UMassAmherst

Embedding-based Query Language Models

Hamed Zamani, W. Bruce Croft

Center for Intelligent Information Retrieval (CIIR)
University of Massachusetts Amherst

{zamani, croft}@cs.umass.edu

Introduction

- Vocabulary mismatch problem in IR
  - e.g., query: “football”, document: “soccer” ✖
- Possible solutions:
  - **Query expansion**
  - Translation models
  - Document expansion
  - …

zamani@cs.umass.edu

Contributions (overview)

Expanding query with terms that are **semantically similar** to the query.

- How to calculate semantic similarities?
  - Sigmoid transformation of word embedding similarities
- How to do expansion?
  - Embedding-based query expansion (EQE1 and EQE2)
  - Embedding-based relevance model (ERM)

zamani@cs.umass.edu

Related Work

Query Expansion

- Global vs. local analysis [Xu & Croft, SIGIR' 96]
- Relevance Feedback [Rocchio, 1971]
- Pseudo-Relevance Feedback (PRF) [Croft & Harper, J. Doc.' 79]
- The "semantic effect" axiom for PRF [Montazeralghaem et al., SIGIR '16]

zamani@cs.umass.edu

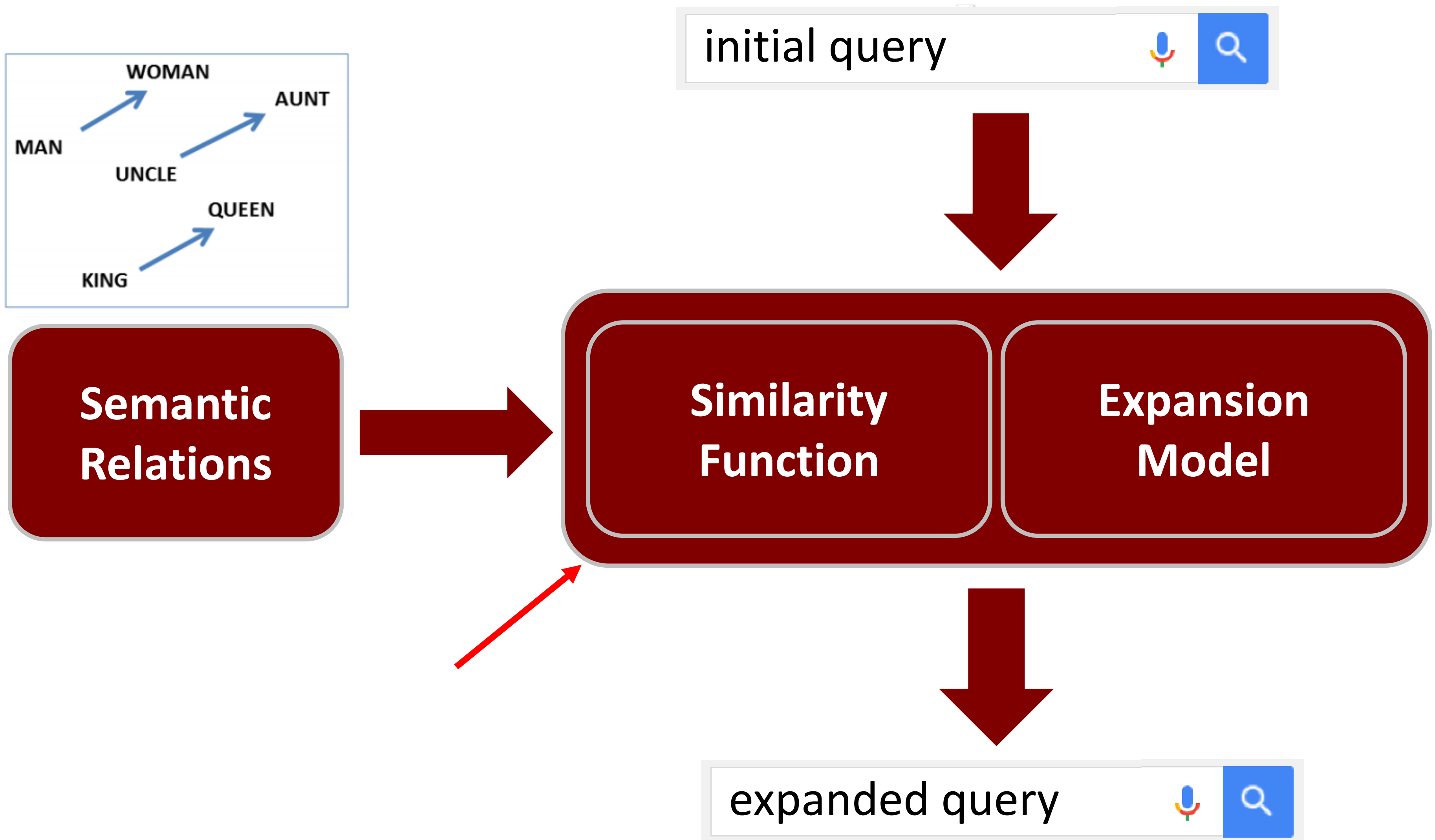
Related Work

Word Embedding for IR

- Supervised term re-weighting [Zheng & Callan, SIGIR '15]
- Word mover's distance [Kusner et al., ICML '15]
- Short texts similarities [Kenter & de Rijke, CIKM '15]
- Bilingual word embedding [Vulic & Moens, SIGIR '15]
- Embedding-based LM smoothing [Ganguly et al., SIGIR '15]
- Embedding-based translation model [Zuccon et al., ADCS '15]
- Heuristic embedding-based query expansion [ALMasri et al., ECIR '16]
- Locally-trained word embedding [Diaz et al., ACL '16]

