Journal of Engineering Science and Technology
Vol. 13, No. 11 (2018) 3808 - 3820

SENTIGRADE: A SENTIMENT BASED USER PROFILING STRATEGY FOR PERSONALISATION

SIPRA SAHOO*, BIKRAM KESARI RATHA

SOA Deemed to be University, Bhubaneswar, Odisha, India
*Corresponding Author: sipracse@gmail.com

Abstract

Nowadays, the availability of folksonomy data is increased to make importance for user profiling approaches to provide results of the retrieval data or personalized recommendation. The approach is used for detecting the preferences for users and can be able to understand the interest of the user in a better way. In this approach, the incorporation of information with numerous data which depends upon sentiment is implemented using a framework SentiGrade by User Profiles (UP) and Resource Profiles (RP) for user Personalized Search (PS). From the folksonomy data, the discovery of User Preference (UsP) is presented by a rigorous probabilistic framework and relevance method are proposed for obtaining Sentiment-Based Personalized (SBP) ranking. According to the evaluation of the approach, the proposed SBP search is compared with the existing method and uses the two datasets namely, Movielens and FMRS databases. The experimental outcome of the research proved the effectiveness of the framework and works well when compared to the existing method. Through user study, the evaluation of approaches and developed systems are made which shows that considering information such as relevance and probabilistic data in Web Personalization (WP) systems can able to offer better recommendations and provide much effective personalization services to users.

Keywords: Folksonomy data, Personalization, Probabilistic framework, Recommendation, User Profiling.

1. Introduction

The web services for personalizing and also for building the recommendation systems, user profiling described as a vital procedure. The aggregation of the data associated with the user contain the information such as name, age, address, preferences, records of his/her past activities, in UP. On the contrary, focusing on building the interest and preferences of users, a contextual profile can be used which contains no identification information about user personal [1]. The UPs contains the record of topical categories or keywords that resembles the interest score of the users were replaced. In the place of keywords, the weighted ontologies (i.e., ontological profiles) are used in the UPs that allows users for exploiting the reference ontologies structure and able to model the interests of the user. The invasive methods (user-centered approach) may contain survey, questionnaires and interviews are created by the construction of UP that are recently shifted by non-invasive methods (session-based, ontology-based approach). Those methods depend upon the investigation of a large amount of user data [2]. The numerous online applications (e.g., Periscope, Glide, Vine, etc.), social networks such as Twitter, Facebook, and online publishing platforms like YouTube, Wikipedia, provides relevance feedback, social recommendation, and personalization as dominant mechanisms. The feedback provided by the personalized services with limited explicitly requested by the ultimate user for improving their Experience Quality (EQ) [3].

A research area has acquired more consideration in personalization, which is a key enabling technology of such system. The aim of the personalization is to provide information exchange information with users for their specific interests [4]. The variations of online behaviours face many issues are dealing the appealing approach known as personalization. There are also many individual differences appeared in information needs, user interests, query context, search goals and others. The research on the behaviour of user search is implemented by many methods and the profile of interest of the user are built by these methods depends upon the interactions of users on the Web [5].

The web experience of a personalized system of the user makes an action according to their taste is defined as WP. The 'firm initiated' is used by WP and includes preferences of individual rather than group interest, which is in contrast to the mass customization that is initiated by the customer. According to the browsing behaviour of the user, the WP provides the content and navigation information to each user and also able to provide demographic data to the related user [6]. The content description such as classification of products or items is involved as a major component of the systems. The system tried to match the best for UP by providing information to the users, profiling the user and filtering such as the content selection that is derived from collecting information about user behaviours and interests [7]. The system lacks adequate knowledge for providing the new user/items preferences with relevant suggestions, the problem of cold-start will arise.

The problem occurs because of the system's inability for gathering enough user information and the items as they have used in the past [8]. Because of prevailing of Web 2.0 communities, the generation of the user data by collaborative tagging system attains rapid growth in recent years. The understanding of user behaviours and preferences is critical because of assisting the customers in finding their desired resources effectively. In folksonomy-based systems, the vector of relevant tags modelling the users and resources are widely employed by a technique called Tag-based profile techniques [9]. The measurement of relevance between UP and RP is provided by a recommendation system and PS. The sentiment aspects of user-generated tags are neglected by the conventional measurements. The perceptions and feelings of resources are expressed by the users, which can be carried out by tags that are subjective and very emotional. Hence, sentiment relevance is considered in recent years for measuring the user-generated tags [10].

