Visualizing and Quantifying Vocabulary Learning During Search

Nilavra Bhattacharya[a], Jacek Gwizdka[a]

[a]*School of Information, The University of Texas at Austin, USA*

Abstract

We report work in progress for visualizing and quantifying learning during search. Users initiate a search session with a Pre-Search Knowledge state. During search, they undergo a change in knowledge. Upon conclusion, users attain a Post-Search Knowledge state. We attempt to measure this dynamic knowledge-change from a stationary reference point: Expert Knowledge on the search topic Using word-embeddings of searchers' written summaries, we show that w.r.t. Expert Knowledge, there is observable and quantifiable difference between the Pre-Search knowledge (Pre-Exp distance) and Post-Search knowledge (Post-Exp distance).

Keywords

search as learning, quantifying learning, expert knowledge, word embedding

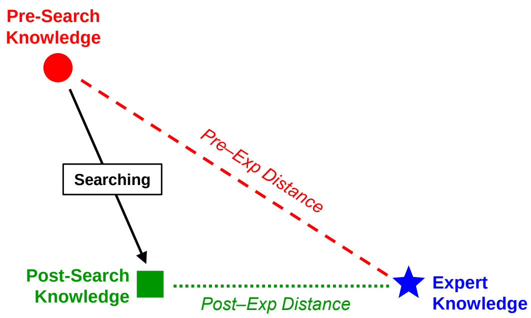

Figure 1: Conceptual framework of Search-as-Learning.

1. Introduction

An important aspect of understanding learning during web search is to measure and quantify learning, possibly in an automated fashion. Recent literature adopts three broad approaches for this purpose. The **first approach** asks searchers to rate their self-perceived pre-search and post-search knowledge levels [1, 2]. This approach is the easiest to construct, and can be generalized over any search topic. However, self-perceptions may not objectively represent true learning. The **second approach** tests searchers' knowledge using factual multiple choice questions (MCQs). The answer options can be a mixture of fact-based responses (*TRUE*, *FALSE*, or *I DON'T KNOW*), [3, 4] or recall-based responses (*I remember / don't remember seeing this information*) [5, 6]. Constructing topic-dependant MCQs may take time and effort, which may be aided by automated question generation techniques[7]. For evaluation, this approach is the easiest, and often automated. However, MCQs allow respondents to answer correctly by guesswork. The **third approach** lets searchers write natural language summaries or short answers, before and after the search [8, 2]. Depending on experimental design, prompts for writing such responses can be generic (least effort) [9] or topic-specific (some effort) [7]. While this approach can provide the richest information about the searcher's knowledge state, evaluating such responses is the most challenging, and requires extensive human intervention.

We report progress on extending work by [9], and take the third approach mentioned above. We attempt to visualize and quantify vocabulary learning during search, using natural language Pre-Search and Post-Search responses. The previous authors used sentence embedding models, and reported not finding strong associations between search interactions and knowledge change measures. A possible reason is that sentence embedding approaches are yet to attain maturity, and typically employ average pooling operation to generate sentence vectors from individual word vectors. Devising effective strategies to obtain vectors for compound units (phrases / sentences) from individual word vectors is always a challenge [10]. Differently from [9], we use word embedding vectors and max-pooling operations (taking element wise maximum of individual word vectors to form sentence vectors), which experi-

Proceedings of the CIKM 2020 Workshops, October 19-20, 2020, Galway, Ireland
EMAIL: nilavra@ieee.org (N. Bhattacharya); iwilds2020@gwizdka.com (J. Gwizdka)
URL: https://nilavra.in (N. Bhattacharya); http://gwizdka.com (J. Gwizdka)
ORCID: 0000-0001-7864-7726 (N. Bhattacharya); 0000-0003-2273-3996 (J. Gwizdka)

Example Pre-Search Knowledge:
health benefits vitamin consumption is highly debated
I know nothing about Vitamin A specifically

*Participant: **P03***

Example Post-Search Knowledge:
Vitamin A deficiency can led to blindness
Vitamin A is not toxic if over ingested
if over consumed
vitamin A can decrease vitamin B absorption
and increase likelihood of hip fractures
Vitamin A can be found in leafy green vegetables
organ meats
and broccoli
Vitamin A contents can be found on nutritional

Expert Knowledge (Excerpt):
Health benefits of using vitamin A: Vision, Breast cancer, Catarats, measles, Malaria, Diahrrea related to hiv, lower risk of complications during and after pregnancy, Retinitis pigmentosa, Ensures Healthy Eyes, soft skin, strong bones and teeth, acne, prevents muscular dystrophy, slow the aging process, lower risk of leukemia, good vision, Can prevent cancer, antioxidant, protects cells, maintain healthy skin, healthy immune system, healthy skeletal and soft tissue...

