A Study of Personalised Medical Literature Search

Richard McCreadie[1], Craig Macdonald[1], Iadh Ounis[1], and Jon Brassey[2]

[1] University of Glasgow
{firstname.lastname}@glasgow.ac.uk
[2] TRIPDatabase.com
{jon.brassey@tripdatabase.com}

Abstract. Medical search engines are used everyday by both medical practitioners and the public to find the latest medical literature and guidance regarding conditions and treatments. Importantly, the information needs that drive medical search can vary between users for the same query, as clinicians search for content specific to their own area of expertise, while the public search about topics of interest to them. However, prior research into personalised search has so far focused on the Web search domain, and it is not clear whether personalised approaches will prove similarly effective in a medical environment. Hence, in this paper, we investigate to what extent personalisation can enhance medical search effectiveness. In particular, we first adapt three classical approaches for the task of personalisation in the medical domain, which leverage the user's clicks, clicks by similar users and explicit/implicit user profiles, respectively. Second, we perform a comparative user study with users from the TRIPDatabase.com medical article search engine to determine whether they outperform an effective baseline production system. Our results show that search result personalisation in the medical domain can be effective, with users stating a preference for personalised rankings for 68% of the queries assessed. Furthermore, we show that for the queries tested, users mainly preferred personalised rankings that promote recent content clicked by similar users, highlighting time as a key dimension of medical article search.

1 Introduction

Medical practitioners access recent articles and case studies published online in peer-reviewed medical literature to inform the treatments that they recommend to patients [1]. Moreover, increasingly, patients are independently researching their own conditions using resources available on the Web. Indeed, a study by the Pew Research Centre indicated that over 80% of Internet users use search tools to find medical information [2]. As a result, it is critical that the search technologies used to access medical resources are effective, such that medical practitioners and patients can access the most informative and up-to-date information available.

However, the user-base of a medical search engine can be very diverse, and may express very different information needs for the same query. For instance, for the query 'AAO', an ophthalmologist (a specialist focusing on diseases of the eye) might be looking for articles published by the American Association of Ophthalmology, while otolaryngologists (clinicians specialising in ear, nose, and throat disorders) might be searching for new articles from the American Academy of Otolaryngology instead.

E. Kanoulas et al. (Eds.): CLEF 2014, LNCS 8685, pp. 74–85, 2014.

In a Web search environment, this type of problem has been tackled using personalised search approaches (e.g. [3–17]), where the search results for a user are altered based upon an explicit or implicit representation of what the user is interested in.

However, to-date, personalisation has not been examined within the context of medical search. Moreover, due to the differences in the Web and medical search domains, it is unclear whether personalised search approaches that have been shown to be effective for Web search will remain effective. For instance, consider a doctor who consults a series of patients with different illnesses over time. Personalisation approaches that mine contextual information from a user's long-term search history (e.g. [18]) could potentially harm search effectiveness by promoting documents relevant to the wrong patient. Hence, in this paper, we investigate whether search personalisation remains effective when applied to the medical domain.

In particular, we adapt three classical personalisation approaches from the literature to the task of medical search, namely P-Click [19] that uses historical clicks by the user to suggest new documents to promote; G-Click [19], which leverages between-user similarity to identify documents clicked by similar users that are relevant; and a keyword vector-based approach [17] that uses both document and user-level evidence to identify relevant documents clicked by similar users. Using click data from TRIP-Database.com, in addition to a user preference study with volunteers from the user-base of that same provider, we evaluate whether personalisation can increase the effectiveness of medical search. Via automatic evaluation based upon click data, we show that personalised approaches can outperform a baseline production system that does not personalise by a statistically significant margin. Moreover, our user-study showed that for 68% of queries, users preferred the personalised rankings, illustrating that personalisation is remains an effective tool for use in the medical domain.

The remainder of this paper is structured as follows. Section 2 discusses prior works in the fields of medical article search and personalisation. In Section 3, we describe the three personalised medical search approaches that we examine later. Section 4 describes our experimental setup, including our dataset and measures, while in Section 5 we discuss our experimental results. We summarise our conclusions in Section 6.