In this proposed method, the incorporation of a variety of sentiment-based information for PS, which is implemented using SentiGrade by embedding UP and RPs. In the framework, relevance framework and probabilistic modelling are proposed by ranking the sentiment using SBP ranking. The paper is composed of recent works and techniques related to our approach in Section 2. Section 3 described the UPs and RPs framework for PS by incorporating various sentiment-based techniques. Section 4 contains experimental setup and the evaluation of results, which is conducted by public database. At last, the conclusions made on Section 5.

2.Literature Review

The below section contains a number of interrelated areas which resemble like our work, and the section also discussed the recommendation system, modelling of UP and semantic similarity that are related to our paper are as follow.

Yau and Hristova, [11] presented a learning app for Java based on context-aware with personalized UP. The users by his own interest come from different nationalities, ages and knowledge of Java were evaluated through this app for a week. The app suggested learning materials for users according to their preferences and knowledge level, moreover, the app received positive feedback from the voluntary participants. The experiments were conducted on the app and the results proved that the minimum number of participants liked the app because of the Learning Objects (LO) that are depended on their contexts and profile. The app was deal with the single topic like Java and the method did not consider other topics. Therefore, only users who were having interested in Java alone would be allowed to use the app approach.

Vicente-Lopez et al. [12] implemented a weighting scheme that was developed for improving the personalization process by comparing six generic UP representations. The method was used to build the profiles moreover join the benefits of some of the existing techniques. The method was introduced to address the privacy problem of data and also improved to overcome the existing method problems such as reliability. The experimental results provided good performance on personalization when comparing with the original non-personalized retrieval system. The limitation of the method was the generic profiles used in this paper did not represent the real users.

Zhou et al. [13] constructed an enriched UPs for query expansion in a personalized manner by using external corpus. The word embedding was represented as the current state-of-the-art learning framework that was integrated with this model. Between documents and user annotations of user from the external corpus, the method was integrated with the topic model in documents, which is aligned by pseudo code. The two query expansions techniques were implemented by the method which depended upon UP. These two techniques were developed by topical relevance between the weights-enhanced word embedding, term and query. The approach performed well when compared to the traditional techniques such as both personalized query expansion methods. The approach captured fewer UPs because of the incorporation of information was less.

Xie et al. [14] proposed a SenticRank for incorporating sentiment information to numerous information based on sentiment by UPs and RPs for personalization search. The semantic-based personalized ranking was obtained by implementing the sentiment ranking methods such as content-based and collaborative-based approach. The drawback of the existing method was addressed in collaborative tagging system by incorporated this method with sentiment information. Various experiments were used for evaluating the performance of the approach in folksonomy dataset, moreover, the results of the experiments verified the effectiveness of the framework. The method needs more techniques for sentiment analysis to overcome the issues of ambiguity in tags.

Bansal et al. [15] planned to implement the UPs for content-driven for by extracting these latent interests of users and recommending the articles by using these interests. The method proposed a Collaborative Correspondence Topic Models (CCTM), that generated the UPs were leveraged for providing a personalized ranking of articles, which were comment-worthy. The CCTM method solved the problem of cold-start through these content-driven UPs without any need of the additional meta-data. The model affected by the inference problem was intractable with no off-the-shelf solution, in addition, the method developed an algorithm called Monte Carlo EM algorithm for handling the problem. The performance of the approach was less effective because of the time-consuming process to generate the reviews for the articles.

3. Proposed Methodology

In this section, the SentiGrade framework is proposed for incorporation of sentiments into tag-based profiles for searching the personalization of users. At most, the drawback of PS generated by profile tag is described by the approach. Then, the explanation of the basic approaches is discussed, after that the implementation of SentiGrade is also detailed and followed by them, the sub-processes like extraction of profiles based on tags, sentiment spaces are mapped and the resources of candidate are ranked. The basic architecture of the proposed method is described in Fig. A-1 (Appendix A).

