Exploring External Knowledge Base for Personalized Search in Collaborative Tagging Systems

Dong Zhou[1(✉)], Xuan Wu[1], Wenyu Zhao[1], Séamus Lawless[2], and Jianxun Liu[1]

[1] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, China
dongzhou1979@hotmail.com

[2] ADAPT Centre, Knowledge and Date Engineering Group, School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland

Abstract. Alongside the enormous volume of user-generated content posted to World Wide Web, there exists a thriving demand for search personalization services, especially those utilizing collaborative tagging data. To provide personalized services, a user model is usually required. We address the setting adopted by the majority of previous work, where a user model consists solely of the user's past information. We construct an augmented user model from a number of tags and documents. These resources are further processed according to the user's past information by exploring external knowledge base. A novel generative model is proposed for user model generation. This model leverages recent advances in neural language models such as Word Embeddings with latent semantic models such as Latent Dirichlet Allocation. We further present a new query expansion method to facilitate the desired personalized retrieval. Experiments conducted by utilizing real-world collaborative tagging data show that the methods proposed in the current paper outperform several non-personalized methods as well as existing personalized search methods by utilizing user models solely constructed from usage histories.

Keywords: Personalized search · Collaborative tagging systems · Latent semantic models · Word embeddings · Query expansion

1 Introduction

The amount of digital content online has increased exponentially recently. The use of personalized Web search systems has become crucial in retrieving relevant information. Such system fetches relevant information that are most correlated to an individual user rather than only to the issued query [1, 2]. Recording of the individual's interests and past behaviors in user models has been widely adopted. Subsequently the information inside the user model can be used for query and/or results personalization.

With this increasing volume of digital content comes an increasing number of social tagging Websites for Web pages and documents. Collaborative tagging systems like

S. Wang and A. Zhou (Eds.): CollaborateCom 2016, LNICST 201, pp. 408–417, 2017.
DOI: 10.1007/978-3-319-59288-6_37

del.icio.us[1] and *BibSonomy*[2],etc., have become more and more popular. The tags and documents added by different users to the platforms are closely linked to that individual and their interests, providing abundant information for constructing more rigid and characteristic user models. Therefore, constructing user models from collaborative tagging systems has the potential to be instrumental for personalized search. In the collaborative tagging platform, users are freely to choose whatever words/terms to be used in tagging. This behavior makes the information search process even more difficult than normal web search systems. To deal with this problem, personalized search results re-ranking [3–6] and personalized query expansion [7–9] have been widely adopted.

However, there are several drawbacks in the process of searching in such type of systems, including the following. (i) In the approaches that a user model contains potential expansion terms, past researchers used relationships between tags and lexical matching methods between terms and queries. For the most of the cases, tags can not be viewed as accurate summarization of documents, henceforth the search experiences are somewhat depressed [10]. Moreover, lexical matching may miss some latent semantic information exhibited in the user model. (ii) All the previously proposed personalized search methods require a user's past click/browse information stored in a user model. However, we argue that using this information alone is not sufficient. In some cases, a user may have clicked and/or browsed only a few documents. It is relatively hard to personalized search with this little usage information to hand.

To handle these limitations, we construct an augmented user model from a number of tags and documents. These resources are further processed according to the user's past information by exploring external knowledge base. A novel generative model is proposed for user model generation. This model leverages recent advances in neural language models such as Word Embeddings (WE) [11] with the traditional Latent Dirichlet Allocation (LDA) model [12]. We learn latent topics that generate word embeddings and words simultaneously. Based on the topics learnt, we further propose a novel topical query expansion model to be used in collaborative tagging systems. In this model, queries are expanded not only by their lexical similarity with the potential terms, but are also based on their topical relevancy. Observing the obtained evaluation results from a collaborative dataset sourced from a real-world platform, we can find that the personalized search methods provide very good performance improvements over various baseline methods.

Our contribution in the current paper are: (i) We introduce augmented user profiles by exploiting an external knowledge base to perform personalized search in a novel way. (ii) We suggest and evaluate a novel generative model that leverages recent advances in neural language models with latent semantic models.

[1] http://www.delicious.com/.
[2] http://www.bibsonomy.org.

2 Related Work

There exist sufficient researches in personalized search [1, 2]. These can be roughly allocated into two categories. The first one is known as results processing. This is usually done with results re-ranking by re-ordering retrieved results using the information from a user model [13]. Another is query expansion [14]. In this category, new terms selecting from a user model can be utilized to expand the original query, or terms inside the initial query can be re-weighted according to the user model [15].

