Estimating Personalization using Topical User Profile

Sara Abri[1], Rayan Abri[1] and Salih Cetin[2]
[1]*Department of Computer Engineering, Hacettepe University, Ankara, Turkey*
[2]*Mavinci Informatics Inc., Ankara, Turkey*

Keywords: Personalized Web Search, User Search Behaviour, Topical User Model, Latent Dirichlet Allocation.

Abstract: Exploring the effect of personalization on different queries can improve the ranking result. There is a need for a mechanism to estimate the potential for personalization for queries. Previous methods to estimate the potential for personalization such as click entropy and topic entropy are based on the prior clicked document for query or query history. They have limitations like unavailability of the prior clicked data for new/unseen queries or queries without history. To alleviate the problem, we provide a solution for the queries regardless of query history. In this paper, we present a new metric using the topic distribution of user documents in the topical user profile, to estimate the potential for personalization for all queries. Using the proposed metric, we can achieve more performance for queries with history and solve the cold start problem of queries without history. To improve personalized search, we provide a personalization ranking model by combining personalized and non-personalized topic models where the proposed metric is used to estimate personalization. The result reveals that the personalization ranking model using the proposed metric improves the Mean Reciprocal Rank and the Normalized Discounted Cumulative Gain by 5% and 4% respectively.

1 INTRODUCTION

In the context of personalized web search, lots of research and applications based on a user's interest have been done (Abri et al., 2020a). According to some research, personalization should not be used for all queries in the same manner because it varies in effectiveness for different queries. For less ambiguous queries, the current web search ranking might be sufficient, and thus, personalization is unnecessary. On the other hand for other queries with a more clear and specific meaning, the ranking methods without any personalization are more effective (Teevan et al., 2005; Abri et al., 2020c). A measure able to estimate the potential for personalization can enable the selective application of personalization and improve the overall effectiveness of the search system.

Different measures are used to determine the potential for the personalization of queries (Dou et al., 2007; Yano et al., 2016). Click entropy using the query history and documents clicked by the users is one such measure (Dou et al., 2007). This method is recently improved by a topic model-based extension (Yano et al., 2016) and referred to as topic entropy. In this paper, we improve topic entropy by measuring how each user's topical profile differentiates from the query words' topics. Using the topic distributions of clicked documents for each user as a feature, the potential for personalization is modeled on a fine-grained level.

Furthermore, a personalization ranking model by combining personalized and non-personalized topic models is proposed. In this model, the presented metric which we will refer to as unified topic user entropy(UTUE) is used as a metric to estimate the potential for personalization. Then the ranking model will be applied on built topical user profiles to improve personalized search results. The experiments show that the proposed personalization model can process queries without any history and is more effective for queries with history. This allows the system to alleviate the cold-start problem and allows determining the ambiguity of new queries.

We have implemented the proposed models on the data set. The experimental results show that a clear improvement over the baseline methods is achieved. The organization of the paper is as follows. Section 2 discusses the related work on personalized search approaches. We present our methodology consist of proposing a new metric unified topic user entropy and its correlation with another potential for personalization metrics in Section 3. In this section also a new

Abri, S., Abri, R. and Cetin, S.
Estimating Personalization using Topical User Profile.
DOI: 10.5220/0010015201450152
In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020) - Volume 1: KDIR*, pages 145-152
ISBN: 978-989-758-474-9

proposed personalization ranking model is presented. Evaluation methodology and results are given in Section 4. Section 5 includes the concluding remarks.

2 RELATED WORK

Due to the importance of the query in the personalized search process, more recent research in personalization has focused on the potential of query for personalization.

