Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modelling

SIGIR'18, July 8-12, 2018, Ann Arbor, MI, USA

Presented by: Pallavit Aggarwal

MS – Intelligent Systems - 22333721

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

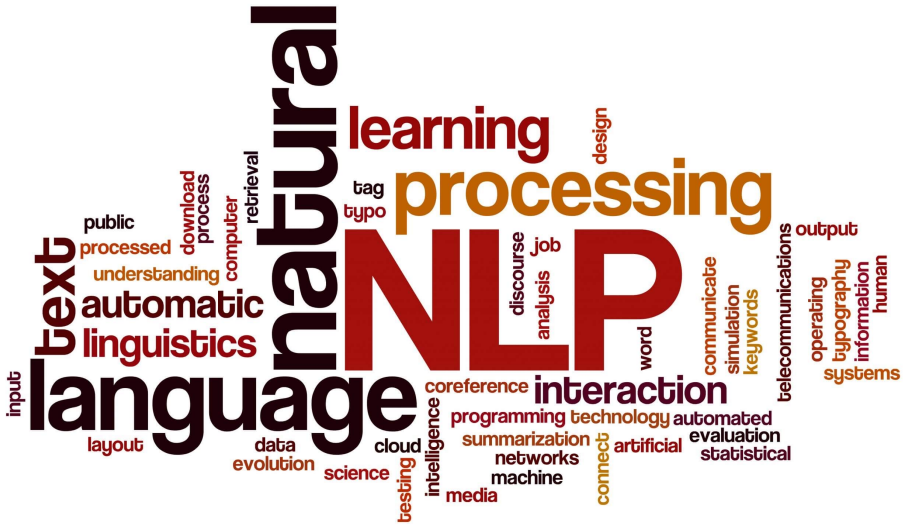
AIM

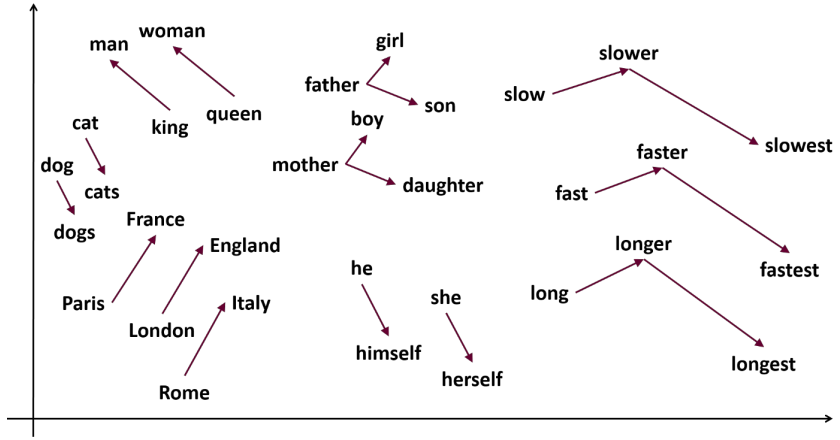
- This paper focuses on improving text understanding and retrieval by better estimating entity salience in documents and presents KESM (Kernel Entity Salience Model)

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Background

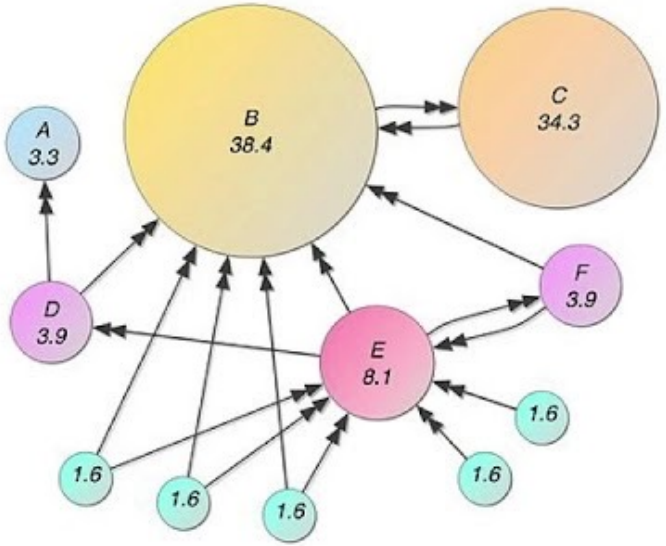
NLP in IR
search engines – text understanding approaches!





1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Background



"Bag-of-Terms":
☺ Effective & Efficient
😕 Mostly Frequency Signals

Can meaning improve text search?
Can we do better than TF-IDF?

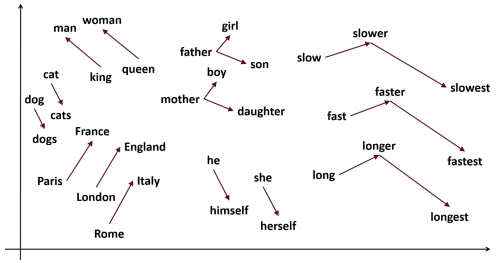
Frequency != Importance

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Related Work

1. End-to-End Neural Ad-hoc Ranking with Kernel Pooling ; SIGIR'17, August 7-11, 2017, Shinjuku, Tokyo, Japan

(mentioned in footer below as 2.)



Word2Vec may be error prone! Eg: "Ireland" "Northern-Ireland"

Can we do better?

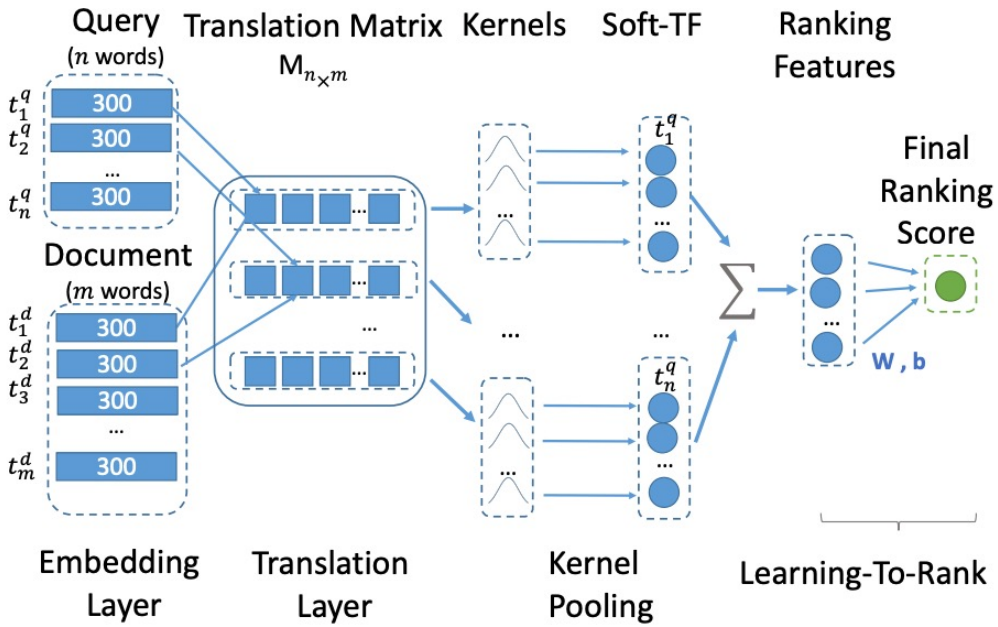
1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Related Work