2 Related Work

Search in the Medical Domain: Prior works in textual medical search have focused on how end-users make use of Web search engines to explore health-related topics. For instance, Cartright et al. [20] examined how general medical search differs from diagnosis seeking intents, showing that users follow distinct patterns during search, e.g. starting with symptoms and generalising to causes. Ayers and Kronenfeld [21] examined the relationship between chronic medical conditions and how often users search the Web on related topics. Their results indicate that a user's search behaviour changes based on the number of chronic conditions they suffer from and that the type of information they find can alter their subsequent behaviour. Meanwhile, White and Horvitz [22] investigated the related topic of search intent escalation by users from symptom search to finding related medical professionals or hospitals. Moreover, there are also a series of dedicated medical article search systems available on the Web, such as TRIPDatabase.com,

PubMed and Health on the Net. In general, this shows that end users commonly make use of search engines to satisfy medical information needs and that the results they find can impact their behaviour, hence there is a need for effective medical article search approaches.

However, there have been few approaches proposed for the dedicated searching of medical articles. Early work into medical article search using the MEDLINE database was examined during the Genomics track at the Text REtrieval Conference (TREC) [23], but focused only on core IR techniques such as field-based retrieval [24]. Later research into medical-related search has targeted the retrieval of semi-structured e-health records. For instance, the CLEF eHealth Evaluation Lab[3] examined search and exploration of eHealth data, while the TREC Medical Records track [25] examined cohort identification.

Relatedly, ImageCLEF currently runs a medical task that examines (among other topics) ad-hoc retrieval of medical images.[4] Indeed, a variety of medical image search engines such as Goldminer and Yottalook exist to enable medical practitioners to find relevant medical images for specified medical conditions. However, these systems are concerned with tagged medical images rather than medical articles examined here. Indeed, to the best of our knowledge, no prior research has examined how search personalisation can be used for medical article search, which is the focus of this paper.

Personalised Search Approaches: In contrast, there is a large volume of literature relating to personalisation in the Web search domain. Classical personalisation approaches involved users providing explicit feedback to the search system in the form of a user profile [8, 10]. However, while these approaches can be effective, it has been shown that users are reluctant to provide explicit profiles when searching [26]. Instead, a number of approaches that use previous user search interactions (queries and clicks) have been proposed, since such data can be collected easily and automatically [4, 5, 7, 11–17]. For instance, Sriram *et al.* [27] described a search engine that used only the current user session for personalisation, although session data proved to be too sparse to perform well for some queries. Personalisation using longer periods of user history has also been investigated [11, 14, 28]. Speretta and Gauch [14] and Qiu and Cho [11] leveraged the user's click history to classify them into topic hierarchies, using these hierarchies to re-rank the search results. Another popular approach to personalise search results is to train a general ranking function using personalised data from query and click logs [7, 12, 15, 16]. Of note is that Dou *et al.* [19] proposed two effective approaches named P-Click and G-Click. P-Click promotes documents previously clicked on by the same user for the same query. G-Click promotes documents clicked on by other users for the same query. In a similar manner to [29] and [17], we extend P-Click to consider all previously clicked documents by the user. We then use both P-Click and G-Click as personalised approaches in our later experiments. However, due to sparsity in the query/click logs for some queries, other approaches incorporate information from additional historical sources, rather than just the previous interactions by the current user [15, 17]. For example, Teevan *et al.* [17] created a combined model of user interests by generating keyword indices for queries issued, web-pages viewed

[3] http://clefehealth2014.dcu.ie/

[4] http://www.imageclef.org/2013/medical

and available local content such as emails. They then personalised the ranking produced for a query by up-weighting terms identified as personally relevant, identified from the user's interest model. Expanding on this concept for the medical domain, we test a similar approach that combines implicit click evidence with an interest model for each clinician. We summarise each of the three personalisation approaches that we use in the next section.

3 Personalisation Approaches

We experiment with three approaches from the personalisation literature for use in the medical domain, each representing different ways to describe the user's interests. In particular, we experiment with the P-Click and G-Click personalisation approaches proposed by Dou *et al.* [19], and propose a similar approach to that used by Teevan *et al.* [17], which builds an interest profile from all of the click and profile data we have on each user. We detail each of these approaches below.

