

Motivation

- Word Embeddings (WE) exploit the co-occurrences of terms, appearing in *short-window contexts*.
- Previous studies show the effectiveness of using related terms, achieved from a WE model, while also cases that deteriorate the retrieval performance. (Rekabsaz et al. CIKM 2016, Zamani and Croft ICTIR 2016)
- In this work, we *investigate* such failure cases and analyze their causes.
- We *suggest* using *global (document) context* to effectively filter the set of related terms.

Extended Translation Model

The **Extended Translation Model** expands tf with a translation probability P_T over the related terms $R(t)$

$$\widehat{tf}_d = tf_d + \sum_{t' \in R(t)} P_T(t|t') tf_d(t')$$

In addition to tf , it propagates related terms in other elements of the models such as idf , dl , $avgdl$, and cs .

Rekabsaz et al. CIKM 2016, Rekabsaz et al. ECIR 2017

Experiment Setup

Name	Collection	# Queries	# Documents
TREC Adhoc 1&2&3	Disc1&2	150	740449
TREC Adhoc 6&7&8	Disc4&5	150	556028
Robust 2005	AQUAINT	50	1033461

- We select related terms from a word2vec model where similarity values are greater than a given threshold.
- We use Extended Translation Models of BM25 and LM.

Experiments & Results

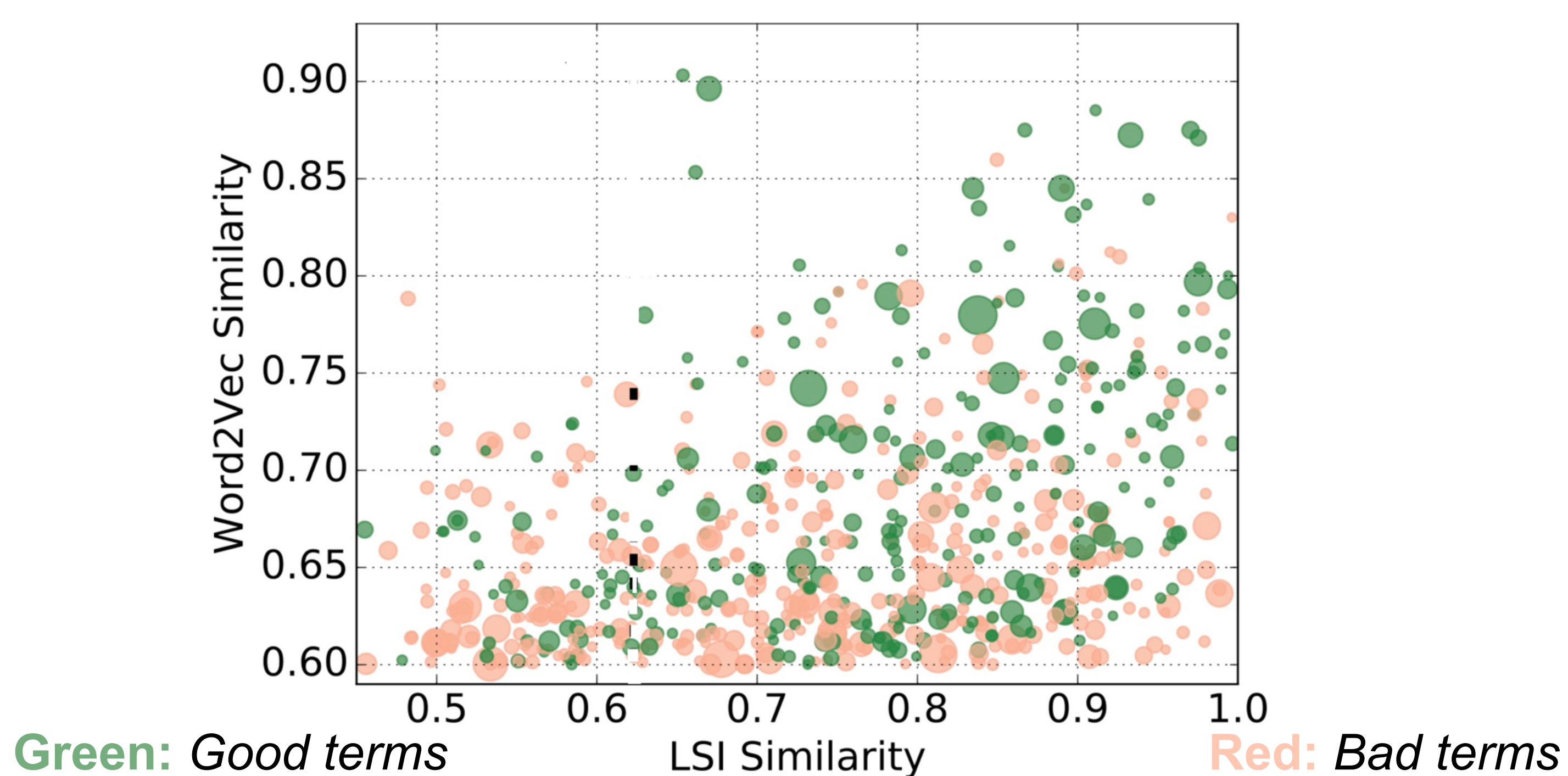
WE Effectiveness Analysis

Similar to Cao et al. (2008), we one-by-one add the related terms and measure the effectiveness in comparison to the effectiveness of the original query.

