Entropy-Based Query Performance Prediction for Neural Information Retrieval Systems

OLEG ZENDEL[1], BINSHENG LIU[2], J. SHANE CULPEPPER[1] AND FALK SCHOLER[1]

[1]RMIT UNIVERSITY

[2]SEEK

THE ILLUSTRATION IMAGES WERE GENERATED WITH BING IMAGE CREATOR (POWERED BY DALL-E)

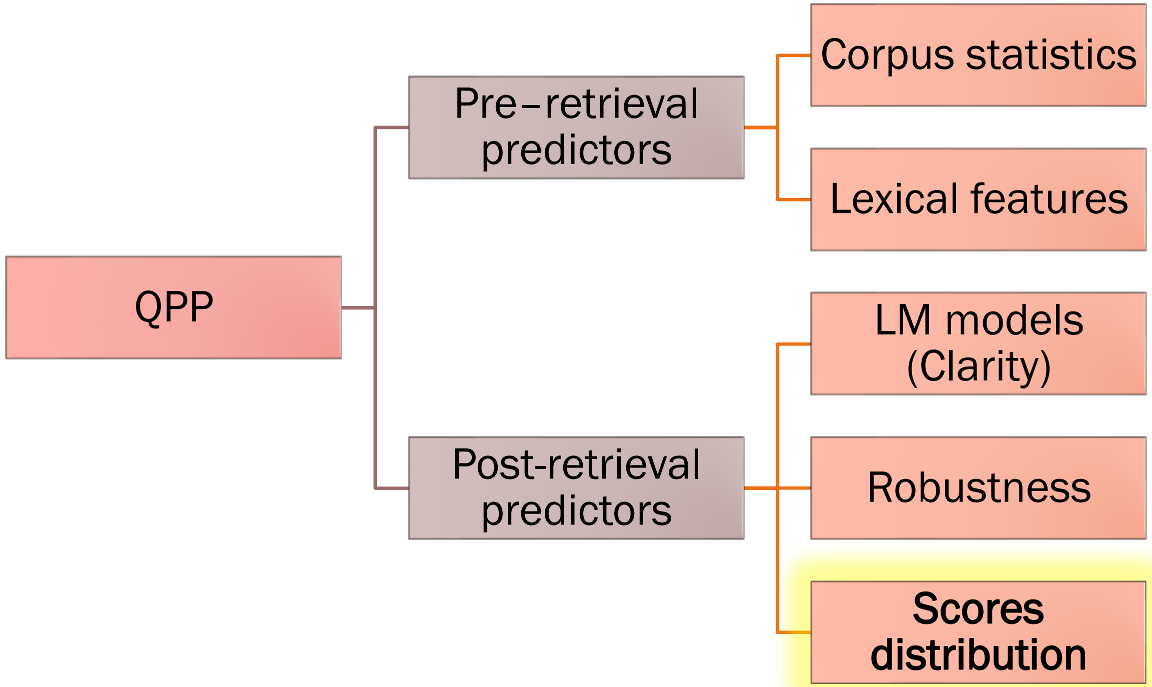
Query Performance Prediction (QPP)

Estimating the effectiveness of a search result, without relying on human assessors.