3.1 P-Click

The idea underpinning the P-Click approach is that if a user has clicked on a document before, then they are likely to click on that document again later. Hence. when a user enters a query that they have previously entered, then documents they have previously clicked should be up-weighted. Hence, P-Click defines a document re-scoring function $score_{p-click}(Q,d,u)$ as follows:

$$score_{p-click}(Q,d,u) = \frac{|P_{clicks}(Q,d,u)|}{|P_{clicks}(Q,u)| + \beta} \quad (1)$$

where Q is a query, d is a document to be re-scored, u is the user, $|P_{clicks}(Q,d,u)|$ is the number of previous clicks on document d by user u for query Q and $|P_{clicks}(Q,u)|$ is the total number of documents that the user clicked on for query Q. β is a normalisation factor that is set to 0.5, as per the original paper [19].

3.2 G-Click

One of the issues identified with P-Click was that the rankings produced will not be personalised the first time a user enters a new query, since they will have not have clicked on any documents for that query previously. The G-Click approach attempts to solve this issue by performing group-based personalisation. In particular, in the original paper, G-Click represented each user as a weighted vector of 67 pre-defined topic categories [19]. This weighted vector enables similar users to be identified. Under G-Click, each document is re-scored based upon the number of times similar users have clicked each document and the degree of similarity between the users. In this way, if similar users click a document, then that document is promoted in the ranking. The score for a document is calculated as follows:

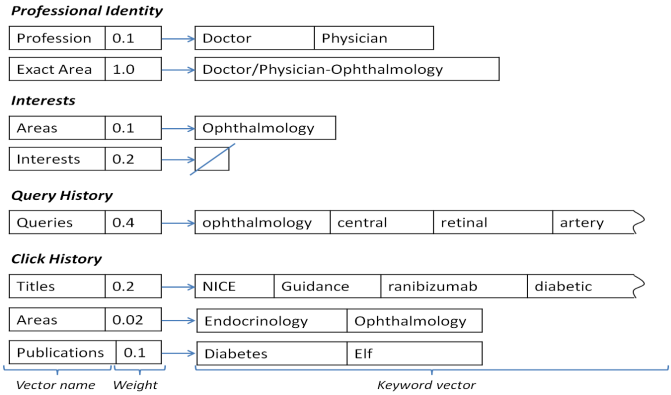


Fig. 1. Representation of a sample user comprised of eight keyword vectors

$$score_{g-click}(Q, d, u) = \frac{\sum_{u_i \in U} sim(u, u_i) \cdot |P_{clicks}(Q, d, u_i)|}{\sum_{u_i \in U} |P_{clicks}(Q, u_i)| + \beta} \tag{2}$$

where U is the set of the k most similar users to u and $sim(u, u_i)$ is the similarity between the users u and u_i. However, in the medical domain, we do not have predefined topic categories for each user. Instead, we build an interest profile based upon the titles of the documents that each user has previously clicked, where each profile is represented as a term vector, denoted ρ_u. User similarity is calculated as the cosine similarity between the two user profiles:

$$sim(u, u_i) = cosine(\rho_u, \rho_{u_i}) \tag{3}$$

3.3 Medical Interest Profiling

As described earlier in Section 2, Teevan *et al.* [17] previously proposed an effective approach that built rich interest profiles for each user using any and all information that the system had access to. Under this approach multiple keyword vectors are constructed, one per information source (e.g. queries or clicked documents), representing the different interests of a user. The interest profiles are used to personalise the document ranking by promoting those documents that share terms with the user's interest profile. We propose a similar approach for medical search personalisation that uses enriched user interest profiles as described below.

First, we represent each user as a weighted series of eight keyword vectors, spread over four aspects, namely: Professional Identity, Interests, Query History and Click History. An example user is illustrated in Figure 1. As can be seen from the figure, each of the four aspects contains one or more keyword vectors. The Professional Identity and Interests aspects come from an explicit profile created when the user registers with the medical search provider. The Query/Click History aspects are generated from an associated query and click log. For example, the 'Areas' vector contains the medical domains that any articles clicked by the user were from, while the 'Publications' vector contains the titles of the publications those clicked articles were published within.