1. End-to-End Neural Ad-hoc Ranking with Kernel Pooling ; SIGIR'17, August 7-11, 2017, Shinjuku, Tokyo, Japan

(mentioned in footer below as 2.)

1. Translation Model ->
2. Kernel Pooling ->
3. Learning to Rank Model ->



$$f(q,d) = \tanh(w^T \phi(M) + b) \quad \text{Learning to Rank} \quad (1)$$

$$\phi(M) = \sum_{i=1}^{n} \log \vec{K}(M_i) \quad \text{Soft-TF Features} \quad (2)$$

$$\vec{K}(M_i) = \{K_1(M_i), \ldots, K_K(M_i)\} \quad \text{Kernel Pooling} \quad (3)$$

$$K_k(M_i) = \sum_j \exp\left(-\frac{(M_{ij} - \mu_k)^2}{2\sigma_k^2}\right) \quad \text{RBF Kernel} \quad (4)$$

$$M_{ij} = \cos(\vec{v}_{t_i^q}, \vec{v}_{t_j^d}) \quad \text{Translation Matrix} \quad (5)$$

$$t \Rightarrow \vec{v}_t. \quad \text{Word Embedding} \quad (6)$$

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Related Work

1. End-to-End Neural Ad-hoc Ranking with Kernel Pooling ; SIGIR'17, August 7-11, 2017, Shinjuku, Tokyo, Japan

(mentioned in footer below as 2.)

- 35 Million Search Sessions
- 96,229 distinct queries
- Avg. 12 documents in a query

Evaluation

- DCTR and TACM click model to generate relevance scores
- NDCG {1,3,10}

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Related Work

1. End-to-End Neural Ad-hoc Ranking with Kernel Pooling ; SIGIR'17, August 7-11, 2017, Shinjuku, Tokyo, Japan

(mentioned in footer below as 2.)

Method	NDCG@1		NDCG@3		NDCG@10	
Lm	0.1261	−20.89%	0.1648	−26.46%	0.2821	−20.45%
BM25	0.1422	−10.79%	0.1757	−21.60%	0.2868	−10.14%
RankSVM	0.1457	−8.59%	0.1905	−14.99%	0.3087	−12.97%
Coor-Ascent	0.1594‡§¶	–	0.2241‡§¶	–	0.3547‡§¶	–
Trans	0.1347	−15.50%	0.1852	−17.36%	0.3147	−11.28%
DRMM	0.1366	−14.30%	0.1902	−15.13%	0.3150	−11.20%
CDSSM	0.1441	−9.59%	0.2014	−10.13%	0.3329‡§	−6.14%
K-NRM	**0.2642**†‡§¶	+65.75%	**0.3210**†‡§¶	+43.25%	**0.4277**†‡§¶	+20.58%

<- Testing-SAME Metrics (DCTR)

Testing-DIFF Metrics (TACM) ->

Method	NDCG@1		NDCG@3		NDCG@10	
Lm	0.1852	−11.34%	0.1989	−17.23%	0.3270	−13.38%
BM25	0.1631	−21.92%	0.1894	−21.18%	0.3254	−13.81%
RankSVM	0.1700	−18.62%	0.2036	−15.27%	0.3519	−6.78%
Coor-Ascent	0.2089‡¶	–	0.2403‡	–	0.3775‡¶	–
Trans	0.1874	−10.29%	0.2127	−11.50%	0.3454	−8.51%
DRMM	0.2068	−1.00%	0.2491‡	+3.67%	0.3809‡¶	+0.91%
CDSSM	0.1846	−10.77%	0.2358‡	−1.86%	0.3557	−5.79%
K-NRM	**0.2984**†‡§¶	+42.84%	**0.3092**†‡§¶	+28.26%	**0.4201**†‡§¶	+11.28%

Evaluation

- DCTR and TACM click model to generate relevance scores
- NDCG {1,3,10}

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Related Work

2. **A New Entity Salience Task with Millions of Training Examples** (mentioned in footer below as 3.)

Jesse Dunietz
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213, USA
jdunietz@cs.cmu.edu

Dan Gillick
Google Research
1600 Amphitheatre Parkway
Mountain View, CA 94043, USA
dgillick@google.com

Dataset Used:
The New York Times

3. **Crowdsourced Corpus with Entity Salience Annotations** (mentioned in footer below as 4.)

Milan Dojchinovski[1], **Dinesh Reddy**[2], **Tomáš Kliegr**[3], **Tomas Vitvar**[1], **Harald Sack**[2]
[1] Web Intelligence Research Group
Faculty of Information Technology, Czech Technical University in Prague
firstname.lastname@fit.cvut.cz

Semantic Technologies Group
[2] Hasso Plattner Institute, University of Potsdam, Germany
firstname.lastname@hpi.de

[3] Department of Information and Knowledge Engineering
Faculty of Informatics and Statistics, University of Economics, Prague
firstname.lastname@vse.cz