Figure 2: Example of Pre-Search and Post-Search knowledge assessment responses from a participant, for Task T3 (Vitamin A), alongside Expert Knowledge .

mentally showed better results than average-pooling.

2. Experimental Design

We analyze data from the user-study reported in [8, 9]. Participants (N = 30, 16 females, mean age 24.5 years) searched for health-related information on the web, over two search-tasks, T3 (topic: Vitamin A) and T4 (topic: Hypotension). Each search task began (Pre-Search) and ended (Post-Search) with a knowledge assessment, to gauge the participants' initial and final knowledge states. Participants entered natural language responses from free-recall, as answers. A vocabulary of Expert Knowledge was also created for each topic, in consultation with a medical doctor. Example participant responses, and an excerpt from the Expert Knowledge are shown in Fig. 2. After data cleaning, we obtained data from 49 participant-task pairs ($N_{T3} = 26; N_{T4} = 23$). Due to space limitations, please see [9] for more details about the study.

3. Data Analysis & Preliminary Results

We hypothesize that participants' learning during search can be assessed from the 'difference' in their Pre-Search and Post-Search responses. Since different participants may have different initial and final knowledge states, we measured it from a stationary reference-point: the expert knowledge. Calculating such differences between pieces of natural language texts is challenging, and is an active research topic. Word embedding is a popular method of computing semantic similarity (or distances) between two pieces of natural language texts. A word embedding algorithm produces a numeric, high-dimensional vector for each word, which is assumed to encapsulate the 'meaning' of the word. In this work, we leverage two popular pre-trained word-embedding models: word2vec [11], and GloVe [12], to compute 'differences' or 'distances' between Pre-Search, Post-Search, and Expert Knowledge (Fig. 1). word2vec contains 300 dimensional vectors for about 100 billion words (tokens) from the Google News dataset, and is claimed to be the most stable word-embedding [13]. GloVe offers multiple pre-trained word embeddings; we ran experiments with 50, 100, and 300 dimensional versions.

Word embedding algorithms produce vectors for individual words. To obtain vectors for phrases and sentences, the individual word vectors are usually pooled or aggregated. As discussed in Sec. 1, we performed max pooling, to produce a single high dimensional vector for a participant response (or expert knowledge). We employed two distance metrics – euclidean, and angular (cosine) – to compute distances between vectors of Pre-Search responses, Post-Search responses, and Expert's Knowledge (Fig. 1). The euclidean distance is unbounded, while the angular distance (Eqn. 1) ranges from 0 (no distance) to 1 (maximum distance).

$$\text{angular distance}(\mathbf{u}, \mathbf{v}) = \arccos\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}\right) / \pi \qquad (1)$$

We manually set the angular distance to be 1 (i.e, maximum) if one of the input vectors was a zero vector. This makes sense because zero vectors are obtained only if participants' responses do not contain any signs of knowledge (e.g., "none" or "i dont know").

To visualize the high-dimensional vectors of various knowledge states, we employed the t-SNE algorithm. This algorithm projects a set of high-dimensional objects on a 2D plane in such a way that similar objects are modelled by nearby points, and dissimilar objects are modelled by distant points. Using this algorithm, we obtained 2D representations of the Pre-Search, Post-Search, and Expert Knowledges (Fig. 3, left column). The visualization shows an almost clear separation between the Pre-Search (red circle) and Post-Search (green square) knowledge states, with Expert Knowledge (blue star) residing near the Post-Search knowledge states. This is a visual confirmation and support to the hypothesis that participants gain knowledge dur-