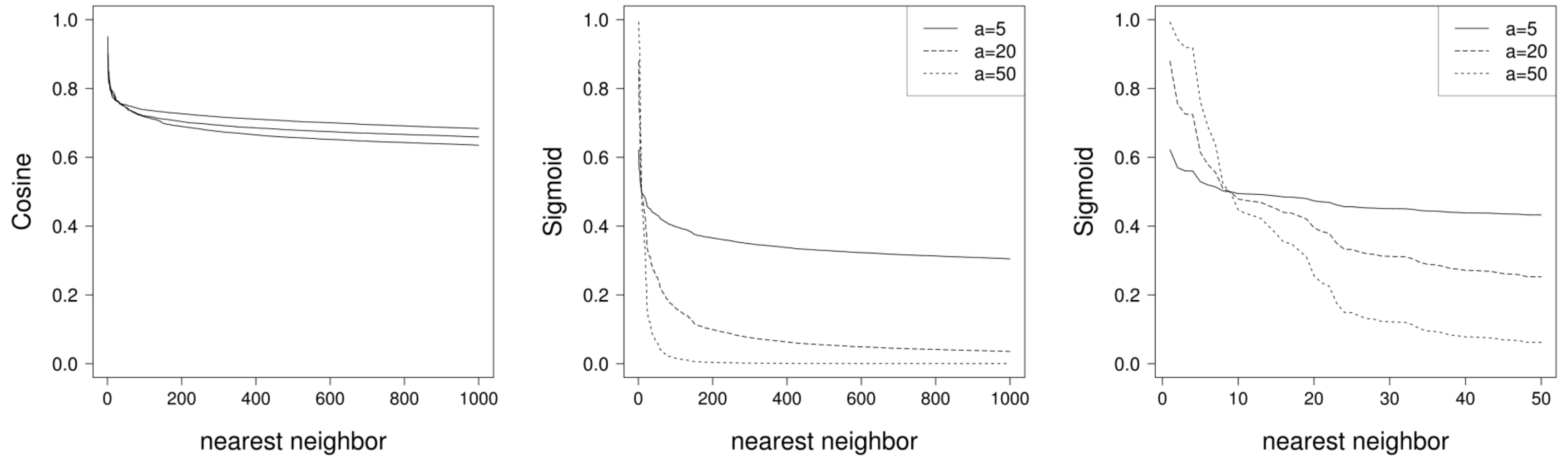
zamani@cs.umass.edu

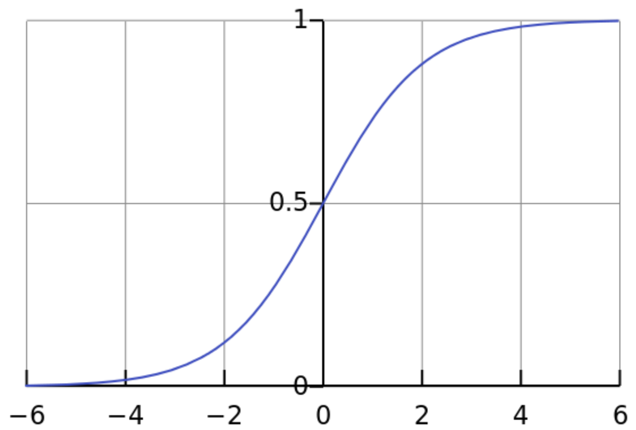
Methodology



zamani@cs.umass.edu

Sigmoid Transformation

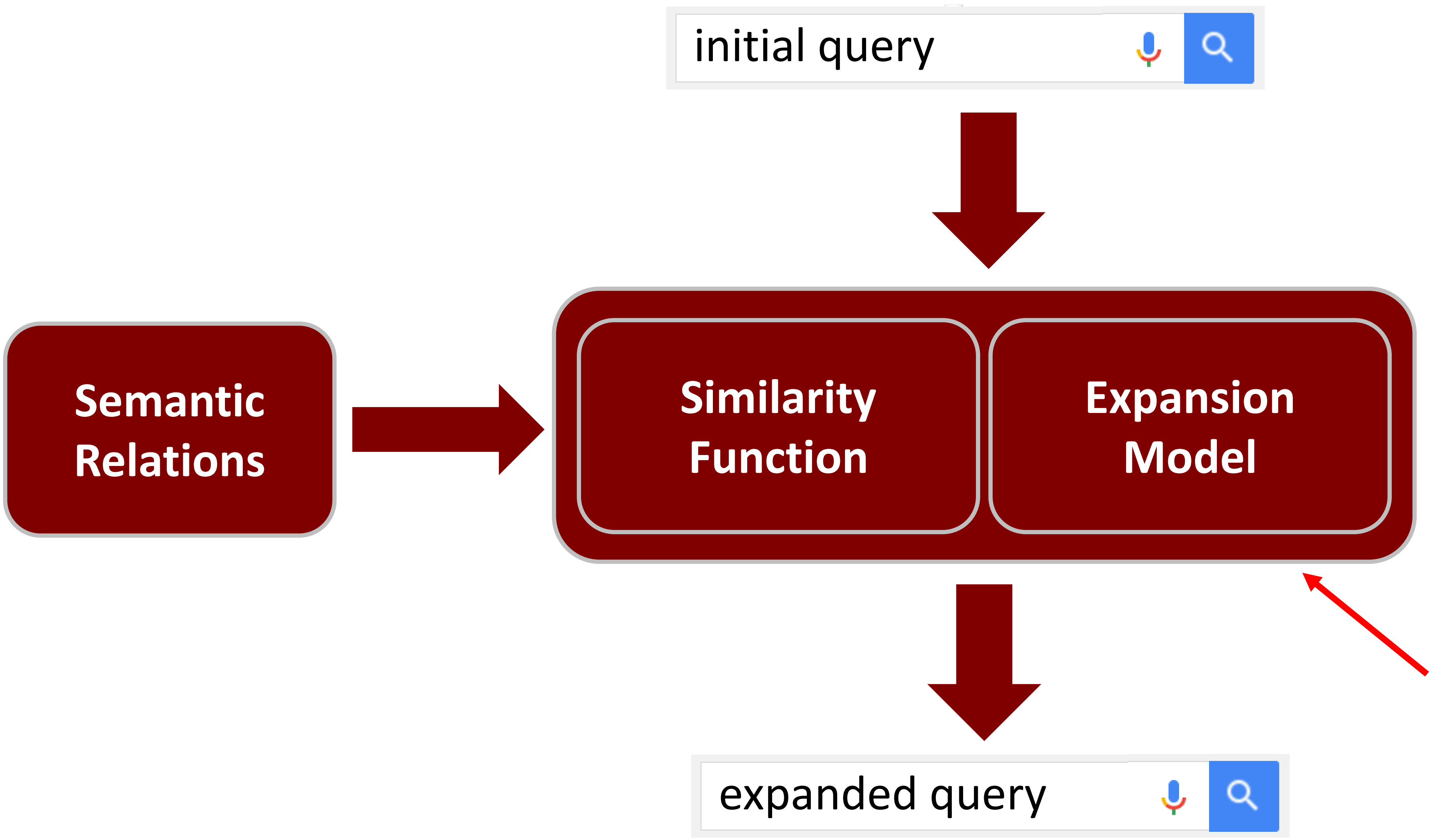




$$S(x) = \frac{1}{1 + \mathrm{e}^{-\mathrm{a}(\mathrm{x}-\mathrm{c})}}$$

zamani@cs.umass.edu

Methodology



zamani@cs.umass.edu

Embedding-based Query Expansion

Idea: expanding the query using the terms that are semantically similar to the query.

We propose two models:

- Conditional independence of query terms
- Query-independent term similarities

zamani@cs.umass.edu

Embedding-based Query Expansion

EQE1

$$p(w|\theta_Q) = \frac{p(w)p(\theta_Q|w)}{p(Q)}$$
$$\propto p(w)p(\theta_Q|w)$$
$$\approx p(w)\boxed{\prod_{i=1}^{k} p(q_i|w)}$$

Conditional independence of query terms

$$p(q_i|w) = \frac{\delta(\overrightarrow{q_i}, \vec{w})}{\sum_{w' \in V} \delta\left(\vec{w}, \overrightarrow{w'}\right)}, p(w) = \sum_{w' \in V} p(w, w') \propto \sum_{w \in V} \delta(\vec{w}, \overrightarrow{w'})$$