3.1. Formulating the problem

The aim of the research in the relevance framework is finding the mapping function θ_m, that a collection of queries, resources, and users are mapped to a set of ranking scores, which is represented in Eq. (1),

$$\theta_m: U_s \times R_\epsilon \times Q_u \to S_r \quad (1)$$

where U_s, R_ϵ and Q_u are the set of users, resources, and queries respectively, S_r is the set of ranking scores.

This is a major stream of models to achieve PS, i.e., the method examines if the resource contents are relevant to intentions of query and UsPs or not. The mapping function θ_m involved in this category are divided into three sub-functions by the existing approaches that are described as in Eqs. (2), (3) and (4) as:

$$\theta_{m_1}: R_\epsilon \times Q_u \to S_{r_1} \quad (2)$$

$$\theta_{m_2}: R_\epsilon \times U_s \rightarrow S_{r_2} \quad (3)$$

$$S_{m_3}: S_{r_1} \times S_{r_2} \rightarrow S_r \quad (4)$$

where S_{r_1} describes the score of relevance between resources and queries, S_{r_2} represents the resource and users relevance score, θ_{m_1} and θ_{m_2} are the mapping function which is used for measuring the above scores, and θ_{m_3} is used to find the final scores S_r for resource ranking. The method uses a single equation for representing the following Eq. (5).

$$S_r \alpha \theta_m (U_s \times R_\epsilon \times T_l) \alpha \theta_{m_3} \left(\theta_{m_1}(Q_u \times R_\epsilon), \theta_{m_2}(U_s \times R_\epsilon) \right) \quad (5)$$

The research follow the above described model, after that the incorporation of relevance scores into function θ_m , where T represents the tag corpus and l describes the length of tag corpus.

3.2. Tag-based profiles extraction

Tags have represented the perceptions on the resources and also expressing the feelings by the user annotate resources in tagging systems. The ternary relationship between the users, resources, and tags are also created. Basically, the folksonomy data is formed by the collection of users, tags, resources and the relationship between ternary data. The four elements with tuple are represented as in Eqs. (6) and (7) as,

$$F_d = (U_s, R_\epsilon, T_l, K_c) \quad (6)$$

$$K_c \subseteq U_s \times R_\epsilon \times T_l \quad (7)$$

where K_c describes the collection of resources, tags, and users in ternary relationships.

The contents of UsPs (annotator) and resources (annotatee) to some extent is reflected by tags of user. Based on the above assumption, the major concept of profiles based on tags are extracted, which is represents as bag-tags that are adopted by the Vector Space Model (VSM). The relationship of ternary among tags, users and resources extracted the tag-based UP and resources profiles that are characterized by a vector of tags. The suitable mapping function θ_m is used for obtaining the relevance scores that are found out by the method after getting the profiles of user and resources.

3.3. Mapping to sentiment spaces

The incorporation of the sentiment relevance is introduced by the framework to solve the problem stated in the paper by mapping the sentiment spaces to tag spaces. The similarity measure is used to evaluate the relevance between UPs and RPs in sentiment space. In other words, from the tag space to the sentiment space, for each user the method maps the tag-based profiles. Thus, the approach can map all tag to sentiment spaces, and defined the sentiment-based UPs and RPs.

3.4. Ranking candidate resources

As said earlier, the method has two approaches such as relevance framework and probabilistic method. In this below part, paper validates the probabilistic approach in sentiment versions in details.

3.4.1. Probabilistic modelling framework of preference discovery

In this section, the method uses U_s represents as the all users, D_i describes all items that are favoured by the users in U_s, and W_t denotes the all tags collected which are associated with items inD_i. $\Lambda_{tt} \subseteq U_s \times D_i \times W_t$represented the favouring and tagging relation, which is collected from all observations. The anunified probabilistic framework is proposed by the method for discovering UsP depends upon the ternary tuples in $(u_s, d_i, w_t) \in \Lambda_{tt}$. The characterization of dependency of u_s and $d_i$is done by tagging $w_t \in W_t$ and the paper modelled the UsP $u_s \in U_s$and the favours relations of the user to items $d_i \in D_i$in a probabilistic way. A latent topic variable $z_v \in Z_V$ and $z_v = \{1, 2, .., K_c\}$ are introduced after the triple variables $(u_s, d_i, w_t) \in \Lambda_{tt}$ are observed. The gaps of semantic between users and tags are filled by the formation of hidden topical representation structure. The admixture of topics contains users and items by using the components such as topical tag distributions.

