Modeling Age-related Differences in Information Search

Saraschandra Karanam[1], Herre van Oostendorp[1]

Department of Information and Computing Sciences, Utrecht University[1]

Abstract

A number of cognitive processes are involved in the process of information search: memory, attention, comprehension, problem solving, executive control and decision making. Several cognitive factors such as aging-related cognitive abilities in turn influence either positively or negatively the above cognitive processes. We argue that the traditional click models from information retrieval community that predict user clicks do not take into account the effects of aging. We propose to exploit the capabilities of computational cognitive models that can simulate the effects of cognitive factors on information search behavior. In this direction, we present some ideas to incorporate effects of aging into a computational cognitive model called CoLiDeS+. Preliminary analysis of our ideas on predicting individual differences due to aging effects show promising outcomes.

1   Introduction

Internet is transforming how we communicate, shop, get health-related advice, plan a vacation, entertain ourselves and all in all how we do just about anything. It is pervasive, ubiquitous and becoming an indispensable part of human life today. However, a number of barriers still exist preventing large scale adoption and optimal usage of the Internet. One of the reasons for the slow adoption is that a number of cognitive processes are involved during an information search process such as memory (keeping track of previously viewed information), attention (understanding the visual layout of the website or search engine result pages), comprehension (evaluating the relevance of search results, understanding the content of websites), problem solving (information can be found in multiple locations and there could be multiple paths leading to them), decision making (choosing a relevant search result) and executive control (reformulating unfruitful queries, backtracking to earlier pages, comparing new information with what was found earlier). These cognitive processes, in turn, are known to be affected by one or more cognitive factors such as age (Chevalier *et al.,* 2015).

Modeling and simulation of user behavior during information search has been an active area of research in the information retrieval community. Many click models to simulate and predict user behavior during information search have been proposed (Chucklin *et al.,* 2015). However, except for a few models (Xing *et al.,* 2013, Shen *et al.,* 2012, Hu *et al.,* 2011), most of them do not consider variations caused by cognitive factors and do not treat search behavior as a sequence of interdependent steps. Moreover, they provide only limited process description. Our focus therefore in this paper, will be on computational cognitive models. The focus of computational cognitive models is on describing and explaining the process that leads to arriving at the target information given an information need and they are therefore more capable of providing opportunities to incorporate behavioral differences due to variations in cognitive factors. In this paper, we propose some preliminary ideas that can be incorporated into computational cognitive models to simulate the behavioral differences in information search performance due to aging as cognitive factor.

The remainder of this paper is organized as follows. Section 2 introduces the computational cognitive model CoLiDeS+ (Juvina & Van Oostendorp, 2008), which is based on well tested theories of cognitive psychology and cognitive science on information search and gives a process view of information search. Section 3 describes the effects of aging on information search performance and Section 4 gives details on how age-related differences in information search behavior can be incorporated into CoLiDeS+. Section 5 concludes the paper.

2   Cognitive Model CoLiDeS+

CoLiDeS+ (Juvina & Van Oostendorp, 2008) shares the main theoretical foundations (Construction-Integration theory of text-comprehension (Kintsch, 1998) and Information Foraging Theory (Pirolli & Card, 1999) on which it is based with its predecessor CoLiDeS (Comprehension-based Linked model of Deliberate Search) proposed by Kitajima *et al.,* 2000. The notion of *information scent* is central to both models. It is defined as the estimate of the value or cost of information sources represented by proximal cues (such as hyperlinks). Information scent is operationalized as the semantic similarity between the user goal and each of the hyperlinks. Information scent has been found to be one of the driving factors steering navigation. The higher the information scent of a cue is, the higher the probability of a user clicking on it is.

