Weibo Recommendation Method Based on User Extended Interests (Postprint)

Authors: Xu Jianmin, Mingyan Liu, Wang Miao

Date: 2018-07-23T00:00:00+00:00

Abstract

To address the issue of inaccurate interest extraction for Weibo users, this paper proposes a Weibo recommendation method based on user extended interest. This method combines user individual interest and associated interest as user extended interest for Weibo recommendation. Specifically, user individual interest is extracted from user tags, posted microblogs, and interacted microblogs; user associated interest is obtained through the strength of follow relationships, interaction frequency, and individual interest similarity between users and their followed users. Finally, the similarity between user extended interest and candidate microblogs is calculated, and a recommendation list is generated by sorting the similarities in descending order. Experimental results demonstrate that the proposed method is more effective and accurate than traditional methods.

Full Text

Preamble

Microblog Recommendation Method Based on Extended Interest of Users

Xu Jianming, Liu Mingyan, Wang Miao†

(School of Cyber Security & Computer, Hebei University, Baoding Hebei 071002, China)

Abstract: To address the problem of inaccurate extraction of microblog user interests, this paper proposes a microblog recommendation method based on extended user interest. This method combines individual interest and associated interest to form extended interest for microblog recommendation. Individual interest is extracted from user tags, posted microblogs, and interaction microblogs. Associated interest is obtained through the strength of following relationships, interaction frequency, and individual interest similarity between users and their

followees. Finally, the similarity between user extended interest and candidate microblogs is calculated, and a recommendation list is generated by sorting the similarity in descending order. Experimental results demonstrate that the proposed method is more effective and accurate than traditional methods.

Keywords: individual interest; associated interest; extended interest; microblog recommendation

0 Introduction

With the popularity of emerging social media, microblogging has become an important platform for people to share, disseminate, and acquire information [**?**]. The explosive growth of users has led to exponential increases in information on microblog platforms, exacerbating the problem of information overload [**?**]. Therefore, recommending microblogs that match users' potential interests has become particularly important.

The key to implementing microblog recommendation lies in extracting user interests [**?**]. Gao et al. [**?**] utilized the LDA topic model to infer the topic distribution of users' posted microblogs to obtain user interests. Wang et al. [**?**] employed the TextRank ranking method to extract keywords from users' posted microblogs as user interests. Zhou et al. [**?**] represented user interests by constructing user tag graphs. While these methods extracted user interests from users' own information and achieved certain recommendation effects, they failed to consider inter-user following relationships. Theoretically, following behavior can directly reflect users' interest orientation [**?**], making it feasible for mining user interests. For instance, Tan et al. [**?**] used the K-means clustering method to cluster microblogs posted by "special attention" users and extracted keywords from each category as user interests. Ma et al. [**?**] extracted keywords from users' posted microblogs, combined them with tags selected by users in the microblog system to construct a user tag matrix, and updated the matrix using tag correlation and user following relationships to obtain final user interests.

However, these methods only considered static following relationships between users and could not accurately measure the relationship strength between users and their followees. Moreover, followees may have different interests from the user, leading to inaccurate extraction of user interests. In addition to following behavior, microblog users frequently perform dynamic interaction behaviors such as liking, forwarding, and commenting on favored microblogs. Utilizing these behaviors for interest extraction can more accurately reflect user interests and better demonstrate the correlation degree between users compared to static following relationships.

To address these issues, this paper considers both following relationships and interaction behaviors between users, proposing a microblog recommendation method based on extended user interest. This method introduces switch vari-

ables and tuning parameters to fuse individual interest and associated interest to obtain extended interest. Individual interest is extracted from user self-information, including user tags, posted microblogs, and interaction microblogs. Associated interest is calculated through the correlation degree between users and their followees' individual interests.

1.1 User Individual Interest

User individual interest is generally described as the degree of preference for various interest words [**?**]. Keywords can be extracted from user tags, posted microblogs, and interaction microblogs to represent individual interest words, with keyword weights indicating preference degrees.