zamani@cs.umass.edu

Embedding-based Query Expansion

EQE2

$$p(w|\theta_Q) = \sum_{w' \in V} p(w, w'|\theta_Q)$$

$$= \sum_{w' \in V} p(w|w', \theta_Q) p(w'|\theta_Q)$$

$$\approx \sum_{w' \in V} \boxed{p(w|w')}\, p(w'|\theta_Q)$$

Query-independent term similarities

$$p(w|w') = \frac{\delta\left(\vec{w}, \overrightarrow{w'}\right)}{\sum_{w'' \in V} \delta\left(\overrightarrow{w'}, \overrightarrow{w''}\right)},$$

$$p(w'|\theta_Q) = \frac{\mathrm{count}(w', Q)}{|Q|}$$

Embedding-based Relevance Model

Idea: semantic similarity in addition to term matching for pseudo-relevance feedback.

We extend relevance model (RM3) [Lavrenko & Croft, SIGIR '01] by adding a semantic similarity term.

zamani@cs.umass.edu

Embedding-based Relevance Model

$$p(w|\theta_F) = \sum_{D \in F} p(w, Q, D)$$
$$= \sum_{w' \in V} p(Q|w, D)p(w|D)P(D)$$

$$p(Q|w, D) = \beta p_{tm}(Q|w, D) + (1 - \beta)p_{sem}(Q|w, D)$$