Teevan et al. (Teevan et al., 2008; Teevan et al., 2010) evaluate different metrics to predict the ambiguity of a query and its potential for personalization. They evaluated intrinsic features like query length, click entropy introduced by Duo et al. (Dou et al., 2007), clarity measure which compares the language model of the retrieved result set to a background language model (Cronen-Townsend et al., 2002) and result entropy for predicting the potential for personalization. Wang et al. (Wang and Agichtein, 2010) proposed user entropy which averages click entropy by each user and discussed that the user entropy is useful for low-frequency queries. They report click entropy as a reliable method for predicting the potential when a history for the query is available. Click-entropy models the ambiguity using only the user interactions, ignoring the contents of the documents.

Instead of just relying on the click information, augmenting click-entropy with the content of the documents is also investigated (Yano et al., 2016; Song et al., 2007). Song et al. (Song et al., 2007) discuss the relationship between query ambiguity and topic distributions. They use the latent topic model variable to model the clicked documents' content and improve the click-entropy model for predicting the ambiguity of queries. The topic model based approach proposed in this research is motivated in a similar way but extends the model proposed by Yano et al. (Yano et al., 2016) so that the newly proposed metric can handle new queries and perform better for low-frequency queries.

In the other side, more solution in ranking comes to the process of profiling of user interest and preferences. In the process of personalization, user interest models are created by user specific content, user behaviour, and user context. A personalization system first models the user profile and re-ranks the results using this profile. A natural source for building a user profile is the user's browsing history. Matthijs and Radlinski (Matthijs and Radlinski, 2011) use the words in titles, full text, metadata of the browsed web pages to construct a user profile composed of terms. External sources like Open Directory Project (ODP) are also used as an external knowledge source for modeling user profiles (Siegg et al., 2007; Chirita et al., 2005; Karimi-Mansoub and Abri, 2016).

Topic model based personalization methods exist (Harvey et al., 2013; Vu et al., 2015a; Vu et al., 2015b). Harvey et al. (Harvey et al., 2013) use Latent Dirichlet Allocation (LDA) and builds latent topic models to represent the document sets. The users are modeled by the topic distributions of the documents that they have clicked. Vu et al. (Vu et al., 2015a) use a time-aware topic model for personalization with a motivation to capture the dynamic nature of users' interests. Since user interests and search intentions are changing during a search session, long term and short term profiles were also discussed in some papers such as (Vu et al., 2017; Bennett et al., 2012). Vu et al. (Vu et al., 2017) create a temporal user profile using the user's clicked documents and uses these profiles for ranking the results. Bennet et al. (Bennett et al., 2012) splits the user profile into three based on different temporal periods and builds a long-term profile, a daily profile, and a session profile. In their experiments, they show that using these profiles is more effective than click entropy and query position in a search session.

Probabilistic topic models are also used for personalization (Hofmann, 1999). They use pLSI (Wei et al., 2010) and Kullback-Leibler Divergence to estimate a query model. In a similar method (Shao and Qin, 2014), a text similarity algorithm using LDA is proposed for personalization. They use the topic model and word co-occurrence analysis to calculate topics in the text. More recently topic models are used for query suggestion (Momtazi and Lindenberg, 2016) and modeling the semantic relationships on the AOL query log. They report unseen queries as an important shortcoming for their method. Also, Amer et al. (Amer et al., 2016) used word embeddings as opposed to topic models for the user profiles, however, their model failed to improve search effectiveness.

3 PERSONALIZATION FOR A GLOBAL QUERY

Personalization is not appropriate for all user queries and may even yield worse results than generic ranking methods. For example, the query "myspace" is usually a navigational query for the social networking website regardless of the user issuing this query. For such a query, trying to personalize can produce an inferior ranking. In this section, we divide the process into two steps. In the first step, appropriate and effective metrics to estimate personalization in queries are investigated. In the second step, we propose a personalization model using investigated metrics and evaluate our approach by using experiments.

3.1 Metrics to Estimate Potential for Personalization

We explore known metrics used to estimate the potential for personalization in queries. To find these metrics we consider the conducted research in this field. Click entropy in (Teevan et al., 2008) is defined by Teevan et al. as an effective variable of the clicked results for each query. Click entropy measures the query's personalization potential using the clicked documents for the same query. If the click entropy for a query is high, it means that different users click on different documents and the potential for personalization in query is high. The relationship between query frequency and click entropy can help to explain this relationship. This relationship is illustrated in Figure 1 for all queries in the AOL data set.