Each keyword vector is assigned a weight, indicating the contribution of that vector to the overall similarity calculation. In our example, the 'Professional Identity' aspect contains two keyword vectors, 'Profession' and 'Exact Area', where the 'Profession' vector has weight 0.1 and contains the keywords 'doctor' and 'physician'. Notably, by weighting each keyword vector, we are able to emphasise aspects of the profile that are important when identifying similar users in the medical domain, e.g. by focusing on the Professional Identity aspects, the algorithm will rank users that come from the same medical background more highly. Furthermore, due to sparsity in the data available, one or more of the keyword vectors may be empty for a given user. For instance, in Figure 1, the 'Interests' vector is empty, indicating that this user did not specify any interests when they registered. Notably, since the search engine we have access to only logs of users who have logged in before querying, the query and click history aspects can also be sparse.

In a similar manner to G-Click described earlier, we define a similarity function between two users $sim(u, u_i)$, such that we can better identify other users with similar interest profiles. We calculate similarity between two users as the sum of the similarities between the each of the eight keyword vectors as shown below:

$$sim(u, u_i) = \sum_{0 \leq j \leq |V|} weight(v_j) \cdot sim(v_j^u, v_j^{u_i}) \tag{4}$$

where u and u_i are users, v_j^u is the j'th keyword vector for u, $weight(v_j)$ is the weight assigned to that vector, where $0 \leq weight(v_j) \leq 1$ and $|V|$ is the number of vectors (eight). We use the cosine measure to calculate the similarity between vectors. In this way, we are able to arrive at a combined estimate of user similarity between medical search engine users that combines both implicit query/click information with explicit information about the user's interests and profession. The weights for each keyword vector are trained using volunteer clinicians on a separate training set (see Section 4).

Using this similarity function, we then re-score each document based both the user's interest profile and the profiles of similar users. In particular, we define a re-scoring function $score_{MIP}(Q, d, u)$ as follows:

$$score_{MIP}(Q, d, u) = \sum_{u_i \in U} score(d, Q) \cdot \begin{cases} \lambda \cdot sim(u, u_i) & \text{if } |P_{clicks}(Q, d, u_i)| > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where d is a document to be scored, Q is the user query, u is the user, U is the set of top k similar users ranked for u, $score(d, Q)$ is the original score assigned to document d for the query Q and $|P_{clicks}(Q, d, u_i)|$ is the number of times user u_i clicked on document d for the query Q. For our experiments, we use the PL2 [30] weighting model to calculate $score(d, Q)$ as it was the most effective based on prior testing on a hold-out training set. We refer to this approach in our later experiments as medical interest profiles (MIP). In the next section, we describe our experimental setup for evaluation of these three personalisation approaches within a medical search scenario.

4 Experimental Setup

Methodology. We evaluate our proposed personalisation approach to medical article search in two manners, both based upon a medical document collection containing

1,418,996 medical articles from the period of June 1855 to April 2013 (although over 90% of these articles were published during the last 10 years) provided by the TRIP-Database.com medical article search engine. We use this dataset since it comes with an associated query/click log and user profiling information from that search engine. First, inspired by prior literature [19], we attempt to evaluate in a fully automatic manner using the aforementioned query/click log. Second, we evaluate as part of a preference user study with volunteers from the aforementioned search engine. We describe the preparation of each evaluation below.

