Inoculating Relevance Feedback Against Poison Pills

Mostafa Dehghani[1] Hosein Azarbonyad[2] Jaap Kamps[1]
Djoerd Hiemstra[3] Maarten Marx[2]
[1]Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands
[2]Informatics Institute, University of Amsterdam, The Netherlands
[3]University of Twente, The Netherlands
{dehghani,h.azarbonyad,kamps,maartenmarx}@uva.nl,d.hiemstra@utwente.nl

1. INTRODUCTION

Relevance Feedback (RF) is a common approach for enriching queries, given a set of explicitly or implicitly judged documents to improve the performance of the retrieval. Although it has been shown that on average, the overall performance of retrieval will be improved after relevance feedback, for some topics, employing some relevant documents may decrease the average precision of the initial run. This is mostly because the feedback document is partially relevant and contains off-topic terms which adding them to the query as expansion terms results in loosing the retrieval performance. These relevant documents that hurt the performance of retrieval after feedback are called "poison pills" [2, 4]. In this paper, we discuss the effect of poison pills on the relevance feedback and present *significant words language models (SWLM)* as an approach for estimating feedback model to tackle this problem.

Significant words language models are family of models [1, 3] aiming to estimate models for a set of documents so that all, and only, the significant shared terms are captured in the models. This makes these models to be not only distinctive, but also supported by all the documents in the set. To do so, SWLM assumes that terms in the each document in the set are drawn from mixture of three models: 1. *General model*, representative of common observation, 2. *Specific model*, representative of partial observation, and 3. *Significant Words model*, latent model representing the significant characteristics of the whole set. Then, it tries to extract the significant words model.

2. POISON PILLS AND ANTIDOTES

We investigated the effect of poison pills on relevance feedback. To do so, for each topic with more than ten relevant documents, we add them one by one, based on their ranking in the initial run, to the feedback set and keep the track of the change in the performance of the feedback run after adding each relevant document to the feedback set compared to the feedback run without its presence.

To evaluate the robustness of different systems against bad relevant documents, we define a variant of *Robustness Index (RI)* to be applicable in the document level instead of topic level. For a set of relevant documents, D_r, the RI measure is defined as: $RI(D_r) = N_r^+ - N_r^-/|D_r|$ where N_r^+ and N_r^- denote number of helpful and harmful relevant documents, respectively. $|D_r|$ is total number of relevant documents. Higher values of $RI(D_r)$ means more robustness. Table 1 presents the $RI(D_r)$ of different systems on different datasets. As can be seen, SWLM is strongly robust against the effect of bad relevant documents in all datasets.

Employing SWLM enables the feedback system to control the contribution of feedback documents and prevents their specific or general terms affect the feedback model. Figure 1 shows how using SWLM empowers the feedback system to deal with the poison pills.
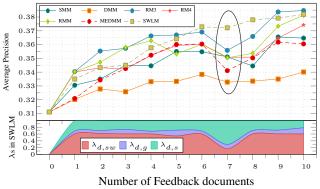
---

[*]This is an extended abstract of Dehghani et al. [3].

Figure 1: Dealing with poison pills: Effectiveness of different feedback systems facing with a bad relevant document in topic 374 of TREC Robust04.

Table 1: Robustness of different systems against bad relevant documents based on $RI(D_r)$ measure

Dataset	SMM	DMM	RM3	RM4	RMM	MEDMM	SWLM
Robust04	0.8663	0.7841	0.8716	0.8681	0.8843	0.8914	**0.9319**
WT10G	0.8504	0.8190	0.8783	0.8961	0.8990	0.9082	**0.9583**
GOV2	0.8456	0.8062	0.8809	0.8519	0.8910	0.8801	**0.9386**

In this figure, the performance of different systems in topic 374 on Robust04 dataset are illustrated. As can be seen, adding the seventh relevant document to the feedback set leads to a substantial decrement in the performance of the feedback in all the systems. The query is "Nobel prize winners" and the seventh document is about one of the Nobel peace prize winners, Yasser Arafat, but at the end, it has a discussion concerning Middle East issues, which contains some highly frequent terms that are non-relevant to the query. However, SWLM is able to distinguish this document as a poison pill and by reducing its contribution to the feedback model, i.e. learning a low value for $\lambda_{d_7,sw}$, they prevent the severe drop in the feedback performance.

3. CONCLUSIONS

So, SWLM inoculates the feedback model against poison pills by automatically determining whether adding a specific relevant document to the feedback set hurts the retrieval performance for a specific topic or not and controls its effect in the feedback model.

References

[1] M. Dehghani. Significant words representations of entities. In *SIGIR '16*, pages 1183–1183, 2016.
[2] M. Dehghani, S. Abnar, and J. Kamps. The healing power of poison: Helpful non-relevant documents in feedback. In *CIKM '16*, 2016.
[3] M. Dehghani, H. Azarbonyad, J. Kamps, D. Hiemstra, and M. Marx. Luhn revisited: Significant words language models. In *CIKM '16*, 2016.
[4] E. Terra and R. Warren. Poison pills: Harmful relevant documents in feedback. In *CIKM '05*, pages 319–320, 2005.