Poison Pills: Harmful Relevant Documents in Feedback

Egidio Terra
PUC/RS – Brazil
egidio@inf.pucrs.br

Robert Warren
University of Waterloo – Canada
rhwarren@uwaterloo.ca

Abstract

Feedback performance of retrieval systems depends on many factors and a great variability in precision occurs from topic to topic. We investigate the existence of poison pills: relevant documents that, when used in relevance feedback, deteriorate the retrieval system effectiveness. We present numbers that demonstrate that under some circumstances there are poison pills.

Systems

- cmu: Carnegie Mellon University used Lemur to implement a language modeling approach based on a uni-gram KL-divergence. This system differs slightly in that its query expansion engine would expand a query up to several hundred terms in blind feedback. Most systems at the RIA workshop would only use several dozen terms.
- waterloo: The University of Waterloo used its own Multitext system, which functions using passage retrieval and hotspot word extraction for blind feedback. The system also uses its own implementation of the Okapi BM25 algorithm to rank the final documents.
- sabir: Sabir Corporation used the SMART system Version 14 that uses a vector space model and lnu-ltu weighting. The system supplements its blind feedback function with selected statistical phrases.
- clarit: Clairvoyance Corporation also participated using a Java implementation of their system. It uses additional indexing of sub-document contents to perform blind feedback and was optimized for the use of a small number of feedback documents to select precise feedback terms.

Other systems that did not use blind feedback were used to contrast the value of the poison pill documents:

- city: City University in London provided results from their own Okapi probabilistic system using the BM25 algorithm. While all of the other systems used the topic description exclusively, the city system also uses the title field.
- albany: SUNY Albany also provided results using the SMART IR system Version 11.

Hypothesis

- Document are heterogeneous, multi-topic or simply too broad in scope: In this case, the feedback mechanism may be selecting a term from another topic in the document.
- Binary relevance judgments: Some documents may be marginally relevant, but emphasize another aspect of it .
- System parameters: These may be optimized for single document feedback and some adjustment may be necessary. However, preliminary tests on multi-document feedback runs have indicated that while the problem is attenuated, the presence of bad documents will still lower the average precision.
- Feedback with (non)-relevant documents: In general pseudo-relevance feedback doesn't assume relevancy for all top-k documents. Dunlop et al. further suggest that documents that do not match the query could be used.
- Number of feedback documents: The feedback mechanisms may not be getting enough information to expand queries from single documents.

Conclusions and Future Work

This is a problem that seems to affect a number of IR systems, suggesting a generic problem not related to an implementation. We present numerical evidence of this problem and the need to investigate further.

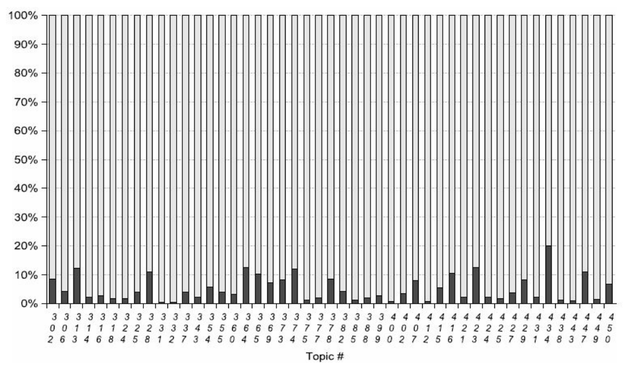
From further experiments, we know that is especially a problem when a low number of feedback documents are in use and the high number of these poison pills in the top ranks of the baseline runs are worrisome We note that the Clairvoyance system seems to have an affinity to these documents; this could be explained by its sub-document indexing technique which could tolerate poor documents.

At this early stage, we conclude that when selecting small sets of documents for pseudo-relevance feedback, careful selection should occur.

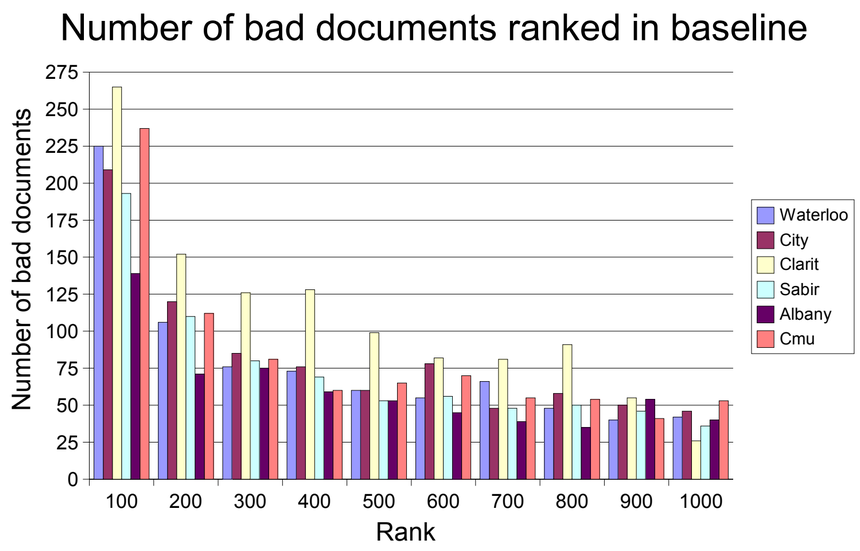
Experiment

- Compared the Average Precision of a baseline query against the same query expanded using pseudo-relevance feedback.
- Used 150 topic from TREC 6, 7 and 8 (ad hoc) with the clarit, cmu, city and sabir systems.
- All 14,000+ relevant documents were individually used in the feedback runs.

Empirical Evidence



Rank	# bad docs	# relevant	%
5	31	326	8.68
10	43	568	7.04
15	51	763	6.27
20	60	919	6.13
40	85	1381	5.80
100	148	2137	6.48
1000	269	3940	6.39
Overall	299	5315	5.33



Poison Pills Examples

Topic 314 : Marine Vegetation
Document: FR940511-0-00055
Problem: The document topic is broader than the information required. It is relevant, but concerns itself about aquaculture in general, including both vegetation and animals (broad topic).

Topic 328: Pope Beatifications
Document: LA050389-0062
Problem: The document talks about a visit to Zambia for most of its length, but in the end cites that the pope came from a beatification on the Indian Ocean island of Reunion (multi-topic).

Topic 313: Magnetic Levitation-Maglev
Document: LA110689-0059
Problem: The relevance of this document is ambiguous (binary judgments).

Acknowledgments

This work was sponsored by the Northeast Regional Research Center which is funded by ARDA, a U.S. Government entity which sponsors and promotes research of import to the Intelligence Community which includes, but is not limited to the CIA, DIA, NSA, NIMA and NRO.
The authors gratefully acknowledge travel assistance from the CIKM conference and the University of Waterloo.

References

Full references are provided in the conference proceedings.