Query-Log Evaluation. For our query-log evaluation, we use medical search queries and clicks issued to TRIPDatabase.com since 2010. In particular, for each user that registered with the search engine, a query log lists that user's search queries and the documents they clicked, grouped into search sessions. Each user also has a profile containing their profession and clinical areas of interest. To create our test topics, we sampled 968 users who had used the search engine for more than one month and had issued one or more queries during the month of April 2013 (our chosen test month). For each of these users, we selected the last search session they made. The first query in that session for each user is used as a topic query, forming a test set comprised of 968 topics. The document(s) that the user clicked upon during the last session are considered to be relevant, similarly to [5]. All queries and clicks made by the user prior to the last session are used as training data, i.e. to calculate P_{clicks} in Equation 1,2 and 5, as well as generate the user interest profiles under G-Click and MIP. However, for 897 of the 968 test topics, the click data available in the test session was sufficiently sparse that none of the documents promoted by the personalised approaches had clicks. This is to be expected, since the personalised approaches (particularly G-Click and MIP that use other similar user's interaction history to re-rank the results) are likely to promote documents that the user did not originally see/click on. As a result, we further down sample the test set to the 71 topics where the performance of the baseline and personalised approaches differ.

User-Study Evaluation. To mitigate the limitations of the query-log evaluation, we also perform a user-study using volunteers from the user-base of TRIPDatabase.com. We recruited 17 users and then had each suggest queries that we then use to test personalisation. A total of 90 queries were suggested, which we use as topics. These topics do not overlap with the 71 used in the query-log evaluation. For each query and the user that suggested it, we produce both personalised and unpersonalised rankings. We then performed a blind side-by-side evaluation [31] to determine whether the users preferred the personalised rankings over the unpersonalised ones. Our evaluation interface is shown in Figure 2. As can be seen, we render each ranking side-by-side in a pair-wise manner per-query, the user selects their preferred ranking using the buttons at the top of each. The positioning of the personalised and unpersonalised rankings (left or right position) are randomised. This ordering is hidden from the assessors. To further investigate why users preferred one ranking approach over another, for each ranking pair, we also had the volunteer fill a questionnaire, where they select zero or more reasons for choosing that ranking as follows:

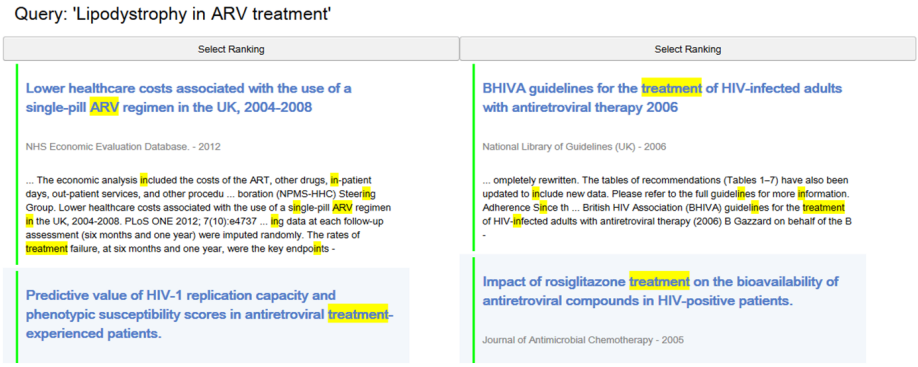


Fig. 2. User study preference assessment interface

- The selected ranking was more relevant to the query.
- Documents within the selected ranking were more informative.
- The selected ranking provided better coverage of the topic.
- The documents within the selected ranking were more recent.

Parameter Training. The MIP approach requires weights for the λ parameter and each keyword vector ($weight(v_j)$ in Equation 4). To generate these weights, we had volunteers from the user-base of TRIPDatabase.com score each document ranked by MIP on a separate set of 18 topics (this does not overlap with either the two other topic sets used above). Volunteers labelled each document as relevant or not to the topic for a given user profile. The optimal value for the λ parameter and each keyword vector weight was optimised via a parameter scan over each (where $0 \leq weight(v_j) \leq 1$, $0 \leq \lambda \leq 10$ and increments of size 0.1 were tested).

Measures. For evaluation using the query log, we report the classical IR ranking metrics mean average precision (MAP), Precision at rank 5 (P@5) and the Rank Scoring metric from the collaborative filtering literature [32].

