Contextualizing Trending Entities in News Stories

Marco Ponza, Diego Ceccarelli, Paolo Ferragina, Edgar Meij, Sambhav Kothari

Bloomberg Engineering

Università di Pisa

{mponza, dceccarelli4, emeij, skothari44}@bloomberg.net, paolo.ferragina@di.unipi.it

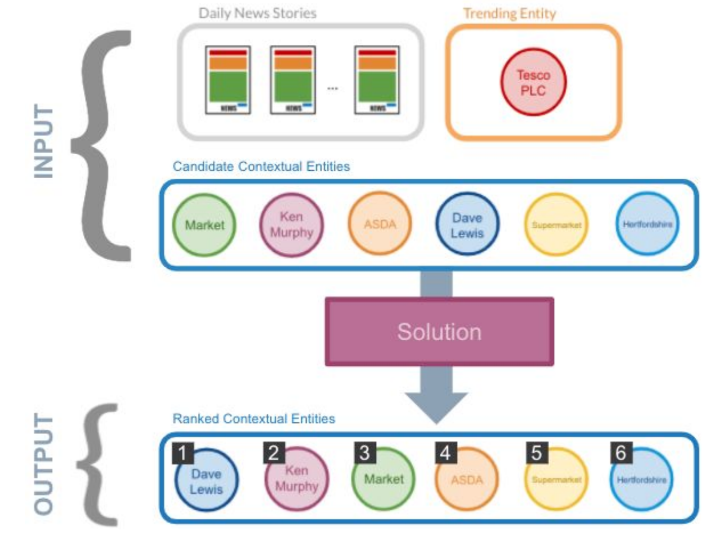
Our Contributions

New research problem that aims to contextualize trending entities by ranking related entities

Unsupervised solution based on Personalized PageRank
Supervised solution based on feature engineering and learning to rank

Creation of a test collection built with crowdsourcing
Available at https://doi.org/10.5281/zenodo.4422044

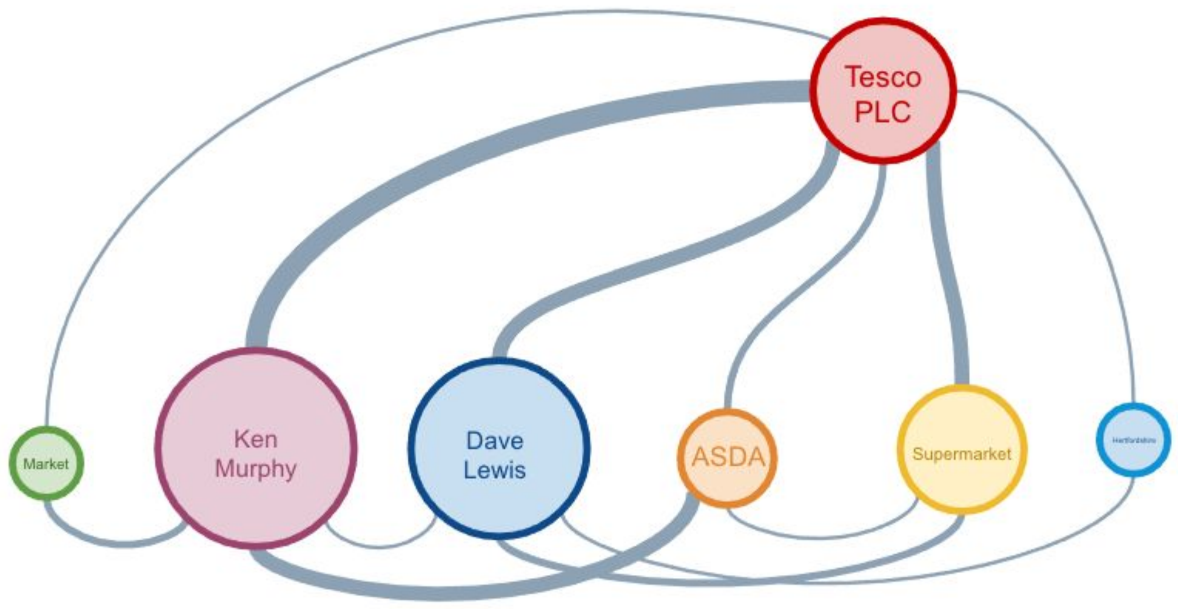
Problem Formulation



The CEO of Tesco Dave Lewis resigned and Ken Murphy is supposed to become the new CEO