1.1.1 User Tag Representation

Users add tags based on their fields or interests [**?**], either by selecting from system-provided tags or by custom input. Tag content includes phrases that identify interests and identities, such as "travel," "music fan," and "basketball enthusiast." User u_i' s tags are represented as $P_{t_i} = \langle (t_{i1}, f_{t_{i1}}), (t_{i2}, f_{t_{i2}}), ..., (t_{ik}, f_{t_{ik}}) \rangle$, where t_{ik} is the k-th tag of user u_i and $f_{t_{ik}}$ is its weight. For the same user, all tags are considered equally important, so $f_{t_{ik}} = 1/k$, where k is the total number of tags for user u_i.

1.1.2 User Posted Microblogs Representation

Microblog users record daily life and express opinions through posted microblogs, which can indicate individual interests to some extent [**?**]. Since microblogs are typical short texts, keyword frequency statistics are not ideal. Therefore, this paper concatenates all recently posted microblogs of a user into a long text before extracting keywords.

User u_i' s concatenated long text from recently posted microblogs is represented as $P_{p_i} = \langle (p_{i1}, f_{p_{i1}}), (p_{i2}, f_{p_{i2}}), ..., (p_{ik}, f_{p_{ik}}) \rangle$, where p_{ik} is the k-th keyword extracted from the long text, $N_{p_{ik}}$ is the total number of keywords, and $f_{p_{ik}}$ is its weight calculated using Equation (1): $f_{p_{ik}} = \frac{N_{p_{ik}}}{\sum_{k=1}^{N_{p_{ik}}} N_{p_{ik}}}$, where $N_{p_{ik}}$ is the occurrence count of keyword p_{ik} in the long text.

1.1.3 User Interaction Microblogs Representation

User interaction microblogs refer to microblogs that users have liked, forwarded, or commented on. Different interaction behaviors reflect varying degrees of preference for microblogs. This paper concatenates recently liked, forwarded, and commented microblogs into three separate long texts P_{z_i}, P_{r_i}, and P_{c_i}, represented as: $P_{z_i} = \langle (z_{i1}, f_{z_{i1}}), (z_{i2}, f_{z_{i2}}), ..., (z_{ik}, f_{z_{ik}}) \rangle$, $P_{r_i} = \langle (r_{i1}, f_{r_{i1}}), (r_{i2}, f_{r_{i2}}), ..., (r_{ik}, f_{r_{ik}}) \rangle$, and $P_{c_i} = \langle (c_{i1}, f_{c_{i1}}), (c_{i2}, f_{c_{i2}}), ..., (c_{ik}, f_{c_{ik}}) \rangle$,

where z_{ik}, r_{ik}, and c_{ik} are the k-th keywords extracted from P_{z_i}, P_{r_i}, and P_{c_i}, respectively, and N_{z_i}, N_{r_i}, and N_{c_i} are the total numbers of keywords. The weights $f_{z_{ik}}$, $f_{r_{ik}}$, and $f_{c_{ik}}$ are calculated using Equation (1).

Merging keywords from P_{z_i}, P_{r_i}, and P_{c_i} with different weights yields user u_i' s interaction microblogs $P_{b_i} = \langle (b_{i1}, f_{b_{i1}}), (b_{i2}, f_{b_{i2}}), ..., (b_{is}, f_{b_{is}}) \rangle$, where b_{is} is the s-th keyword in user u_i' s interaction microblogs and $f_{b_{is}}$ is its weight calculated by Equation (2): $f_{b_{is}} = \alpha_1 \times w_l \times f_{z_{ik}} + \alpha_2 \times w_r \times f_{r_{ik}} + \alpha_3 \times w_c \times f_{c_{ik}}$, where w_l, w_r, and w_c are the weights for like, forward, and comment behaviors, respectively, with $w_l + w_r + w_c = 1$. The variables α_1, α_2, and α_3 are switch variables: $\alpha_1 = 1$ if b_{is} is extracted from user u_i' s liked microblogs, otherwise 0; $\alpha_2 = 1$ if b_{is} is from forwarded microblogs, otherwise 0; $\alpha_3 = 1$ if b_{is} is from commented microblogs, otherwise 0.