CoLiDeS+ further augments CoLiDeS and makes it more consistent with its theoretical assumptions by drawing inspiration from work on text comprehension that lays emphasis on the role of context (Budiu & Anderson, 2004). Analogously, when interacting with a search engine or navigating on a website, users often encounter information that is varying in its degree of ambiguity. CoLiDeS+ retains in memory the selected links which are used to compute the navigation path and *path adequacy* (PA) in addition to information scent. The navigation path is the sequence of hyperlinks clicked by a user at any given moment and path adequacy is defined as the semantic similarity between the user goal and the navigation path. Only if the information from an incoming hyperlink increases in information scent (i.e., the semantic similarity with the user goal), it is considered for selection. If it does not increase in

information scent, path adequacy is checked. If path adequacy increases, then the incoming hyperlink is selected even when it does not increase in information scent.
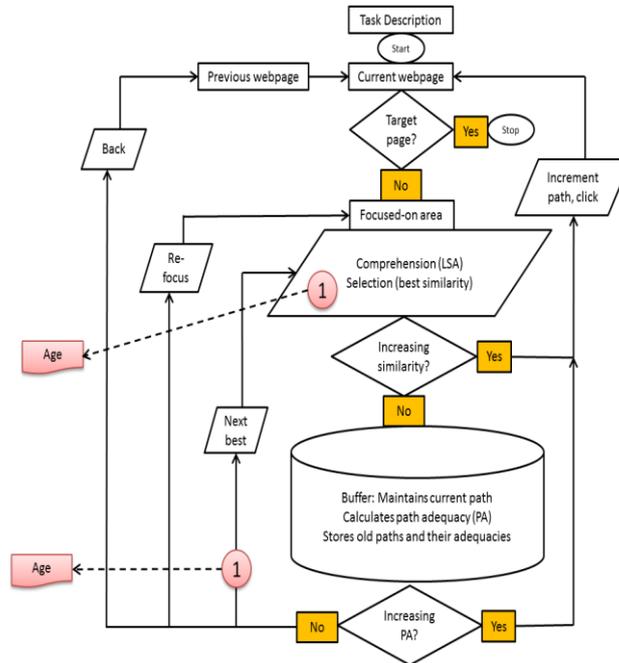

Figure 1: Schematic diagram of steps involved in CoLiDeS+ (reproduced from Van Oostendorp & Juvina, 2008). Shaded circles marked by (1) indicate the locations where individual differences due to aging are involved. See also later in text references to these locations.

In other words, first semantic similarity is locally evaluated based on information scent, and only when it is not satisfying, a more effortful evaluation of the context is performed by checking the path adequacy. If path adequacy also does not increase, a latent impasse is said to have occurred and CoLiDeS+ invokes backtracking strategy i.e., backtracking to other regions within the same page and eventually to the previously visited pages and the next-best strategy. CoLiDeS+ stops when the user declares the current page is the page with the target information. CoLiDeS+ uses Latent Semantic Analysis (LSA, henceforth) (Landauer *et al.,* 1998), to compute the semantic similarities corresponding to path adequacy and information scent. LSA is an unsupervised machine learning technique that builds a high dimensional semantic space using a large corpus of documents that represent a given user population's knowledge and understanding of words. The meaning of a word or sentence is represented as a vector in that high dimensional space. The degree of similarity between a link and the goal of the user is measured by the cosine value (correlation) between the corresponding vectors (Martin & Berry, 2007). Each cosine value lies between +1 and -1. The closer the value to +1 is, the higher the similarity between the two words is. Figure 1 shows a schematic diagram of the steps involved in CoLiDeS+. See Karanam, Van Oostendorp & Fu, (2015) for a

detailed process description of both models. The CoLiDeS model is developed to describe the navigation path within websites but we have shown it can also be applied to the interaction with search engines (Karanam *et al.,* 2015). Both applications offer a process description of the interaction of the participants with the information environment and the system.

