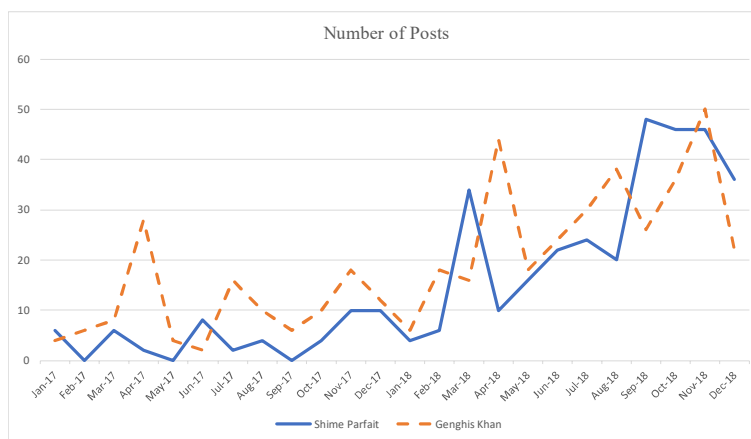
Figure 5.   Number of posts about "Shime Parfait" and "Genghis Khan".