1.1.4 User Individual Interest Representation

Definition 1. User individual interest is interest mined from user self-information, represented as a binary tuple vector of individual interest words and their preference degrees: $P_{e_i} = \langle (w_{i1}, f_{w_{i1}}), (w_{i2}, f_{w_{i2}}), ..., (w_{ik}, f_{w_{ik}}) \rangle$, where w_{ik} is user u_i' s k-th individual interest word and $f_{w_{ik}}$ is its weight calculated by Equation (3): $f_{w_{ik}} = \beta_1 \times f_{t_{ik}} + \beta_2 \times f_{p_{ik}} + \beta_3 \times f_{b_{ik}}$, where β_1, β_2, and β_3 are switch variables indicating whether w_{ik} is extracted from user u_i' s tags, posted microblogs, or interaction microblogs, respectively. After normalization, the final weight $f'_{w_{ik}}$ for user u_i' s individual interest word w_{ik} is calculated by Equation (4): $f'_{w_{ik}} = \frac{f_{w_{ik}}}{\max\{f_{w_{i1}}, f_{w_{i2}}, ..., f_{w_{ik}}\}}$.

1.2 User Associated Interest

User associated interest is influenced by followees, with the degree of influence quantified through the correlation degree between users.

1.2.1 Inter-user Correlation Degree

The correlation degree between user u_i and user u_j reflects their mutual association, determined jointly by relationship closeness and individual interest similarity, calculated by Equation (5): $F_{ij} = \frac{1}{2}(G_{ij} + I_{ij})$, where G_{ij} is the relationship closeness between users u_i and u_j, determined by attention degree and interaction degree as shown in Equation (6): $G_{ij} = \frac{1}{2}(A_{ij} + S_{ij})$, and I_{ij} is the individual interest similarity between users u_i and u_j, calculated using cosine similarity.

The attention degree A_{ij} reflects the strength of following relationships, calculated by Equation (7): $A_{ij} = \begin{cases} 0, & \text{if } u_i \text{ and } u_j \text{ have no mutual attention} \\ 0.5, & \text{if } u_i \text{ and } u_j \text{ have one-way attention} \\ 1, & \text{if } u_i \text{ and } u_j \text{ have mutual attention} \end{cases}$.

Intuitively, among these three following relationships, the attention degree increases from no mutual attention to one-way attention to mutual attention. Accordingly, this paper sets the values of A_{ij} to 0, 0.5, and 1 for these three cases.

The interaction degree S_{ij} represents the frequency of likes, forwards, and comments between users, calculated by Equation (8): $S_{ij} = w_l \times SL_{ij} + w_r \times SR_{ij} + w_c \times SC_{ij}$, where SL_{ij}, SR_{ij}, and SC_{ij} represent the like interaction degree, forward interaction degree, and comment interaction degree between users u_i and u_j, respectively, calculated by Equations (9)-(11): $SL_{ij} = \frac{1}{2}\left(\frac{NL_{ij}}{NL_i} + \frac{NL_{ji}}{NL_j}\right)$, $SR_{ij} = \frac{1}{2}\left(\frac{NR_{ij}}{NR_i} + \frac{NR_{ji}}{NR_j}\right)$, and $SC_{ij} = \frac{1}{2}\left(\frac{NC_{ij}}{NC_i} + \frac{NC_{ji}}{NC_j}\right)$. Here, NL_{ij}, NR_{ij}, and NC_{ij} are the counts of user u_i liking, forwarding, and commenting on user u_j, respectively; NL_i, NR_i, and NC_i are the total numbers of users that user u_i has liked, forwarded, and commented on; and NL_j, NR_j, and NC_j are the corresponding totals for user u_j.

1.2.2 User Associated Interest Representation

Using the method in Section 1.2.1, we calculate the correlation degree between user u_i and their followees, filtering followees u_j with correlation degree greater than a threshold δ. **Definition 2.** User associated interest is interest mined from followees with correlation degree greater than threshold δ, represented as a binary tuple vector of associated interest words and their preference degrees: $P_{q_i} = \langle (q_{i1}, f_{q_{i1}}), (q_{i2}, f_{q_{i2}}), ..., (q_{ik}, f_{q_{ik}}) \rangle$, where q_{ik} is user u_i' s k-th associated interest word and $f_{q_{ik}}$ is its weight calculated by Equation (12): $f_{q_{ik}} = \sum_{j=1}^{n} \gamma_{ik} \times f_{w_{jk}}$, where γ_{ik} is a switch variable: $\gamma_{ik} = 1$ if q_{ik} and w_{jk} represent the same word, otherwise 0. After normalization, the final weight $f'_{q_{ik}}$ for user u_i' s associated interest word q_{ik} is calculated by Equation (13): $f'_{q_{ik}} = \frac{f_{q_{ik}}}{\max\{f_{q_{i1}}, f_{q_{i2}}, ..., f_{q_{ik}}\}}$.