The problem studied in the current paper falls in one of the above two strategies. Tags and documents crawled from a collaborative tagging system can be used to construct a test collection. This collection can be further utilized to advance the research in search personalization. For example, tags and documents can be employed to automatically learn an individual's preferences. The retrieval results can be personalized based on the topical relevance between documents and information from the user model [3]. Signals from multiple rather than single collaborative systems can be used for search personalization [16]. Bouadjenek et al. [4] presented an enhanced document representation based on user relationships to re-rank documents on a social platform. Cai et al. [17] treated the relevance as fuzzy satisfaction between users and queries. However, if the relevant documents cannot be returned in the retrieval list, the strategy has no way to fetch more relevant results.

Another strategy expands a user's issued query with potential terms from user models. Relationships between tags have been used for personalized query expansion. If a tag appears in a query, the most related tags will be selected from the user model to expand the query [10]. Researchers also considered using lexical matching methods such as co-occurrence-based method to expand the query based on a user's past information [15]. Recently, a query expansion method for personalization has been proposed [8], which is state-of-the-art. This method captures term relationships through a Tag-Topic model. Mutual information between the terms is then utilized to choose the potential expansion terms.

All of the above systems consider building user models from his/her past information only. In contrast, in this paper we explore an external knowledge base to build augmented user models. We also propose a novel topical query expansion model for personalization.

3 User Model Generation

We now define the research problem studied in the current paper. Subsequently we describe the user model generation process. Formally, data in collaborative tagging systems can be represented by $\mathcal{P}: =(\mathcal{U}, \mathcal{D}, \mathcal{T}, \mathcal{A})$. The elements in the ternary relation $\mathcal{A} \subseteq \mathcal{U} \times \mathcal{D} \times \mathcal{T}$ are called annotations, or bookmarking activities performed by different users. $\mathcal{A}^u := \{(t, d) | u \in \mathcal{U}, d \in \mathcal{D}, t \in \mathcal{T}\}$ represents a user's annotations. $\mathcal{D}^u := \{d | (t, d) \in \mathcal{A}^u\}$ represents all documents annotated by a user u. $\mathcal{T}^u := \{t | (t, d) \in \mathcal{A}^u\}$ represents the total tags used by a user. $term^{\mathcal{D}^u} := \{w | w \in \mathcal{D}^u\}$ represents all terms from $\mathcal{D}^u$, w is a word/term extracted from $\mathcal{D}^u$. Similarly,

$term^{\mathcal{D}^u_{exter}} := \{w | w \in \mathcal{D}^u_{exter}\}$ represents all terms extracted from $\mathcal{D}^u_{exter}$.$\mathcal{D}^u_{exter}$ represents all external documents extracted from an corpus $\mathcal{D}_{exter}$ different from $\mathcal{D}$.

Algorithm 1. Generative process for Enhanced User Model Generation

Require: the total tags used by a user $\mathcal{T}^u$
Require: all documents annotated by a user $\mathcal{D}^u$
Require: a user's set of external documents $\mathcal{D}^u_{exter}$
Require: word embeddings calculated by Skip-Gram for all words in $\mathcal{T}^u \cup \mathcal{D}^u \cup \mathcal{D}^u_{exter}$ $(\mathcal{D}')$
1. **for** $k \in [1, K]$
2. draw mixture components $\varphi_k \sim Dirichlet(\beta)$
3. **for** d_j in $\mathcal{D}'$
4. draw mixture proportion $\theta_j \sim Dirichlet(\alpha)$
5. **for** $w_i \in [1, N_{d_j}]$
6. draw topic index $z_{j,i} \sim Multnomial(\theta_{d_j})$
7. draw term for word $w_{j,i} \sim Multnomial(\varphi_{z_{j,i}})$
8. for each dimension of the embedding of $w_{j,i}$, draw $f^e_{j,i} \sim \mathcal{N}(\mu^e_{z_{j,i}}, \sigma^e_{z_{j,i}})$

In the problem studied here, we construct a user model contains a number of words/terms $\{w_1, w_2 \ldots w_n\} \in term^{\mathcal{D}^u} \cup term^{\mathcal{D}^u_{exter}} \cup \mathcal{T}^u$. If a user submits a query, potential expansion terms will be selected from a sorted list of terms.