The conditional probability $p_c(d_i|u_s)$ formulate the favours of user u_s on item p_c, i.e., predictive probability, as follows in Eq. (8),

$$p_c(d_i \vee u_s) = \prod_{w_t \in W_{t_{d_l}}} p_c(w_t \vee u_s) = \prod_{w_t \in W_{t_{d_l}}} \sum_{k_c-1}^{k_c} p_c(w_t \vee z_{vk_c}) p_c(z_{vk_c} \vee u_s) = \prod_{w_t \in W_{t_{d_l}}} \phi_{w_t} . \theta_{u_s} \tag{8}$$

where $W_{t_{d_l}}$ is the tags set associated with item $d_i \in D_i$. Then, assume the annotated tags represents the items.

The two families of categorical conditional distributions such as $\Theta_\star = \{\theta_{u_x}\}$ and $\Phi_\circ = \{\phi_{k_c}\}$are derived by the key inference steps and the complete data $D_i = \{\Lambda_{tt}, z_v\} = \{u_s, d_i, w_t, z_v\}$. These families used for the generation of data D within this probabilistic framework, which is a preference discovery. In concrete models, the Bayesian and frequentist ways are inferred by these probabilities.

From the frequentist viewpoint, $\Theta_\star$ and $\Phi_\circ$are treated as the fixed unknown parameters. The optimal distributions are finding the transmission problem-based on the maximum data likelihood principle in the parameterized distribution families Ψ that are represented in Eq. (9) as

$$\Theta_\star, \Phi_\circ = \underset{\theta_{u_s}, \phi_{k_c} \in \Psi p_c(\Lambda_{tt}, z_v \vee \Theta_\star, \Phi_\circ)}{arg\,max}, (u_s \in U_s, k_c \in K_c) \tag{9}$$

In the Bayesian approach, the method treats $\Theta_\star and\, \Phi_\circ$ as probability distributions and can impose appropriate priors on them. The estimation of posterior distributions by the given the latent variables and the discovered data Λ_{tt} are integrated with z_v, the following Eq. (10) as,

$$p_c(\Theta_\star, \Phi_\circ \vee \Lambda_{tt}) \alpha \int_{-z}^{+z} p_c(z_v, \Lambda_{tt} \vee \Theta_\star, \Phi_\circ) p_c(\Theta_\star, \Phi_\circ \vee \pi, \delta) dz \tag{10}$$

where π, δ are the hyper-parameters.

3.4.2. User preference for probabilistic model

For processing the text, the context of latent topic model contains three probabilistic graphical models namely Author-Topic Model (ATM), Probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet Allocation (LDA) [16, 17]. As concrete examples, the method adjust and adopts an LDA model for modelling the dependencies among used tags, interesting items and users into our approach. In the following section, this model is reformulated as tagging process based on modelling assumptions as well as the method provides the inference approaches.

Latent Dirichlet Allocation (LDA)

Essentially, LDA is a full Bayesian version of pLSA, which imposes the Dirichlet prior on the user topical preference distributions $\Theta_{\star}$, and the tagging topical semantics distributions $\Phi_{\circ}$, with the Dirichlet-categorical conjugate properties. As applying LDA to the preference discovery problem, we assume the same probabilistic admixture assumption as pLSA, i.e., the mixing proportions are user specified. The corresponding generative process of the personomies, $W_{t_{u_x}}$ for each $u_s \in U_s$, is formulated as follows:

I. Each $u_s \in U_s$ generates $\theta_{mu_x}: Dir(\pi)$, and each topic $k_c = 1, 2, \ldots, K_c$ generates $\phi_{k_c}: Dir(\delta)$.

