Bayesian Optimization in Interactive Scientific Search

Tuukka Ruotsalo[1,2], **Jaakko Peltonen**[1,2,4], **Manuel J.A. Eugster**[1,2], **Dorota Gł owacka**[1,3], **Ksenia Konyushkova**[1,3], **Kumaripaba Athukorala**[1,3], **Ilkka Kosunen**[1,3], **Aki Reijonen**[1,3], **Petri Myllymäki**[1,3], **Giulio Jacucci**[1,3], **Samuel Kaski**[1,2,3]

[1] Helsinki Institute for Information Technology HIIT
[2] Aalto University
[3] Department of Computer Science, University of Helsinki
[4] School of Information Sciences, University of Tampere

Abstract

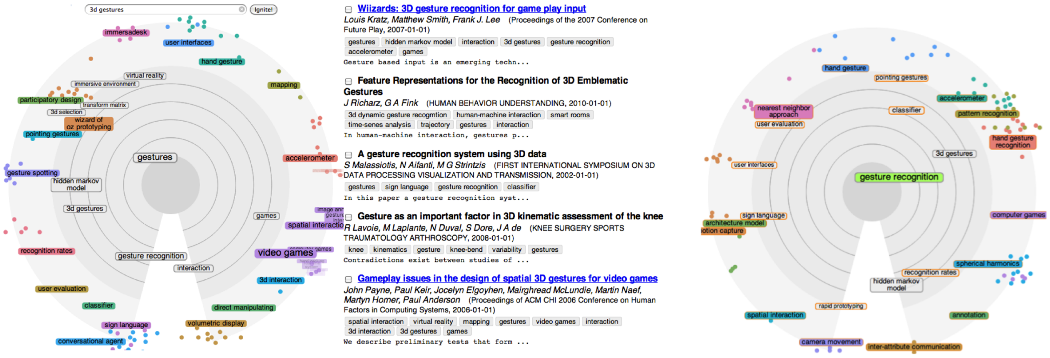
Researchers must navigate big data. Current scientific knowledge includes 50 million published articles. How can a system help a researcher find relevant documents in her field? The key is to model the researcher's information need, and use Bayesian optimization to interactively improve the model. We introduce *IntentRadar*, an interactive search user interface and search engine that anticipates user's search intents by estimating them form user's interaction with the interface. The system emphasizes feedback options according to Bayesian optimization based estimates, to encourage user feedback on the most promising options. To do this, the estimated intents are visualized on a radial layout that organizes potential intents as directions in the information space. The intent radar assists users to direct their search by allowing feedback to be targeted on keywords that represent the potential intents. Users can provide feedback by manipulating the position of the keywords on the radar. The system then learns and visualizes improved estimates and corresponding documents. *IntentRadar* has been shown to significantly improve users' task performance and the quality of retrieved information without compromising task execution time. This is a short version of our previous work [6], focusing on use of Bayesian optimization within the system.

1 Introduction

Exploration and search for relevant scientific literature are main tasks of a researcher. In big data traditional search solutions become increasingly insufficient, and machine learning based assistance is needed. A main problem in exploratory search is that it can be hard for users to formulate queries precisely, since information needs evolve throughout the search session as users gain more information. In a common search strategy, the user issues a quick, imprecise query, hoping to get into approximately the right part of the information space, and then directs the search to obtain the information of interest around the initial entry-point in the information space [7].

Given the vast number of possible interests of the user, gathering information on user intent in an efficient interactive manner is crucial. Current methods to support users to explore are either based on suggesting query terms, or allowing faster access to the search result set by faceted browsing or search result clustering [9, 3]. Such feedback mechanisms can trap the user to the initial query context and cause cognitive burden to the user [4]. We propose that better support for exploration can be provided by visualizing the relevant information space using higher level representations of the data, namely keywords extracted from documents and using Bayesian optimization strategies to choose and organize keywords for feedback on the interface [6, 2]. Our system improves interactive search of 50 million scientific articles from Thomson Reuters, ACM, IEEE, and Springer. [1]

[1] This work was recently presented at CIKM 2013. [6]



□ Wiizards: 3D gesture recognition for game play input
Louis Kratz, Matthew Smith, Frank J. Lee (Proceedings of the 2007 Conference on Future Play, 2007-01-01)
gestures | hidden markov model | interaction | 3d gestures | gesture recognition | accelerometer | games
Gesture based input is an emerging techn...