Grand Takeaway

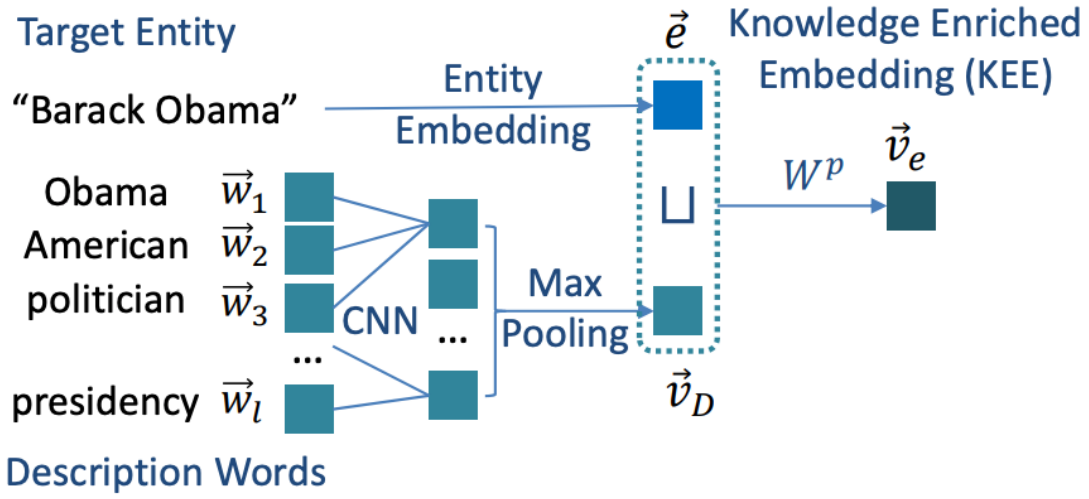
Frequency != Importance

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

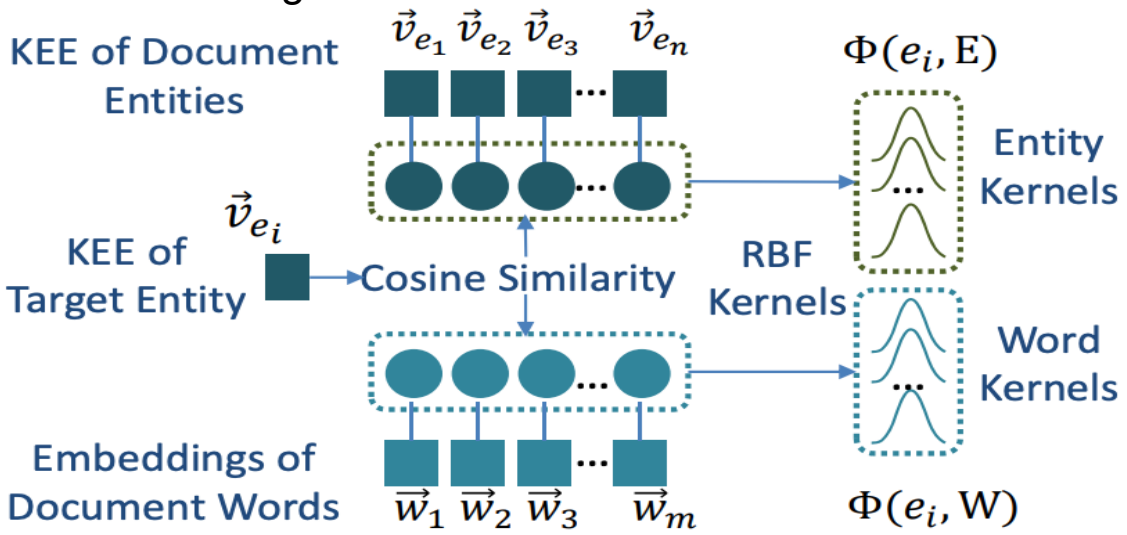
Brief Overview

Fig 1



- Learn to represent entities using embeddings
- Integrate knowledge graph semantics

Fig 2



- Model term interactions in the embedding space
- Capture multi-level interactions using kernels

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KEE

KNOWLEDGE-ENRICHED EMBEDDING

- Learn to represent entities using embeddings
- Integrate knowledge graph semantics

Map entities to embeddings (to be learned)

- $e_i \rightarrow \vec{e}_i$

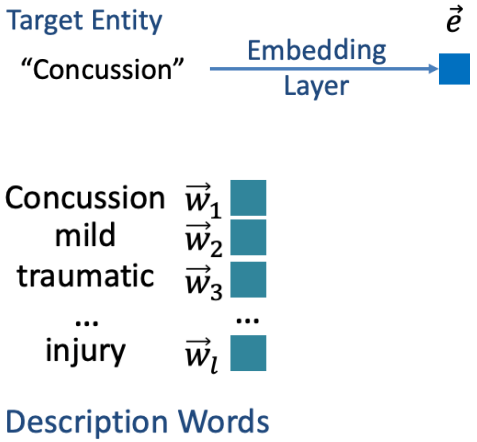
Entity Embeddings

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KEE

Introduce words in the entity description



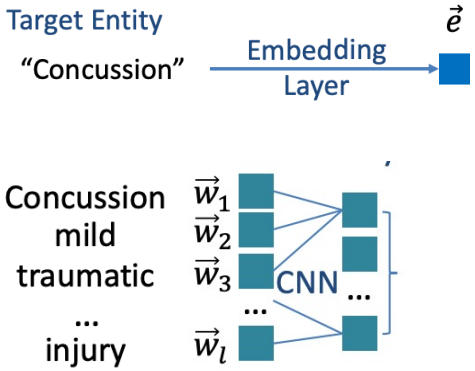
- $e_i \rightarrow \vec{e}_i$ Entity Embeddings
- $w_p \rightarrow \vec{w}_p$ Word Embeddings

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KEE

Compose words by CNN filters



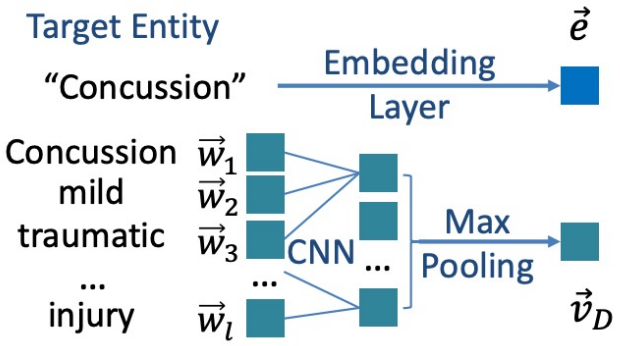
- $e_i \rightarrow \vec{e}_i$ Entity Embeddings
- $w_p \rightarrow \vec{w}_p$ Word Embeddings
- $C_p = W^c \cdot \vec{w}_{p:p+h}$ CNN Filter