II. For each words $w^j_{tu_x} \in W_{tus}$ of $u_s \in U_s$, and $W_t = U_s W_{tu_x}$,

- select an assignment of topics $z_{vu^j_x}$ for $w^j_{tu} and z^j_{vu_x}: Cat\left(\theta_{mu_x}\right)$.
- According to this topic, draw the word $w^j_{tu_x}: Cat\left(\phi_{z^t_{vu_x}}\right)$.

The model corresponding to this process express the following joint distribution of the latent and observed variable in Eq. (11),

$$p_c(w_t, z_v, \Theta_{\star}, \Phi_{\circ} \vee \pi, \delta) = \prod_{k_c-1}^{k_c} p_c\left(\phi_{k_c} \vee \delta\right) \prod_{u_s-1}^{U_s} p_c\left(\phi_{u_s} \vee \delta\right) \prod_{i-1}^{W_{tu}} p_c\left(z^i_{vu_x} \vee \theta_{mu_x}\right) p_c\left(w^i_{tu_x} \vee \phi^t_{z_{iux}}\right) \quad (11)$$

By integrating out $\theta_m\ and \phi$, the approach can obtain the collapsed posterior probability, $p_c(z_v | w_t; \pi, \delta)$, of the latent topic variables.

4. Experimental Outcome

The experiments conducted on FRMS and Movielens datasets to calculate the performance of the approach is described in the below sections. The proposed method implemented on Java Netbeans IDE 8.2 on a system with Windows 7, i5 processor and 4 GB Ram. The proposed method achieved better results than the existing method by using tag-based extraction, mapping the sentiment for better profiling.

4.1. Dataset description

The method uses the two different kinds of datasets such as FMRS and Movielens. In both datasets, there are 203 and 71,567 users whereas the resources are 500 and 10,681. Moreover, the tags for FMRS and Movielens are 7889 and 95,580, whereas the domain represents the recipes and movies. The datasets are split into 20% for testing and 80% for training the data that is used to evaluate the approach.

4.2. Performance measures

The methods consider the two main performance measures like P/No. (Precision/Number) and Mean Reciprocal Rank (MRR) for experiments to validate the effectiveness of the approach. The measurement of the search strategy accuracy is defined by the P/No., whereas MRR metric defines how fast a user can able to find desired resources that are described below in Eq. (12),

$$MRR = \frac{1}{n} \cdot \sum_{i-q}^{n} \frac{1}{rank(r_i^q)} \tag{12}$$

where q_i query for the target positions $p(r_i^q)$and in the testing set, $n$is the overall number of tuples.

4.3. Overall performance

Figures 1 and 2 represent the achievement of P/No. by all methods on the Movielens datasets, whereas the conduct of P/No. by all relevance methods are illustrated by Fig. 1 and the act performed by all the probabilistic methods in terms of P/No. is represented by Fig. 2. In addition to this, Figs. 3 and 4 describe the P/No. by all methods on the FMRS dataset.

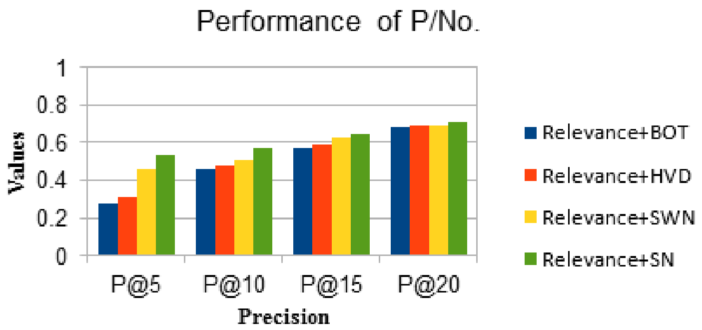

Fig. 1. The outcome of P/No. by relevance method on Movielens.

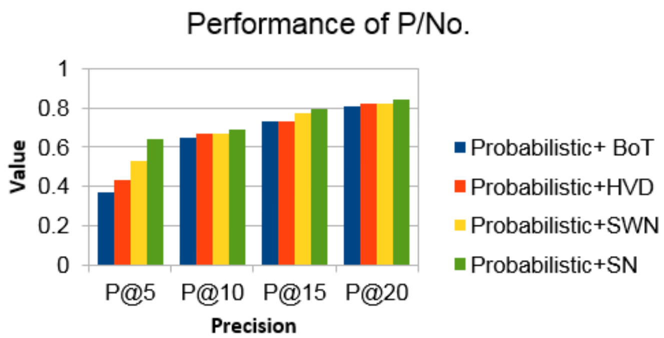

Fig. 2. The achievement of P/No. by probabilistic method on Movielens.