1.3 User Extended Interest Representation

Definition 3. User extended interest is the harmonized result of individual interest and associated interest, represented as a binary tuple vector of extended interest words and their preference degrees: $P_{d_i} = \langle (d_{i1}, f_{d_{i1}}), (d_{i2}, f_{d_{i2}}), ..., (d_{ik}, f_{d_{ik}}) \rangle$, where d_{ik} is user u_i' s k-th extended interest word and $f_{d_{ik}}$ is its weight calculated by Equation (14): $f_{d_{ik}} = \gamma_1 \times \lambda \times f'_{w_{ik}} + \gamma_2 \times (1 - \lambda) \times f'_{q_{ik}}$. Here, γ_1 and γ_2 are switch variables indicating whether d_{ik} is user u_i' s individual interest word or associated interest word, respectively. λ is a tuning parameter with $\lambda \in [0, 1]$. As shown in Equation (14), when $\lambda = 0$, the weight of extended interest words equals the weight of associated interest words, representing user u_i' s associated interest. When $\lambda = 1$, $f_{d_{ik}} = f'_{w_{ik}}$, meaning the weight of extended interest words

equals the weight of individual interest words, representing user u_i' s individual interest.

---

2 Microblog Recommendation Method

For newly published microblogs, we calculate the cosine similarity between user extended interest and microblogs, sort microblogs by similarity in descending order, and recommend the top-N microblogs to users. A microblog is represented as $P_{m_t} = \langle (m_{t1}, f_{m_{t1}}), (m_{t2}, f_{m_{t2}}), ..., (m_{tk}, f_{m_{tk}}) \rangle$, where m_{tk} is the k-th keyword extracted from the microblog, $f_{m_{tk}}$ is its weight calculated using Equation (1), and k is the total number of keywords.

The specific process of our microblog recommendation method (UEI) is shown in Algorithm 1.

Algorithm 1: Recommending TOP-N Microblogs for Users

Input: User tag vector P_{t_i}, posted microblog vector P_{p_i}, interaction microblog vector P_{b_i}, candidate microblog vectors P_{m_t}, threshold δ, total number of followees n.

Output: User' s TOP-N recommendation list.

1. for $i = 1$ to n do
2. Calculate user u_i' s individual interest P_{e_i}
3. for $j = 1$ to n do
4. `if $F_{ij} > \delta$ then`
5. `  Calculate user $u_i$'s associated interest $P_{q_i}$`
6. `end if`
7. end for
8. Calculate user u_i' s extended interest P_{d_i}
9. end for
10. for $i = 1$ to n do
11. for $t = 1$ to m do
12. `Calculate similarity between $P_{d_i}$ and candidate microblog $P_{m_t}$`
13. end for
14. Generate recommendation list for user u_i
15. end for
16. return recommendation lists

3.1 Experimental Data

As there is currently no unified, authoritative microblog dataset available, this paper' s experimental data was collected using a crawler tool. First, 40 verified microblog users from eight domains (movie, game, music, food, finance, real estate, sports, and automobile) were selected as target users. Using a snowball sampling crawling strategy, we expanded one layer along the target users' followee chains, ultimately obtaining 3,087 microblog users and their microblog data (posted, liked, forwarded, and commented) from March 22, 2018 to June 21, 2018, totaling 137,945 microblogs. Additionally, the experimental data includes these users' tags and following relationships among them. To enable final microblog recommendations, we also collected microblogs published from June 22, 2018 to June 24, 2018.

3.2 Evaluation Standards

This paper employs Mean Reciprocal Rank (MRR) and Precision (P) as evaluation metrics for recommendation performance.