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KEE

Max-pool to description embedding



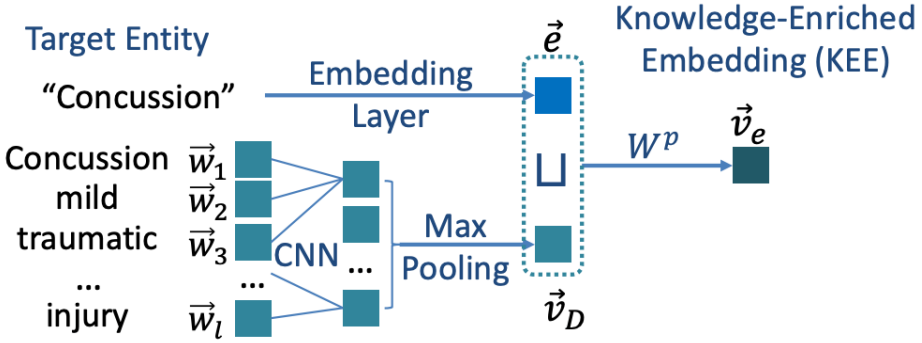
- $e_i \rightarrow \vec{e}_i$ Entity Embeddings
- $w_p \rightarrow \vec{w}_p$ Word Embeddings
- $C_p = W^c \cdot \vec{w}_{p:p+h}$ CNN Filter
- $\vec{v}_D = \max(C_1, \ldots, C_{l-h})$ Description Embeddings

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KEE

Combine to the Knowledge Enriched Embedding (KEE)



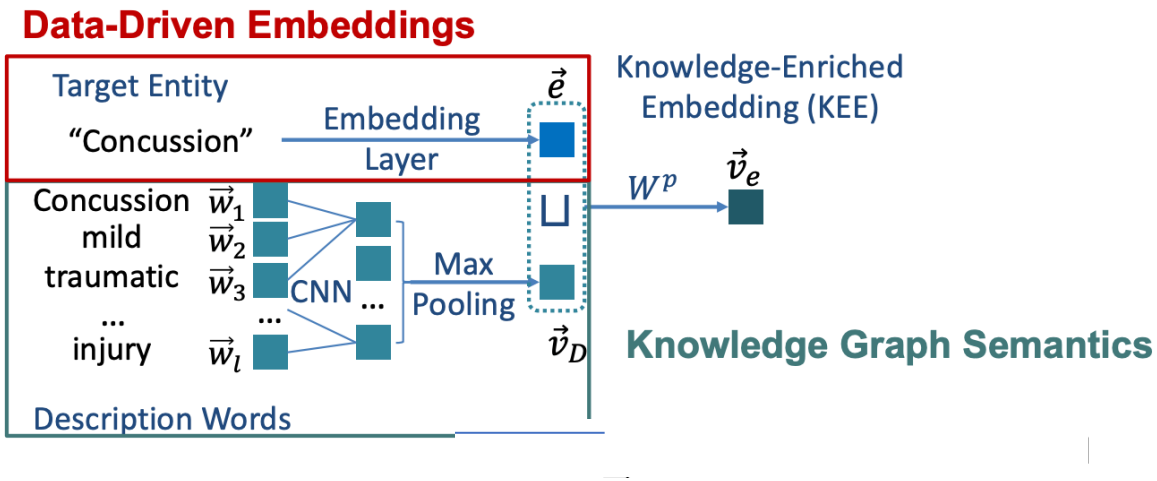
- $e_i \rightarrow \vec{e}_i$ Entity Embeddings
- $w_p \rightarrow \vec{w}_p$ Word Embeddings
- $C_p = W^c \cdot \vec{w}_{p:p+h}$ CNN Filter
- $\vec{v}_D = \max(C_1, \ldots, C_{l-h})$ Description Embeddings
- $\vec{v}_{e_i} = W^P \cdot (\vec{e}_i \sqcup \vec{v}_D)$ KEE Embeddings

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KEE

Combine to the Knowledge Enriched Embedding (KEE)



- $e_i \rightarrow \vec{e}_i$ Entity Embeddings
- $w_p \rightarrow \vec{w}_p$ Word Embeddings
- $C_p = W^c \cdot \vec{w}_{p:p+h}$ CNN Filter
- $\vec{v}_D = \max(C_1, \ldots, C_{l-h})$ Description Embeddings
- $\vec{v}_{e_i} = W^P \cdot (\vec{e}_i \sqcup \vec{v}_D)$ KEE Embeddings

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

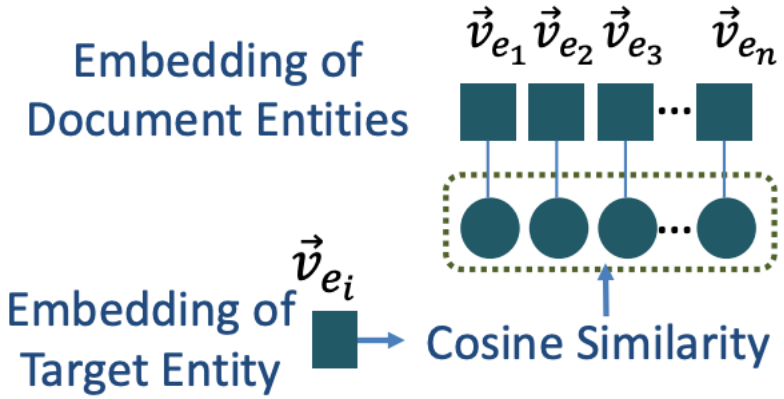
Methodology

In-Depth Overview - KIM

KERNEL INTERACTION MODEL

- Model term interactions in the embedding space
- Capture multi-level interactions using kernels