The user model generation has two steps: external documents fetch and user model construction. We augment a user's past information in step 1. All tags t in $\mathcal{T}^u$ are joined to form a query $q^{\mathcal{T}^u}$. We iterate through d in $\mathcal{D}^u$ to extract high weighted terms to form a number of queries $Q^{\mathcal{D}^u}$ (inverted document frequency used here and we extract top λ terms). Then we issue queries in $q^{\mathcal{T}^u} \cup Q^{\mathcal{D}^u}$ to an external corpus $\mathcal{D}_{exter}$ to fetch γ number of documents to form $\mathcal{D}^u_{exter}$.

In step 2, we integrate $\mathcal{T}^u$, $\mathcal{D}^u$ and $\mathcal{D}^u_{exter}$ into a novel generative model. In this model, multinomial distribution of topics of each single document can be easily acquired. The procedure is described in the remaining of this section.

It is well known that the LDA model can mine the thematic structure of documents. Recently, WE has played an increasingly vital role in building continuous word vectors based on their context in a corpus. There are also some attempts to integrate LDA with WE for different purposes [18, 19]. Inspired by those works, a novel generative model for user model generation is presented in this paper. We named this model enhanced user model generation (EUMG).

To jointly model words and word embeddings produced by WE, EUMG learns a shared latent topic space to generate words in documents and corresponding word embeddings. The WE are all pre-trained, together with documents as input to the model. Skip-Gram model [11] is utilized here before running our model to learn WE. A normal distribution is used for WE to learn latent topics from the documents as well as the words. With the WE and documents trained by the Skip-Gram model, the generation process of the EUMG model can be summarized as in Algorithm 1.

In Algorithm 1, the mean and deviation of the normal distribution are defined as μ and σ respectively. The parameters of a topic Dirichlet prior and word Dirichlet prior are defined as α and β. θ_j is the multinomial topic distribution of document d_j. $f^e_{j,i}$ are

word embeddings. The number of latent topics and dimensions for WE are both fixed. Both the words and WE determine posterior distribution of topics.

In the proposed model, Gibbs Sampling is used to solve the intractable inference problem. A conjugate prior is used here. We also integrate out θ and φ. The conditional distribution $p\left(z_{j,i}=k\right)$ has to be calculated in sampling. Specifically, the topic is chosen from the following equation for each word:

$$p\left(z_{j,i}=k\right) \propto \frac{n_{j,k,\neg i}+\alpha}{n_{j,\cdot,\neg i}+K\cdot\alpha} \times \frac{v_{k,w_{j,i},\neg}+\beta}{v_{k,\cdot,\neg}+V\cdot\beta} \times \prod_{e=1}^{E} \frac{1}{\sqrt{2\pi}\sigma_{z_{j,i}}} \exp\left(-\frac{\left(f_{j,i}^{e}-\mu_{z_{j,i}}\right)^{2}}{2\sigma_{z_{j,i}}^{2}}\right) \tag{1}$$

The amount of times that topic k is added up in $n_{j,k,\neg i}$ (from multinomial distribution of the document j). Note that the present $z_{j,i}$ is not included. The amount of times $w_{j,i}$ is generated by topic k is added up in $v_{k,w_{j,i},\neg}$. The present $w_{j,i}$ is not included. The summation over all values of the variable is denoted by a dot. E is the dimensions of word embeddings. The posterior estimate of θ and φ can then be easily obtained.

4 Topical Query Expansion

Next, The output from step 2 of the last section can be utilized to build a query expansion model that ranks terms from the user model to be added to the query. We only layout the key steps in this section because of space constrains.

We assume there exists a query $q=\{w_a\}_{a=1}^{n}$, where $\{w_a\}_{a=1}^{n}$ denotes n independent query words. We approximate the probability of q generating w from a hidden model H is by: (see also [20, 21]):

$$P(w|H) \approx P(w|q) \tag{2}$$

We further define a number of relevant documents $\{d_b\}_{b=1}^{M}$. These documents have relationships with both the query and the words in a user model.M represents total number of documents. Associate $\{d_b\}$ into Eq. (2) leads to:

$$P(w|q)=\sum_{b=1}^{M} P\left(w|d_b\right)P\left(d_b|q\right) \tag{3}$$

The uniform prior of documents is also put outside of the summation. $P(q)$ has been eliminated here as a uniform prior.