□ **Feature Representations for the Recognition of 3D Emblematic Gestures**
J Richarz, G A Fink (HUMAN BEHAVIOR UNDERSTANDING, 2010-01-01)
3d dynamic gesture recogmtion | human-machine interaction | smart rooms | time-senes analysis | trajectory | gestures | interaction
In human-machine interaction, gestures p...

□ **A gesture recognition system using 3D data**
S Malassiotis, N Aifanti, M G Strintzis (FIRST INTERNATIONAL SYMPOSIUM ON 3D DATA PROCESSING VISUALIZATION AND TRANSMISSION, 2002-01-01)
gestures | sign language | gesture recognition | classifier
In this paper a gesture recognition syst...

□ **Gesture as an important factor in 3D kinematic assessment of the knee**
R Lavoie, M Laplante, N Duval, S Dore, J A de (KNEE SURGERY SPORTS TRAUMATOLOGY ARTHROSCOPY, 2008-01-01)
knee | kinematics | gesture | knee-bend | variability | gestures
Contradictions exist between studies of ...

□ Gameplay issues in the design of spatial 3D gestures for video games
John Payne, Paul Keir, Jocelyn Elgoyhen, Mairghread McLundie, Martin Naef, Martyn Horner, Paul Anderson (Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems, 2006-01-01)
spatial interaction | virtual reality | mapping | gestures | video games | interaction | 3d interaction | 3d gestures | games
We describe preliminary tests that form ...

Figure 1: **Left: IntentRadar interface.** The system uses a radar metaphor. The current intent estimate for which the results on the right-side list are retrieved, is visualized as keywords on a radar screen (inner dark grey area). Predicted alternative intents, that help users explore their information need, are shown as as keywords in the outer (light gray) area, organized as directions on the radar. In both areas, radius corresponds to relevance of keywords (closer to center = more relevant) and angular distance between keywords corresponds to directions of search (technically: similarity of how keyword relevance would be affected by potential feedbacks). The keywords can be enlarged with a fisheye lens that follows the mouse cursor The user can give feedback by dragging keywords closer to (or away from) the center of the radar. **Right: after feedback.** The user increased relevance of "gesture recognition" by dragging it to the center of the radar. The system computed and visualized new estimated relevant intents, such as "pointing gestures", "recognition rates", and "hidden Markov models". Documents are retrieved for the new intent estimate. The user can continue exploring the new intents.

2 Search User Interface

The *IntentRadar* interface is shown in Figure 1. It assists users in exploring information related to a given topic effectively by allowing rapid feedback loops and helping make sense of available information around the initial query context. We represent the user's interests by weights on keywords, and display them on a radial layout for feedback. Choice and locations of keywords are optimized by Bayesian optimization, both for the the inner circle (representing current intent) and for the outer circle (representing future intents) as discussed in the next section. Figure 1 (left) shows *IntentRadar*'s response to an initial query "3d gestures" with a list of retrieved documents and intents visualized as a radar for feedback. Potential feedback, chosen by Bayesian optimization, includes "video games", "user interfaces", "gesture recognition" and "virtual reality". In Figure 1 (right) the user gave positive feedback to "gesture recognition", directing the search towards it. The user is offered further options to continue the exploration towards topics estimated as potentially relevant, including specific topics such as "hidden Markov models" and general topics such as "spatial interaction".

3 Bayesian Optimization for Learning Search Intents

Learning the user's search intents is based on two models: the *retrieval model* which estimates the probability of relevant documents based on the estimates of the intent model, and the *intent model* which estimates the present and potential future intents of the user based on the interaction history.