5 Results

To determine whether personalisation is effective when applied to the task of medical article search, we investigate the following two research questions, each in a separate section.
- Can personalisation approaches more effectively rank medical documents than the unpersonalised baseline? (Section 5.1)
- Do end-users prefer the (MIP) personalised rankings to the unpersonalised baseline? (Section 5.2)

5.1 Evaluating Using Click-Logs

We begin by investigating our first research question, i.e. can personalisation approaches more effectively rank medical documents than the unpersonalised baseline? To do so,

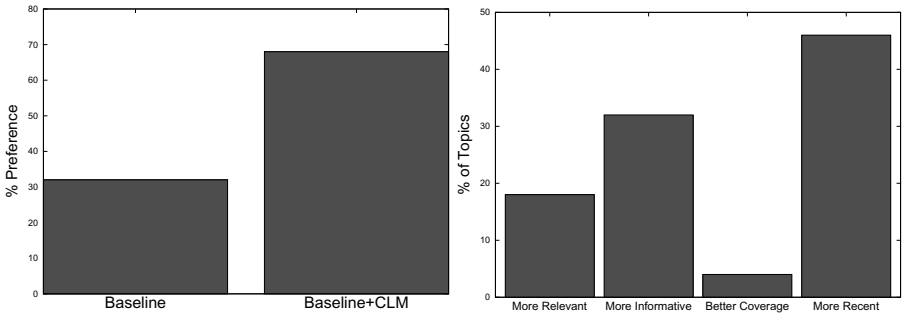
Table 1. Ranking performance of the medical search approaches tested using the 71 topics where a personalised approach promoted one or more documents that also received a click. Statistical significance ($p<0.05$) over the Baseline approach using the paired t-test is denoted ▲.

Approach	Number of Documents Promoted	Performance Measures		
		P@5	MAP	Rank Scoring
Baseline	n/a	0.1250	0.1651	0.3101
Baseline + P-Click	1,291	0.1361	0.1769	0.3334
Baseline + G-Click	8,459	0.1694	**0.2731** ▲	0.3851
Baseline + MIP	2,063	**0.1951** ▲	0.2571 ▲	**0.4282**

we measure the ranking performance of the three personalised ranking approaches (P-Click, G-Click and MIP) against the baseline production system (Baseline) using the 71 topics sampled from our medical query-log. Table 1 reports the performance of the baseline and personalised approaches in terms of the MAP, P@5 and Rank Scoring metrics. Higher scores indicate that the approaches are ranking the documents clicked by users higher in the ranking.

From Table 1, we observe the following. First, all of the personalised approaches outperform the production system. Indeed, P-Click provides the smallest increase in performance of 1.21% absolute P@5, while the largest increase in performance was observed from the MIP approach, with a statistically significant +6.94% absolute P@5. Second, we see that under the precision-orientated metrics, P@5 and Rank Scoring, MIP personalisation outperformed the unpersonalised baseline and both the personalised P-Click and G-Click approaches. This shows that in the high ranks, using rich user interest profiles to find similar users is more effective than using the user-profile or user-click data alone when identifying additional documents that the user might be interested in. For instance, one topic where MIP outperformed G-Click was for the query 'personality disorder' when the user was interested in 'antisocial behaviour' and had selected the clinical area 'Psychiatry'. For this query, the MIP approach used its user interest profiles to find a similar clinician who had previously clicked on two documents relevant to the current user, i.e. 'Antisocial personality disorder, treatment, management and prevention' and 'Psychological interventions for antisocial personality disorder', which were then ranked higher. For the same topic, G-Click failed to find these two documents, because the current user had not clicked on documents about antisocial behaviour before.

However, we also see from Table 1 that G-Click evidences higher performance than MIP under MAP. This result is to be expected, since MIP is by its nature more conservative in how it promotes documents from similar users, since its more granular between-user similarity function (Equation 4) limits the contribution of users where only a sub-set of the interest profile vectors match. This contrasts with G-Click, where only the (cosine) similarity between the clicked document titles are is used to weight the emphasis placed on each previously clicked document. Indeed, from the second column of Table 1, we can see that G-Click promoted more than 4 times the number of documents than MIP did. Hence, MIP can be characterised as a more precision-orientated approach than G-Click. To answer our first research question, based upon the click-based assessments available, personalised approaches appear to be effective, significantly outperforming the baseline production system in the case of MIP. Indeed,



(a) User Preference (b) Questionnaire Answers

Fig. 3. Proportion of queries that users prefer the MIP personalised ranking in comparison to the baseline and the reasons provided for that preference

the MIP approach that uses richer user interest profiles is able to markedly outperform both other personalised approaches in terms of precision at rank 5 (P@5) and provides close-to G-Click's performance under mean average precision (MAP).