Model entity interactions in the embedding space



$$KIM(e_i, d) = \Phi(e_i, \mathbb{E}) \sqcup \Phi(e_i, \mathbb{W}).$$

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

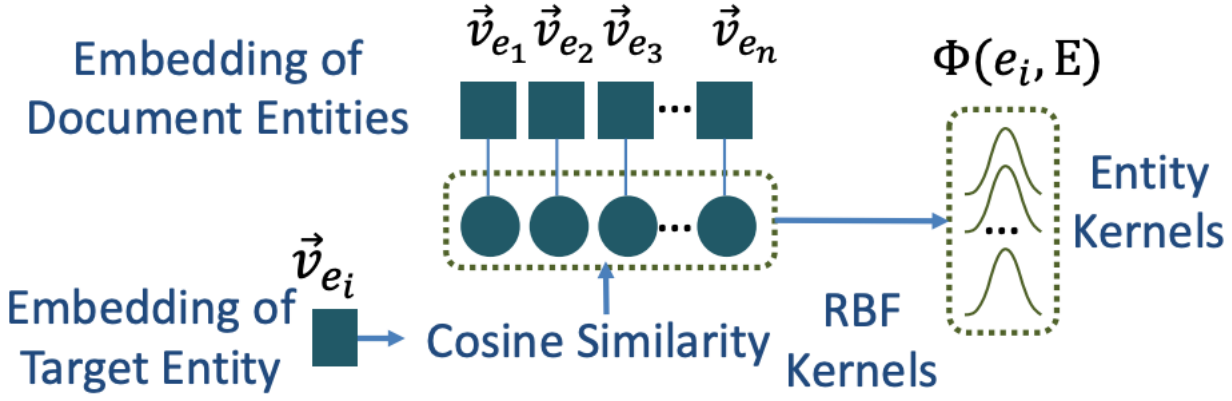
Methodology

In-Depth Overview - KIM

KERNEL INTERACTION MODEL

- Model term interactions in the embedding space
- Capture multi-level interactions using kernels

Use kernels to capture multi-level interaction (K-NRM)



$$\Phi(e_i, \mathbb{E}) = \{\phi_1(e_i, \mathbb{E}), \ldots \phi_k(e_i, \mathbb{E}) \ldots, \phi_K(e_i, \mathbb{E})\},$$

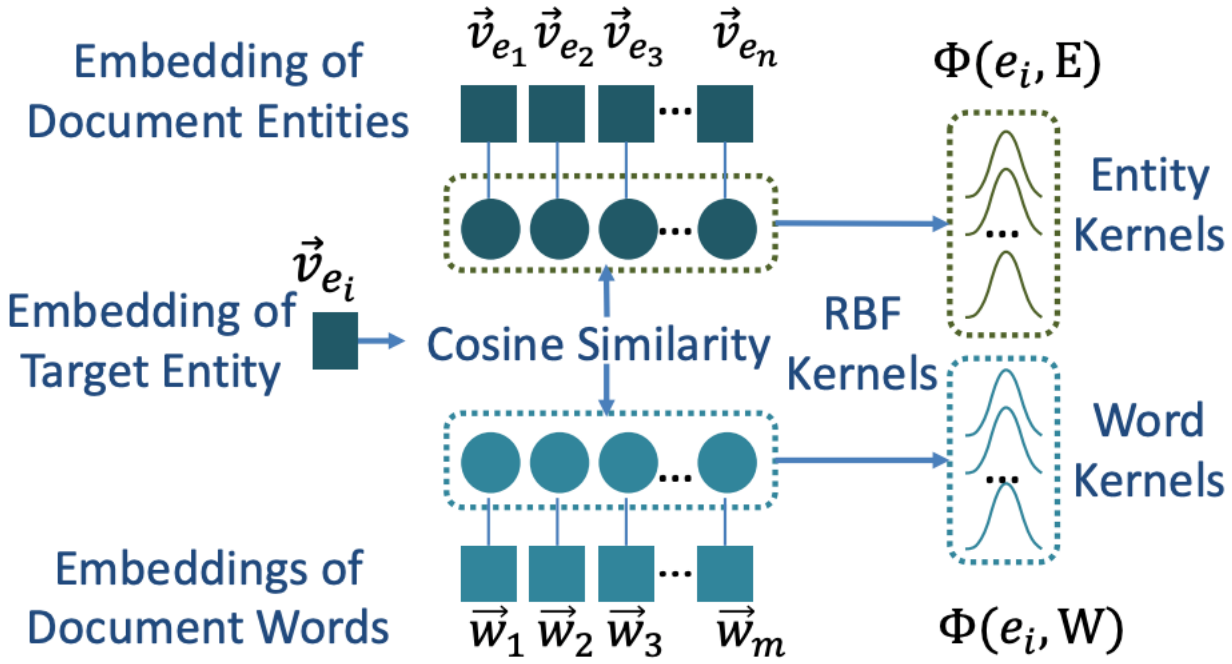
$$\phi_k(e_i, \mathbb{E}) = \sum_{e_j \in \mathbb{E}} \exp\left(-\frac{\left(cos(\vec{v}_{e_i}, \vec{v}_{e_j}) - \mu_k\right)^2}{2\sigma_k^2}\right).$$

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview - KIM

Kernel scores as features



$$\Phi(e_i, \mathbb{W}) = \{\phi_1(e_i, \mathbb{W}), \ldots \phi_k(e_i, \mathbb{W}) \ldots, \phi_K(e_i, \mathbb{W})\},$$