Document retrieval model. We use a language modeling approach [10] to estimate the relevance ranking of documents d_j. The intent model yields a weight vector $\hat{\mathbf{v}}$ having a weight $\hat{v}_i$ for each keyword k_i. On the first iteration, we use the typed query with weight 1 as the intent model. We use a multinomial unigram language model. The $\hat{\mathbf{v}}$ is treated as a sample of a desired document, and the d_j are ranked by probability to observe $\hat{\mathbf{v}}$ as a random sample from the language model M_{d_j} of d_j; with maximum likelihood estimation $\hat{P}(\hat{\mathbf{v}}|M_{d_j}) = \prod_{i=1}^{|\hat{\mathbf{v}}|} \hat{v}_i \hat{P}_{mle}(k_i|M_{d_j})$. To improve estimation we use Bayesian Dirichlet smoothing so that $\hat{P}_{mle}(k_i|M_{d_j}) = \frac{c(k_i|d_j)+\mu p(k_i|C)}{\sum_k c(k|d_j)+\mu}$ where $c(k_i|d_j)$ is the count of keyword k_i in d_j, $p(k_i|C)$ is the proportion of k_i in the document collection, and we set $\mu = 2000$ as suggested in the literature [10]. The documents d_j are ranked by $\alpha_j = \hat{P}(\hat{\mathbf{v}}|M_{d_j})$. We could just show the top ranked documents, but to expose the user to more novel documents, we sample documents from the list and show them in ranked order: we use Dirichlet Sampling, where $f_j \sim Gamma(\alpha_j, 1)$ is sampled for each document d_j, and the d_j with highest f_j are shown to the user. We favor documents whose keywords get positive feedback: at each iteration, α_j is increased by 1 for d_j where at least one keyword got positive user feedback, and the α_j are then renormalized.

Estimation of search intent. We use Bayesian optimization to build the current estimate of search intent. By dragging keywords, the user gives feedback as relevance scores $r_i \in [0, 1]$ to a subset of J keywords $k_i, i = 1, \ldots, J$. Here $r_i = 1$ denotes keyword k_i is highly relevant and the user would like to direct her search in that direction, and $r_i = 0$ denotes the keyword is of no interest. Let $\mathbf{k}_i$ be binary $n \times 1$ vectors telling which of the n documents keyword k_i appeared in. To boost documents with rare keywords, we convert the $\mathbf{k}_i$ into *tf-idf* representation.We assume the relevance score r_i is a random variable with expected value $\mathbb{E}[r_i] = \mathbf{k}_i^\top \mathbf{w}$. The weight vector $\mathbf{w}$ is estimated from the user's feedback by the LinRel algorithm [1]. Let the column vector $\mathbf{r}^{feedback} = [r_1, r_2, \ldots, r_p]^\top$ contain the p relevance scores received so far from the user for keywords $k_1, \ldots, k_p$, and let $\mathbf{K} = [\mathbf{k}_1, \ldots, \mathbf{k}_p]^T$ be the matrix of their feature vectors. LinRel estimates $\hat{\mathbf{w}}$ by solving the linear regression $\mathbf{r}^{feedback} = \mathbf{K}\mathbf{w}$, and calculates an estimated relevance score $\hat{r}_i = \mathbf{k}_i^\top \hat{\mathbf{w}}$ for each keyword k_i.