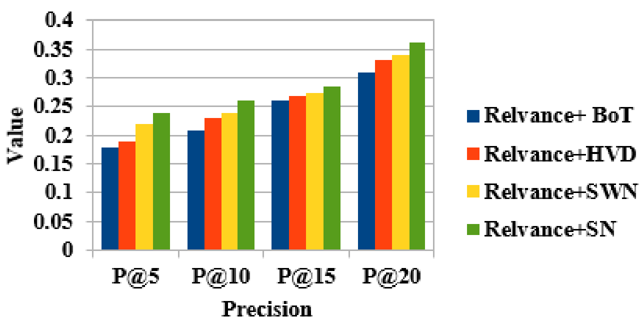

Fig. 3. The performance of P/No. by relevance method on FMRS dataset.

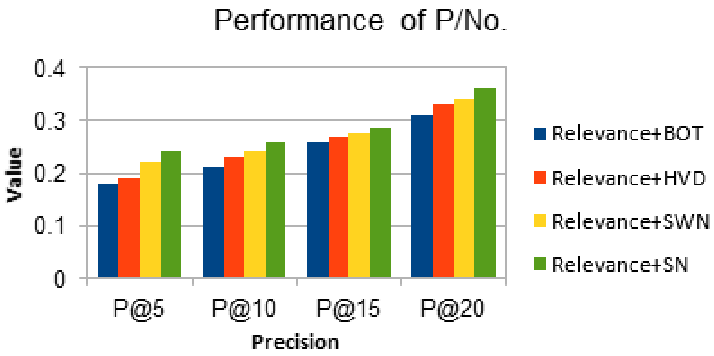

Fig. 4. The performance of P/No. by probabilistic method on FMRS dataset.

Based on these results, the method can make the following observations.

- The graph results provide needful for incorporating the information in tag-based profiles. A support for this method comes from other incorporating methods such as SentiWordNet 3:0 (SWN), Harvard IV-4 sentiment dictionary (HVD) and SenticNet 3:0 (SN) yields good performance than non-sentiment methods like Bag of Tags (BOT). The significance of these methods to give the sentence polarity based on the synonmy and similar words.
- The performance of the PS is affected by the different sentiment dictionaries, among them, the best performance in P/No. is given by the SN, while HVD and SWN provide the next best and worst performance.
- The reason for providing the best performance by the SN is that the number of sentiment dimensions is fewer dimensions in number so that the problems like under-fitting and over-fitting does not occur.
- The SWN may suffer from the problem like an under-fitting problem because of having only two dimensions. The performance of HVD was affected by the problem of over-fitting due to the more number of sentiment such as 185 dimensions.
- The results identified the similar relations in both datasets like FMRS and Movielens databases, and these observations are applicable in different scales of datasets also.

Tables 1 and 2 represent the observations made in the metric MRR on Movielens and FMRS dataset.

The users show a preference towards systems that can provide such context-specific personalized services with the help of above results. Figure 5 explains the performance of MRR on Movielens dataset. The graph clearly stated that the proposed method provides better results than the existing methods. The graphical results of Fig. 6 describe the performance of MRR on FMRS datasets. All the significance methods in proposed method clearly explains the quality of search from the sentiment approach. The ability of the recommender and personalization systems is to provide hidden semantic information from UsPs.

Table 1. Comparison of MRR on Movielens by proposed with existing method.

	Methodology	BoT	HVD	SWN	SN
Existing Method [14]	Content-based	0.109	0.125	0.138	0.151
	Collaborative	0.213	0.232	0.249	0.256
Proposed Method	Relevance	0.318	0.327	0.382	0.396
	Probabilistic	0.376	0.368	0.425	0.462

Table 2. MRR comparison between proposed and existing method on FMRS dataset.

