Multimed Tools Appl
DOI 10.1007/s11042-018-5876-x

Sentiment-aware personalized tweet recommendation through multimodal FFM

Ryosuke Harakawa · Daichi Takehara · Takahiro Ogawa · Miki Haseyama

Received: 8 July 2017 / Revised: 18 February 2018 / Accepted: 8 March 2018

Abstract For realizing quick and accurate access to desired information and effective advertisements or election campaigns, personalized tweet recommendation is highly demanded. Since multimedia contents including tweets are tools for users to convey their sentiment, users' interest in tweets is strongly influenced by sentiment factors. Therefore, successful personalized tweet recommendation can be realized if sentiment in tweets can be estimated. However, sentiment factors were not taken into account in previous works and the performance of previous methods may be limited. To overcome the limitation, a method for sentiment-aware personalized tweet recommendation through multimodal Field-aware Factorization Machines (FFM) is newly proposed in this paper. Successful personalized tweet recommendation becomes feasible through the following three contributions: (i) sentiment factors are newly introduced into personalized tweet recommendation, (ii) users' interest is modeled by deriving multimodal FFM that enables collaborative use of multiple factors in a tweet, *i.e.*, publisher, topic and sentiment factors, and (iii) the effectiveness of using sentiment factors as well as publisher and topic factors is clarified from results of experiments using real-world datasets related to worldwide hot topics, "#trump", "#hillaryclinton" and "#ladygaga". In addition to showing the effectiveness of the proposed method, the applicability of the proposed method to other tasks such as advertisement and social analysis is discussed as a conclusion and future work of this paper.

Keywords Twitter · Recommendation · User modeling · Sentiment analysis · Field-aware Factorization Machines (FFM)

R. Harakawa · T. Ogawa · M. Haseyama
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Japan
Tel.: +81-11-706-6078
Fax: +81-11-706-7369
E-mail: {harakawa, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

D. Takehara
NTT DATA Corporation, Tokyo, Japan
E-mail: daichi.takehara0730@gmail.com

1 Introduction

In Twitter[1], more than 300 million active users are posting and sharing information daily through tweets with multimedia contents [31]. In this situation, there is a need for personalized tweet recommendation that enables recommendation of tweets corresponding to a user's interest. Both users and providers, who respectively receive and deliver information, would benefit from personalized tweet recommendation. For users, if successful personalized recommendation can be realized, quick and accurate access to the desired information would become possible [42, 48]. For providers, it would become possible to effectively decide target users for advertisements or election campaigns [44, 46].

Sharing actions, *i.e.*, re-tweets, are important clues to realize personalized tweet recommendation since they explicitly represent users' interest in the observed tweets. However, a traditional collaborative filtering technique, which is effective for a typical recommendation problem, does not work well due to the sparseness of user-tweet interactions in the large amount of generated tweets [21]. To overcome this difficulty, several methods [1, 5, 7, 13, 16, 45] use contextual information of tweets as well as user-tweet interactions. Publisher factors that represent users' interest in a publisher who posts a tweet are considered in the methods [5, 16]. Publisher factors come from an intuition that a user tends to prefer tweets posted by an authoritative publisher [5]. Publisher factors are defined on the basis of the profile information of a publisher such as the numbers of followees and followers. The methods proposed in previous papers [7, 13] use topic factors that represent topics found in the content of a tweet. Topic factors are defined by the frequency of words and hashtags in the text [13] and objects and contexts in images [7].

In addition to publisher and topic factors, users' interest is also induced by sentiment factors that represent positiveness/negativeness of each tweet. In fact, tweets that strongly represent a positive or negative state in the content tend to be re-tweeted more than neutral tweets by users. As shown in Fig. 1, the percentage of re-tweeted tweets in such tweets is 2.7% greater than that in neutral tweets in our crawled dataset, the details of which are shown in Section 4.1. Since sentiment factors were not taken into account in previous works [1, 5, 7, 13, 16, 45], there are cases in which the performance of the methods may be limited. For example, let us consider the tweets "Trump's policy is good." and "Trump's policy is bad." Although they are related to the same topic, "Trump's policy", their sentiment polarities are opposite. In such a case, since the difference between positive polarity and negative polarity was not considered in previous works, recommendation cannot be realized successfully.

To overcome this limitation, in this paper, we attempt to model users' interest by considering sentiment factors in tweets for personalized tweet recommendation. The main contributions of this paper are threefold:

- We have newly introduced sentiment factors into personalized tweet recommendation.
- We have successfully modeled users' interest by deriving multimodal Field-aware Factorization Machines (FFM) that enables collaborative use of multiple factors in a tweet, *i.e.*, publisher, topic and sentiment factors.

[1] https://twitter.com/

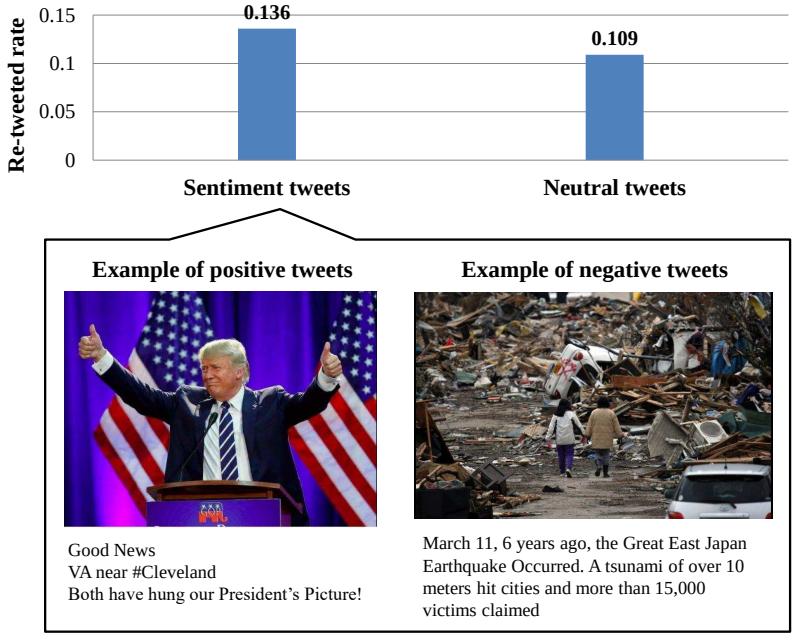


Fig. 1 Re-tweeted rate (re-tweeted tweets / not re-tweeted tweets) of sentiment and neutral tweets. We define sentiment tweets as ones that include a value of positive or negative polarity obtained by using Valence Aware Dictionary for sEntiment Reasoning (VADER) [17].

– We have conducted experiments using real-world datasets including tweets and users related to specific topics, "#trump", "#hillaryclinton" and "#ladygaga".