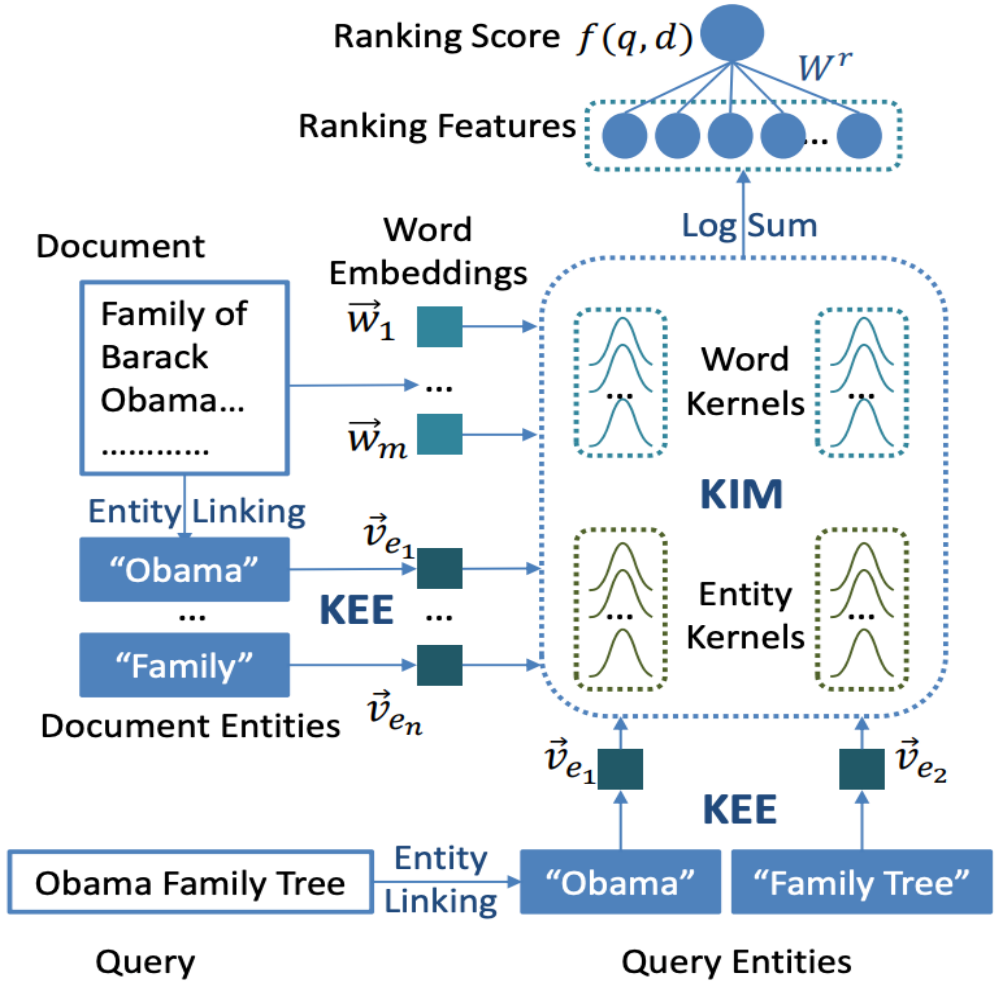
$$\phi_k(e_i, \mathbb{W}) = \sum_{w_j \in \mathbb{W}} \exp\left(-\frac{\left(cos(\vec{v}_{e_i}, \vec{w}_j) - \mu_k\right)^2}{2\sigma_k^2}\right).$$

$$f(e_i, d) = W^s \cdot KIM(e_i, d) + b^s.$$

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

In-Depth Overview – Ranking Model



$$f(q,d) = W^r \cdot \Psi(q,d),$$

$$\Psi(q,d) = \sum_{e_i \in \mathbb{E}^q} \log\left(\frac{KIM(e_i,d)}{|\mathbb{E}^d|}\right).$$

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Methodology

Learning and Training KESM

1. **Ad-Hoc Search Task**
   - **Labels available**
   - **Labels not available**

2. **Entity Salience Task**

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

1. **Entity Salience Task**
2. **Modeling Salience vs Counting Frequency**
3. **Ad-hoc search**

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

1. Datasets Used

The New York Times

Semantic Scholar

2. Baselines Established

- **Frequency** - estimates the salience of an entity by its term frequency
- **PageRank** - estimates the salience score of an entity using its PageRank score
- **LeToR** - is a feature-based learning to rank (entity) model.

4. Evaluation Metrics:

- Precision 1 and Precision 5
- Recall 1 and Recall 5

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

3. Feature Extraction

Name	Description	Source
Frequency	The frequency of the entity	Entity Linking
First Location	The location of the first sentence that contains the entity	Entity Linking
Head Word Count	The frequency of the entity's first head word in parsing	Dependency Parsing
Is Named Entity	Whether the entity is considered as a named entity	Named Entity Recognition
Coreference Count	The coreference frequency of the entity's mentions	Entity Coreference Resolution
Embedding Vote	Votes from other entities through cosine embedding similarity	Entity Embedding (Skip-gram)

4. Evaluation Metrics:

- Precision 1 and Precision 5
- Recall 1 and Recall 5

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

RESULT OF ENTITY SALIENCE TASK

New York Times

Method	Precision@1		Precision@5		Recall@1		Recall@5	
Frequency	0.5840	−8.53%	0.4065	−11.82%	0.0781	−11.92%	0.2436	−14.44%
PageRank	0.5845†	−8.46%	0.4069†	−11.73%	0.0782†	−11.80%	0.2440†	−14.31%
LeToR	0.6385	–	0.4610	–	0.0886	–	0.2848	–
KESM (E)	0.6470†‡§	+1.33%	0.4782†‡§	+3.73%	0.0922†‡§	+4.03%	0.3049†‡§	+7.05%
KESM (EK)	0.6528†‡§¶	+2.24%	0.4769†‡§	+3.46%	0.0920†‡§	+3.82%	0.3026†‡§	+6.27%
KESM (EW)	0.6767†‡§¶	+5.98%	0.5018†‡§¶	+8.86%	0.0989†‡§¶	+11.57%	0.3277†‡§¶	+15.08%
KESM	**0.6866**†‡§¶	+7.53%	**0.5080**†‡§¶	+10.21%	**0.1010**†‡§¶	+13.93%	**0.3335**†‡§¶	+17.10%

Semantic Scholar

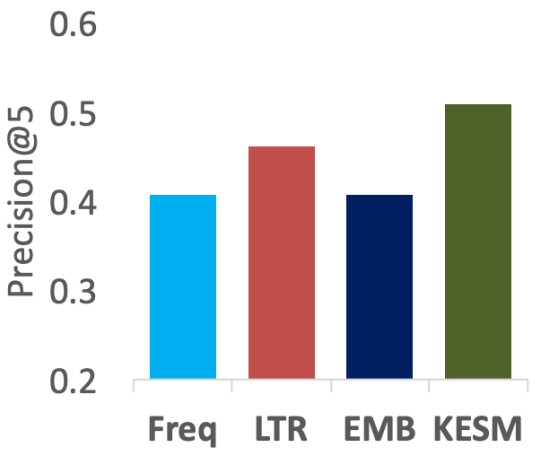
Method	Precision@1		Precision@5		Recall@1		Recall@5	
Frequency	0.3944	−9.99%	0.2560	−11.38%	0.1140	−12.23%	0.3462	−13.67%
PageRank	0.3946†	−9.94%	0.2561†	−11.34%	0.1141†	−12.11%	0.3466†	−13.57%
LeToR	0.4382	–	0.2889	–	0.1299	–	0.4010	–
KESM (E)	0.4793†‡§	+9.38%	0.3192†‡§	+10.51%	0.1432†‡§	+10.26%	0.4462†‡§	+11.27%
KESM (EK)	0.4901†‡§¶	+11.84%	0.3161†‡§	+9.43%	0.1492†‡§¶	+14.91%	0.4449†‡§	+10.95%
KESM (EW)	0.5097†‡§¶	+16.31%	0.3311†‡§¶	+14.63%	0.1555†‡§¶	+19.77%	0.4671†‡§¶	+16.50%
KESM	**0.5169**†‡§¶	+17.96%	**0.3336**†‡§¶	+15.47%	**0.1585**†‡§¶	+22.09%	**0.4713**†‡§¶	+17.53%