The output form step 2 of the last section (i.e. the documents in the user model) can be treated as $\{d_b\}_{b=1}^{M}$ in the above equation. In Eq. (3), w has a direct dependency on d_b and w_a also has a direct dependency on d_b, the assumption is too simplistic. Through the EUMG model, we obtain latent topics. These topics can be used to re-calculate the probability of q generating w:

$$P(w|q) \propto \frac{1}{M}\sum_{b=1}^{M}(\sum_{k=1}^{K} P(w|topic_k)P(topic_k|d_b))(\prod_{a=1}^{n}\sum_{k=1}^{K} P(w_a|topic_k)P(topic_k|d_b)) \quad (4)$$

$topic_k$ represents a particular topic learnt. After we obtain the probability scores, we sort all the terms and select the top δ terms as the final expansion query terms.

5 Experiments

5.1 Evaluation Setup

To examine the performance of the user model generation and query expansion methods, we perform the experiments in a dataset which merges two real-world sub-datasets from a collaborative tagging system *del.icio.us*: *socialbm0311* and *deliciousT140*. Please refer to [22, 23] for details about the two datasets. There are 5,153,720 annotation activities, 259,511 users, 137,870 tags and 131,283 documents in our merged dataset. An external knowledge base is constructed from Wikipedia[3]. This knowledge base contains 4,634,369 documents. It was crawled on 14/08/2014. To evaluate the effects of augmented user models, foursets of users with different size are chosen as test users. This includes: users with no more than 50 annotation activities (**User50**), users with 50–100 annotation activities (**User100**), users with 100–500 annotation activities (**User500**) and users with more than 500 annotation activities (**UserG500**). These sets represent users with small, moderate and rich amounts of past usage information respectively. We randomly choose 200 users from each set. We select 25% of each user's annotations for evaluation, and the remaining 75% are utilized to build the user model.

We follow the evaluation procedure of previous research [3, 8, 16]. If a user u issues a query t, this query is viewed as a personalized query. In this case, relevant documents are the documents annotated by u with t.

We use the evaluation metrics that are typically utilized in Web search evaluation: mean average precision (MAP), which is usually employed to report search accuracy;normalized discounted cumulative gain (NDCG), a natural choice for search engine evaluation; as well as another commonly adopted evaluation metric mean reciprocal rank (MRR). Paired t-test is used for significance evaluation. We set the confidence level at 95%. The average performance is computed for all users in the same set.

The methods proposed in this paper are compared with several query expansion methods. This includes non-personalized methods and personalized baselines. We now describe them in detail.

LMA language model based retrieval method. This model is quite popular and produces good results before. We use the model described in [24].

LM + RM-wiki This is a relevance model which uses Wikipedia to obtain the relevance documents, as in [25]. It is a strong non-personalized baseline for comparison because that we also used external corpus in our approach.

[3] http://www.wikipedia.org.

Cooccur + QE This is a personalized baseline method, utilized by many previous researchers [6, 11]. The method calculates co-occurrence scores between terms from a user model and terms from a query [15].

Tag-topic + QE This method captures term relationships through a Tag-Topic model. Mutual information between the terms is then utilized to choose the potential expansion terms [8]. The highest performing method from [8] is selected here as a strong personalized baseline.

EUMG + TQE finally, our approach uses the EUMG model and the query expansion method proposed in Sect. 4 for personalized search in collaborative tagging systems.

5.2 Experimental Results

Results are now fully examined in this section. The overall performance is demonstrated in Table 1, including our new approach presented in the paper together with baseline methods on the test users in four sets. Statistically significant differences between a method with the best performing non-personalized baseline (***LM + RM-wiki***) and the best performing personalized baseline (***Tag-topic + QE***) are indicated by *, $^{\sharp}$ respectively. From the results, we learn that ***LM*** model performs the worst in all different sets of users by using all evaluation metrics. ***LM + RM-wiki*** method performed steadily better than ***LM***. This illustrates the effectiveness of utilizing an external corpus. The results are consistent with previous research [25]. These two methods are surpassed by the three personalized methods with statistically significance. This includes the method ***EUMG + TQE*** proposed in this paper. This shows that terms in the user models can improve the effectiveness of search significantly. Non-personalized query expansion methods only select terms from top documents. There are only limited improvements observed.