Specifically, first, we extract sentiment factors from the text and images of a tweet. From the text of a tweet, we estimate sentiment polarity (positive, negative or neutral) on the basis of Valence Aware Dictionary for sEntiment Reasoning (VADER) [17]. VADER is a sentiment analysis scheme that is suitable for microblog text. By applying VADER [17] to the texts of tweets, we can estimate valence scores, *i.e.*, the degree of positiveness/negativeness of each tweet, even if tweets include abbreviations, emoticons, acronyms and internet slang words. From images of a tweet, we extract visual sentiment concepts, which provide mid-level representation of sentiment included in images, by DeepSentiBank [6]. DeepSentiBank is a visual sentiment concept detection scheme based on deep convolutional neural networks. DeepSentiBank [6] enables prediction of the degree of each tweet including visual sentiment concepts, *i.e.*, adjective-noun pairs such as "beautiful flower" or "sad eyes". Second, in order to jointly utilize sentiment factors with publisher and topic factors, we adopt FFM [18]. FFM is a state-of-the-art prediction model that can incorporate rich features of items (tweets) by considering the differences in types of features. By realizing multimodal FFM that enables collaborative use of publisher, topic and sentiment factors via text and image analysis, successful recommendation based on users' interest becomes feasible. To the best of our knowledge, this work is the first work to combine multiple factors including sentiment factors via text and image analysis, although tweet recommendation has

been developed through text analysis [1, 5, 13, 16, 45] or image analysis [7]. Third, we clarify the effectiveness of using sentiment factors as well as publisher and topic factors through multimodal FFM by presenting results of experiments using real-world datasets related to worldwide hot topics, "#trump", "#hillaryclinton" and "#ladygaga".

The rest of this paper is organized as follows. Works related to the proposed method are described in Section 2. In Section 3, we explain our proposed sentiment-aware recommendation method through multimodal FFM. In Section 4, experimental results for real-world Twitter datasets are shown to verify the effectiveness of our method. Section 5 shows a conclusion and future work.

2 Related Works

We explain previously reported methods for personalized recommendation and ranking that are closely related to the proposed method.

Methods for personalized tweet recommendation using features of texts such as tweets and re-tweets have been proposed [1, 5, 7, 13, 16, 45]. Methods for understanding image tweets [8] and recommending tweets using visual features as well as textual features [7] have also been developed. To the best of our knowledge, however, methods for tweet recommendation that utilize other types of features have not been proposed.

In the field of personalized tag recommendation and ranking, multimodal analysis has been intensively studied. Shah *et al.* [37] proposed a method for ranking Flickr photo tags by voting relevance scores between tags and photos. They showed that voting with consideration of recall-based weights of multiple modalities including textual, visual and spatial features enabled successful tag annotation and ranking. Shah *et al.* [34] also proposed a method for personalized user tag recommendation. By fusing tag co-occurrence, neighbor voting and random walk-based approaches with visual, textual and spatial features and users tagging behavior, successful recommendation becomes feasible. Comprehensive reviews on personalized tag recommendation and ranking can be found in a textbook [40]. In this way, multimodal analysis [32, 33] using various types of features has improved the performance of applications such as recommendation and ranking [34, 37, 39, 40] as well as basic theories including temporal segmentation [36] and event recognition [35, 38].

Meanwhile, users' interest in tweets is induced by not only textual and visual features but also sentiment features since multimedia contents are tools for users to convey their sentiment [47]. Therefore, performance of personalized tweet recommendation will be improved if sentiment features are adopted. However, most of the existing methods for sentiment analysis for multimedia contents such as images, audio and videos are fundamental methods to predict sentiment in multimedia contents [4, 6, 17, 19, 29, 47]. In contrast, there has been a pioneer work [14] that used sentiment features for successfully realizing an application, *i.e.*, Web video retrieval. However, personalized tweet recommendation using sentiment features has not been proposed to the best of our knowledge.

Unlike the previously reported methods mentioned above, this work is the first work on collaborative use of visual, textual and sentiment features for successful

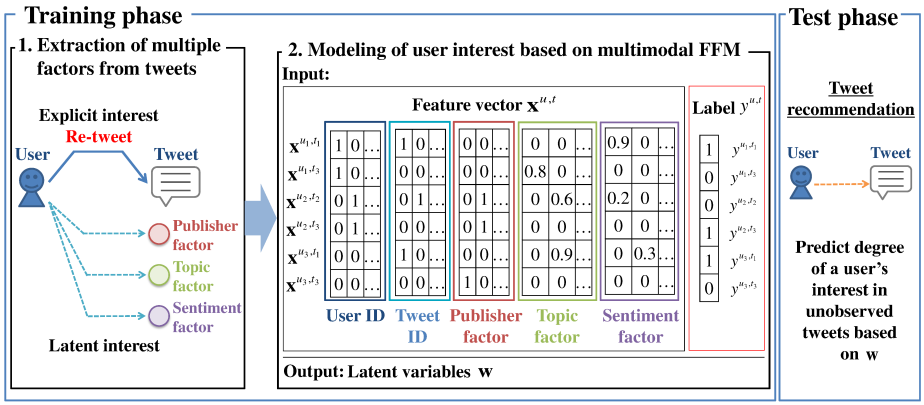


Fig. 2 Overview of the proposed method. In the training phase, we extract publisher, topic and sentiment factors, which induce users' interest in a tweet, from a tweet. By using the extracted factors, we model users' interest in tweets on the basis of multimodal FFM. In the test phase, we predict whether a user will be interested in an unobserved tweet or not based on the obtained model.

personalized tweet recommendation.

3 Sentiment-aware Personalized Tweet Recommendation through Multimodal FFM

In this section, we present the proposed method for sentiment-aware personalized tweet recommendation through multimodal FFM.

3.1 Problem Settings

In the personalized tweet recommendation problem, we model users' interest from the timelines observed by users including their re-tweet history in order to predict unobserved tweets in which each user would be interested. An overview of the proposed method is shown in Fig. 2. Table 1 shows the main symbols used in this paper.

In the training phase, we model users' interest from a dataset $\mathcal{D} = \{(\boldsymbol{x}^{u,t}, y^{u,t}) | u \in \mathcal{U}, t \in \mathcal{O}(u)\}$, where $\mathcal{U}$ is a set of users and $\mathcal{O}(u)$ is a set of tweets observed by u. Note that $\mathcal{O}(u)$ includes not only re-tweeted tweets but also un-retweeted tweets. Also, $\boldsymbol{x}^{u,t} \in \mathbb{R}^n$ is a feature vector obtained from an event that u observes t, and $y^{u,t}$ is a binary indicator that equals 1 if u re-tweets t and 0 otherwise. A tweet generally consists of its publisher, text and images, though text or images may be lacking. The proposed method extracts $\boldsymbol{x}^{u,t}$ from a user, a publisher, text and images (see Section 3.2). Moreover, from the dataset, we model users' interest to predict unobserved tweets in which each user would be interested (see Section 3.3).