$$p_{tm}(Q|w, D) = \prod_{i=1}^{k} p(q_i|D)$$
$$p_{sem}(Q|w, D) = \prod_{i=1}^{k} p_{sem}(q_i|w, D) = \prod_{i=1}^{k} \frac{\delta(\overrightarrow{q_i}, \overrightarrow{w})c(q_i, D)}{Z}$$

If $\boldsymbol{\beta} = \mathbf{1}$ then ERM=RM3

zamani@cs.umass.edu

Retrieval Collections

AP (Associated Press 1988-1989)

- 165K news articles
- 146 out of 150 queries*

Robust (TREC Robust Track 2004)

- 528K news articles
- 241 out of 250 queries*

GOV2 (TREC Terabyte Track 2004-2006)

- 25M web pages
- 147 out of 150 queries*

* We only consider the queries where the embedding vector of all query terms are available.

zamani@cs.umass.edu

Embedding Vectors

Wikipedia 2004 & Gigawords 5

- 6B tokens
- 400K vocabulary terms
- Learning method: Glove [Pennington et al., EMNLP ‘14]
- Dimension: 200

zamani@cs.umass.edu

Experimental Setup

Evaluation Metrics:

- MAP (Mean Avg. Prec.)
- P@5
- P@10
- RI (Robustness Index)

$$RI = \frac{N_+ - N_-}{|Q|}$$

Parameter Setting:

- 2-fold cross-validation over queries of each collection

zamani@cs.umass.edu

UMASS AMHERST

Query Expansion Results

Dataset	Metric	MLE	GLM	VEXP	AWE	EQE1	EQE2
AP	MAP	0.2236	0.2254	0.2338	0.2304	0.2388^{1234}	$\mathbf{0.2391}^{1234}$
	P@5	0.4260	0.4369	0.4412	0.4356	0.4397	**0.4466**
	P@10	0.4014	0.4051	0.4038	0.4058	**0.4075**	0.4014
	RI	–	0.10	0.18	0.14	**0.32**	**0.32**
Robust	MAP	0.2190	0.2244	0.2253	0.2224	$\mathbf{0.2292}^{124}$	0.2257^{1}
	P@5	0.4606	0.4523	0.4722	0.4680	**0.4739**	0.4622
	P@10	0.3979	0.3929	0.4133	0.4066	0.4162	**0.4183**
	RI	–	0.22	0.17	0.14	**0.30**	0.22
GOV2	MAP	0.2696	0.2684	0.2687	0.2657	$\mathbf{0.2745}^{1234}$	0.2727^{4}
	P@5	0.5592	0.5537	0.5932	0.5537	**0.5959**	0.5810
	P@10	0.5531	0.5483	0.5537	0.5503	**0.5660**	0.5517
	RI	–	-0.14	0.10	-0.18	**0.20**	0.08

- EQE1 in general performs better than EQE2.
- EQE1 significantly outperforms the query expansion baselines.

zamani@cs.umass.edu

Pseudo-Relevance Feedback Results

Dataset	Metric	MLE	MLE+RM1 (RM3)	EQE1+RM1	EQE2+RM1	MLE+ERM	EQE1+ERM	EQE2+ERM
AP	MAP	0.2236	0.3051	0.3118^{12}	0.3115^{12}	0.3102^{12}	$\mathbf{0.3178}^{12}$	0.3140^{12}
	P@5	0.4260	0.4644	0.4808	0.4795	0.4699	**0.4822**	0.4644
	P@10	0.4014	0.4500	0.4500	0.4452	0.4521	**0.4568**	0.4479
	RI	–	0.47	0.45	0.41	**0.52**	0.47	**0.52**
Robust	MAP	0.2190	0.2677	0.2712^{12}	0.2710^{12}	0.2711^{12}	0.2731^{12}	$\mathbf{0.2750}^{12}$
	P@5	0.4606	0.4581	0.4747	0.4722	0.4639	**0.4797**	0.4730
	P@10	0.3979	0.4191	0.4241	0.4295	0.4241	0.4307	**0.4369**
	RI	–	0.31	**0.39**	0.35	0.31	0.32	0.36
GOV2	MAP	0.2696	0.2938	0.2987^{12}	0.2922^{1}	0.3005^{12}	$\mathbf{0.3012}^{12}$	0.2957^{1}
	P@5	0.5592	0.5592	0.5687	0.5673	0.5823	**0.5850**	0.5782
	P@10	0.5531	0.5599	0.5816	0.5714	0.5830	**0.5844**	0.5782
	RI	–	0.15	**0.22**	0.14	**0.22**	0.20	0.20

- ERM performs better than the original relevance model.
- EQE1+ERM in general outperforms other methods.