Collection	Threshold 0.60				Threshold 0.80			
	#Rel	Good	Neutral	Bad	#Rel	Good	Neutral	Bad
TREC 123	8.2	7%	84%	9%	1.3	19%	68%	13%
TREC 678	8.8	9%	78%	14%	1.2	34%	48%	18%
Robust 2005	10.3	8%	77%	15%	1.1	39%	44%	17%
ALL	8.1	8%	81%	11%	1.2	27%	58%	15%

Result: Most of the terms are neutral. Frequent incidents of topic shifting: Latvia to Estonia, Boeing to Airbus, and Alzheimer to Parkinson.

Results with Global-Context: Retrieval effectiveness on word2vec (window-context) and LSI (global-context)



Global-Context Post Filtering

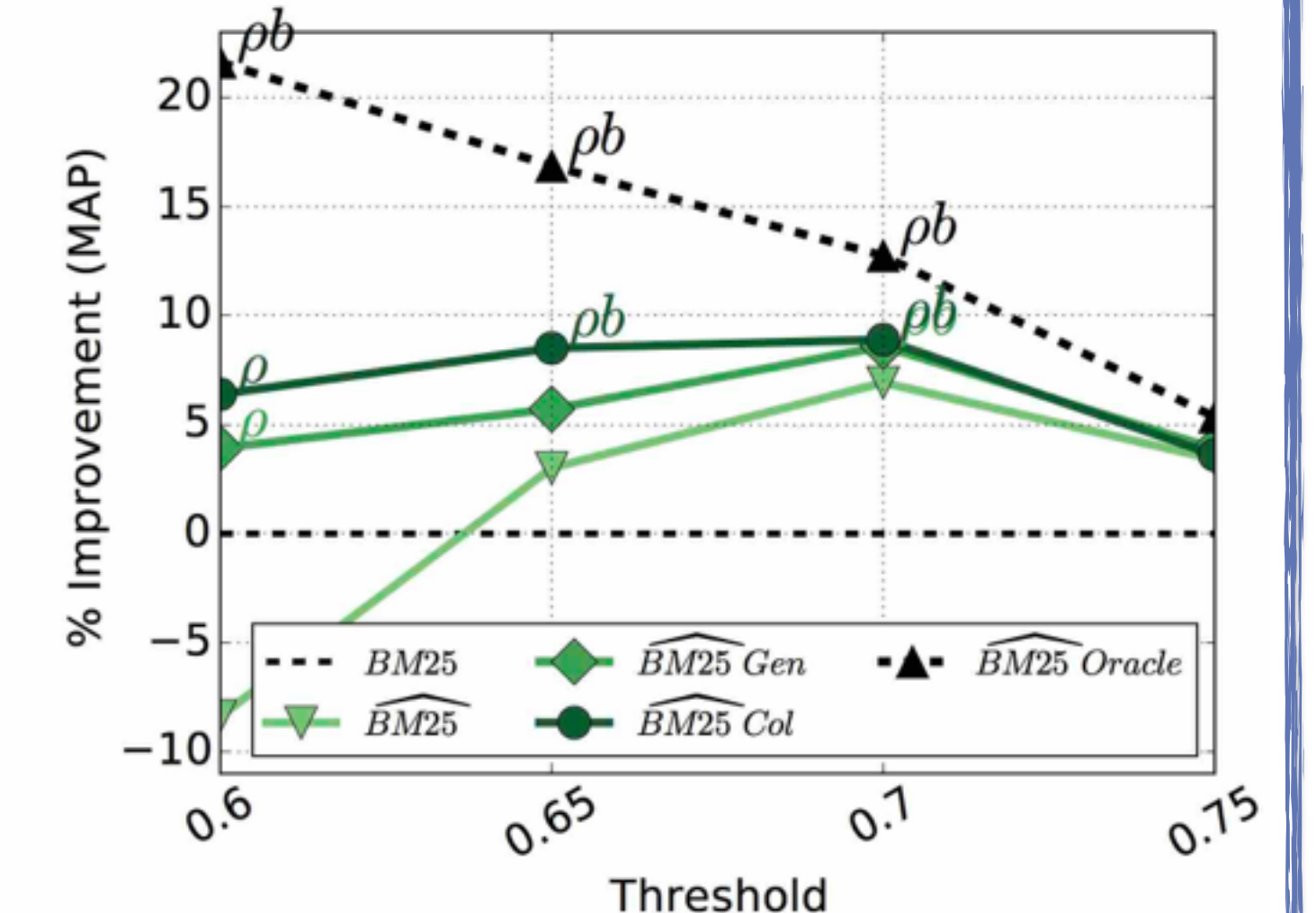
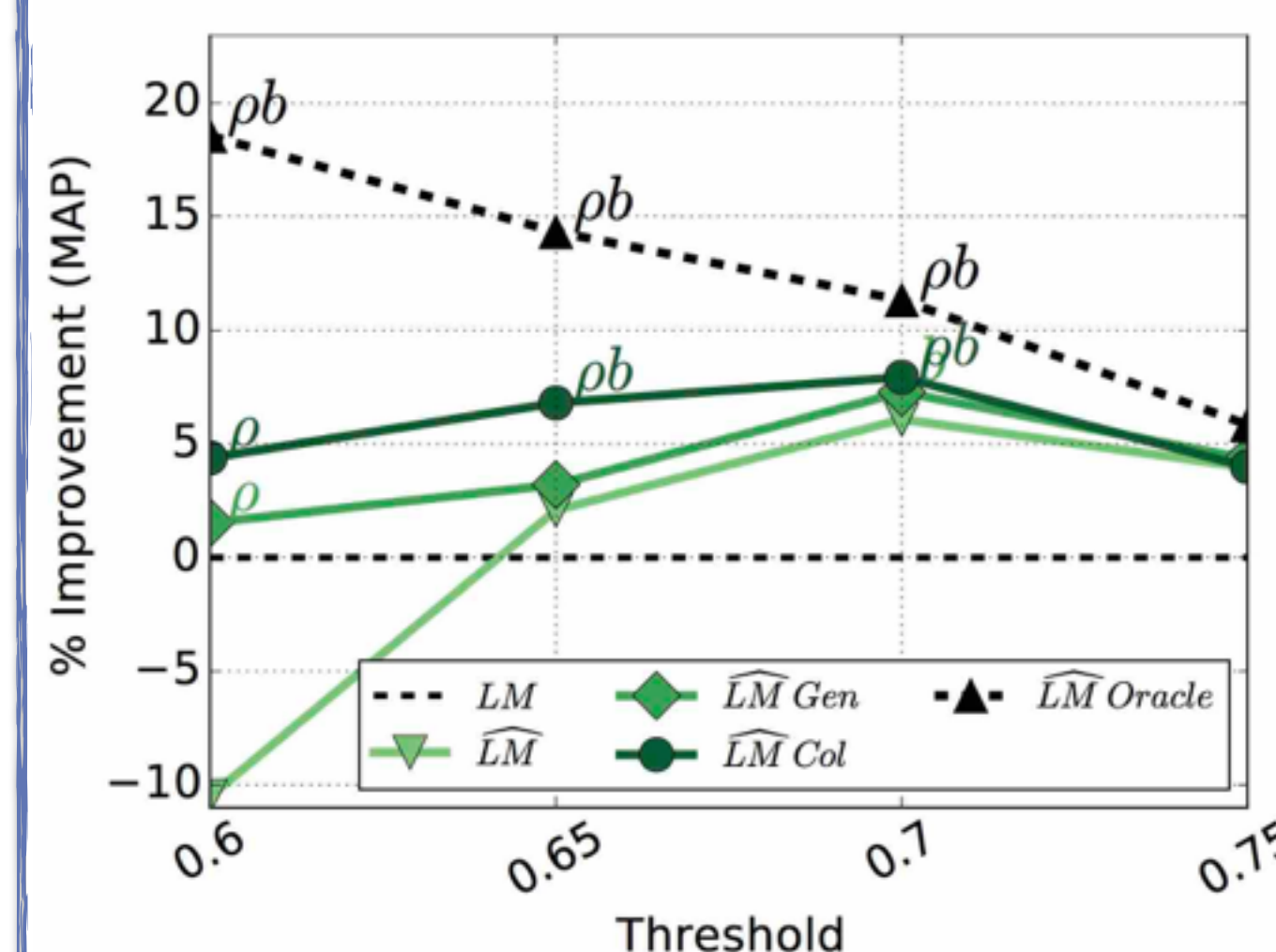
To avoid topic shifting of window-context related terms, we post-filter the terms based on their global-context similarities using a threshold.

To find this threshold, we use the term-based retrieval effectiveness as train data and search for the best overall performance in two ways:

Col: per collection using 5-fold cross validation.

Gen: the best-fit for all the data; (a general threshold).

Collection	Model	Basic	Tran.	Tran.+Gen	Tran.+Col
TREC 123	LM	0.275	0.283	0.290	0.295 $b\rho$
	BM25	0.273	0.285	0.288	0.290 b
TREC 678	LM	0.252	0.259	0.262	0.261
	BM25	0.243	0.255	0.257	0.256 b
Robust 2005	LM	0.183	0.204	0.208	0.209 $b\rho$
	BM25	0.181	0.203	0.207	0.209 $b\rho$



Result: (1) Post-filtering window-context similarities with global-context significantly improves the performance. (2) collection-based threshold is more effective.

Conclusion

- We investigate the effectiveness of each related term, from a window-context embedding (word2vec) and observed: (1) more 'good terms' in higher similarities (2) frequent topic shifting.
- We demonstrate the benefit of combining global-context similarity (LSI) with window-context (word2vec) for document retrieval using the Extended Translation model.