In the test phase, from the input $\boldsymbol{x}^{u,\hat{t}}$ ($\hat{t}$ being an unobserved tweet), we predict the degree that u re-tweets $\hat{t}$, $y^{u,\hat{t}}$, on the basis of the learnt model. Finally, we recommend tweets to each user in descending order of the predicted degree.

Table 1 Main symbols used in this paper.

U	Set of users
$O(u)$	Set of tweets observed by a user $u \in U$
$\boldsymbol{x}^{u,t}$	Feature vector obtained from an event that $u \in U$ observes $t \in O(u)$
$y^{u,t}$	Binary indicator that equals 1 if u re-tweets t and 0 otherwise
$\boldsymbol{w}$	Latent variables obtained by our multimodal FFM to model users' interest
$\phi(\boldsymbol{w}, \boldsymbol{x}^{u,t})$	Prediction values by our multimodal FFM, *i.e.*, degree of u re-tweeting t when $\boldsymbol{w}$ and $\boldsymbol{x}^{u,t}$ are given

Table 2 Components of a feature vector and its label. $\boldsymbol{x}^{u,t}$ consists of five fields, *i.e.,* h_1, h_2, h_3, h_4 and h_5, and their elements. The fields and their elements are shown below.

Field	Element
h_1	User ID
h_2	Tweet ID
h_3	Publisher ID Indicator that shows whether a publisher is a verified account or not Tweet, listed, followee and follower counts
h_4	TF-IDF weights of words TF-IDF weights of hashtags Visual objects
h_5	Sentiment polarity TF-IDF weights of sentiment words Visual sentiment concepts

3.2 Extraction of Multiple Factors from Tweets

We extract multiple factors, which induce users' interest in a tweet, from a tweet t and represent them as a feature vector $\boldsymbol{x}^{u,t}$ for an event that a user u observes t. Since users' interest is induced by various factors, we extract multiple factors, *i.e.*, publisher, topic and sentiment factors. As shown in Table 2, we define different types of elements in $\boldsymbol{x}^{u,t}$ as fields, *i.e.*, h_1, h_2, h_3, h_4 and h_5, to model users' interest by jointly using multiple factors based on the multimodal FFM in Section 3.3. The details of publisher, topic and sentiment factors are shown below.

Publisher factors (h_3)

Publisher factors are based on users' interest in a publisher of a tweet. We represent the authority of a publisher as features. Publisher factors are effective for predicting whether a user is interested in a tweet or not, since a user tends to prefer a tweet posted by an authoritative publisher [5]. A part of the profile information becomes evidence of the authority of a publisher. Concretely, we use the tweet count (the number of tweets posted by a publisher), listed count (the number of public lists of which a publisher is a member), and followee and follower counts. Also, we use publisher ID and an indicator that shows whether

a publisher is a verified account or not.

Topic factors (h_4)
Topic factors are based on users' interest in topics included in the text and images of a tweet. Specifically, we calculate TF-IDF weights [30] of words and hashtags from the text of a tweet. Also, we extract visual objects, which are the top five prediction values of ImageNet [10] categories calculated by GoogLeNet [43], from images. The reason for using only the top five prediction values is to maintain sparseness and avoid noise. It should be noted that the combined use of textual information and visual objects has been shown to be useful for image tweet recommendation [7].

Sentiment factors (h_5)
Users' interest in a tweet will be induced by taking account of opinions, thoughts and feelings included in the text and images of a tweet. Therefore, we newly introduce sentiment factors into personalized tweet recommendation. Concretely, we calculate sentiment polarity from the text of a tweet by using VADER [17]. VADER is a sentiment analysis scheme attuned to microblog-like contexts and enables estimation of valence scores, *i.e.*, degree of positiveness/negativeness, from the text. To represent its sentiment state in more detail, we also use TF-IDF weights of sentiment words that are included in high-quality sentiment lexicon including abbreviations, emoticons, acronyms and internet slang words of VADER. We also extract visual sentiment concepts, which are the top five prediction values calculated by DeepSentiBank [6], from images of a tweet. DeepSentiBank enables prediction of the degree of each object in an image to sentiment concepts, *i.e.*, adjective-noun pairs such as "beautiful flower" or "sad eyes". It has been reported that images in a tweet are also related to opinions, thoughts and feelings in the tweet [8]. Therefore, we collaboratively use them to extract sentiment factors.

By using the above factors with user ID (h_1) and tweet ID (h_2), we represent a feature vector $\boldsymbol{x}^{u,t}$ of an event that a user u observes t. Furthermore, a label $y^{u,t}$, which represents whether u re-tweets t or not, is obtained together with $\boldsymbol{x}^{u,t}$. The details of $(\boldsymbol{x}^{u,t}, y^{u,t})$ are shown in Table 2.

3.3 Modeling of User Interest Based on Multimodal FFM

To predict whether a user will be interested in an unobserved tweet or not, we model users' interest in tweets through multimodal FFM, which enables collaborative use of publisher, topic and sentiment factors. Our work is the first attempt to combine multiple factors including sentiment factors via text and image analysis. We adopt FFM [18] due to the following two merits:

- FFM can incorporate rich features of tweets as well as user-tweet interactions unlike a traditional collaborative filtering technique. Therefore, the problem caused by sparseness in tweet recommendation can be solved.
- FFM can learn latent variables of each element in features with consideration of the difference between fields. Therefore, the proposed method can consider the difference of users' interest in each factor and accurately model interest when users observe tweets.

Training and test phases of our multimodal FFM to model users' interest are explained below.

Training phase

We learn the latent variables $\boldsymbol{w} \in \mathbb{R}^{n \times F \times K}$, where n is the number of elements of a feature vector, F is the number of fields and K is the model complexity, by solving the following optimization problem:

$$\min_{\boldsymbol{w}} \frac{\lambda}{2}||\boldsymbol{w}||_2^2 + \sum_{(\boldsymbol{x}^{u,t}, y^{u,t}) \in D} \log(1 + \exp(-\phi(\boldsymbol{w}, \boldsymbol{x}^{u,t}) y^{u,t})). \tag{1}$$

Here, λ is the regularization parameter and the predictor $\phi(\boldsymbol{w}, \boldsymbol{x}^{u,t})$ is defined as follows:

$$\phi(\boldsymbol{w}, \boldsymbol{x}^{u,t}) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\boldsymbol{w}_{i,f_j}^{\mathrm{T}} \boldsymbol{w}_{j,f_i}) x_i^{u,t} x_j^{u,t}. \tag{2}$$

Here, $f_i \in \{h_1, h_2, h_3, h_4, h_5\}$ are the fields of the ith element and $x_i^{u,t}$ is the ith element of $\boldsymbol{x}^{u,t}$. Also, $\boldsymbol{w}_{i,f_j} \in \mathbb{R}^k$ represents the latent vector of the ith element to learn the latent effect with the elements belonging to the field f_j. For example, when the ith and jth elements belong to h_3 and h_4, respectively, the dot product in Eq. (2) becomes $\boldsymbol{w}_{i,h_4}^{\mathrm{T}} \boldsymbol{w}_{j,h_3}$. In this way, it becomes possible to learn as many latent vectors as the number of fields per element of $\boldsymbol{x}^{u,t}$ unlike conventional FM [27] that learns only one latent vector per element of $\boldsymbol{x}^{u,t}$. Thus, our multimodal FFM enables calculation of rich features to model users' interest accurately.