3 Aging and Information Search

Aging leads to a natural decline in motor skills and fluid intelligence involving processing speed, cognitive flexibility or ability to switch processing strategies, attentional control and visuospatial span (Horn, 2012, Wang & Kaufman, 1993). However, crystallized knowledge (vocabulary skills and knowledge in a specialized domain such as health) seems stable or even increases with age.

Some of these cognitive abilities directly influence the cognitive processes underlying information search resulting in lower efficiency of older adults on information search tasks. Many studies have shown that older adults generate less queries, use less keywords per query, reformulate less, spend longer time evaluating the search results, spend more time evaluating the content of websites opened from search engine result pages (SERPs, henceforth), switch less often between SERPs and websites, find it difficult to reformulate unsuccessful queries (Queen *et al.,* 2012, Pak & Price, 2008, Dommes *et al.,* 2011) and produce semantically less relevant queries as they reformulate (Karanam & Van Oostendorp, 2016). Older adults were found to allocate less resources to exploration (fewer keywords, fewer clicks on search results etc.) and more resources to exploitation (longer time on a search result page, deeper navigation into websites opened from the search results) compared to younger adults (Chin *et al.,* 2015).

4 Modeling Aging Effects on Information Search

In this section, we will describe how we incorporate age-related differences in information search behavior into the CoLiDeS+ model. We present here some preliminary ideas to simulate the differences in the number of search results clicked by younger and older adults on a SERP. First, we vary during running the model the number of times the next-best strategy is applied (ranging from 0 to 9) by CoLiDeS+. This measure indicates how often a participant, after clicking on a search result and exploring the content of the corresponding website, comes back to the search result page to select one or more of the other search results. A higher value indicates more exploration from the side of the participant. Secondly, we also vary the minimum LSA value of a search result (computed in relation to the query, ranging from 0 to 0.9) which is an estimate of its relevancy. A low LSA value indicates that a search result is considered for clicking even when it has a low similarity value with the query. The steps involved in CoLiDeS+ that get affected by these variations have been marked with (1) in Figure 1. To demonstrate the effects of these variations on the match with

user behavior, we ran simulations using CoLiDeS+ under all possible combinations of the two parameters on twelve information search tasks and computed the frequency of matches of the model predictions with actual user clicks (true positives) from a behavioral experiment in which 24 younger adults and 24 older adults solved these tasks (Karanam & Van Oostendorp, 2016).

As younger adults are more impulsive and follow an explore-more and exploit-less strategy, we expect that they apply more often the next-best strategy and click more often on search results with a lower LSA value compared to older adults. Therefore, we expect the difference in the mean number of matches between younger and older adults to increase as the number of times the next-best strategy is applied, is increased and the minimum LSA value of a search result is decreased.
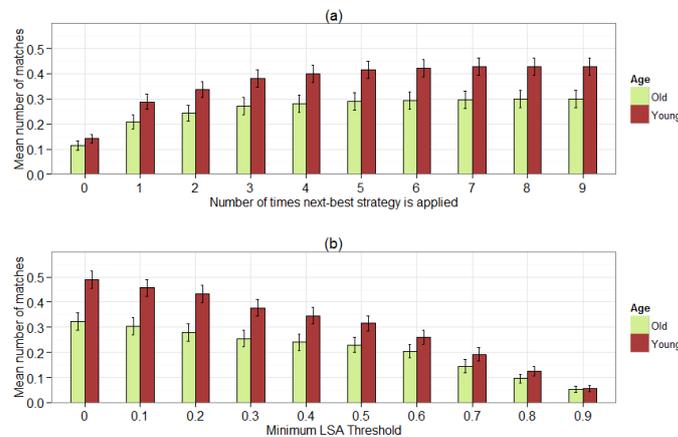

Figure 2: Mean number of matches in relation to variations in (a) number of times next-best strategy is applied and (b) minimum LSA value of a search result.