There is a relationship between query frequency and click entropy that with increasing the query frequency, click entropy is also increased. But as the graph shows, there is some irregularity and it is not a strictly increasing graph. For queries with low query frequency (for example less than one hundred), the graph is ascending but after reaching a certain extent, click entropy does not increase and remains relatively constant or in some cases even decreases. Figure1 illustrates this issue well. This stable mode is somewhat related to the navigational queries.

Click entropy has drawbacks like unavailability prior click data for new or unseen queries. In addition, click entropy is purely based on documents but not their contents. When different documents with similar contents are clicked by users for a query q, click entropy will be high signaling a false ambiguous query. In addition to click entropy, topic entropy introduced in reference (Yano et al., 2016) is proposed as a natural extension of click entropy with more accuracy. Topic entropy models $P(d|q)$ using the topic model distribution of the documents, able to account for documents with similar contents. The topic set Z is obtained using Latent Dirichlet Allocation (LDA).

Topic entropy is the weighted sum of Kullback-Leibler divergences of query and document topic distributions and Yano et al. (Yano et al., 2016) model the topic entropy as the center of gravity for the topic distribution divergences. While this measure incorporates document similarities, the users' behavioral differences are only modeled through the $P(d|q)$ component. Topic entropy is still not defined (its value is zero) for the new queries the same as the click entropy. Although Yano et al. (Yano et al., 2016) also propose topic user entropy (TUE) as in Equation 3 to incorporate the users' behavioral differences, in their experiments the correlation of topic user entropy results with human judgments is low compared to Topic Entropy.

$$TUE(q,U_q,D_{u,q}) = \tag{1}$$

$$\sum_{u\in U_q} \frac{1}{|U_q|} \sum_{d\in D_{u,q}} P(d|u,q) KL(P(z|d)||P(z|q)) = \tag{2}$$

$$\sum_{u\in U_q} \frac{1}{|U_q|} \sum_{d\in D_{u,q}} P(d|u,q) \sum_{z\in Z} P(z|d) log(\frac{P(z|d)}{P(z|q)}) \tag{3}$$

Where $D_{u,q}$ is the documents clicked by the user u for the query q, U_q is the user set issuing the query q. It is assumed that the probability of each user issuing the query is equally likely. Note that TUE weights the divergence of document model from query model by $P(d|u,q)$ which is the number of times the user u clicks document d for query q, divided by the total number of clicks of u for q. For a user who did not issue q previously, TUE is not defined since no document is clicked. To solve this cold start problem, we tried to benefit from extracted topics of topical user model $P(u|q)$. We define $P(u|q)$ in Equation 5 and it is the probability distribution of the query on the users using the LDA topic model.

$$P(u|q) \propto P(u)P(q|u) = P(u) \prod_{w\in q} P(w|u) \tag{4}$$

$$= P(u) \prod_{w\in q} \sum_{z\in Z} P(w|z)P(z|u) \tag{5}$$

Where $P(u)$ is the probability of the user u and it is estimated by the proportion of queries submitted by user u to the total number of queries. $P(w|z)$ is the probability of the word w of the query for the topic z and $P(z|u)$ is the probability of the topic z for the given user u. Using $P(u|q)$ as the weighting factor instead of $P(d|q,u)$, we define our new metric called as the unified topic user entropy (UTUE) as in Equation 6. This metric unifies all users who have or have not issued the query in the past.

$$UTUE(q,U_q,D_u) =$$
$$\frac{1}{|U_q|} \sum_{u\in U_q} P(u) \sum_{d\in D_u} \prod_{w\in q} \sum_{z\in Z} P(z|u)P(w|z)P(z|d) log(\frac{P(z|d)}{P(z|w)}) \tag{6}$$