To solve the optimization problem shown in Eq. (1), we use stochastic gradient descent [28] based on AdaGrad [11] in the same manner as in the paper [18]. At each step, we update $\boldsymbol{w}_{i,f_j}$ and $\boldsymbol{w}_{j,f_i}$ in Eq. (2) from a sample $(\boldsymbol{x}^{u,t}, y^{u,t})$. Concretely, we first calculate the sub-gradients as follows:

$$\boldsymbol{g}_{i,f_j} = \lambda \boldsymbol{w}_{i,f_j} + \kappa x_i^{u,t} x_j^{u,t} \boldsymbol{w}_{j,f_i}, \tag{3}$$

$$\boldsymbol{g}_{j,f_i} = \lambda \boldsymbol{w}_{j,f_i} + \kappa x_i^{u,t} x_j^{u,t} \boldsymbol{w}_{i,f_j}, \tag{4}$$

where

$$\kappa = \frac{-y^{u,t}}{1 + \exp(y^{u,t} \phi(\boldsymbol{w}, \boldsymbol{x}^{u,t}))}. \tag{5}$$

Next, we accumulate the sum of squared gradients for each coordinate $k = 1, \cdots, K$ as follows:

$$(G_{i,f_j})_k \leftarrow (G_{i,f_j})_k + (g_{j,f_i})_k^2, \tag{6}$$

$$(G_{j,f_i})_k \leftarrow (G_{j,f_i})_k + (g_{i,f_j})_k^2. \tag{7}$$

Finally, we update $(w_{i,f_j})_k$ and $(w_{j,f_i})_k$ as follows:

$$(w_{i,f_j})_k \leftarrow (w_{i,f_j})_k - \frac{\eta}{\sqrt{(G_{i,f_j})_k}} (g_{i,f_j})_k, \tag{8}$$

$$(w_{j,f_i})_k \leftarrow (w_{j,f_i})_k - \frac{\eta}{\sqrt{(G_{j,f_i})_k}} (g_{j,f_i})_k, \tag{9}$$

where η is the learning rate. The initial values of $\boldsymbol{w}$ are randomly sampled from a uniform distribution between $[0, 1/\sqrt{K}]$, and the initial values of G are set to one.

Algorithm 1 : Solving the optimization problem of multimodal FFM

Input: Training dataset $\mathcal{D} = \{(\boldsymbol{x}^{u,t}, y^{u,t}) \,|\, u \in \mathcal{U}, t \in \mathcal{O}(u)\}$
Output: Latent variables $\boldsymbol{w} \in \mathbb{R}^{n \times F \times K}$

```
Initialize the values of w from a uniform distribution between [0, 1/√K]
Let G ∈ R^{n×F×K} be a tensor of all ones
Run the following loop for m epochs
for (x^{u,t}, y^{u,t}) ∈ D do
  Calculate κ by Eq. (5)
  for i ∈ non-zero elements in {1, ··· , n} do
    for j ∈ non-zero elements in {i, ··· , n} do
      Calculate the sub-gradients by Eqs. (3) and (4)
      for j ∈ non-zero elements in {i, ··· , n} do
        Update the gradient sum by Eqs. (6) and (7)
        Update the latent variables by Eqs. (8) and (9)
      end for
    end for
  end for
end for
return w
```

In this way, we solve the optimization problem in Eq. (1). For detailed procedures, refer to Algorithm 1.

Test phase
In the test phase, we predict the degree that a user u re-tweets an unobserved tweet $\hat{t}$ by $\phi(\boldsymbol{w}, \boldsymbol{x}^{u,\hat{t}})$ through the latent variables $\boldsymbol{w}$ learnt from the training dataset. We recommend tweets to each user in descending order of the predicted degree.

It thus becomes feasible to model users' interest and recommend tweets to users on the basis of their own interest. By newly adopting sentiment factors, our method can consider the influence of opinions, thoughts and feelings on users' interest such as the difference between "Trump's policy is good." and "Trump's policy is bad." Also, we realize accurate modeling of users' interest by the combined use of publisher, topic and sentiment factors through the multimodal FFM.

4 Experimental Results

In this section, we verify the effectiveness of the proposed method through results of experiments using real-world datasets.

4.1 Datasets

First, we constructed datasets by collecting real-world Twitter data. In order to crawl users and their timelines, we took the following three steps:

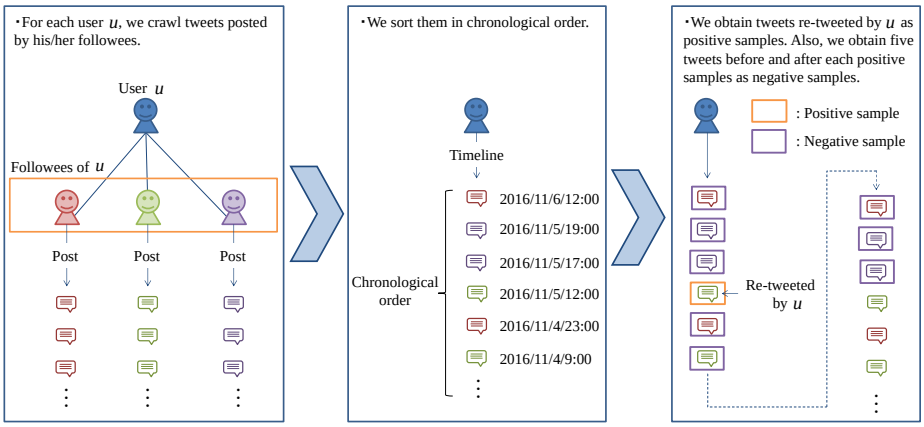


Fig. 3 Illustration of Step 3 "Collection of the timeline of each user" in the crawling procedures. In the right side of this figure, positive samples denote tweets corresponding to ground truth for the recommendation task, while negative samples represent tweets that are not the ground truth.

Step 1: Crawling of users and tweets from public streams

First, we crawled tweets with hashtags "#trump", "#hillaryclinton" and "#ladygaga" and their publishers over a period of time by Twitter Streaming API[2]. Note that we selected worldwide hop topics, which can presume divergence in sentiment polarity, in order to accurately verify the effectiveness of using sentiment factors.