A repeated measures ANOVA with age as between-subjects variable and the number of times the next-best strategy is applied as within-subjects variable and mean number of matches as dependent variable was conducted. The main effect of age was significant $F(1,46) = 5.8$, $p<.05$, indicating that the model matched behavior of younger participants significantly better than that of older participants (See Figure 2(a)). The main effect of number of times the next-best strategy is applied was significant $F(1,46) = 108.34$, $p<.001$. As the number of times the next-best strategy is applied increased, the match between the model and the actual user behavior also increased. Also, the interaction of the number of times the next-best strategy is applied and age was significant $F(1,46) = 5.08$, $p<.05$. Post-hoc tests show that the rate of increase in match was higher for young participants compared to old participants.

A similar ANOVA with the minimum LSA value of a search result as within-subjects variable was conducted. The main effect of age was significant $F(1,46) = 6.2$, $p<.05$ (See Figure 2(b)). The model matched behavior of younger participants significantly better than

that of older participants. The main effect of the minimum LSA value of a search result was significant $F(1,46) = 159.39$, $p<.001$. As the minimum LSA value of a search result decreased, the match between the model and the actual user behavior increased. Also, the interaction of the minimum LSA value of a search result and age was significant $F(1,46) = 9.45$, $p<.001$. Post-hoc tests show that the rate of increase in match was higher for young participants compared to old participants when LSA decreased.

These outcomes are in-line with our expectations. Most importantly, these results imply that the match between the model and the actual behavior varies significantly with the number of times the next-best strategy is applied and the minimum LSA value of a search result that is used as a threshold in the model. We are currently performing more analyses to understand the optimum parameter values for younger and older adults.

5 Conclusions and Discussion

In this paper, using a computational cognitive model called CoLiDeS+, we presented preliminary ideas to simulate variations in information search behavior due to the effects of aging. By varying the number of times the next-best strategy is applied by CoLiDeS+ and the minimum LSA value of a search result, we were able to simulate the variations in the number of search results clicked by younger and older adults on SERPs and compute their matches with the model predictions. It appeared that the mean number of matches between the model-predicted clicks and the actual user clicks for younger and older adults increased as the number of times the next-best strategy is increased (Figure 2a) and as the minimum LSA value of a search result decreased (Figure 2b) with the rate of increase being higher for younger adults in both cases. It is also clear from the shape of the curves in figure 2 that the 'optimum' value for the next-best strategy for young participants is higher than for older participants, while it is lower for the minimum LSA value. The optimum value for the number of next-best strategies is between 3 and 5 for younger adults and 2 and 4 for older adults. Similarly, the optimum value for the minimum LSA threshold is between 0.2 and 0 for younger adults and 0.5 and 0.3 for older adults. We can imagine that a third factor is involved at the same time – the cost of running the model – measured in terms of the number of false positives. A deeper analysis which takes into account the cost factor could give a better insight into the optimum values for younger and older adults.

The influence of aging as discussed in this paper on the psychological processes during information search can be simulated in more than one location in the model (Figure 1). It is not clear, at this moment, whether the influence on one location in the model is more significant than the other. Our long term goal is to understand this issue. Also, there exist other cognitive factors (not discussed in this paper) such as prior domain knowledge, spatial ability, need for cognition, internet experience, gender, complexity of a task and interface characteristics which can have an influence on information search behavior and can be located in the model.

References

Budiu, R. & Anderson, J. R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science*, 28(1), 1-44.

Chevalier, A., Dommes, A. & Marquié, J. C. (2015). Strategy and accuracy during information search on the Web: Effects of age and complexity of the search questions. *Computers in Human Behavior*, 53, 305-315.

Chin, J., Anderson, E., Chin, C. L. & Fu, W. T. (2015). Age differences in information search An exploration-exploitation tradeoff model. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), 85-89. SAGE Publications.

Chuklin, A., Markov, I. & Rijke, M. D. (2015). Click Models for Web Search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 7(3), 1-115.