	Methodology	BoT	HVD	SWN	SN
Existing Method [14]	Content-based	0.183	0.191	0.219	0.225
	Collaborative	0.240	0.245	0.257	0.263
Proposed Method	Relevance	0.261	0.297	0.284	0.320
	Probabilistic	0.304	0.317	0.336	0.374

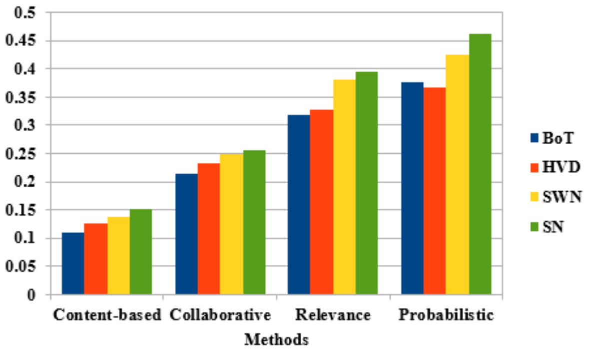

Fig. 5. Performance of MRR on Movielens datasets.

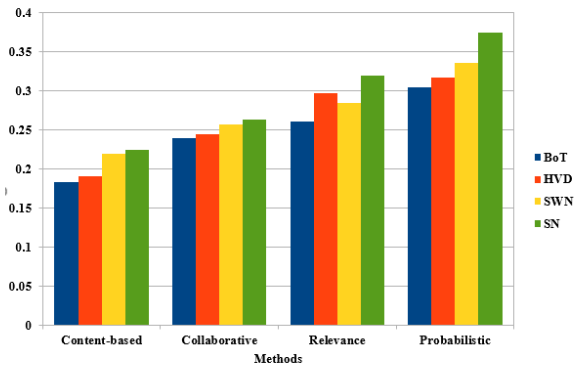

Fig. 6. Performance of MRR on FMRS datasets.

5. Conclusion

The main objective of this work is to develop techniques and methods that can be as part of a system like personalization or recommendation for providing services to users. In this work, an approach SentiGrade is developed for incorporating the PS from sentiment by tag-based UP. In folksonomy data, the drawback of the PS is addressed by the proposed method called sentiment-based ranking approach. Moreover, the proposed SBP approach is validated by comparing the existing methods such as a content-based and collaborative method. The experimental results examined the approach on FMRS and Movielens dataset and the outcomes proved the quality of search benefits from the sentiment approach. In future, the method will extend to process polysemy and synonmy of the tags identified with reduced overlapping terms.

Nomenclatures

Cat(.)	Categorical distribution
D_i	Items
Dir(·)	Dirichlet distribution
F_d	Folksonomy data
K_c	Relationship between ternary data
p_c	predictive probability
q_u	Query
r_ε	Resources
S_r	Resource ranking
T_l	Tags
U_s	User
W_t	Tags collected which are associated with items
w_u^j, w_d^j	jth tag’s label associated with user and item
z_u^j, z_d^j	jth component of z
Z_v	latent topic variable

Greek Symbols

θ_m	Mapping function
θ_u	Kc-dimensional vector; topics distribution given the user.
θ_*	Kc ×|Us| matrix
Λ_t	favouring and tagging relation
π, δ	Symmetric prior hyperparameters of Dirichlet distribution
Φ_o	|Wu|× K c matrix
ϕ_k	|Wt|-dimensional vector; labels distribution given topic kc.
Ψ	parameterized distribution family

Abbreviations

ATM	Author-Topic Mode
BOT	Bag of Tags
CCTM	Collaborative Correspondence Topic Models

EQ	Experience Quality
HVD	Harvard IV-4 sentiment dictionary
LDA	Latent Dirichlet Allocation
LO	Learning Objects
MRR	Mean Reciprocal Rank
P/No.	Precision/Number
pLSA	Probabilistic Latent Semantic Analysis
PS	Personalized Search
RP	Resource Profiles
SBP	Sentiment-Based Personalized
SN	SenticNet 3:0
SWN	SentiWordNet 3:0
UP	User Profiles
UsP	User Preference
VSM	Vector Space Model
WP	Web Personalization