Unsupervised Solution: Personalized PageRank with Embeddings

- Entities are nodes in the graph, all connected to the trending entity
- More edges are drawn by stories co-occurrences
- Edge weights are calculated from the cosine similarities of the entities' embeddings
- The teleport vector is instantiated with scores produced via entity salience
- The ranking of entities is eventually produced by running Personalized PageRank

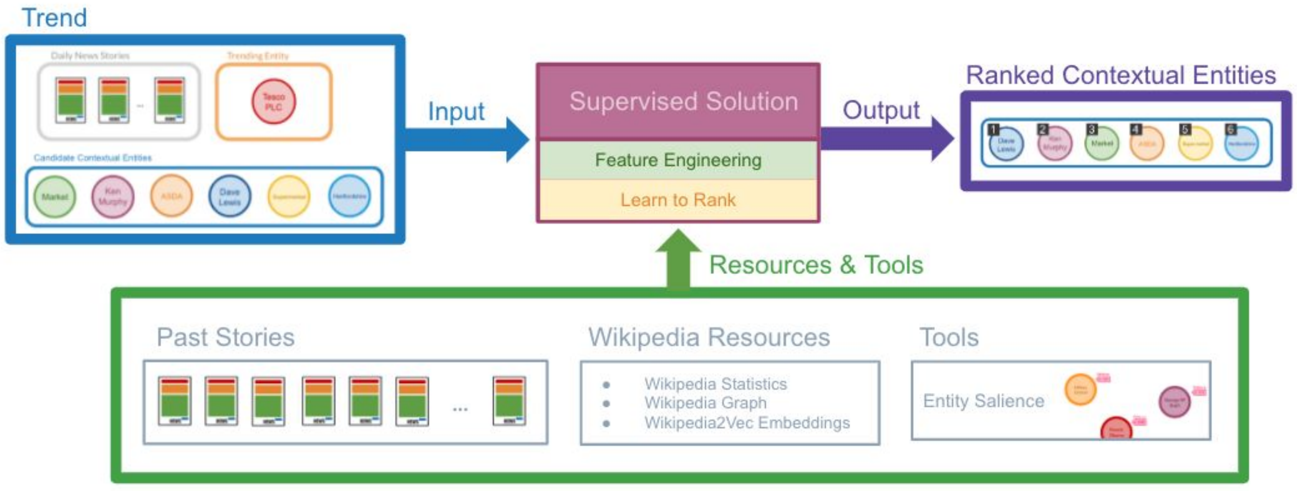


Experimental Results

Method	"Relevant" & "Somewhat Relevant" as Gold Labels						"Relevant" as Gold Label					
	MAP	P@1	P@3	NDCG@5	NDCG@10	MRR	MAP	P@1	P@3	NDCG@5	NDCG@10	MRR
Frequency	0.098	0.262	0.224	0.168	0.233	0.448	0.097	0.208	0.177	0.179	0.242	0.382
Co-Occurrence	0.359	0.477	0.295	0.441	0.479	0.604	0.441	0.416	0.221	0.486	0.515	0.528
Stories Embeddings	0.210	0.208	0.161	0.238	0.287	0.373	0.237	0.148	0.110	0.253	0.299	0.295
Reciprocal Rank	0.418	0.523	0.291	0.460	0.508	0.630	0.488	0.430	0.219	0.501	0.542	0.541
Salience (max)	0.497	0.570	**0.394**	0.556	0.612	0.727	0.555	0.456	**0.286**	0.593	0.640	0.622
PPR	**0.519**	**0.644**	0.391	**0.586**	**0.637**	**0.773**$^{\triangle}$	**0.605**$^{\blacktriangle}$	**0.564**$^{\blacktriangle}$	0.282	**0.639**$^{\triangle}$	**0.678**$^{\triangle}$	**0.686**$^{\triangle}$