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

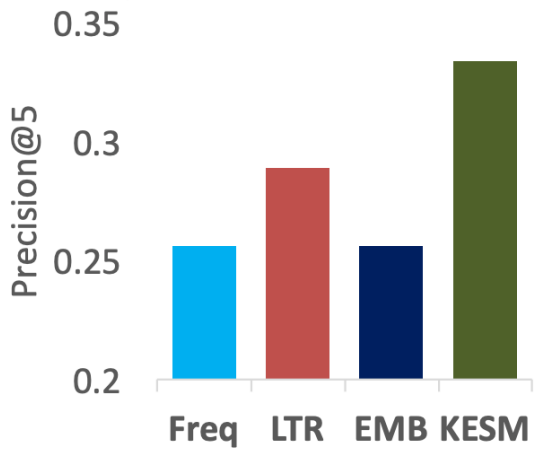
Evaluation and Results

RESULT OF ENTITY SALIENCE TASK - Simplified

The New York Times



Semantic Scholar



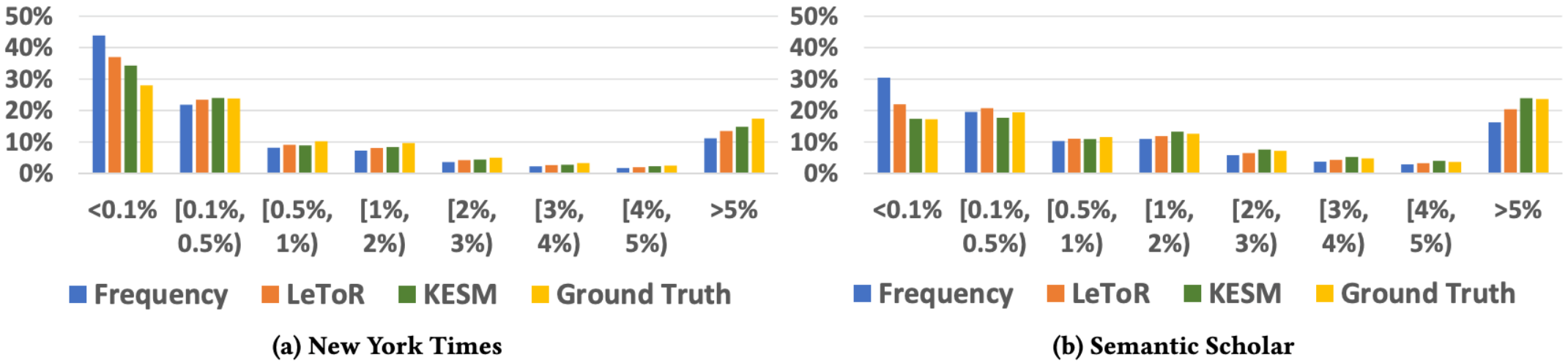
Freq: Frequency Count.
EMB: Raw embeddings.

LTR: Feature-based learning to rank.
KESM: Kernel Entity Salience Model.

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

Modelling Salience VS. Counting Frequency



(a) New York Times

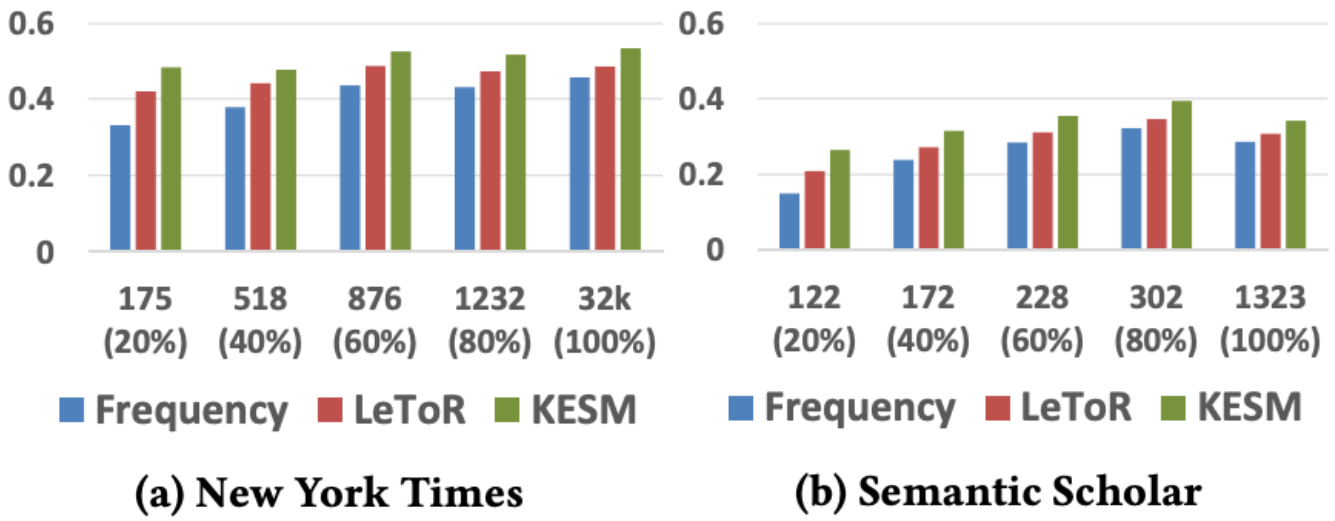
(b) Semantic Scholar

Ability to Model Tail Entities

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

Modelling Salience VS. Counting Frequency



Reliable on Short Documents

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

Ad-hoc Search

1. Datasets

- ClueWeb09 corpus
- ClueWeb12 corpus

2. Baselines

- BOW - base retrieval model, which is SDM on ClueWeb09-B
- BOE is the frequency-based retrieval with bag-of-entities
- IRFusion (BM25, and TFIDF, applied to body and title fields
- ESR (exact and soft match signals in the entity space)

3. Evaluation Metrics

- NDCG @ 20
- ERR @ 20

* Pretrained KESM using New York Times Dataset labels

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

Ad-hoc Search

1. Datasets

- ClueWeb09 corpus
- ClueWeb12 corpus

2. Baselines

- BOW - base retrieval model, which is SDM on ClueWeb09-B
- BOE is the frequency-based retrieval with bag-of-entities
- IRFusion (BM25, and TFIDF, applied to body and title fields
- ESR (exact and soft match signals in the entity space)