The personalized baselines ***Cooccur + QE*** and ***Tag-topic + QE*** expand the queries by using the user's historical information only while our approach explores an external knowledge base. We now analyze the results for these three methods. As seen from Table 1, several conclusions can be drawn. First, ***EUMG + TQE*** out performs the two personalization methods ***Cooccur + QE*** and ***Tag-topic + QE***, in all foursets of users by using different metrics. The differences are consistently significant. The possible reason is that we use an external knowledge base in addition to the user's past information to build an augmented user model for personalized search. Secondly, ***EUMG + TQE*** achieves consistent improvements over baseline approaches across four sets of users. The improvements in **User50** are more remarkable than in other sets of test users. In reality, we often face the situation where a user has little interactions with the search platform. Under such circumstances, personalized search experience is usually unsatisfactory. However, with enhanced content, our method can obtain reasonably better results for this set of users. This result also confirms that our approach performs well for users with small or moderate volumes of past information and those with a rich set of historical data. Third, using Wikipedia seems a good choice of the external corpus. The possible reason, as pointed out in [25], is that if an external knowledge base has good coverage of topics, it is more likely that good expansion terms can be selected from it.

Finally, we examine the optimum number of latent topics and dimensions for WE. The number of topics is chosen from [5, 50]. The dimension numbers is chosen from [10, 100]. Because of the space constraints, we only report the results here. When there are 15 latent topics and 50 dimensions for WE, we obtain the highest performance. When both numbers grow beyond 15 and 50, we have a lower performance. However, comparing to baseline models, ***EUMG + TQE*** always performs better than them, even the lowest performed runs.

Table 1. Overall performance

	User50				User500		
	MAP	NDCG	MRR		MAP	NDCG	MRR
LM	0.0163	0.0309	0.0184	*LM*	0.0167	0.0283	0.0203
LM + RM-wiki	0.0211	0.0501	0.0232	*LM + RM-wiki*	0.0242	0.0468	0.0263
Cooccur + QE	0.0674*	0.0975*	0.0779*	*Cooccur + QE*	0.0886*	0.1195*	0.0993*
Tag-topic + QE	0.1525*	0.1924*	0.2009*	*Tag-topic + QE*	0.1655*	0.2036*	0.203*
EUMG + QE	0.2440$^{*,\sharp}$	0.2868$^{*,\sharp}$	0.2980$^{*,\sharp}$	*EUMG + QE*	0.2154$^{*,\sharp}$	0.2592$^{*,\sharp}$	0.2424$^{*,\sharp}$
	User100				UserG500		
	MAP	NDCG	MRR		MAP	NDCG	MRR
LM	0.0125	0.0314	0.0136	*LM*	0.019	0.0349	0.0193
LM + RM-wiki	0.0225	0.0384	0.0238	*LM + RM-wiki*	0.0319	0.0674	0.0333
Cooccur + QE	0.0843*	0.1216*	0.0897*	*Cooccur + QE*	0.0916*	0.1246*	0.1015*
Tag-topic + QE	0.1586*	0.1647*	0.1721*	*Tag-topic + QE*	0.2004*	0.2405*	0.2528*
EUMG + QE	0.2117$^{*,\sharp}$	0.2476$^{*,\sharp}$	0.2377$^{*,\sharp}$	*EUMG + QE*	0.2385$^{*,\sharp}$	0.2897$^{*,\sharp}$	0.2802$^{*,\sharp}$

6 Conclusions

We study the problem of personalized search utilizing collaborative tagging data in the current paper. In particular, we investigated augmented user models and query expansion methods. We construct an augmented user model from a set of tags and documents, together with an external knowledge base. A novel generative model is proposed, which leverages word embeddings with Latent Dirichlet Allocation for user model generation. Based on the user models constructed, we further present a query expansion model to facilitate the desired personalized retrieval based on topics learnt. The proposed method performed well on a real-world collaborative tagging dataset. It demonstrates statistically significant improvements over several baseline systems including non-personalized and personalized methods. In future research, automatically determination of the number of topics and dimensions will be studied. The effectiveness of different external knowledge bases will also be examined in our subsequent experiments.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (61300129, 61572187 and 61272063), Scientific Research Fund of Hunan Provincial Education Department of China (16K030), Hunan Provincial Innovation Foundation For Postgraduate (CX2016B575). This research was also supported by the ADAPT Centre for Digital

Content Technology, which is funded under the Science Foundation Ireland Research Centres Programme (13/RC/2106) and is co-funded under the European Regional Development Fund.