As a new query will only be submitted by a single user and will not have any clicked documents, $D_{u,q}$ will be an empty set. As a result, $TUE(q,U_q,D_{u,q})$ will be equal to zero. Instead of depending on the

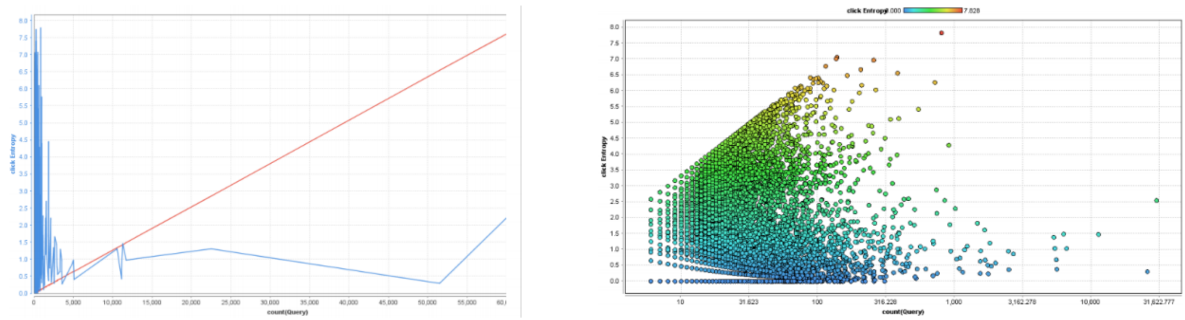

Figure 1: The relationship between query frequency and click entropy in AOL data set.

clicked documents for the specific query q, the documents clicked by the user D_u for all queries are used to compare the user profile with the query. Furthermore, instead of using $P(z|q)$ which depends on the clicked document set for the query q, the topic distribution of words in the query is used. With these two approximations, the proposed method can estimate the potential for personalization for a query without any history.

3.2 Creating a Personalization Model

We build a model to help personalization services to prevent useless personalization. To calculate the probability distribution of users, we need to create a topical user model for each user. Then a list of documents produced by the search engine for the query is re-ordered using the user profile. While this task on its own is independent of the potential for personalization tasks, we try to create a re-ranking model that can yield better results.

Three ranking methods are used in our evaluations, where the first one uses a generic document scoring function as Equation 9 introduced by Harvey at al. (Harvey et al., 2013; Carman et al., 2010), based on topic models without any personalization. In this Non-Personalized Topic Model called NonPTM, documents and words are associated with topics in the document set using LDA topic models. The $P(d|q)$ is estimated using the Bayes rule and the LDA generative model as follows. Since NonPTM is a method without any personalization, comparisons with this baseline method will reveal the improvement of personalization over generic ranking with topic models.

$$NonPTM(d,q) = P(d|q) \propto P(d)P(q|d) \tag{7}$$

$$= P(d)\prod_{w\in q} P(w|d) \tag{8}$$

$$= P(d)\prod_{w\in q}\sum_{z} P(w|z)P(z|d) \tag{9}$$

Where $P(d)$ is the prior document probability and z is the topic latent variable estimated using LDA. $P(w|z)$ and $P(z|d)$ are obtained from the LDA topic model.

The second model uses the personalization factor for the user profile built using the documents clicked by the user. A user topical profile is modeled by the set of documents D_u which the user clicked on. Using the topic distributions of the user's documents that are associated with topics, the user profile can be considered as the vector of posterior probabilities of topics given the user. The personalization based ranking function is defined as in Equation 11 which will be referred to as Personalized Topic Model (PTM). The λ parameter weighs the effect of user topical profile on the ranking process and it is equal to 0.175 similar to Harvey et al. (Harvey et al., 2013).

$$PTM(d,q,u) = P(d|q,u) \propto P(d)\prod_{w\in q} P(w,u|d) \tag{10}$$

$$= P(d)\prod_{w\in q}\sum_{z} P(w|z)P(u|z)^{\lambda}p(z|d) \tag{11}$$

Although user profiles are indicative of the user's interests, they can be incomplete and in more times it is needed to use a combination of both non-personalized and personalized models. here we propose a personalization ranking model by combining NonPTM and PTM models using the proposed UTUE metric as a threshold to identify personalization. This model is described in detail in the next subsection.