3. Evaluation Metrics

- NDCG @ 20
- ERR @ 20

*** Pretrained KESM using New York Times Dataset labels**

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

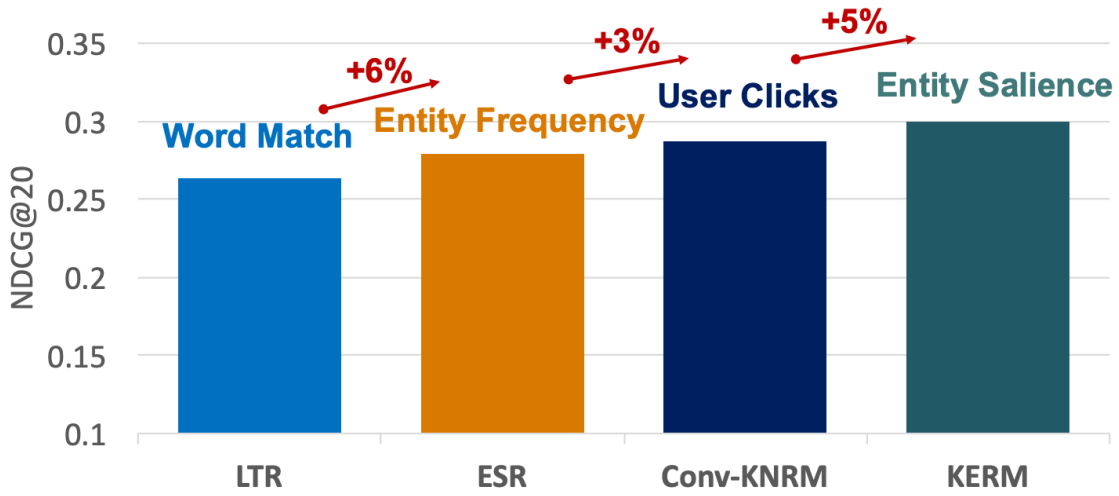
Ad-hoc Search - Results

	ClueWeb09-B					ClueWeb12-B13				
Method	**NDCG@20**		**ERR@20**		**W/T/L**	**NDCG@20**		**ERR@20**		**W/T/L**
BOW	0.2496	−5.26%	0.1387	−10.20%	62/38/100	0.1060	−12.02%	0.0863	−6.67%	35/22/43
BOE	0.2294	−12.94%	0.1488	−3.63%	74/25/101	0.1173	−2.64%	0.0950	+2.83%	44/19/37
IRFusion	0.2635	–	0.1544	–	–/–/–	0.1205	–	0.0924	–	–/–/–
ESR	$0.2695^{\dagger}$	+2.30%	0.1607	+4.06%	80/39/81	0.1166	−3.22%	0.0898	−2.81%	30/23/47
KESM	$0.2799^{\dagger}$	+6.24%	0.1663	+7.68%	85/35/80	$0.1301^{\dagger\S}$	+7.92%	$\mathbf{0.1103}^{\ddagger\S\P}$	+19.35%	43/25/32
ESR+IRFusion	$0.2791^{\dagger\ddagger}$	+5.92%	0.1613	+4.46%	91/34/75	0.1281	+6.30%	0.0951	+2.87%	45/24/31
KESM+IRFusion	$\mathbf{0.2993}^{\dagger\ddagger\S\P}$	+13.58%	$\mathbf{0.1797}^{\dagger\ddagger\S\P}$	+16.38%	98/35/67	$\mathbf{0.1308}^{\dagger\S}$	+8.52%	$0.1079^{\ddagger\S\P}$	+16.77%	43/23/34

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

Ad-hoc Search – Results - Simplified



simplified view of NDCG@20 scores

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Evaluation and Results

Ad-hoc Search – Results - Simplified

	ClueWeb09-B					ClueWeb12-B13				
Method	**NDCG@20**		**ERR@20**		**W/T/L**	**NDCG@20**		**ERR@20**		**W/T/L**
IRFusion-Title	0.2584	−3.51%	0.1460	−5.16%	83/48/69	0.1187	+6.23%	0.0894	+3.14%	41/23/36
ESR-Title	0.2678	–	0.1540	–	–/–/–	0.1117	–	0.0867	–	–/–/–
KESM-Title	**0.2780**†	+3.81%	**0.1719**†‡	+11.64%	91/46/63	**0.1199**	+7.36%	**0.0923**	+6.42%	35/28/37
IRFusion-Body	0.2550	+0.48%	0.1427	−3.44%	80/46/74	0.1115	+4.61%	0.0892	−3.51%	36/30/34
ESR-Body	0.2538	–	0.1478	–	–/–/–	0.1066	–	0.0924	–	–/–/–
KESM-Body	**0.2795**†‡	+10.13%	**0.1661**†‡	+12.37%	96/39/65	**0.1207**‡	+13.25%	**0.1057**†‡	+14.44%	43/24/33

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

Conclusion

1. Experiments show that KESM successfully transfers the text understanding ability from entity salience estimation to search.

2. Interaction and consistency between query entities with the document entities and words—actually improves the ranking accuracy

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.

My Views

1. Thorough evaluations
2. “Search with meanings” outperforms “search by matching”
3. NDCG and ERR are official and proper metrics for evaluations
4. Case study comparing ESR and KESM also provided
5. Code for KESM not available, but for K-NRM available - https://github.com/AdeDZY/K-NRM
6. KESM compiled faster than LTR with a typical GPU
7. Search performance is slow, due to feature richness
8. Multiple new variants that improve KESM are available now

1. Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 575–584. https://doi-org.elib.tcd.ie/10.1145/3209978.3209982
2. Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-End Neural Ad-hoc Ranking with Kernel Pooling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 55–64. https://doi-org.elib.tcd.ie/10.1145/3077136.3080809
3. Dojchinovski, Milan & Reddy, Dinesh & Kliegr, Tomáš & Vitvar, Tomas & Sack, Harald. (2016). Crowdsourced Corpus with Entity Salience Annotations.
4. Dunietz, Jesse & Gillick, Dan. (2014). A New Entity Salience Task with Millions of Training Examples. EACL 2014. 205. 10.3115/v1/E14-4040.