Step 2: Sampling of active users

Next, we sampled active users from the crawled users and tweets. As shown in previous works [7, 45], sampling active users is a general procedure for removing users who hardly observe the timeline. Concretely, we sampled users who posted at least 100 tweets (including re-tweets) and had 100–1000 followees and followers. It should be noted that each user had at least one re-tweet.

Step 3: Collection of the timeline of each user

Finally, for each user u, we crawled tweets posted by his/her followees and sorted them in chronological order. We also obtained tweets that were re-tweeted by u as positive samples. We selected only un-retweeted tweets from five tweets before and after a re-tweeted one, and we defined the selected un-retweeted tweets as negative samples. In a previous work [7], it was also assumed that the user is interested in re-tweeted tweets and that he/she is not interested in several tweets before and after a re-tweeted one. In this way, we simulated the scanned timeline of each user. An illustration of this step is shown in Fig. 3.

The details of the constructed datasets are shown in Table 3. The statistics of sentiment tweets are shown in Table 4. It should be noted that we define sentiment tweets as tweets that include a value of positive or negative polarity based on

[2] https://dev.twitter.com/streaming/public/

Table 3 Details of the datasets.

Dataset ID	1	2	3
Hashtag	#trump	#hillaryclinton	#ladygaga
Crawling date	5–7 Nov. 2016	5–7 Nov. 2016	6–11 Mar. 2017
Num. of users	368	141	97
Num. of re-tweets in the training dataset	508,649	163,080	104,916
Num. of un-retweeted tweets in the training dataset	3,528,083	1,185,748	664,778
Num. of re-tweets in the test dataset	124,884	40,073	27,193
Num. of un-retweeted tweets in the test dataset	884,483	297,209	165,274

Table 4 Statistics of sentiment tweets in our crawled dataset. We define tweets that include a value of positive or negative polarity obtained via VADER [17] as sentiment tweets, and other tweets are defined as neutral tweets. The column "Re-tweeted rate" represents the rate of re-tweeted tweets in sentiment or neutral tweets.

	Re-tweets	Others	Re-tweeted rate
Sentiment tweets	654,034	4,817,401	0.136
Neutral tweets	316,114	2,886,953	0.109

the sentiment analysis scheme VADER [17]. From this table, we can see that re-tweeted tweets tend to include a sentiment state compared with others, supporting the importance of considering sentiment factors in tweet recommendation.

In this experiment, we applied our method to the users and their timelines in each dataset. The latest one fifth of tweets in the timeline of each user was used as a test dataset and the other tweets were used for a training dataset. As preprocessing, we applied stemming, lemmatization and stop word removal to the text in the datasets by using NLTK [3]. Also, some statistical counts of publishers, *i.e.*, tweet, listed, followee and follower counts, were standardized to make the mean and standard deviation 0 and 1, and we normalized them between 0 and 1, respectively.

4.2 Results

We first explain a metric used to evaluate our method (Section 4.2.1). In this experiment, we investigated the effectiveness of introducing sentiment factors into personalized tweet recommendation and the contribution of each factor (Section 4.2.2). We verified the effectiveness of utilizing multimodal FFM for personalized tweet recommendation compared with other modeling methods (Section 4.2.3).

Table 5 Comparison of incorporating factors. Evaluation metric is MAP. Comparative method "Baseline" used only user-tweet interactions and did not use publisher, topic and sentiment factors. Maximum values in each column are written with bold style.

Dataset ID	1	2	3	Mean
Baseline	0.154	0.275	0.363	0.264
Publisher	0.328	0.410	0.542	0.427
Topic	0.252	0.321	0.456	0.343
Sentiment	0.179	0.283	0.400	0.287
Publisher+Topic	0.362	0.408	0.554	0.441
Publisher+Sentiment	0.351	**0.425**	0.559	0.445
Topic+Sentiment	0.258	0.331	0.477	0.355
Publisher+Topic+Sentiment	**0.373**	0.414	**0.569**	**0.452**

4.2.1 Evaluation Metric

To evaluate our method, we use Mean Average Precision (MAP) [2] as an evaluation metric, which is widely used for a recommendation task. MAP is defined as follows:

$$\mathrm{MAP} = \frac{\sum_{u \in \mathcal{U}} \mathrm{AP}(u)}{|\mathcal{U}|}, \tag{10}$$

where

$$\mathrm{AP}(u) = \frac{\sum_{r=1}^{|\mathcal{O}(u)|} \mathrm{P@}r \times l_{u,r}}{|\mathcal{R}(u)|}. \tag{11}$$

In the above equations, $\mathcal{U}$, $\mathcal{O}(u)$ and $\mathcal{R}(u)$ are sets of users, tweets observed by u and tweets re-tweeted by u, respectively. Also, P@r is precision at rank r and $l_{u,r}$ is a binary indicator that represents whether u re-tweets the tweet at rank r or not.

4.2.2 Comparison of Each Factor

Table 5 shows MAP values for results via the multimodal FFM including the baseline that used only user-tweet interactions and did not use publisher, topic and sentiment factors. Here, the regularization parameter λ, the model complexity K and the learning rate η of each method were set through a grid search strategy. Also, we set the number of training epochs m for optimizing the model to five. From this result, we can confirm the effectiveness of using publisher, topic and sentiment factors in a tweet since MAP for the baseline is lower than those for the others. Moreover, we can see that the use of sentiment factors in addition to other factors produces a higher level of accuracy. Therefore, the effectiveness of using sentiment factors in personalized tweet recommendation has been confirmed. The proposed method that models users' interest by jointly using publisher, topic and sentiment factors shows the highest performance.

To further analyze the results, we investigated the users that tend to re-tweet sentiment tweets, which include the value of positive or negative polarity obtained from VADER. We ranked users of each dataset in descending order of the degree p_u^{sen} that a user u re-tweets sentiment tweets. Note that p_u^{sen} was defined as follows:

$$p_u^{sen} = \frac{\text{Num. of sentiment tweets re-tweeted by } u}{\text{Num. of tweets re-tweeted by } u}. \tag{12}$$

By using the rank based on p_u^{sen}, we calculated MAP@R_s as follows:

$$\text{MAP@}R_s = \frac{\sum_{r_s=1}^{R_s} \text{AP}(u_{r_s})}{R_s}, \tag{13}$$

where u_{r_s} represents a user at rank r_s based on p_u^{sen}. Note that the smaller r_s is, the higher is the degree p_u^{sen} of u_{r_s}. Thus, when R_s is small, MAP@R_s represents the mean of average precision of users who re-tweet sentiment tweets frequently. In Fig. 4, we show MAP@R_s of our method for higher rank users (R_s being from 1 to 30.). From the results, we can confirm that our method tends to show higher MAP@R_s when R_s is small. In other words, users who re-tweet sentiment tweets frequently show higher average precision. Therefore, our method is especially effective for users that tend to re-tweet sentiment tweets.

