References


Yoshua Bengio, Jean-Sébastien Senécal, and others. 2003. Quick Training of Probabilistic Neural Nets by Importance Sampling,. In AISTATS.


Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li, Vince Gatto, and H Chi. 2018. Latent Cross: Making Use of Context in Recurrent Recommender Systems. In WSDM.


Fernando Diaz, Bhaskar Mitra, and Nick Craswell. 2016. Query expansion with locally-trained word embeddings. In ACL.


References V


Tom Kenter, Llion Jones, and Daniel Hewlett. 2018. Byte-level Machine Reading across Morphologically Varied Languages. In *AAAI.*


References VII


Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In AAAI. 2181–2187.


Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-Value Memory Networks for Directly Reading Documents. In EMNLP.


References VIII


D. Sculley. 2009. Large scale learning to rank. In NIPS Workshop on Advances in Ranking.


D. Sculley. 2009. Large scale learning to rank. In NIPS Workshop on Advances in Ranking.


Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In ACL.


Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-To-End Memory Networks. In NIPS.

Yaming Sun, Lei Lin, Duyu Tang, Nan Yang, Zhenzhou Ji, and Xiaolong Wang. 2015. Modeling Mention, Context and Entity with Neural Networks for Entity Disambiguation. In IJCAI.


References X


Ashish Vaswani, Yinggong Zhao, Victoria Fossum, and David Chiang. 2013. Decoding with Large-Scale Neural Language Models Improves Translation.. In EMNLP. 1387–1392.


Ivan Vulić and Marie-Francine Moens. 2015. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. In SIGIR. ACM, 363–372.


Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.. In AAAI. 1112–1119.


### Notation

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single query</td>
<td>$q$</td>
</tr>
<tr>
<td>Single document</td>
<td>$d$</td>
</tr>
<tr>
<td>Set of queries</td>
<td>$Q$</td>
</tr>
<tr>
<td>Collection of documents</td>
<td>$D$</td>
</tr>
<tr>
<td>Term in query $q$</td>
<td>$t_q$</td>
</tr>
<tr>
<td>Term in document $d$</td>
<td>$t_d$</td>
</tr>
<tr>
<td>Full vocabulary of all terms</td>
<td>$T$</td>
</tr>
<tr>
<td>Set of ranked results retrieved for query $q$</td>
<td>$R_q$</td>
</tr>
<tr>
<td>Result tuple (document $d$ at rank $i$)</td>
<td>$\langle i, d \rangle$, where $\langle i, d \rangle \in R_q$</td>
</tr>
<tr>
<td>Relevance label of document $d$ for query $q$</td>
<td>$rel_q(d)$</td>
</tr>
<tr>
<td>$d_i$ is more relevant than $d_j$ for query $q$</td>
<td>$rel_q(d_i) &gt; rel_q(d_j)$, or $d_i \succ_q d_j$</td>
</tr>
<tr>
<td>Frequency of term $t$ in document $d$</td>
<td>$tf(t, d)$</td>
</tr>
<tr>
<td>Number of documents that contain term $t$</td>
<td>$df(t)$</td>
</tr>
<tr>
<td>Vector representation of text $z$</td>
<td>$\vec{z}$</td>
</tr>
<tr>
<td>Probability function for an event $\mathcal{E}$</td>
<td>$p(\mathcal{E})$</td>
</tr>
<tr>
<td>The set of real numbers</td>
<td>$\mathbb{R}$</td>
</tr>
</tbody>
</table>

We adopt some neural network related notation from [Goodfellow et al., 2016] and IR related notation from [Mitra and Craswell, 2017]
Acknowledgments

All content represents the opinion of the author(s), which is not necessarily shared or endorsed by their employers and/or sponsors.