4 EVALUATION AND DATASET

4.1 Dataset

For the experiments, AOL, TREC 2013 Session Track[1] and TREC 2014 Session Track[2] of web search

[1] https://trec.nist.gov/data/session2013.html
[2] https://trec.nist.gov/data/session2014.html

engine logs are used. As it is done in Harvey at al. (Harvey et al., 2013), we cleaned the dataset by only retaining queries which resulted in a click on a URL. The Session Track consists of query sessions for different topics along with the clicked documents and user ids. The URLs are manually annotated by judges for the topics. We use the content of the clicked URL to create topic models of user profiles.

4.2 Quantifying Query Personalization

For the evaluation of the method, we make a correlation between metrics to quantify personalization. The results are calculated for all queries in the data set and are presented in Table 1. These correlations are based on 50k queries from the AOL dataset.

The number of topics used for LDA is an important parameter. The relationship between MRR and this parameter is investigated in a small development set. Parameters of the LDA model are trained using the training corpus[3]. Figure 2 shows the MRR for different topic numbers ranging from 10 topics to 100 in the AOL and TREC 2014 Session Track datasets. The results indicate that using 40 and 30 topics yields the best results in the AOL and TREC 2014 Session Track.

The performance of the four potential for personalization metrics is investigated using a similar methodology to Yano et al. (Yano et al., 2016) on different data sets. Table 1 shows the correlation coefficient between four metrics, namely Click Entropy (Dou et al., 2007), Topic Entropy and Topic User Entropy (Yano et al., 2016; Abri et al., 2020b) and $UTUE$ metric along query frequency. In the AOL data set, approximately 11% of the queries have a frequency equal to one and so there is no history for these queries. It means that for these queries click entropy or topic entropy is equal to zero while UTUE can be used to estimate the potential for personalization. This result indicates that UTUE is highly correlated (88%) with the topic and click entropy and it can be used for queries where the other metrics fall short, in queries without a history.

4.3 Personalization Ranking Model

To evaluate the personalized model, we divided the dataset into %95 for training and the last%5 of queries for testing. The personalization is evaluated using the mean reciprocal rank (MRR) up to rank 10 and Normalized Discounted Cumulative Gain (nDCG@k). *NormalizedDCG* is a measure of ranking quality discussed in (Manning et al., 2008) and measures the usefulness, or gain, of a document based on its position in the result list. Queries are sorted according to the potential for personalization metrics and personalization is applied using a combination model of PTM and NonPTM to queries above a threshold.

To investigate the importance of the combination model, different personalization metrics are used to predict the query's potential and they are normalized using the maximum value. Then, for a specified range (for example [0.0-0.2]) if the potential using the discussed metric is in this range it is ranked with personalized $PTM(d,q,u)$, otherwise, it is ranked with the topic model based ranking algorithm $NonPTM(d,q)$. A more accurate personalization metric is expected to yield better performance gains with a combination personalization model as it can identify queries more suitable for personalization.

Figures 3, 4, 5, 6, 7, 8, 9, and 10 report the MRR, and nDCG@10 scores for the four potential for personalization metrics in AOL and Session Tracks 2013 and 2014 datasets. In all figures there are two ranges $[0.0-1.0]$ and $[1.0-1.0]$. The first one ($[0.0-1.0]$ range) shows the MRR result when all queries are reranked using $PTM(d,q,u)$ and the other ($[1.0-1.0]$ range) represents the ranking score when using only $NonPTM(d,q)$, which is no personalization. Naturally, these two cases are independent of the potential metric used and are common for all four metrics. When we consider the results of UTUE, it is evident that it achieves a higher score for all different thresholds. This indicates that it assigns a more accurate prediction for personalization, and the queries with lower UTUE score does not benefit from personalization. A similar result is observed between Topic Entropy and Click Entropy, confirming the experiments in Yano et al. (Yano et al., 2016), Topic entropy performs better than Click entropy.