4.2.3 Comparison of Modeling Methods

We compare MAP values for results obtained by our multimodal FFM and baseline methods. We adopted well-known baseline methods for a latent factor model (FM [27] and SVD++ [20]), a neighborhood model (k-NN [20]) and a matrix factorization model (NMF [22]), whose details are shown as follows:

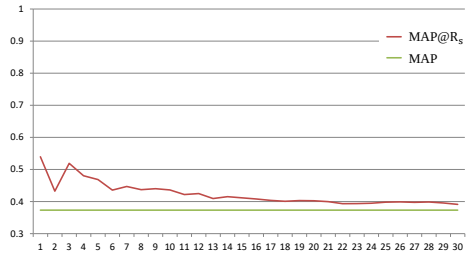
FM [27]
FM can incorporate rich features of tweets as can our multimodal FFM but cannot consider different types of features, *i.e.*, the difference between fields. In other words, FM uses publisher, topic and sentiment factors as the same field.

SVD++ [20]
This is an improved version of a well-known latent factor model SVD. SVD++ can incorporate implicit feedback unlike SVD. For the details, refer to Eq. (15) in the paper [20].

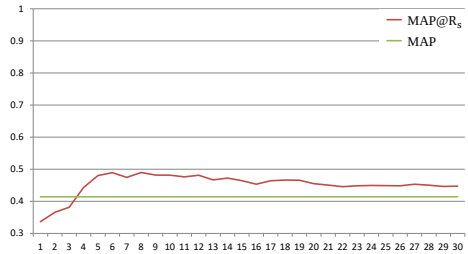
k-NN [20]
This is a neighborhood model that enables prediction of the unobserved rating by using similarities of neighborhoods. This model corresponds to Eq. (3) in the paper [20].

NMF [22]
This is a latent factor model based on non-negative matrix factorization (NMF), which enables accurate recommendation by guaranteeing non-negativity of latent variables.

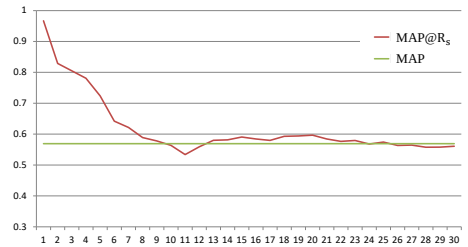
Figure 5 shows MAP values for results obtained by using our multimodal FFM and the above methods. The results show that the multimodal FFM outperforms the FM. In particular, the multimodal FFM tends to work well as the number of incorporating factors increases. As the number of factors used increases, candidates of latent variables that can be estimated from the factors become varied. In such a case, it is beneficial for accurate modeling of users' interest to automatically find optimal latent variables $\boldsymbol{w}$, which is shown in Eq. (1), from many latent effects between different factors by solving the optimization problem of the multimodal FFM. On the other hand, if the number of factors used is small, there may be a case in which optimal latent variables can be found without searching for latent effects between different factors. Thus, we consider that FFM rather than FM is suitable



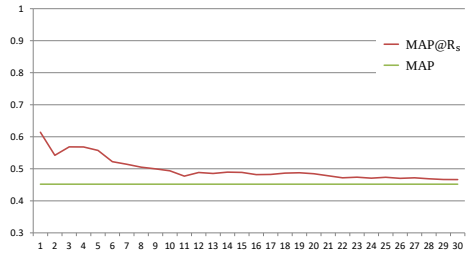
(a) Dataset 1 "#trump"



(b) Dataset 2 "#hillaryclinton"



(c) Dataset 3 "#ladygaga"



(d) Mean of datasets 1 "#trump", 2 "#hillaryclinton" and 3 "#ladygaga"

Fig. 4 MAP@R_s of our method. When R_s is small, MAP@R_s represents the mean of average precision of users who re-tweet sentiment tweets frequently.

for the proposed method since our method uses multiple factors. Meanwhile, there are cases in which FM is better than FFM for dataset 2 and the difference of the performance for dataset 1 is larger than datasets 2 and 3. We consider that this is because we searched the same range of parameters to select a suitable

one for all datasets and optimal parameters might not be selected for dataset 2. Improvement of the average performance was confirmed in the current parameter settings; however, the performance for dataset 2 will be improved if parameters can be adaptively determined according to its characteristics in the future. Also, our multimodal FFM that uses multiple factors outperforms SVD++ [20], k-NN [20] and NMF [22]. If we do not use factors, *i.e.*, content-based features, there is a case in which SVD++, k-NN and NMF are superior to our multimodal FFM and FM (see Fig. 5 (a)). However, since it is difficult to easily introduce the factors into SVD++, k-NN and NMF, performance improvement by adopting the factors may be difficult unlike our method. Therefore, the superiority of our multimodal FFM can be confirmed from the viewpoint of accuracy and applicability.

The above-described experimental results have confirmed that the proposed method enables successful recommendation of tweets to users on the basis of their own interest.

5 Conclusion and Future Work

In this paper, we have proposed a method for sentiment-aware personalized tweet recommendation through multimodal FFM. When users observe tweets, users' interest is strongly influenced by sentiment factors in the tweet, *i.e.*, opinions, thoughts and feelings included in its text and images. To overcome the problem of performance degradation that occurred in previous works without consideration of sentiment factors, our method models users' interest by deriving multimodal FFM that enables collaborative use of multiple factors in a tweet, *i.e.*, publisher, topic and sentiment factors. Experimental results for real-world Twitter datasets verified the effectiveness of our method. Notably, we showed that (1) the use of multiple factors increases the performance of personalized tweet recommendation, (2) our method is especially effective for users that tend to re-tweet sentiment tweets and (3) consideration of the difference between factors via the multimodal FFM is effective.

Finally, future work of this paper is described. For feature extraction, we should consider users' own information such as their profiles and their posted tweets to improve our sentiment-aware personalized tweet recommendation since such information will be useful for modeling users' interest. Also, we have room for improvement of extraction of visual sentiment concepts by DeepSentiBank [6]. It is reported that top-10 annotation accuracy of a 1,200 visual sentiment concepts, *i.e.*, adjective-noun pairs was 44.4% [6]. If the annotation accuracy can be improved, the final recommendation performance of our method will be improved. Thus, we should improve the annotation accuracy by developing a scheme such as construction of a CNN architecture that specializes in Twitter data. Meanwhile, in the experiment, we selected worldwide hot topics, which can presume divergence in sentiment polarity, in order to accurately verify the effectiveness of using sentiment factors. We will verify the effectiveness of our method for contraversive topics, which cannot presume divergence in sentiment polarity. On the other hand, since our method worked well for users that tend to deliver sentiment messages, our method could be used to monitor social issues such as "echo chamber effect" [9]. Furthermore, we are interested in applying our method to advertisement after performing more detailed analysis such as estimation of the best combinations of

images, text and sentiment polarity to maximize the number of re-tweets. Note that the aim of this paper is to propose a new scheme for accurate recommendation and to verify the performance improvement by the proposed multimodal FFM with consideration of sentiment factors. However, we should also consider security and privacy issues in the future. Methods for secure recommendation considering data access permission [26, 49], homomorphic encryption [12, 41], differential privacy [23] and privacy-preserving ratings [15] have been proposed. Furthermore, methods [24, 25] for detecting malicious attacks have been proposed to protect the security of recommendation systems and users privacy. In the future, we should develop a new method including schemes for protecting security and privacy that can preserve recommendation accuracy.