5.2 Evaluating via User-Study

Having shown that the personalised approaches are effective using the click data available to us, we next evaluate our second research question, i.e. do end-users prefer the (MIP) personalised ranking to the production system currently deployed by the search engine?[5] To this end, we perform a user study, where we have users state their preference for either the unpersonalised baseline or the MIP personalised rankings for the 90 query topics that they themselves suggested.

Figure 3 (a) shows the proportion of users that preferred the MIP personalised ranking in comparison to the production system over the 90 topics suggested by volunteer clinicians in our user study. From Figure 3 (a), we see that the personalised ranking was preferred over the unpersonalised baseline for the majority (68%) of queries tested, supporting our earlier observations on the click data indicating that MIP is effective. To illustrate, one example where the MIP ranking was preferred was for the query 'temporomandibular joint dysfunction' (about the jaw joint in humans) where the user had specified the clinical area 'Otolaryngology' (a field dealing with the ear, nose, and throat) and an interest about hearing aids. From this information, the MIP component identified a clinician from the same field who had been searching about anti-inflammatory drugs to treat people with hearing aids, thereby promoting a study about anti-inflammatory drugs that the volunteer clinician found informative.

However, also of interest is why users prefer the MIP ranking. To evaluate this, recall that for each ranking pair, the volunteer clinicians also filled a questionnaire regarding why they preferred that ranking in terms of 4 criteria, namely: relevance, informativeness, topical coverage and recency (see Section 4). Figure 3 (b) reports the proportion

[5] Note that we compare against MIP only here, since we showed previously that MIP outperforms P-Click and G-Click.

of volunteers that selected each reason when the MIP ranking was preferred. From Figure 3 (b), we see that the primary reasons that users preferred the MIP-personalised ranking were because the documents ranked highly were either more recent (46%) or informative (32%). This indicates that MIP is mainly identifying similar users that have recently queried on a similar topic and clicked on useful documents that were not already prominent in the ranking.

To answer our second research question, we conclude that personalisation (using MIP) is effective, since the personalised rankings that it produced are preferred over the unpersonalised baseline for the majority of queries. Furthermore, we have shown that users mainly preferred personalised rankings that promote recent content clicked by similar users, highlighting time as a key dimension of medical article search.

6 Conclusions

In this paper, we examined whether personalisation approaches previously proposed for use in the Web search domain remain effective for the task of medical article search. We adapted three classical personalisation approaches from the literature for medical search that leverage the user's clicks, clicks by similar users and explicit/implicit user interest profiles, respectively. Through experimentation over a medical search query log, we showed that these approaches could outperform an unpersonalised baseline system, and that the approach that used explicit/implicit user interest profiles was the most effective, suggesting that using the affinity between clinicians is a useful source of evidence to use when finding additional relevant content. Moreover, through a user study with volunteer clinicians, we showed that users prefer the rankings produced through personalisation to the baseline in a blind test - showing that personalisation approaches are effective in the medical article search domain. Finally through a questionnaire with our volunteer users, we found that the main reasons that the personalised approach improved over the baseline ranking from the user perspective was that the personalised results contained additional informative and recent content - highlighting the importance of finding the most up-to-date medical content of interest to each clinician. For future work, we aim to investigate how the across-session and within-session search patterns of users can be used to further personalise the medical search results for a user.