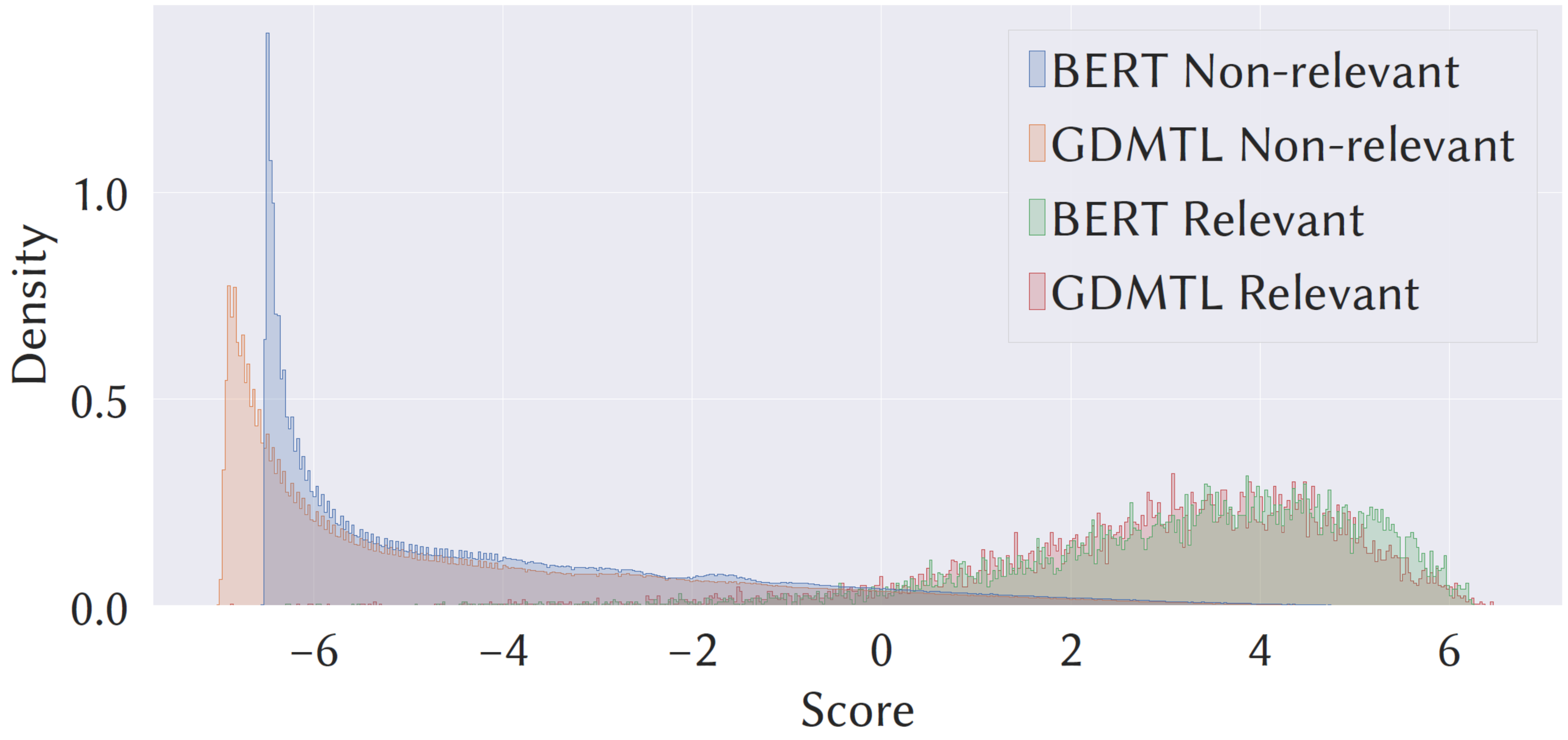
- QPP
  - Pre–retrieval predictors
    - Corpus statistics
    - Lexical features
  - Post-retrieval predictors
    - LM models (Clarity)
    - Robustness
    - Scores distribution

Query Performance Prediction (QPP)

Estimating the effectiveness of a search result, without relying on human assessors.



