Digital Media Laboratory

Weekly Report November 28, 2021

Ali J. Alaee Report No. 18

alialaee98@gmail.com 00/09/1- 00/09/7

Abstract

As discussed last week, I tried to find out the measurements related to the effectiveness of personalization on the performance of the retrieval systems, according to different metrics. Fortunately, [1] contained great helpful information about this. Consequently, Numeric metric results are available for different methods (ad-hoc and personalized) tested on the available AOL dataset.

Description

- **Ranking Metrics**
  - **MAP (mean average precision)**: AP is taking k-rankings, and sum up the fraction of relevant items in the first i items, for i = 1 to k. MAP is just a mean. nDCG advantage is its capability for managing different weights for relevancy.
  - **MRR (mean reciprocal rank)**: RR takes the highest relevant item and 1/Item_Rank is result. MRR is just a mean.
  - **P@1**: Is equal to one, if the first recommendation is relevant and zero if it's not.
  - **P-improve**: more credible (for bias) — not used for AOL, cause it doesn't contain the original recommended list for each query submitted.
- **Baselines**
  - **Ad-hoc Models (Not Personalized)**
    * Original method: the default method that had been used back when the data was collected.
    * KNRM: takes query and document word embeddings, and soft-match them via an LTR algorithm.
  - **User Profile Based Personalization**
    * P-Click: re-ranks documents based on the number of clicks made by the same user under the same query in history, satisfying the user's refinding behaviors.
    * SLTB: extracts 102 features from the user's search history, including click-based features, topic-based features and so on. Then, all the features are combined with the LTR algorithm LambdaMart to generate the personalized ranking list.
    * HRNN: This model dynamically builds short and long- term user interest profiles with a hierarchical RNN and query- aware attention mechanism. Documents are re-ranked based on the similarities with the user profile and the additional SLTB features.
    * PSGAN: It is a personalized framework that applies GAN to generate queries that match the user's query intent better and select document pairs more valuable for learning user interests. In this paper, we take the variant PSGAN-D as our baseline.
  - **Embedding Based Personalization**
    * PPWE: This is a pipeline personalized word embedding based model we implement as a baseline. To re-rank the documents for the current query, we first train personalized word embeddings on the user's search data before this query by word2vec model to obtain the query and document representations, and then compute the relevance scores using the KNRM model with SLTB features.

* PWEBA: This is a personalization model for Twitter search. It first trains personal word embeddings on the user's history and creates a word-synonym table based on word vector similarity. Then, it re-ranks the generic list with cosine similarities between the query's synonyms and documents.
* PEPS: the proposed method

Model	AOL Dataset					
	MAP		MRR		P@1	
Adhoc search model						
Ori.	.2504	-54.3%	.2596	-53.6%	.1534	-68.6%
KNRM	.4291	-21.7%	.4391	-21.6%	.2704	-44.7%
ConvK	.4738	-13.5%	.4849	-13.4%	.3266	-33.2%
User profile based personalized search model						
PClick	.4224	-22.9%	.4298	-23.3%	.3788	-22.6%
SLTB	.5072	-7.5%	.5194	-7.3%	.4657	-4.8%
HRNN	.5423	-1.0%	.5545	-1.0%	.4854	-0.8%
PSGAN	.5480	–	.5601	–	.4892	–
Embedding based personalized search model						
PWEBA	.4284	-21.8%	.4368	-22.0%	.2687	-45.1%
PPWE	.6542‡	19.4%	.6668‡	19.1%	.5613‡	14.7%
PEPS(fix)	.6971‡	27.2%	.7107‡	26.9%	.6153‡	25.8%
PEPS	**.7127**‡	30.1%	**.7258**‡	29.6%	**.6279**‡	28.4%

Figure 1: Results of different retrieving approaches on AOL dataset[1]

- **Implementation Details**
  - The first five weeks log are set as the history, and the remaining data are used for model training, validation and testing with the proportion 6:1:1.
  - Each document corresponds to its title tag, for example for example, this website has a title of Modern Image San Diego — Car Paint Protection, Vinyl Wraps, Tinting.

Dataset	AOL Dataset			Commercial Dataset		
	Train	Valid	Test	Train	Valid	Test
#session	187,615	26,386	23,040	71,731	13,919	12,208
#query	814,129	65,654	59,082	188,267	37,951	41,261
avg query len	2.845	2.832	2.895	3.208	3.263	3.281
avg #click	1.249	1.118	1.115	1.194	1.182	1.202

Figure 2: Some Statistics about AOL dataset[1]

Next Week

- Given the available information, testing our implemented model might not be a bad idea.

References

[1] Jing Yao et al. Employing personal word embeddings for personalized search. *SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020.