References

1. Guyatt, G., Rennie, D., Hayward, R., et al.: Users' guides to the medical literature: A manual for evidence-based. In: Clinical Practice, vol. 706
2. Susannah Fox: Report: Health, Digital Divide - Health Topics (2011), `http://pewinternet.org/Reports/2011/HealthTopics.aspx`
3. Chirita, P.A., Nejdl, W., Paiu, R., Kohlschütter, C.: Using ODP metadata to personalize search. In: Proc. of SIGIR (2005)
4. Daoud, M., Tamine-Lechani, L., Boughanem, M., Chebaro, B.: A session based personalized search using an ontological user profile. In: Proc. of SAC (2009)
5. Dou, Z., Song, R., Wen, J.R.: A large-scale evaluation and analysis of personalized search strategies. In: Proc. of WWW (2007)

6. Gauch, S., Chaffee, J., Pretschner, A.: Ontology-based personalized search and browsing. Web Intelligence and Agent Systems Journal (2003)
7. Joachims, T.: Optimizing search engines using clickthrough data. In: Proc. of SIGKDD (2002)
8. Liu, F., Yu, C., Meng, W.: Personalized Web search by mapping user queries to categories. In: Proc. of CIKM (2002)
9. Liu, F., Yu, C., Meng, W.: Personalized web search for improving retrieval effectiveness. IEEE Transactions on Knowledge and Data Engineering (2004)
10. Pretschner, A., Gauch, S.: Ontology based personalized search. In: Proc. of ICTAI (1999)
11. Qiu, F., Cho, J.: Automatic identification of user interest for personalized search. In: Proc. of WWW (2006)
12. Shen, X., Tan, B., Zhai, C.: Implicit user modeling for personalized search. In: Proc. of CIKM (2005)
13. Sieg, A., Mobasher, B., Burke, R.: Web search personalization with ontological user profiles. In: Proc. of CIKM (2007)
14. Speretta, M., Gauch, S.: Personalized search based on user search histories. In: Proc of WIC (2005)
15. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive Web search based on user profile constructed without any effort from users. In: Proc. of WWW (2004)
16. Sun, J.T., Zeng, H.J., Liu, H., Lu, Y., Chen, Z.: CubeSVD: A novel approach to personalized web search. In: Proc. of WWW (2005)
17. Teevan, J., Dumais, S.T., Horvitz, E.: Personalizing search via automated analysis of interests and activities. In: Proc. of SIGIR (2005)
18. Tan, B., Shen, X., Zhai, C.: Mining long-term search history to improve search accuracy. In: Proc. of SIGKDD (2006)
19. Dou, Z., Song, R., Wen, J.R.: A large-scale evaluation and analysis of personalized search strategies. In: Proc. of WWW (2007)
20. Cartright, M.A., White, R.W., Horvitz, E.: Intentions and attention in exploratory health search. In: Proc. of SIGIR (2011)
21. Ayers, S., Kronenfeld, J.: Chronic illness and health-seeking information on the Internet. Health Journal (2007)
22. White, R.W., Horvitz, E.: Web to world: Predicting transitions from self-diagnosis to the pursuit of local medical assistance in Web search. In: Proc. of AMIA (2010)
23. Hersh, W., Voorhees, E.: TREC Genomics special issue overview. Information Retrieval Journal (2009)
24. Fujita, S.: Revisiting Again Document Length Hypotheses TREC 2004 Genomics Track Experiments at Patolis. In: Proceedings of TREC (2004)
25. Voorhees, E., Hersh, W.: Overview of the TREC 2012 Medical Records Track. In: Proc. of TREC (2012)
26. Carroll, J.M., Rosson, M.B.: Paradox of the active user. The MIT Press (1987)
27. Sriram, S., Shen, X., Zhai, C.: A session-based search engine. In: Proc. of SIGIR (2004)
28. Tan, B., Lv, Y., Zhai, C.: Mining long-lasting exploratory user interests from search history. In: Proc. of CIKM (2012)
29. Matthijs, N., Radlinski, F.: Personalizing web search using long term browsing history. In: Proc. of WSDM (2011)
30. Amati, G.: Probabilistic Models for Information Retrieval based on Divergence from Randomness. PhD thesis, University of Glasgow (2003)
31. Benjamin, C.A.: Low-Cost and Robust Evaluation of Information Retrieval Systems. PhD thesis, University of Massachusetts Amherst (2009)
32. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proc. of UAI (1998)