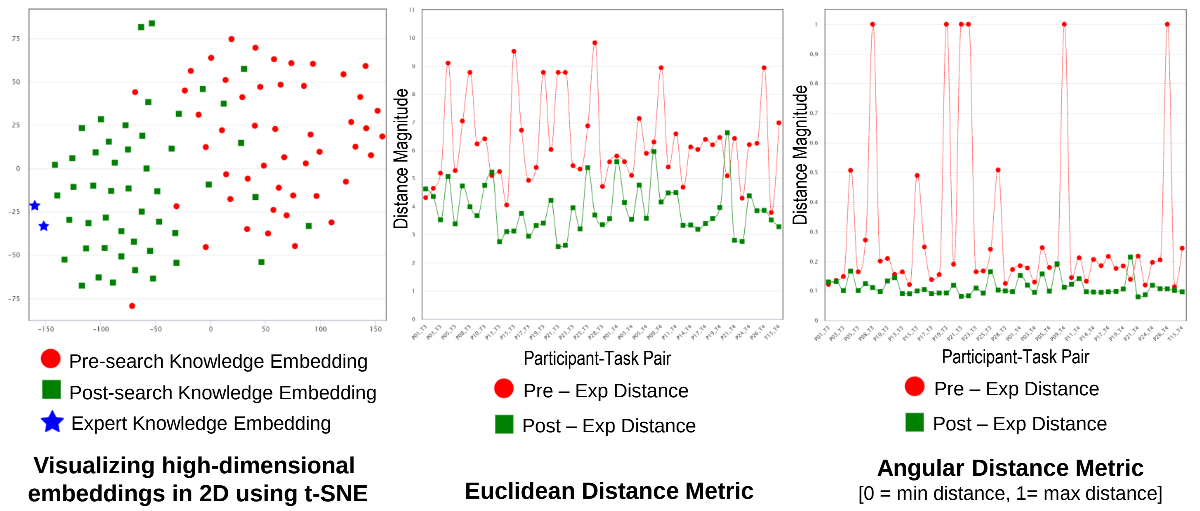

Figure 3: Results using `word2vec 300d` word embeddings, across tasks T3 and T4 combined. A clear separation can be observed between the majority of Pre-Search and Post-Search knowledge states **(left column)**, as well as between Pre-Exp and Post-Exp distances **(middle and right column)**.

ing search, and move 'closer' to the Expert Knowledge state at the end of a search.

The Euclidean and Angular distances between Pre-Search and Expert (Pre-Exp distance), and Post-Search and Expert (Post-Exp distance), are shown in the middle and right columns, respectively, in Fig. 3. For both distance metrics, the majority of the participants have lower Post-Exp distances than Pre-Exp distances (i.e. their Post-Search response is *less distant*, or *more similar* to, Expert Knowledge). These metrics were calculated between the high dimensional embedding vectors, which supports the fact that the 2D visualizations (left column) showing the clear separation between Pre- and Post-Search Knowledge levels is not merely by random chance. Interestingly, for few participants, the Post-Exp distance was higher than the Pre-Exp distance. This possibly demonstrates a 'loss' in knowledge level: users were closer to Expert Knowledge before the search, and moved away from Expert Knowledge after the search.

We further tested whether these visual differences between Pre-Exp and Post-Exp distances were statistically significant. Since the distance values were not normally distributed, we employed the non-parametric Wilcoxon Signed-Rank test, which is used for comparing paired or related samples. The results are presented in Table 1. We can see that across different choices of word embeddings, there were significant differences between the Pre-Exp and Post-Exp distances. Thus, the results are not due to choice of particular word embedding models. The directionalities of the differences in the Wilcoxon Signed-Rank test are expressed using the sum of the positive difference ranks (ΣR_+) and the sum of the negative difference ranks (ΣR_-). Since ΣR_- was greater than ΣR_+ in all the tests, the difference between Pre-Exp and Post-Exp distances is negative. This means that the majority of participants had *lower* Post-Exp distance than Pre-Exp distance (i.e. they moved *closer* to expert knowledge at the end of the task). The magnitude of a phenomenon is measured by effect size, which ranges from 0 (no effect) to 1 (maximum effect). All the tests had effect sizes greater than 0.8, signifying that searching online had a strong effect on minimizing the distance between participants' knowledge level and expert knowledge.

4. Conclusion and Future Work

We showed that word embeddings have promise for visualizing and quantifying vocabulary-based learning during search. Clear separation between user's Pre-Search and Post-Search knowledge states was seen and measured using simple distance metrics. Possible future directions include predicting these learning metrics from search-interactions measures. Another direction is to experiment with contextual embeddings (e.g., BERT). We also plan to investigate individual differences in learning during search.

4.0.1. Acknowledgements

We thank Sudipto Mukherjee, for technical and conceptual mentoring; Dr. Andrzej Kahl, our medical doctor

Table 1
Descriptive values of Pre-Exp and Post-Exp distances, and results of statistical significance tests, using different word-embeddings to model knowledges. As evident from Fig. 3, Pre-Exp and Post-Exp distances are significantly different for all the tested choices of word embedding models.