The results of UTUE for $[0.6-1.0]$ achieves the highest-ranking scores for all measures. This indicates that using personalization only for queries with a potential higher than 0.6 is a better strategy than using other thresholds. When considering the difference between applying personalization to all of the queries and combination personalization model with $[0.6-1.0]$, the performance gain for MRR is as high as 0.264 in the AOL dataset, 0.224 in TREC 2014 and 0.241 in TREC 2013.

[3]Gensim library is used for the LDA estimation https://radimrehurek.com/gensim/

Table 1: Correlation Coefficient between personalizaton metrics.

	Frequency	Click Entropy	Topic Entropy	TUE	UTUE
Frequency	1.0	0.650	0.821	0.852	0.726
Click Entropy	0.650	1.0	0.791	0.884	0.895
Topic Entropy	0.821	0.791	1.0	0.70	0.884
TUE	0.852	0.884	0.70	1.0	0.791
UTUE	0.726	0.895	0.884	0.791	1.0

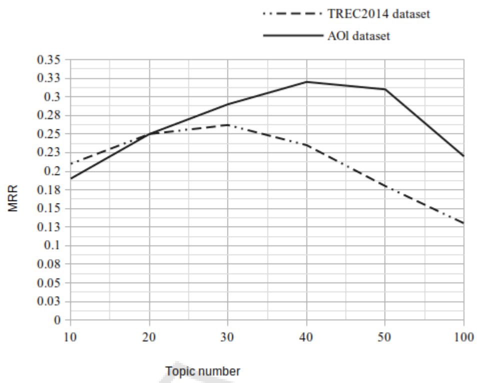

Figure 2: The Changes in MRR with different topic numbers using the LDA model.

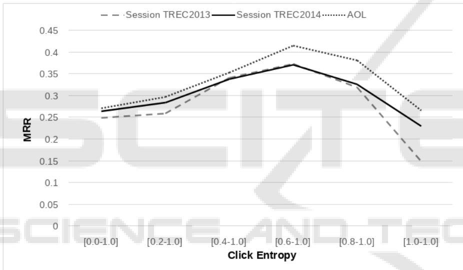

Figure 3: The Changes in MRR with different ranges of click entropy.

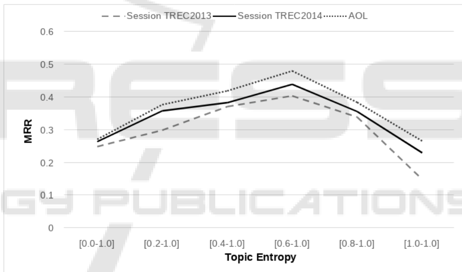

Figure 4: The Changes in MRR with different ranges of topic entropy.

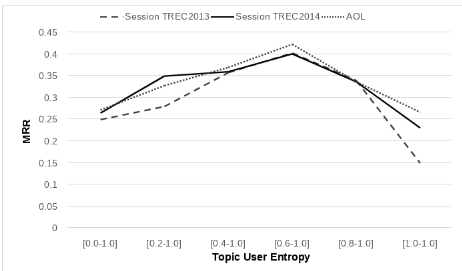

Figure 5: The Changes in MRR with different ranges of topic user entropy.