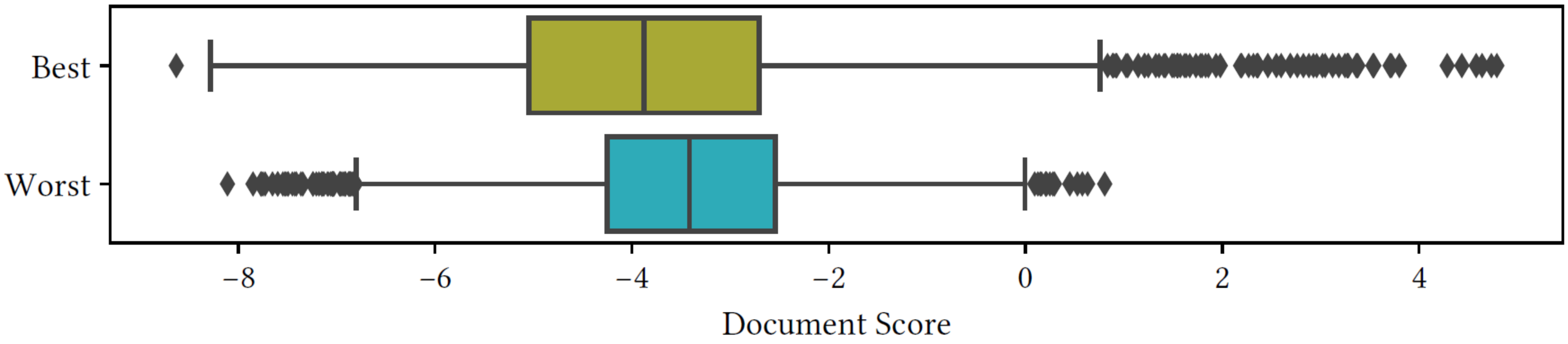
Neural IR Scores Distribution



Score distributions for MS MARCO collection.

B. Liu, H. Zamani, X. Lu, J. S. Culpepper, Generalizing Discriminative Retrieval Models Using Generative Tasks, in: Proc. WWW, 2021, pp. 3745–3756.

Different Score Distributions



Score distributions on Robust04 collection for 5 extreme queries.

Scores Distribution QPP

Use retrieval score values (RSVs) generated by the retrieval system to estimate the success of the search.

- Simple to calculate
- Computationally efficient
- Do not require any additional information
- Explainable (to some extent)

Examples:

- NQC[1], WIG[2], SMV[3], RSD[4]

[1] A. Shtok, O. Kurland, D. Carmel, F. Raiber, G. Markovits, Predicting Query Performance by Query-Drift Estimation, ACM Trans. Inf. Sys. 30 (2012) 1–35.
[2] Y. Zhou, W. B. Croft, Query Performance Prediction in Web Search Environments, in: Proc. SIGIR, 2007, pp. 543–550.
[3] Tao, Y., Wu, S.: Query performance prediction by considering score magnitude and variance together. In: Proceedings of the CIKM, pp. 1891–1894 (2014)
[4] H. Roitman, S. Erera, B. Weiner, Robust Standard Deviation Estimation for Query Performance Prediction, in: Proc. ICTIR, 2017, pp. 245–248.

Standard Deviation (SD) as QPP

Most common scores based QPP methods rely on SD.
For example: NQC[1], σ_{100}[2], $n(\sigma_{50\%})$[3], SMV[4], RSD[5]

$$\sigma = \sqrt{\frac{\sum_{d \in D_q} (S(q,d) - \mu)^2}{|D_q|}}, \text{ where } \mu = \frac{\sum_{d \in D_q} S(q,d)}{|D_q|}.$$

Standard deviation

[1] A. Shtok, O. Kurland, D. Carmel, Predicting Query Performance by Query-Drift Estimation in: Proc. ICTIR, 2009, 305–312.
[2] J. Pérez-Iglesias, L. Araujo, Ranking List Dispersion as a Query Performance Predictor, in: Proc. ICTIR, 2009, pp. 371–374.
[3] R. Cummins, J. Jose, C. O'Riordan, Improved Query Performance Prediction Using Standard Deviation, in: Proc. SIGIR, 2011, pp. 1089–1090.
[4] Tao, Y., Wu, S.: Query performance prediction by considering score magnitude and variance together. In: Proc. CIKM, 2014, pp. 1891–1894.
[5] H. Roitman, S. Erera, B. Weiner, Robust Standard Deviation Estimation for Query Performance Prediction, in: Proc. ICTIR, 2017, pp. 245–248.