Acknowledgements This work was partly supported by JSPS KAKENHI Grant Numbers JP16J02042 and JP17H01744. We are grateful that a publisher, Springer permits us to deposit this accepted manuscript in the open access repository. The final publication is available at "https://link.springer.com/article/10.1007/s11042-018-5876-x".

References

1. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing user modeling on twitter for personalized news recommendations. In: Proc. Int. Conf. User Modeling, Adaptation, and Personalization, pp. 1–12 (2011)
2. Baeza-Yates, R.A., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1999)
3. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python, 1st edn. O'Reilly Media, Inc. (2009)
4. Borth, D., Chen, T., Ji, R., Chang, S.F.: Sentibank: Large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In: Proc. ACM Multimedia Conf., pp. 459–460 (2013)
5. Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., Yu, Y.: Collaborative personalized tweet recommendation. In: Proc. Int. ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 661–670 (2012)
6. Chen, T., Borth, D., Darrell, T., Chang, S.F.: Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. Computing Research Repository **abs/1410.8586**, 1–7 (2014)
7. Chen, T., He, X., Kan, M.: Context-aware image tweet modelling and recommendation. In: Proc. ACM Int. Conf. Multimedia, pp. 1018–1027 (2016)
8. Chen, T., Lu, D., Kan, M.Y., Cui, P.: Understanding and classifying image tweets. In: Proc. ACM Int. Conf. Multimedia, pp. 781–784 (2013)
9. Colleoni, E., Rozza, A., Arvidsson, A.: Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. Journal of Communication **64**(2), 317–332 (2014)
10. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 248–255 (2009)
11. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research **12**, 2121–2159 (2011)
12. Erkin, Z., Veugen, T., Toft, T., Lagendijk, R.L.: Generating private recommendations efficiently using homomorphic encryption and data packing. IEEE Trans. Information Forensics and Security **7**(3), 1053–1066 (2012)
13. Feng, W., Wang, J.: Retweet or not?: Personalized tweet re-ranking. In: Proc. ACM Int. Conf. Web Search and Data Mining, pp. 577–586 (2013)
14. Harakawa, R., Ogawa, T., Haseyama, M.: Extracting hierarchical structure of web video groups based on sentiment-aware signed network analysis. IEEE Access **5**, 16,963–16,973 (2017)

15. Hoens, T.R., Blanton, M., Chawla, N.V.: A private and reliable recommendation system for social networks. In: Proc. IEEE Int. Conf. Social Computing, pp. 816–825 (2010)
16. Hong, L., Doumith, A.S., Davison, B.D.: Co-factorization machines: Modeling user interests and predicting individual decisions in twitter. In: Proc. ACM Int. Conf. Web Search and Data Mining, pp. 557–566 (2013)
17. Hutto, C.J., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proc. Int. AAAI Conf. Weblogs and Social Media, pp. 216–225 (2014)
18. Juan, Y., Zhuang, Y., Chin, W.S., Lin, C.J.: Field-aware factorization machines for ctr prediction. In: Proc. ACM Conf. Recommender Systems, pp. 43–50 (2016)
19. Katsurai, M., Satoh, S.: Image sentiment analysis using latent correlations among visual, textual, and sentiment views. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp. 2837–2841 (2016)
20. Koren, Y.: Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 426–434 (2008)
21. Luo, X., Xia, Y., Zhu, Q.: Incremental collaborative filtering recommender based on regularized matrix factorization. Knowledge-Based Systems **27**, 271–280 (2012)
22. Luo, X., Zhou, M., Xia, Y., Zhu, Q.: An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. IEEE Trans. Industrial Informatics **10**(2), 1273–1284 (2014)
23. McSherry, F., Mironov, I.: Differentially private recommender systems: Building privacy into the netflix prize contenders. In: Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 627–636 (2009)
24. O'Mahony, M.P., Hurley, N.J., Silvestre, G.C.M.: Recommender systems: Attack types and strategies. In: Proc. National Conf. Artificial Intelligence - Volume 1, AAAI'05, pp. 334–339 (2005)
25. O'Mahony, M.P., Hurley, N.J., Silvestre, G.C.M.: Detecting noise in recommender system databases. In: Proc. ACM Int. Conf. Intelligent User Interfaces, pp. 109–115 (2006)
26. Peng, M., Zeng, G., Sun, Z., Huang, J., Wang, H., Tian, G.: Personalized app recommendation based on app permissions. World Wide Web pp. 1–16 (2017). DOI 10.1007/s11280-017-0456-y
27. Rendle, S.: Factorization machines. In: IEEE Int. Conf. Data Mining, pp. 995–1000 (2010)
28. Robbins, H., Monro, S.: A stochastic approximation method. The Annals of Mathematical Statistics **22**(3), 400–407 (1951)
29. Sager, S., Borth, D., Elizalde, B., Schulze, C., Raj, B., Lane, I., Dengel, A.: Audiosentibank: Large-scale semantic ontology of acoustic concepts for audio content analysis. Computing Research Repository **abs/1607.03766**, 1–10 (2016)
30. Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys **34**(1), 1–47 (2002)
31. Seltzer, E.K., Jean, N., Kramer-Golinkoff, E., Asch, D.A., Merchant, R.: The content of social media's shared images about ebola: a retrospective study. Public health **129**(9), 1273–1277 (2015)
32. Shah, R.R.: Multimodal analysis of user-generated content in support of social media applications. In: Proc. ACM Int. Conf. Multimedia Retrieval, pp. 423–426 (2016)
33. Shah, R.R.: Multimodal-based multimedia analysis, retrieval, and services in support of social media applications. In: Proc. ACM Int. Conf. Multimedia, pp. 1425–1429 (2016)
34. Shah, R.R., Samanta, A., Gupta, D., Yu, Y., Tang, S., Zimmermann, R.: Prompt: Personalized user tag recommendation for social media photos leveraging personal and social contexts. In: Proc. IEEE Int. Symp. Multimedia, pp. 486–492 (2016)
35. Shah, R.R., Shaikh, A.D., Yu, Y., Geng, W., Zimmermann, R., Wu, G.: Eventbuilder: Real-time multimedia event summarization by visualizing social media. In: Proc. ACM Int. Conf. Multimedia, pp. 185–188 (2015)
36. Shah, R.R., Yu, Y., Shaikh, A.D., Tang, S., Zimmermann, R.: Atlas: Automatic temporal segmentation and annotation of lecture videos based on modelling transition time. In: Proc. ACM Int. Conf. Multimedia, pp. 209–212 (2014)
37. Shah, R.R., Yu, Y., Tang, S., Satoh, S., Verma, A., Zimmermann, R.: Concept-level multimodal ranking of flickr photo tags via recall based weighting. In: Proc. ACM Workshop on Multimedia COMMONS, pp. 19–26 (2016)
38. Shah, R.R., Yu, Y., Verma, A., Tang, S., Shaikh, A.D., Zimmermann, R.: Leveraging multimodal information for event summarization and concept-level sentiment analysis. Knowledge-Based Systems **108**(Supplement C), 102 – 109 (2016)