zamani@cs.umass.edu

Pseudo-Relevance Feedback Results

Dataset	Metric	MLE	MLE+RM1 (RM3)	EQE1+RM1	EQE2+RM1	MLE+ERM	EQE1+ERM	EQE2+ERM
AP	MAP	0.2236	0.3051	0.3118^{12}	0.3115^{12}	0.3102^{12}	$\mathbf{0.3178}^{12}$	0.3140^{12}
	P@5	0.4260	0.4644	0.4808	0.4795	0.4699	**0.4822**	0.4644
	P@10	0.4014	0.4500	0.4500	0.4452	0.4521	**0.4568**	0.4479
	RI	–	0.47	0.45	0.41	**0.52**	0.47	**0.52**
Robust	MAP	0.2190	0.2677	0.2712^{12}	0.2710^{12}	0.2711^{12}	0.2731^{12}	$\mathbf{0.2750}^{12}$
	P@5	0.4606	0.4581	0.4747	0.4722	0.4639	**0.4797**	0.4730
	P@10	0.3979	0.4191	0.4241	0.4295	0.4241	0.4307	**0.4369**
	RI	–	0.31	**0.39**	0.35	0.31	0.32	0.36
GOV2	MAP	0.2696	0.2938	0.2987^{12}	0.2922^{1}	0.3005^{12}	$\mathbf{0.3012}^{12}$	0.2957^{1}
	P@5	0.5592	0.5592	0.5687	0.5673	0.5823	**0.5850**	0.5782
	P@10	0.5531	0.5599	0.5816	0.5714	0.5830	**0.5844**	0.5782
	RI	–	0.15	**0.22**	0.14	**0.22**	0.20	0.20

- ERM performs better than the original relevance model.
- EQE1+ERM in general outperforms other methods.

zamani@cs.umass.edu

Analysis of Sigmoid Transformation

Dataset	Method	EQE1	EQE2	ERM
AP	Cosine	0.2293	0.2366	0.3038
	Sigmoid	**0.2388***	**0.2391**	**0.3102***
Robust	Cosine	0.2247	0.2233	0.2677
	Sigmoid	**0.2292***	**0.2257**	**0.2711***
GOV2	Cosine	0.2709	0.2654	0.2971
	Sigmoid	**0.2745***	**0.2727***	**0.3005**

- Transforming the embedding similarity scores using the sigmoid function improves the performance in all query expansion models.

zamani@cs.umass.edu

Sensitivity to Embedding Vectors

Wiki

- Wikipedia 2004 & Gigawords 5
- 6b tokens

Web 42b

- Web crawl
- 42b tokens

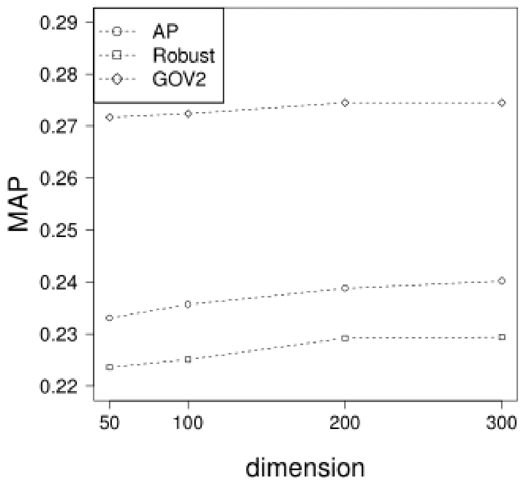
Web 840b

- Web crawl
- 840b tokens

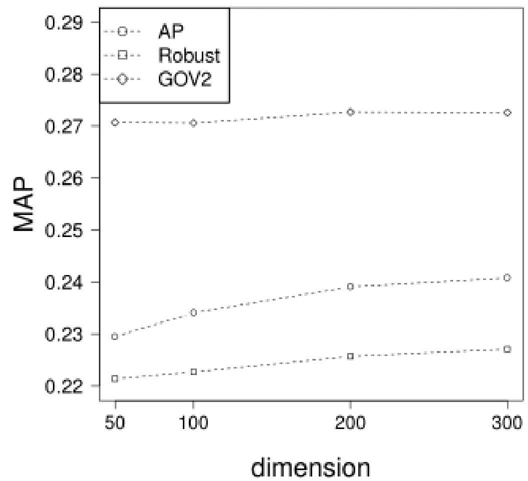
Dataset	Method	Wiki	Web 42b	Web 840b
AP (146 queries)	EQE1	0.2402	0.2356	0.2362
	EQE2	0.2408	0.2352	0.2400
	ERM	0.3106	0.3094	0.3081
Robust (240 queries)	EQE1	0.2294	0.2255	0.2273
	EQE2	0.2271	0.2237	0.2266
	ERM	0.2713	0.2705	0.2683
GOV2 (146 queries)	EQE1	0.2745	0.2729	0.2767
	EQE2	0.2726	0.2713	0.2743
	ERM	0.3013	0.2989	0.3021