Entropy as QPP

$$H(q) = -\sum_{d \in D_q} P(d|q) \log P(d|q)$$

Entropy

$$P(d|q) = \frac{exp(S(q,d))}{\sum_{d' \in D_q} exp(S(q,d'))}$$

Softmax function

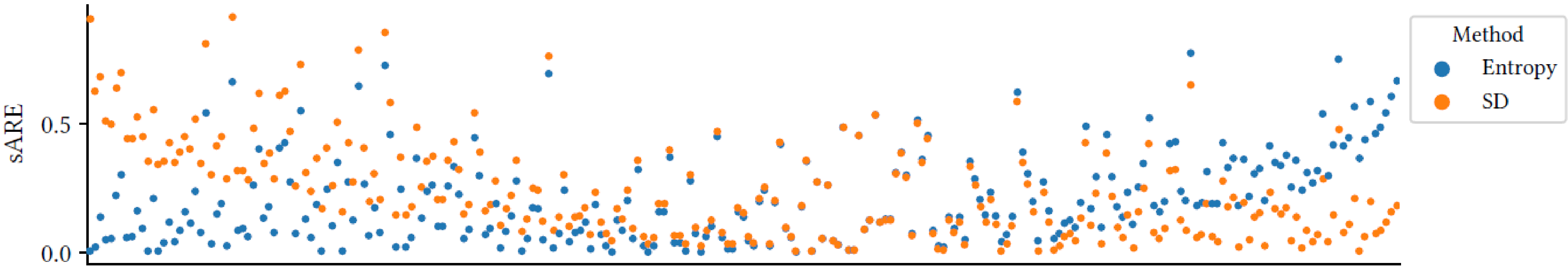
Entropy vs SD – Prediction Quality

		QL			NeuralRanker		
		sMARE	P-r	K-τ	sMARE	P-r	K-τ
Robust04	**SD**	0.224	**0.481**	0.345	0.220	0.526	0.359
	Entropy	**0.221**	0.385*	**0.349**	**0.214***	**0.529**	**0.388***

Prediction quality for SD and entropy as QPP measures

Entropy vs SD – Similarity

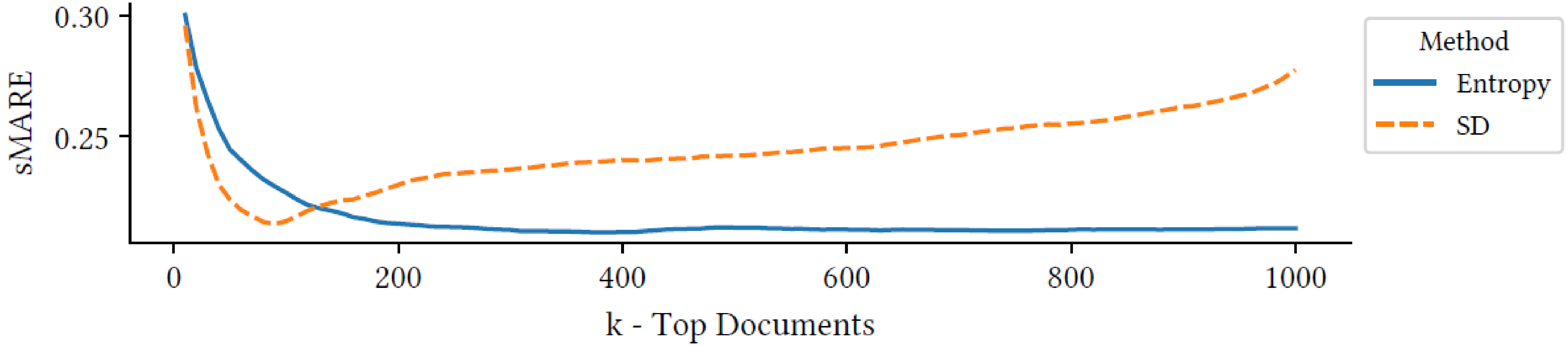
sARE (error) value per query



Queries ordered by difference in sARE (SD – *Entropy)*
The overall correlation is *r*=0.5