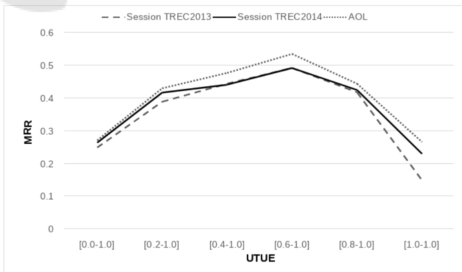

Figure 6: The Changes in MRR with different ranges of UTUE.

5 CONCLUSIONS

In this article, a new metric is proposed to estimate the potential for personalization of new/unseen queries. To do this, the state of the art metrics are investigated and a new metric is proposed based on the correlation between the metrics. When compared to the method proposed by Yano et al. (Yano et al., 2016), our proposed potential for personalization metric is defined in terms of the latent topic models, rather than relying

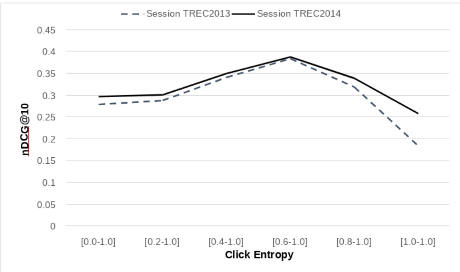

Figure 7: The Changes in nDCG@10 with different ranges of click entropy.

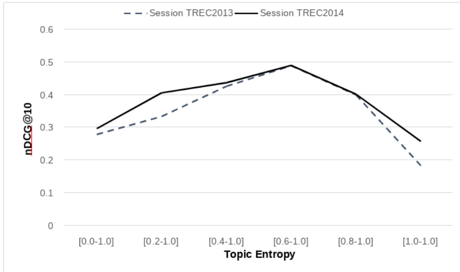

Figure 8: The Changes in nDCG@10 with different ranges of topic entropy.

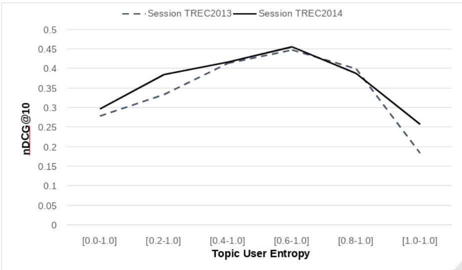

Figure 9: The Changes in nDCG@10 with different ranges of topic user entropy.

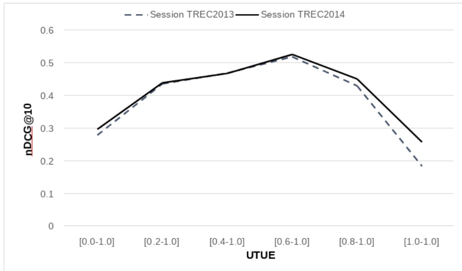

Figure 10: The Changes in nDCG@10 with different ranges of UTUE.

solely on the query history directly. This allows the UTUE to generalize better to rare queries as well as new queries that are not issued previously as it is. Using the topic models, these queries are modeled using similar queries more flexibly. Furthermore, we show that personalization using a combination of PTM and NonPTM improves personalization effectiveness with using UTUE. The mean reciprocal rank and normalized discounted cumulative gain obtained by the personalization ranking model exceeded %53 and %49 respectively. Our results indicate a 4-5% improvement.

ACKNOWLEDGEMENTS

This research is supported by Mavinci Informatics Inc. in Turkey. Mavinci is an R&D company working especially in information and communication technologies, security and defense areas with the capability of software development, artificial intelligence, and machine learning.

REFERENCES

Abri, S., Abri, R., and Cetin, S. (2020a). A classification on different aspects of user modelling in personalized web search. In *4th International Conference on Natural Language Processing and Information Retrieval*.

Abri, S., Abri, R., and Cetin, S. (2020b). Group-based personalization using topical user profile. In *Adjunct Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. ACM.

Abri, S., Abri, R., and Cetin, S. (2020c). Personalized web search using key phrase-based user profiles. In *In 19th International Conference on WWW/Internet (ICWI 2020)*.