Word Embedding	Euclidean Distance Metric			Angular Distance Metric (Normalized) [0=least distance; 1=max distance]		
	Pre – Exp mean (±SD) median	**Post – Exp** mean (±SD) median	**Wilcoxon SR Test** all tests significant at $p < .05$	**Pre – Exp** mean (±SD) median	**Post – Exp** mean (±SD) median	**Wilcoxon SR Test** all tests significant at $p < .05$
word2vec	6.30 (±1.52) 6.12	3.90 (±0.87) 3.68	ΣR_+ = **20.0**, ΣR_- = **1205.0** 95% CI: **-2.76 to -1.82** Effect Size: **0.84**	0.30 (±0.28) 0.18	0.11 (±0.03) 0.10	ΣR_+ = **28.0**, ΣR_- = **1197.0** 95% CI: **-0.13 to -0.06** Effect Size: **0.83**
GloVe 6B 50d	8.67 (±2.39) 8.26	5.12 (±1.29) 4.68	ΣR_+ = **37.0**, ΣR_- = **1188.0** 95% CI: **-4.03 to -2.48** Effect Size: **0.82**	0.27 (±0.28) 0.17	0.10 (±0.03) 0.09	ΣR_+ = **43.0**, ΣR_- = **1182.0** 95% CI: **-0.12 to -0.06** Effect Size: **0.81**
GloVe 6B 100d	9.34 (±2.55) 8.96	5.46 (±1.42) 5.17	ΣR_+ = **30.0**, ΣR_- = **1195.0** 95% CI: **-4.46 to -2.79** Effect Size: **0.83**	0.30 (±0.28) 0.19	0.11 (±0.03) 0.10	ΣR_+ = **32.0**, ΣR_- = **1193.0** 95% CI: **-0.15 to -0.07** Effect Size: **0.82**
GloVe 6B 300d	12.15 (±3.18) 11.97	7.20 (±1.72) 6.81	ΣR_+ = **29.0**, ΣR_- = **1196.0** 95% CI: **-5.79 to -3.65** Effect Size: **0.83**	0.30 (±0.27) 0.20	0.11 (±0.03) 0.10	ΣR_+ = **35.0**, ΣR_- = **1190.0** 95% CI: **-0.14 to -0.07** Effect Size: **0.82**

consultant for expert-vocabulary creation; and Ying-long Zhang, for contributing to experimental data collection. The research was partially funded by IMLS Award #RE-04-11-0062-11 to Jacek Gwizdka.

References

[1] S. Ghosh, M. Rath, C. Shah, Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks, in: Conference on Human Information Interaction & Retrieval (CHIIR), 2018.

[2] H. L. O'Brien, A. Kampen, A. W. Cole, K. Brennan, The role of domain knowledge in search as learning, in: Conference on Human Information Interaction and Retrieval (CHIIR), 2020.

[3] L. Xu, X. Zhou, U. Gadiraju, How does team composition affect knowledge gain of users in collaborative web search?, in: Conference on Hypertext and Social Media (HT), 2020.

[4] U. Gadiraju, R. Yu, S. Dietze, P. Holtz, Analyzing knowledge gain of users in informational search sessions on the web, in: Conference on Human Information Interaction & Retrieval (CHIIR), 2018.

[5] S. Kruikemeier, S. Lecheler, M. M. Boyer, Learning from news on different media platforms: An eye-tracking experiment, Political Communication 35 (2018) 75–96.

[6] N. Roy, F. Moraes, C. Hauff, Exploring users' learning gains within search sessions, in: Conference on Human Information Interaction and Retrieval (CHIIR), 2020.

[7] R. Syed, K. Collins-Thompson, P. N. Bennett, M. Teng, S. Williams, D. W. W. Tay, S. Iqbal, Improving learning outcomes with gaze tracking and automatic question generation, in: The Web Conference (WWW), 2020.

[8] N. Bhattacharya, J. Gwizdka, Relating eye-tracking measures with changes in knowledge on search tasks, in: Symposium on Eye Tracking Research & Applications (ETRA), 2018.

[9] N. Bhattacharya, J. Gwizdka, Measuring learning during search: differences in interactions, eye-gaze, and semantic similarity to expert knowledge, in: Conference on Human Information Interaction and Retrieval (CHIIR), 2019.

[10] D. Roy, D. Ganguly, M. Mitra, G. J. Jones, Representing documents and queries as sets of word embedded vectors for information retrieval, in: ACM SIGIR workshop on neural information retrieval (Neu-IR), 2016.

[11] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in neural information processing systems, 2013, pp. 3111–3119.

[12] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[13] L. Burdick, J. K. Kummerfeld, R. Mihalcea, Factors influencing the surprising instability of word embeddings, in: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 2092–2102.