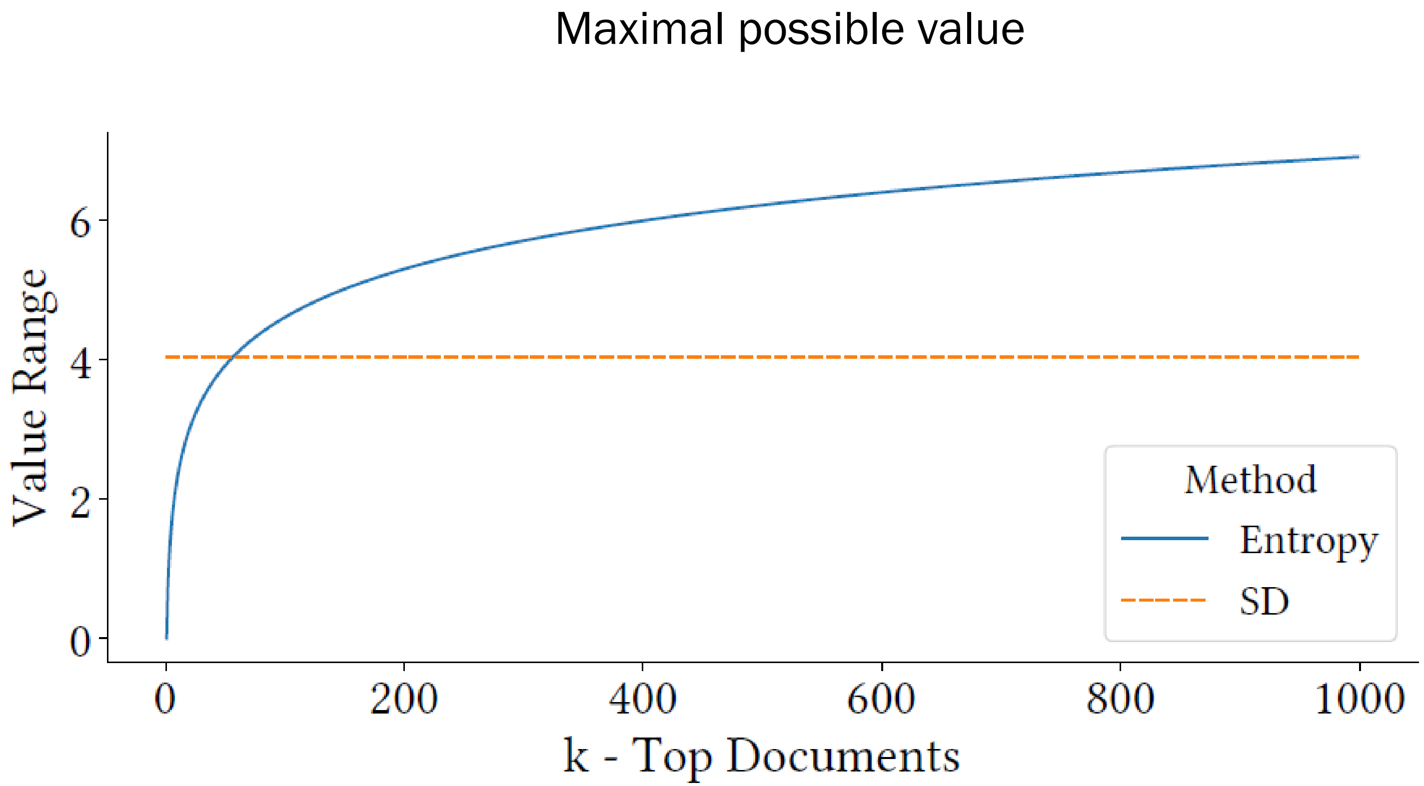
Entropy vs SD – Robustness

sMARE (mean error) for all queries



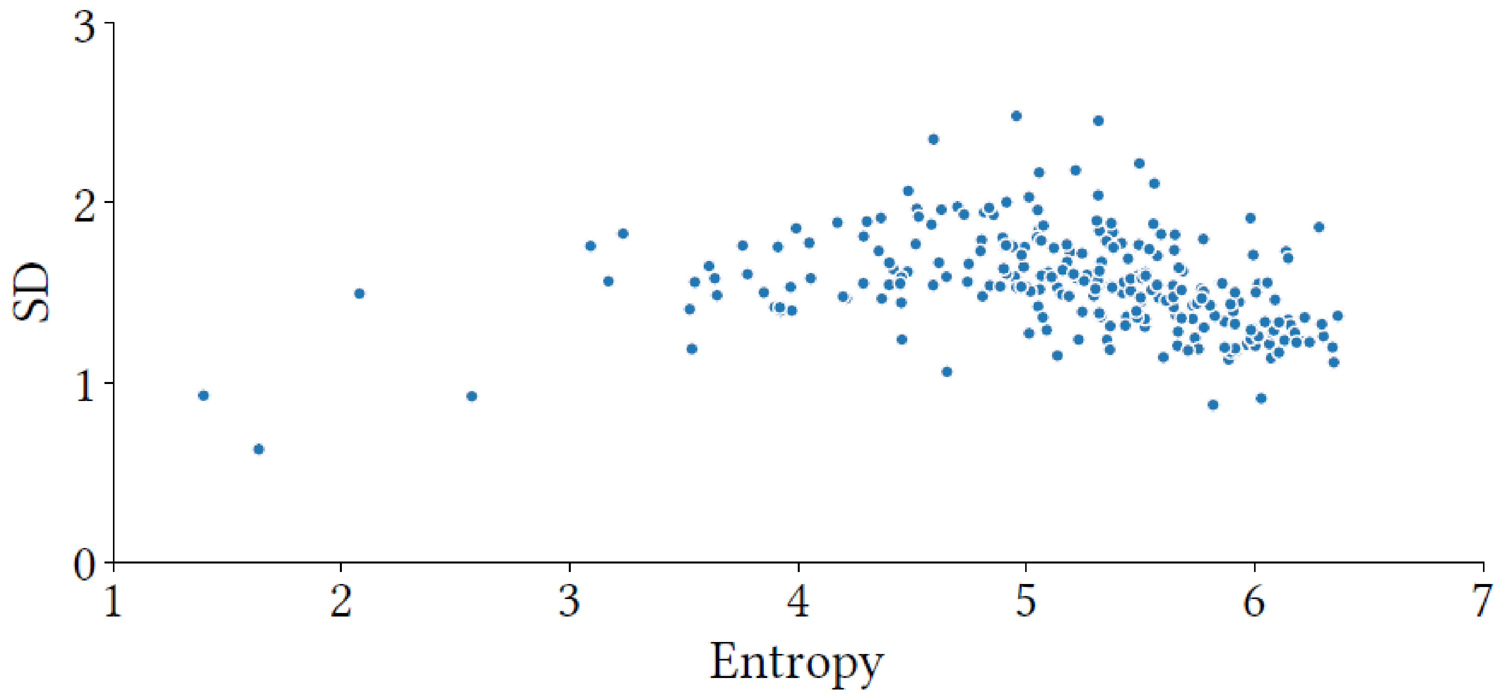
Why Doesn’t SD Keeps Improving?

Entropy vs SD – Discussion



Entropy vs SD – Discussion cont.

Actual (empirical) values for 1000 documents over 250 queries



Conclusions

Introduced *Entropy* as a new QPP method, it is:

- competitive with SD,
- differs from SD;
- more robust than SD.

Future Directions

Test the proposed method

- on other collections and tasks;
- with other retrieval systems.

Test the proposed method within different QPP frameworks:

- with reference queries;
- with reference lists.

Test additional measures of dispersion, e.g.,

- Kurtosis;
- Mean absolute difference.

Questions