Amer, N. O., Mulhem, P., and Géry, M. (2016). Toward word embedding for personalized information retrieval. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*.

Bennett, P. N., White, R. W., Chu, W., Dumais, S. T., Bailey, P., Borisyuk, F., and Cui, X. (2012). Modeling the impact of short- and long-term behavior on search personalization. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 185–194. ACM.

Carman, M. J., Crestani, F., Harvey, M., and Baillie, M. (2010). Towards query log based personalization using topic models. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1849–1852. ACM.

Chirita, P.-A., Nejdl, W., Paiu, R., and Kohlschütter, C. (2005). Using odp metadata to personalize search. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 178–185. ACM.

Cronen-Townsend, S., Zhou, Y., and Croft, B. (2002). Predicting query performance. In *Proceedings of the 25th annual international ACM SIGIR conference on*

Research and development in information retrieval, pages 299–306. ACM.

Dou, Z., Song, R., and Wen, J.-R. (2007). A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*, pages 581–590. ACM.

Harvey, M., Crestani, F., and Carman, M. J. (2013). Building user profiles from topic models for personalised search. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 2309–2314. ACM.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc.

Karimi-Mansoub, S. and Abri, R. (2016). Improvement of semantic search results with providing an updatable dynamic user model. *International Journal of Computer Applications*, 155(4):7–14.

Manning, C. D., Raghavan, P., and Hinrich, S. (2008). Introduction to information retrieval. In *Introduction to Information Retrieval*. Cambridge University Press.

Matthijs, N. and Radlinski, F. (2011). Personalizing web search using long term browsing history. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 25–34. ACM.

Momtazi, S. and Lindenberg, F. (2016). Generating query suggestions by exploiting latent semantics in query logs. In *Journal of Information Science*, pages 437–448. ACM.

Shao, M. and Qin, L. (2014). Text similarity computing based on lda topic model and word co-occurrence. In *2nd International Conference on Software Engineering, Knowledge Engineering and Information Engineering (SEKEIE 2014)*.

Siegg, A., Mobasher, B., and Burke, R. (2007). Web search personalization with ontological user profiles. In *Proceedings of the ACM Conference on information and knowledge management*, pages 525–534. ACM.

Song, R., Luo, Z., Wen, J.-R., Yu, Y., and Hon, H.-W. (2007). Identifying ambiguous queries in web search. In *Proceedings of the 16th international conference on World Wide Web*, pages 1169–1170. ACM.

Teevan, J., Dumais, S., and Horvitz, E. (2005). Beyond the commons: Investigating the value of personalizing web search. In *Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA)*, pages 84–92.

Teevan, J., Dumais, S., and Horvitz, E. (2010). Potential for personalization. In *ACM Transactions on Computer-Human Interaction (TOCHI) TOCHI*, pages 1–31. ACM.

Teevan, J., Dumais, S. T., and Liebling, D. J. (2008). To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 163–170. ACM.

Vu, T., Willis, A., Kruschwitz, U., and Song, D. (2017). Personalised query suggestion for intranet search with temporal user profiling. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 265–268. ACM.

Vu, T. T., Willis, A., and Song, D. (2015a). Modelling time-aware search tasks for search personalisation. In *Proceedings of the 24th International Conference on World Wide Web*, pages 131–132. ACM.

Vu, T. T., Willis, A., Tran, S., and Song, D. (2015b). Temporal latent topic user profiles for search personalisation. In *ECIR:37th European Conference on IR Research*, pages 605–616. ACM.

Wang, Y. and Agichtein, E. (2010). Query ambiguity revisited: clickthrough measures for distinguishing informational and ambiguous queries. In *HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 361–364. ACM.

Wei, S., Yu, Z., Ting, L., and Sheng, L. (2010). Bridging topic modeling and personalized search. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1167–1175. ACM.

Yano, Y., Tagami, Y., and Tajima, A. (2016). Quantifying query ambiguity with topic distributions. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1877–1880. ACM.