Mean Reciprocal Rank (MRR) represents the mean of the reciprocal ranks of the first correct microblog in top-N recommendation lists. Higher MRR values indicate that microblogs of interest appear closer to the top of recommendation lists, suggesting more reasonable recommendation ordering. MRR is calculated by Equation (15): $MRR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{rank_i}$, where n is the total number of target users and $rank_i$ is the position of the first correct microblog in user u_i' s recommendation list.

Precision (P) represents the proportion of microblogs that users are interested in within the recommendation list. Higher Precision values indicate higher accuracy of the recommendation method. Precision is calculated by Equation (16): $P = \frac{\sum_{i=1}^{n} N_{hit}^{i}}{\sum_{i=1}^{n} N_{rec}^{i}}$, where N_{hit}^{i} is the number of microblogs that user u_i is interested in within the recommendation list, and N_{rec}^{i} is the total number of recommended microblogs.

3.3 Parameter and Threshold Settings

1) Interaction Behavior Weights

The weights for like, forward, and comment behaviors are determined using the pairwise comparison matrix and consistency test method from the Analytic Hierarchy Process (AHP). Through pairwise comparison of w_l, w_r, and w_c, the judgment matrix shown in Table 1 is obtained. The maximum eigenvalue of the judg-

ment matrix is 3.0183, with corresponding eigenvector $(0.1862, 0.8527, 0.4881)$. After vector normalization, the standardized vector is $(0.1219, 0.5584, 0.3197)$. Therefore, the weights for like, forward, and comment behaviors are $w_l = 0.1219$, $w_r = 0.5584$, and $w_c = 0.3197$, respectively.

2) Correlation Degree Threshold

The correlation degree threshold δ is set as the minimum correlation degree between users and their followees. The minimum correlation occurs when users have a one-way following relationship, no interaction behaviors (likes, forwards, or comments), and completely different individual interests. First, using Equations (7) and (8), the attention degree is 0.5 and the interaction degree is 0 under these conditions. Substituting these into Equation (6) yields a relationship closeness of 0.25. Finally, substituting the relationship closeness of 0.25 and individual interest similarity of 0 into Equation (5) gives a correlation degree of 0.125. Therefore, the correlation degree threshold δ is set to 0.125.

3) Tuning Parameter

The parameter λ harmonizes the proportions of individual interest and associated interest. Larger λ values give more weight to individual interest, while smaller values emphasize associated interest. Different datasets require different λ values. For our experimental dataset, λ was determined through repeated experiments. Using manual annotation, we labeled microblogs of interest for 40 target users and calculated performance metrics for our recommendation method under top-20 recommendations with $\lambda = 0, 0.1, 0.2, ..., 0.9, 1$. The results are shown in Figure 1 [Figure 1: see original paper].

As shown in Figure 1, MRR and Precision both reach their maximum when $\lambda = 0.7$. Therefore, subsequent experiments set $\lambda = 0.7$. When $\lambda = 0$, user extended interest depends entirely on associated interest. When $0 < \lambda \leq 0.7$, as λ increases, the proportion of individual interest grows, interference from followees' individual interest gradually decreases, and both Precision and MRR increase, indicating improving recommendation performance. When $0.7 < \lambda \leq 1$, as λ increases, insufficient user self-data becomes more apparent, leading to gradually decreasing recommendation performance.

3.4 Experimental Comparison Results

To verify the effectiveness and accuracy of the microblog recommendation method, two comparison experiments were conducted on our dataset.

The first experiment compared the performance of recommendation methods based solely on User Individual Interest (UPI), User Associated Interest (UAI), and their fusion (UEI). Considering that different recommendation list lengths affect performance, we evaluated the methods with list lengths of 5, 10, 15, and 20. Figures 2 [Figure 2: see original paper] and 3 [Figure 3: see original paper]

show the MRR and Precision values for the three methods under top-5, top-10, top-15, and top-20 recommendations.