Dommes, A., Chevalier, A. & Lia, S. (2011). The role of cognitive flexibility and vocabulary abilities of younger and older users in searching for information on the web. *Applied Cognitive Psychology*, 25(5), 717-726.

Horn, J. L. (2012). The Theory of Fluid and Crystallized Intelligence in Relation to Concepts of Cognitive Psychology and Aging in. *Aging and Cognitive Processes*, 8, 237-278.

Hu, B., Zhang, Y., Chen, W., Wang, G. & Yang, Q. (2011). Characterizing search intent diversity into click models. In *Proceedings of the 20th international conference on World Wide Web,* 17-26. ACM.

Juvina, I. & Van Oostendorp, H. (2008). Modeling semantic and structural knowledge in Web navigation. *Discourse Processes*, 45(4-5), 346-364.

Karanam, S. & Van Oostendorp, H. (2016). Age-related differences in the content of search queries when reformulating. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5720-5730. ACM.

Karanam, S., Van Oostendorp, H. & Fu, W. T. (2016). Performance of computational cognitive models of web-navigation on real websites. *Journal of Information Science*, 42(1), 94-113.

Karanam, S., Van Oostendorp, H., Sanchiz, M., Chevalier, A., Chin, J. & Fu, W. T. (2015). Modeling and predicting information search behavior. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, Article Number 7. ACM.

Kintsch, W. (1998*). Comprehension: A paradigm for cognition*. Cambridge university press.

Kitajima, M., Blackmon, M. H. & Polson, P. G. (2000). A comprehension-based model of Web navigation and its application to Web usability analysis. *People and Computers*, 357-374.

Landauer, T. K., Foltz, P. W. & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3), 259-284.

Martin, D.I. & Berry, M.W. (2007). Mathematical Foundation Behind Latent Semantic Analysis. In T.K. Landauer, D.McNamara, S. Dennis & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Mahwah, Nj: L. Erlbaum Ass.

Pak, R. & Price, M. M. (2008). Designing an information search interface for younger and older adults. Human Factors: *The Journal of the Human Factors and Ergonomics Society*, 50(4), 614-628.

Pirolli, P. & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643-675.

Queen, T. L., Hess, T. M., Ennis, G. E., Dowd, K. & Grühn, D. (2012). Information search and decision making: Effects of age and complexity on strategy use. *Psychology and Aging*, 27(4), 817-824.

Sharit, J., Hernández, M. A., Czaja, S. J. & Pirolli, P. (2008). Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1), Article Number 3.

Shen, S., Hu, B., Chen, W. & Yang, Q. (2012). Personalized click model through collaborative filtering. In *Proceedings of the fifth ACM international conference on Web Search and Data Mining*. 323-332. ACM.

Wang, J. J. & Kaufman, A. S. (1993). Changes in fluid and crystallized intelligence across the 20-to 90-year age range on the K-BIT. *Journal of Psychoeducational Assessment*, 11(1), 29-37.

Xing, Q., Liu, Y., Nie, J. Y., Zhang, M., Ma, S. & Zhang, K. (2013). Incorporating user preferences into click models. In *Proceedings of the 22nd ACM international conference on Conference on Information & Knowledge Management*, 1301-1310. ACM.

Autoren

Saraschandra, Karanam

Postdoctoral researcher in the Department of Information and Computing Sciences at Utrecht University with an interdisciplinary background: PhD in Computer Science with a specialization in Cognitive Science and Human-Computer Interaction. He is currently busy adapting cognitive models of information search to simulate information search behavior of older adults. He worked briefly in two industrial research labs: HP Labs and Xerox Research Centre-India.

Herre, van Oostendorp

Holds a PhD in Cognitive Science from the University of Amsterdam and is senior researcher in the field of games, cognition, and the web at Utrecht University. He is working together with dr Karanam on a computational cognitive model for web navigation focused on providing automatic support for web users. He was also leader of the project 'Cognitive Learning Principles in Serious Games' within the successful national GATE-project.