References

1. Ghorab, M.R., Zhou, D., O'Connor, A., Wade, V.: Personalised information retrieval: survey and classification. User Model. User-Adap. Inter. **23**, 381–443 (2013)
2. Zhou, D., Lawless, S., Wu, X., Zhao, W., Liu, J.: A study of user profile representation for personalized cross-language information retrieval. Aslib J. Inf. Manage. **68**, 448–477 (2016)
3. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 155–162. ACM (2008)
4. Bouadjenek, M.R., Hacid, H., Bouzeghoub, M.: Sopra: a new social personalized ranking function for improving web search. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 861–864. ACM, Dublin (2013)
5. Xie, H., Li, X., Wang, T., Chen, L., Li, K., Wang, F.L., Cai, Y., Li, Q., Min, H.: Personalized search for social media via dominating verbal context. Neurocomputing **172**, 27–37 (2016)
6. Xie, H., Li, X., Wang, T., Lau, R.Y.K., Wong, T.-L., Chen, L., Wang, F.L., Li, Q.: Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy. Inf. Process. Manage. **52**, 61–72 (2016)
7. Bouadjenek, M.R., Hacid, H., Bouzeghoub, M., Daigremont, J.: Personalized social query expansion using social bookmarking systems. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1113–1114. ACM, Beijing (2011)
8. Zhou, D., Lawless, S., Wade, V.: Improving search via personalized query expansion using social media. Inf. Retr. **15**, 218–242 (2012)
9. Zhou, D., Lawless, S., Wade, V.: Web search personalization using social data. In: Zaphiris, P., Buchanan, G., Rasmussen, E., Loizides, F. (eds.) TPDL 2012. LNCS, vol. 7489, pp. 298–310. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33290-6_32
10. Bender, M., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J.X., Schenkel, R., Weikum, G.: Exploiting social relations for query expansion and result ranking. In: Proceedings of the IEEE 24th International Conference on Data Engineering Workshop, ICDEW 2008, pp. 501–506. IEEE (2008)
11. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, NIPS 2013, pp. 3111–3119 (2013)
12. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
13. Dou, Z., Song, R., Wen, J.-R.: A large-scale evaluation and analysis of personalized search strategies. In: Proceedings of the 16th International Conference on World Wide Web, pp. 581–590. ACM, Banff (2007)
14. Zhou, D., Lawless, S., Liu, J., Zhang, S., Xu, Y.: Query expansion for personalized cross-language information retrieval. In: Proceedings of the 10th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2015, pp. 1–5. IEEE, Trento (2015)

15. Chirita, P.-A., Firan, C.S., Nejdl, W.: Personalized query expansion for the web. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 7–14. ACM, Amsterdam (2007)
16. Wang, Q., Jin, H.: Exploring online social activities for adaptive search personalization. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 999–1008. ACM, Toronto (2010)
17. Cai, Y., Li, Q.: Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 969–978. ACM, Toronto (2010)
18. Das, R., Zaheer, M., Dyer, C.: Gaussian LDA for topic models with word embeddings. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, pp. 795–804. ACL, Beijing (2015)
19. Liu, Y., Liu, Z., Chua, T.-S., Sun, M.: Topical word embeddings. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI 2015, pp. 2418–2424. AAAI Press, Austin (2015)
20. Lavrenko, V., Croft, W.B.: Relevance based language models. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 120–127. ACM, New Orleans (2001)
21. Ganguly, D., Leveling, J., Jones, G.J.F.: Topical relevance model. In: Hou, Y., Nie, J.-Y., Sun, L., Wang, B., Zhang, P. (eds.) AIRS 2012. LNCS, vol. 7675, pp. 326–335. Springer, Heidelberg (2012). doi:10.1007/978-3-642-35341-3_28
22. Zubiaga, A., Garcia-Plaza, A.P., Fresno, V., Martinez, R.: Content-based clustering for tag cloud visualization. In: Proceedings of the International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009, pp. 316–319. IEEE (2009)
23. Zubiaga, A., Fresno, V., Martinez, R., Garcia-Plaza, A.P.: Harnessing folksonomies to produce a social classification of resources. IEEE Trans. Knowl. Data Eng. **25**, 1801–1813 (2013)
24. Zhai, C., Lafferty, J.: Model-based feedback in the language modeling approach to information retrieval. In: Proceedings of the Tenth International Conference on Information and Knowledge Management, pp. 403–410. ACM (2001)
25. Diaz, F., Metzler, D.: Improving the estimation of relevance models using large external corpora. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 154–161. ACM, Seattle (2006)