Acquisition of feedback. As feedback is given by the user, it is crucial to present (select and organize) keywords so that user feedback will be targeted to relevant options: keywords that are relevant to the user and informative to the system. We apply Bayesian optimization based approaches both for selection and organization of keywords. **Selection of presented keywords:** At each iteration the system might simply show keywords with highest estimated relevance, but with limited feedback this exploitative choice could be suboptimal; or the system could exploratively pick keywords where feedback would improve accuracy of $\hat{\mathbf{w}}$. To deal with the exploration-exploitation tradeoff we show keywords not with the highest relevance score, but with the largest upper confidence bound (UCB) for the score. If σ_i is an upper bound on standard deviation of the relevance estimate $\hat{r}_i$, the upper confidence bound of keyword k_i is computed as $\hat{r}_i + \alpha\sigma_i$, where $\alpha > 0$ is a constant used to adjust the confidence level of the bound. Let $\mathbf{r}^{feedback}$ again denote the vector of all relevance scores received from the user. In each iteration, LinRel computes $\mathbf{s}_i = \mathbf{K}(\mathbf{K}^\top\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{k}_i$ where λ is a regularization parameter, and the keywords k_i that maximize $\mathbf{s}_i^\top \mathbf{r}^{feedback} + \frac{\alpha}{2}\|\mathbf{s}_i\|$ are selected for presentation. This selection lets us both maximize relevance of intent estimates and reduce system uncertainty with limited feedback. **Organization of presented keywords.** To allow users to target their feedback, it is useful to organize keywords intelligently on the radar. Radius of keywords is proportional to their current relevance UCB (closer = more relevant). Angles of keywords are computed so that keywords will form directions in the information space: keywords get similar angles if their UCB would change similarly with respect to a set of additional feedback. In detail we consider a set of potential feedback, L on-screen keywords the user might drag to center of the radar: in turn we give each of them a pseudo-feedback score of 1 added to current feedback, and recompute UCBs of all keywords; for each keyword k_i we thus get a vector $\hat{\mathbf{r}}_i^{future}$ containing L future relevances (UCBs), one for each alternative feedback. We then organize keywords by dimensionality reduction from the L-dimensional $\hat{\mathbf{r}}^{future,l}$ to the one-dimensional angles, by applying a neighbor embedding method [8], with minor adjustments detailed in [6]. This data-driven layout thus represents future intents as directions on the radar.

4 Experiments

Effectiveness of *IntentRadar* has been studied in task-based experiments where users (30 graduate students from two universities) were asked to solve research tasks using a database of over 50 million scientific articles. The comparisons were conducted against 1) within-system baselines [6, 2] of list-based visualization and only typed-query interaction, and 2) Google Scholar [5]. Experts conducted double-blind relevance assessments of articles and keywords presented by any of the sysystems, on binary scales: *relevance*—is this article relevant to the search topic, *obviousness*—is it a well-known overview article, *novelty*—is it uncommon yet relevant to a given topic/subtopic. The assessments were used as ground truth for evaluations of *user task performance* (assessment of their answers