zamani@cs.umass.edu

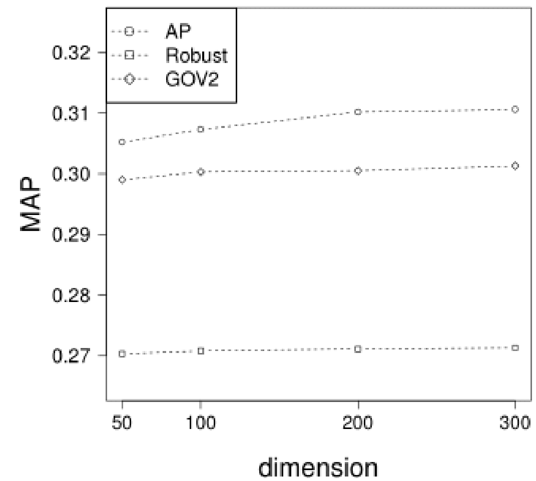
Sensitivity to Embedding Dimension



(a) EQE1



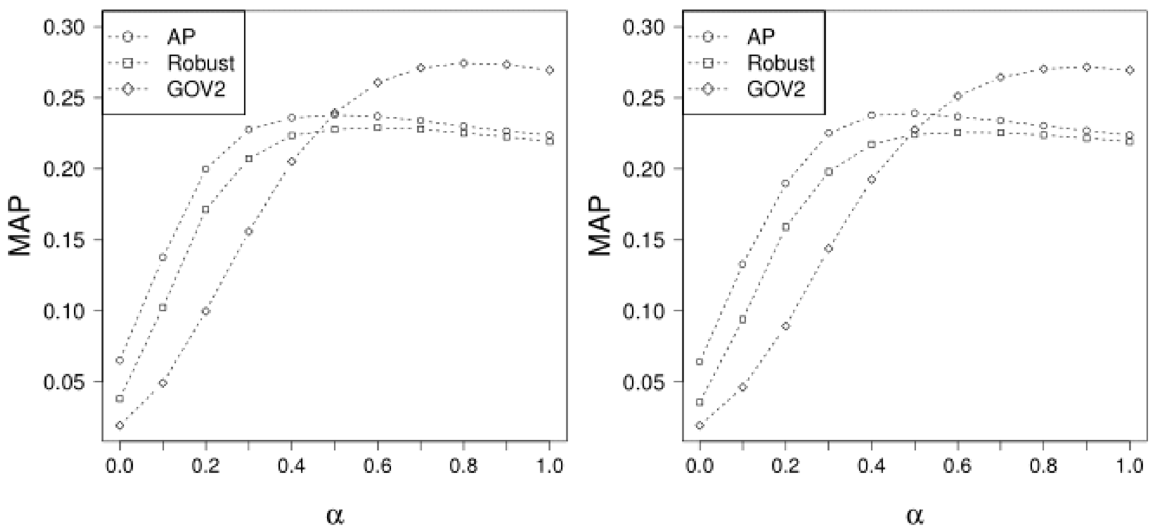
(b) EQE2



(c) ERM

zamani@cs.umass.edu

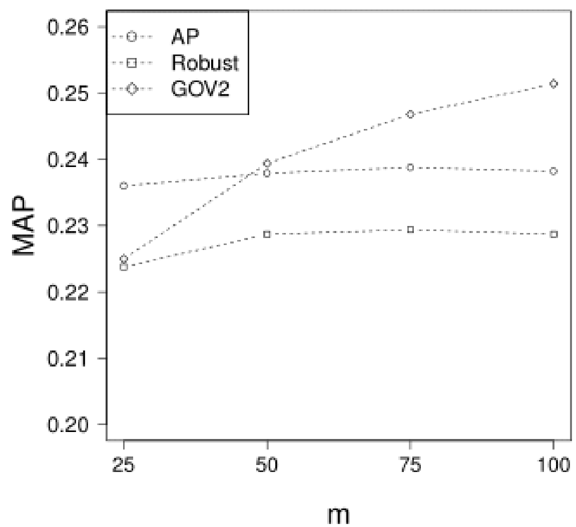
Sensitivity to Hyper-parameters



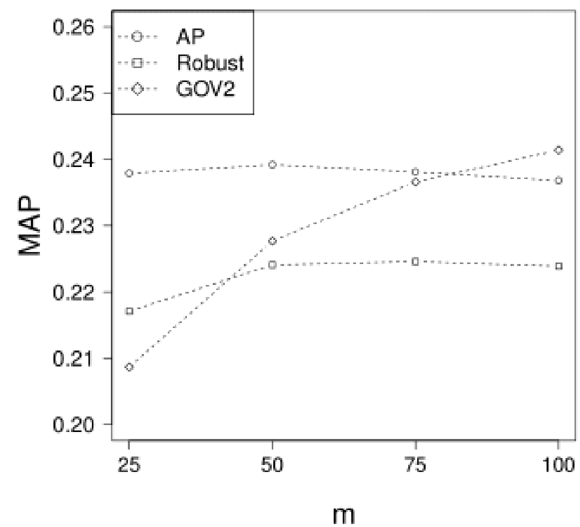
(a) EQE1 (b) EQE2

zamani@cs.umass.edu

Sensitivity to Hyper-parameters



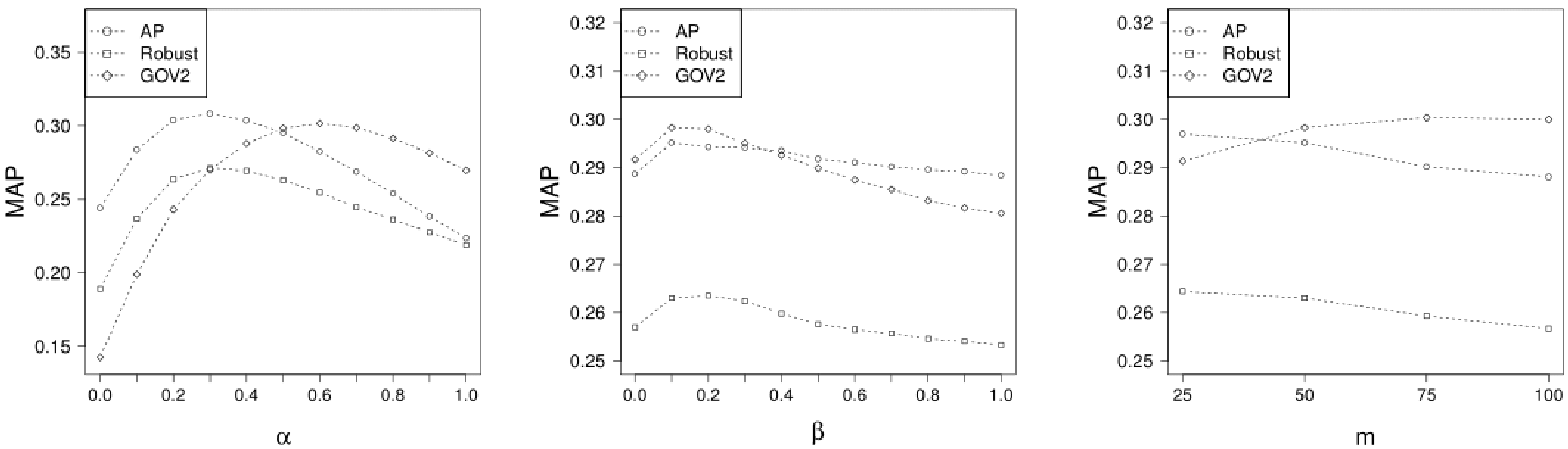
(c) EQE1



(d) EQE2

zamani@cs.umass.edu

Sensitivity to Hyper-parameters



zamani@cs.umass.edu

Conclusions

- We proposed **two query expansion models** as well as **a pseudo-relevance feedback model** based on word embedding similarities.
- We proposed to **transform** semantic similarity scores using the sigmoid function.
- Evaluation over three TREC collections indicated the **effectiveness** of the proposed models.

zamani@cs.umass.edu

Future Work

- Modifying the learning process of embedding vectors instead of transforming the similarity scores.
- Theoretical analysis of similarity score transformation.
- Studying the necessity of using score transformation for other embedding approaches (e.g., word2vec).
- Studying the usefulness of word embedding vectors in other aspects of IR.

zamani@cs.umass.edu

UMASS AMHERST

Thank You!

?

zamani@cs.umass.edu