References

1. Tang, X.; Xu, Y.; and Geva, S. (2018). Factorization-based primary dimension modelling for multidimensional data in recommender systems. *International Journal of Machine Learning and Cybernetics*, 1-20.
2. Kosir, D.; Kononenko, I.; and Bosnic, Z. (2014). Web user profiles with time-decay and prototyping. *Applied Intelligence*, 41(4), 1081-1096.
3. Stai, E.; Kafetzoglou, S.; Tsiropoulou, E. E.; and Papavassiliou, S. (2018). A holistic approach for personalization, relevance feedback and recommendation in enriched multimedia content. *Multimedia Tools and Applications*, 77(1), 283-326.
4. Parra-Arnau, J.; Rebollo-Monedero, D.; and Forne, J. (2014). Measuring the privacy of user profiles in personalized information systems. *Future Generation Computer Systems*, 33, 53-63.
5. Hawalah, A.; and Fasli, M. (2014). Utilizing contextual ontological user profiles for personalized recommendations. *Expert Systems with Applications*, 41(10), 4777-4797.
6. Krishnaraju, V.; Mathew, S.K.; and Sugumaran, V. (2016). Web personalization for user acceptance of technology: An empirical investigation of E-government services. *Information Systems Frontiers*, 18(3), 579-595.
7. Amoretti, M.; Belli, L.; and Zanichelli, F. (2018). UTravel: Smart mobility with a novel user profiling and recommendation approach. *Pervasive and Mobile Computing*, 38(2), 474-489.
8. Otebolaku, A.M.; and Andrade, M.T. (2017). Context-aware personalization using neighborhood-based context similarity. *Wireless Personal Communications*, 94(3), 1595-1618.
9. Kumar, H.; Lee, S.; and Kim, H.-G. (2014). Exploiting social bookmarking services to build clustered user interest profile for personalized search. *Information Sciences*, 281, 399-417.
10. Orso, V.; Ruotsalo, T.; Leino, J.; Gamberini, L.; and Jacucci, G. (2017). Overlaying social information: The effects on users' search and information-selection behaviour. *Information Processing & Management*, 53(6), 1269-1286.

11. Yau, J.Y.-K.; and Hristova, Z. (2017). Evaluation of an extendable context-aware "learning java" app with personalized user profiling. *Technology, Knowledge and Learning*, 23(2), 315-330.
12. Vicente-Lopez, E.; de Campos, L.M.; Fernandez-Luna, J.M.; and Huete, J.F. (2016). Use of textual and conceptual profiles for personalized retrieval of political documents. *Knowledge-Based Systems*, 112, 127-141.
13. Zhou, D.; Wu, X.; Zhao, W.; Lawless, S.; and Liu, J. (2017). Query expansion with enriched user profiles for personalized search utilizing folksonomy data. *IEEE Transactions on Knowledge and Data Engineering*, 29(7), 1536-1548.
14. Xie, H.; Li, X.; Wang, T.; Lau, R.Y.K.; Wong, T.-L.; Chen, L.; Wang, F.L.; and Li, Q. (2016). Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy. *Information Processing and Management*, 52(1), 61-72.
15. Bansal, T.; Das, M.; and Bhattacharyya, C. (2015). Content driven user profiling for comment-worthy recommendations of news and blog articles. *Proceedings of the 9th ACM Conference on Recommender Systems*. New York, United States of America, 195-202.
16. Blei, D.M.; Ng, A.Y.; and Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4-5), 993-1022.
17. Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; and Smyth, P. (2004). The author-topic model for authors and documents. *Proceedings of the 20th Conference in Uncertainty in Artificial Intelligence.* Banff, Canada, 487-494.

Appendix A

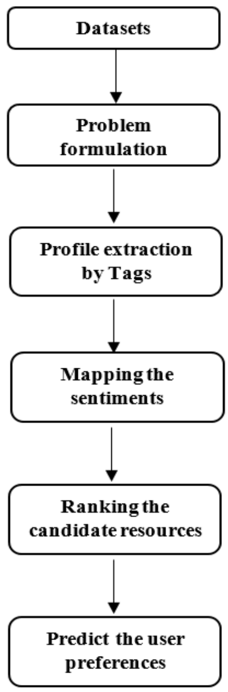

Fig. A-1. Structure of the proposed method.