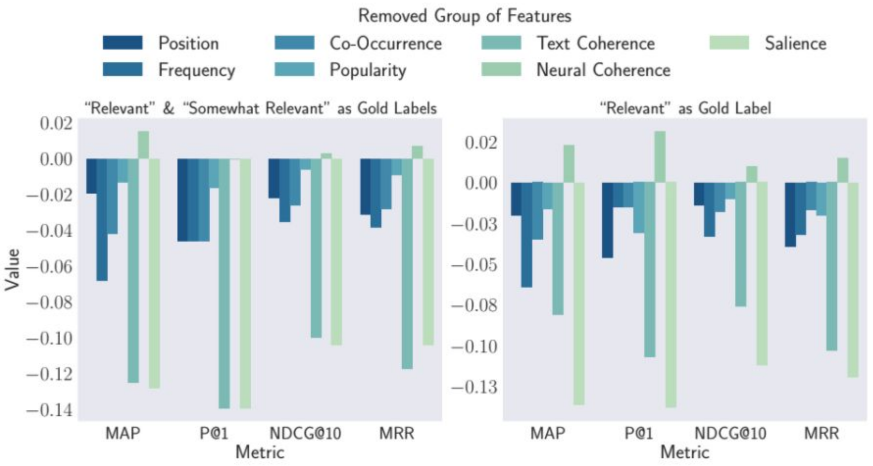
Supervised Solution: Feature Engineering with Learning to Rank

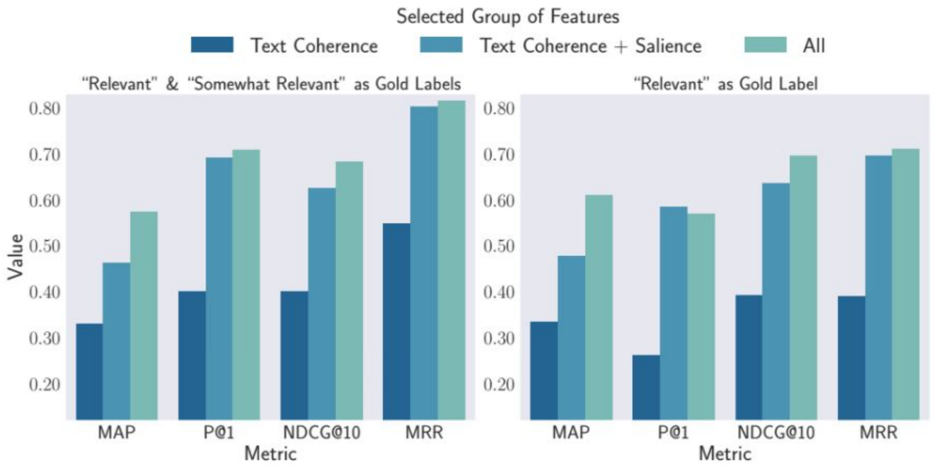
- Entities are transformed into vectors of features
- Features are derived from different signals:
  - Position
  - Frequency
  - Co-Occurrence
  - Popularity
  - Text and Neural Coherence
  - Salience
- Learning to Rank is implemented via LightGBM



Experimental Results

Method	"Relevant" & "Somewhat Relevant" as Gold Labels						"Relevant" as Gold Label					
	MAP	P@1	P@3	NDCG@5	NDCG@10	MRR	MAP	P@1	P@3	NDCG@5	NDCG@10	MRR
Salience (max)	0.474	0.569	0.364	0.526	0.584	0.714	0.534	0.462	0.251	0.566	0.616	0.604
PPR	0.495	0.646	0.364	0.565	0.617	0.767	0.591	0.554	0.256	0.622	0.659	0.665
LTR	**0.574**$^{\blacktriangle\triangle}$	**0.708**	**0.472**$^{\blacktriangle\triangle}$	**0.629**$^{\triangle}$	**0.682**$^{\blacktriangle\triangle}$	**0.815**$^{\triangle}$	**0.609**	**0.569**	**0.308**$^{\blacktriangle\triangle}$	**0.654**$^{\triangle}$	**0.696**$^{\triangle}$	**0.710**$^{\triangle}$





© 2021 Bloomberg Finance L.P. All rights reserved.