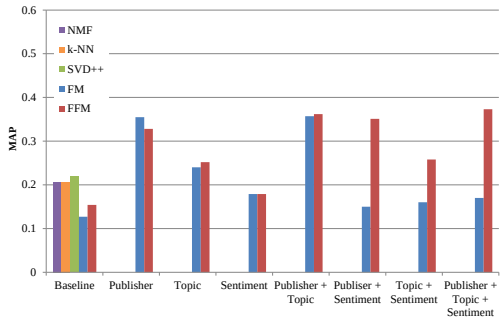
39. Shah, R.R., Yu, Y., Zimmermann, R.: Advisor: Personalized video soundtrack recommendation by late fusion with heuristic rankings. In: Proc. ACM Int. Conf. Multimedia, pp. 607–616 (2014)
40. Shah, R.R., Zimmermann, R.: Tag Recommendation and Ranking. Springer International Publishing (2017)
41. Soni, K., Panchal, G.: Data security in recommendation system using homomorphic encryption. In: Satapathy S, Joshi A. (eds) Information and Communication Technology for Intelligent Systems (ICTIS 2017), vol 1. Springer, Cham, pp. 308–313 (2017)
42. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: Proc. Int. ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 841–842 (2010)
43. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–9 (2015)
44. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. ICWSM **10**(1), 178–185 (2010)
45. Uysal, I., Croft, W.B.: User oriented tweet ranking: A filtering approach to microblogs. In: Proc. ACM Int. Conf. Information and Knowledge Management, pp. 2261–2264 (2011)
46. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: Proc. Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 115–120 (2012)
47. You, Q., Luo, J., Jin, H., Yang, J.: Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Proc. AAAI Conf. Artificial Intelligence, pp. 381–388 (2015)
48. Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., Li, X.: Comparing twitter and traditional media using topic models. In: European Conf. Information Retrieval, pp. 338–349 (2011)
49. Zhu, H., Xiong, H., Ge, Y., Chen, E.: Mobile app recommendations with security and privacy awareness. In: Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 951–960 (2014)

Ryosuke Harakawa received the B.S., M.S., and Ph.D. degrees from Hokkaido University, Japan, in 2013, 2015, and 2016, respectively, all in electronics and information engineering. He is currently a Post-Doctoral Fellow with the Graduate School of Information Science and Technology, Hokkaido University. His research interests include multimedia information retrieval and Web mining. He is a member of the IEEE, IEICE and Institute of Image Information and Television Engineers (ITE).

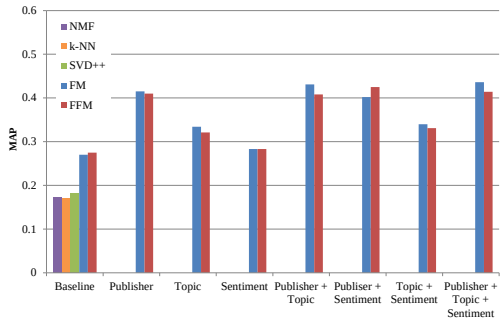
Daichi Takehara received the B.S. and M.S. degrees from Hokkaido University, Japan, in 2015 and 2017, respectively, all in electronics and information engineering. He joined NTT DATA Corporation, Japan in 2017. His research interests include audiovisual processing and Web mining.

Takahiro Ogawa received the B.S., M.S., and Ph.D. degrees from Hokkaido University, Japan in 2003, 2005, and 2007, respectively, all in electronics and information engineering. He is currently an Associate Professor with the Graduate School of Information Science and Technology, Hokkaido University. His research interests are multimedia signal processing and its applications. He has been an Associate Editor of the ITE Transactions on Media Technology and Applications. He is a member of the IEEE, ACM, EURASIP, IEICE, and Institute of Image Information and Television Engineers (ITE).

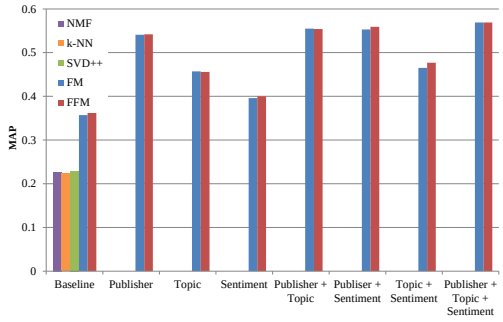
Miki Haseyama received the B.S., M.S., and Ph.D. degrees from Hokkaido University, Japan, in 1986, 1988, and 1993, respectively, all in electronics. She joined the Graduate School of Information Science and Technology, Hokkaido University, as an Associate Professor in 1994. She was a Visiting Associate Professor with Washington University, USA, from 1995 to 1996. She is currently a Professor with the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She is a member of the IEEE, IEICE, Institute of Image Information and Television Engineers (ITE), and Information Processing Society of Japan (IPSJ). She has been a Vice-President of the ITE, the Editor-in-Chief of the ITE Transactions on Media Technology and Applications, the Director of the International Coordination and Publicity of the IEICE.



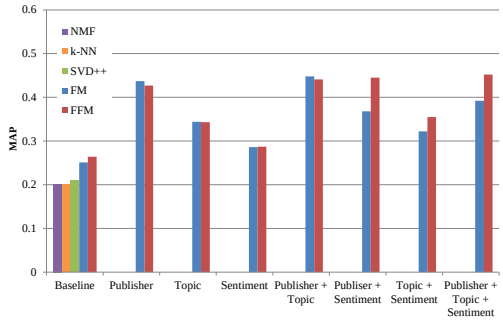
(a) Dataset 1 "#trump"



(b) Dataset 2 "#hillaryclinton"



(c) Dataset 3 "#ladygaga"



(d) Mean of datasets 1 "#trump", 2 "#hillaryclinton" and 3 "#ladygaga"

Fig. 5 Comparison of modeling methods. Our multimodal FFM is denoted by "FFM" and compared with FM [27], SVD++ [20], k-NN [20] and NMF [22]. Since SVD++, k-NN and NMF are methods that use only user-tweet interactions, their results are shown only in "Baseline".