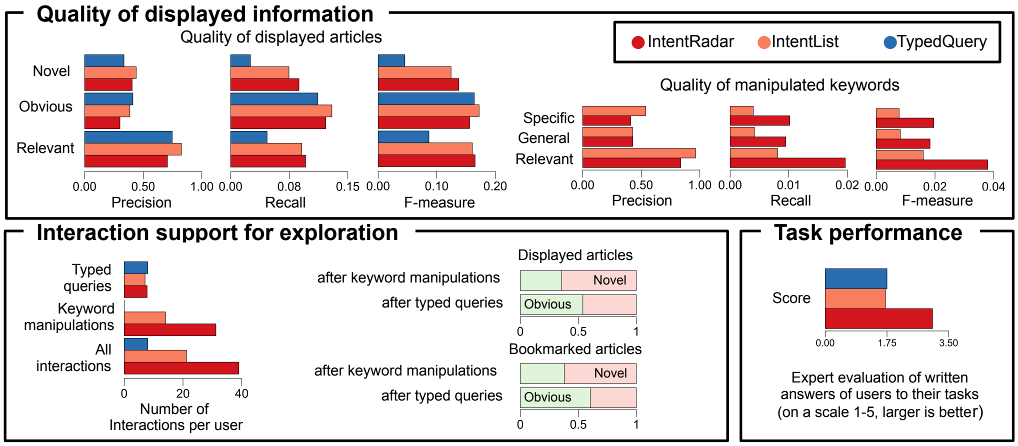


Figure 2: Results. Comparison methods: IntentList (simplified interface representing intents only as a list of top keywords) and TypedQuery (traditional search interface based on typing queries). **Improved task performance:** *IntentRadar* improves users' task performance (answers submitted for research tasks) compared to state-of-the-art retrieval methods and the prominent commercial search engine Google Scholar [6, 5]. **Quality of retrieved information:** *IntentRadar* helps users to move away from the initial query context, thus allowing to substantially increase recall while preserving precision in particularly for novel information [6, 2]. **Enhanced interaction:** Despite more complex visualization, users interacted with the *IntentRadar* interface twice to nearly four times more than the comparison systems without compromising the task execution time. [6].

to tasks), *quality of displayed information* (precision, recall, F-measure), *interaction support for directing exploration* (numbers of interactions, information received in response). Full details of the procedures are in [6, 2, 5].

Fig. 2 summarizes the results. *IntentRadar* improved user's task performance: answers that users provided in response to the given search tasks were graded higher by experts. The interface also enhanced interaction: users of *IntentRadar* initiated up to three times more interaction and the interface reduces users' scanning time with respect to the available option space. *IntentRadar* also yielded higher-quality retrieved information (precision and recall of novel information returned by the search engine in response to user interactions).

5 Conclusions

We introduced IntentRadar for directing exploratory search, based on interactive intent modeling where user feedback is targeted to keywords selected and organized through Bayesian optimization, and demonstrated its usefulness in task-based user experiments. The interactive intent modeling and visual interface let users direct their search; results show it can significantly improve users' performance in exploratory search tasks. The improvements can be attributed to improved quality of displayed information in response to interaction, better targeted interaction, and improved support for directing search to achieve novel information. Interaction with the visualization does not replace query-typing, but offers a complementary way to direct search towards novel, but still relevant information. The improved quality of information displayed on the IntentRadar interface also transfers to improved task performance. Our findings suggest that interactive intent modeling and Bayesian optimization can significantly improve effectiveness of exploratory search.

Acknowledgements

This work has been partly supported by the Academy of Finland (Multivire and the COIN Center of Excellence, and 252845) and TEKES (D2I and Re:KnoW). Certain data included herein are derived from the Web of Science prepared by THOMSON REUTERS, Inc., Philadelphia, Pennsylvania, USA: Copyright THOMSON REUTERS, 2011. All rights reserved. Data is also included from the Digital Libraries of the ACM, IEEE, and Springer.

References

[1] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3:397 – 422, 2002.

[2] Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Kumaripaba Athukorala, Samuel Kaski, and Giulio Jacucci. Directing exploratory search: reinforcement learning from user interactions with keywords. In *Proc. IUI'13*, pages 117–128. ACM, 2013.

[3] Marti A. Hearst and Jan O. Pedersen. Reexamining the cluster hypothesis: scatter/gather on retrieval results. In *Proc. SIGIR'96*, pages 76–84. ACM, 1996.

[4] Diane Kelly and Xin Fu. Elicitation of term relevance feedback: an investigation of term source and context. In *Proc. SIGIR'06*, pages 453–460. ACM, 2006.

[5] T. Ruotsalo, K. Athukorala, D. Głowacka, K. Konuyshkova, A. Oulasvirta, S. Kaipiainen, S. Kaski, and G. Jacucci. Supporting exploratory search tasks with interactive user modeling. In *Proc. ASIST' 13*, 2013.

[6] Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing exploratory search with interactive intent modeling. In *Proc. CIKM'13*, pages 1759–1764. ACM, 2013.

[7] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proc. of SIGCHI*, pages 415–422, 2004.

[8] Jarkko Venna, Jaakko Peltonen, Kristian Nybo, Helena Aidos, and Samuel Kaski. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *J. Mach. Learn. Res.*, 11:451–490, 2010.

[9] Ka-Ping Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proc. CHI'03*, pages 401–408. ACM, 2003.

[10] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.