The results show: (a) The UPI-based method outperforms the UAI-based method, indicating that user associated interest serves only as a supplement to individual interest, and ignoring individual interest leads to suboptimal recommendation performance. (b) The proposed UEI method outperforms both other methods, demonstrating that combining individual and associated interests can more accurately mine user interests, alleviate the difficulty of extracting interests for inactive users, and produce more relevant recommendations.

The second experiment compared the performance of traditional Content-Based (CB) microblog recommendation, the tag-based microblog information recommendation method (ITCAUSR) proposed in [**?**], and our method (UEI). The CB method builds user interest vectors from posted microblogs. ITCAUSR constructs an initial user tag matrix from user tags and posted microblogs, then updates the matrix using tag correlation and inter-user following relationships to obtain user interests. Figures 4 [Figure 4: see original paper] and 5 [Figure 5: see original paper] show the MRR and Precision values for the three methods under top-5, top-10, top-15, and top-20 recommendations.

The results demonstrate: (a) ITCAUSR outperforms CB, indicating that considering inter-user following relationships improves recommendation effectiveness. (b) UEI outperforms ITCAUSR, showing that incorporating user interaction behaviors can more accurately extract user interests and distinguish correlation degrees between users and followees, achieving better recommendation performance.

---

4 Conclusion

This paper studied microblog recommendation methods and proposed a microblog recommendation method based on extended user interest by combining individual interest and associated interest. This method enables more accurate extraction of user interests and alleviates the difficulty of extracting interests for inactive users to some extent. Experiments demonstrate that our method achieves superior performance compared to previous CB and ITCAUSR algorithms. However, this paper only incorporates interests from followees to obtain extended interest, without considering the influence of users' fans and non-followed users with whom they interact. Therefore, determining whether these users' interests can be incorporated as part of extended interest will be the focus of future research.

References

[1] Ye Peng, Wang Changbo, Liu Yuhua, et al. Visual analysis of micro-blog retweeting using an information diffusion function [J]. Journal of Visualization, 2016, 19 (4): 823-838.

[2] Xu Yan, Zhou Meilin, Han Siyao. Feature representation for microblog followee recommendation in classification framework [C]// Proc of the 7th International Conference on Advanced Computational Intelligence. Piscataway, NJ: IEEE Press, 2015: 318-322.

[3] Zhang Junjie, Lei Yongmei. Improving content recommendation in social streams via interest model [J]. Computer and Information Science, 2015, 566 (1): 57-70.

[4] Gao Ming, Jin Cheqing, Qian Weining, et al. Real-time and personalized recommendation on microblogging system [J]. Chinese Journal of Computers, 2014, 37 (4): 963-975.

[5] Wang Ningning, Lu Ran, Wang Zhihao, et al. Microblog recommendation algorithm based on users' tag [J]. Application Research of Computers, 2017, 34 (1): 58-61.

[6] Zhou Xianke, Wu Sai, Chun Chun, et al. Realtime recommendation for microblog [J]. Information Sciences, 2014, 279: 301-325.

[7] Zhao Ling, Zhang Jing. Multi-dimensional analysis of microblog behavior research [J]. Information and Documentation, 2013 (5): 65-70.

[8] Tan Jinxiu, He Yue. Study on sina microblog personalized recommendation based on K-means text clustering [J]. Information Science, 2016, 34 (4): 74-79.

[9] Ma Huifang, Jia Meihuizi, Zhang Di, et al. Combining tag correlation and user social relation for microblog recommendation [J]. Information Sciences, 2017, 385-386: 325-337.

[10] Zhong Zhaoman, Guan Yan, Hu Yun, et al. Mining user interest on microblog based on profile and content [J]. Journal of Software, 2017, 28 (2): 278-291.

[11] Zhu Peisong, Qian Tieyun, Zhong Ming, et al. Inferring users' gender from interests: A tag embedding approach [J]. Neural Information Processing, 2016, 9950: 86-94.

[12] Liu Zhiyuan, Chen Xinxiong, Sun Maosong. Mining the interests of Chinese microblogs via keyword extraction [J]. Frontiers of Computer Science, 2012, 6 (1): 76-87.

[13] Jia Meihuizi. Research on microblog recommendation based on tags [D]. Lanzhou: Northwest Normal University, 2016.

Note: Figure translations are in progress. See original paper for figures.