Entity-Level Event Impact Analytics

Govind Govind

Entity-Level Event Impact Analytics

Govind
govind@unicaen.fr
Université de Caen Normandie, France

Objective:
Automatically Predict the Event Diffusion into Foreign Language Communities

Conceptual Approach

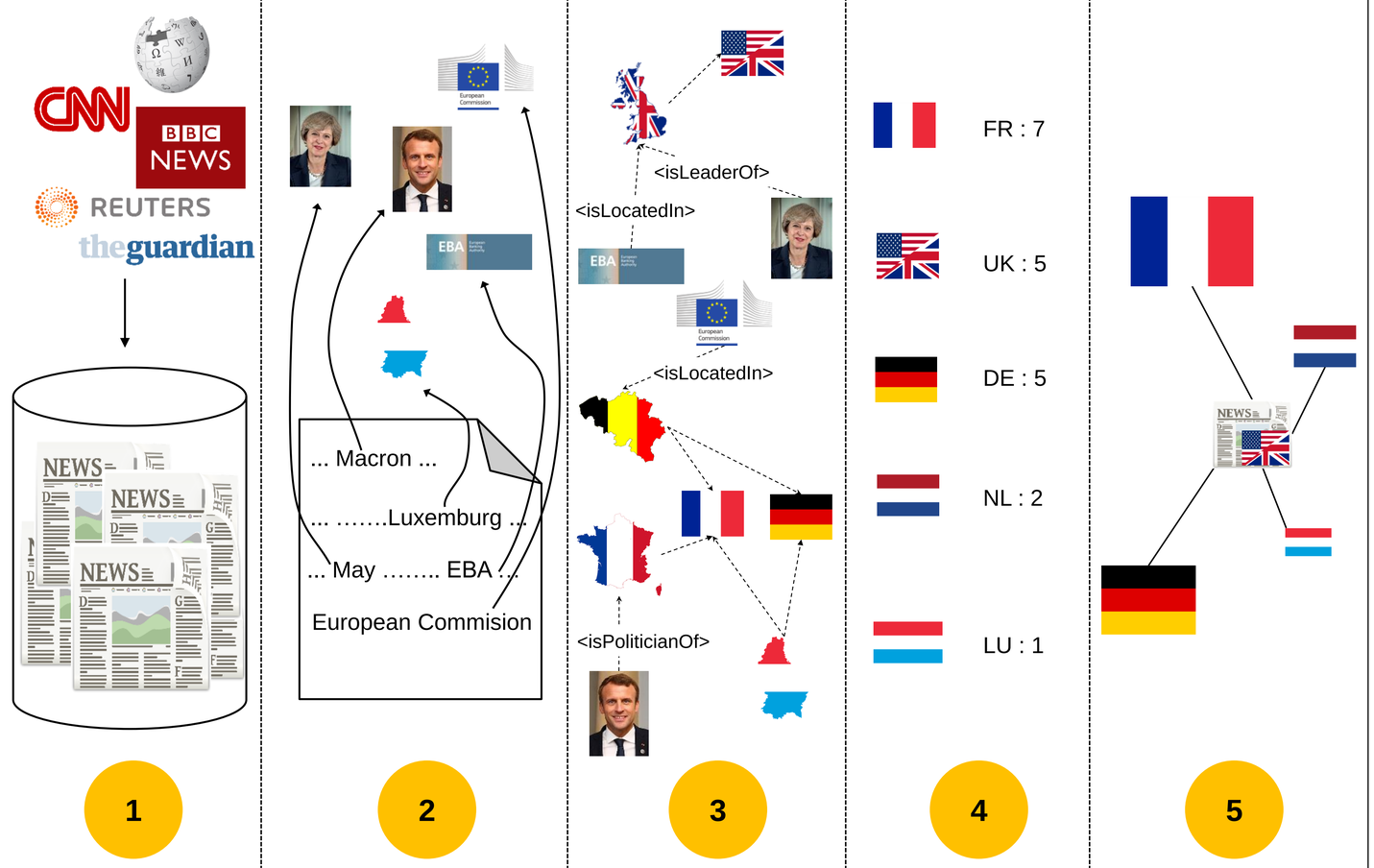
"Discovery of semantic connections to languages from named entities in the article"

ELEVATE Framework [1]
- Recursive exploration of yago relations
- **Country centric**:
  <isCitizenOf>, <diedIn>, <isLocatedIn>, <isLeaderOf>, <isPoliticianOf>, <wasBornIn>, <livesIn>
- **Organization centric**:
  <owns>, <created>, <worksAt>
- Linking of entities to languages

Breadth-first-search (BFS):
Stopping the exploration after the discovery of first language for a named entity

Depth-first-search (DFS):
Revealing all languages associated with a named entity exhaustively

ELEVATE Pipeline



Spread Prediction

Task: Pick the best candidates from all the scored languages

Adjusted Thresholding
- Threshold(θ) = average spread in the ground truth
- k-fold cross-validation
- Risk of picking the irrelevant languages

Multi-label Classification
- Output labels as the languages in event spread
- Candidate language scores as feature vectors
- Classifiers decide the spread

Other Research Works

ELEVATE-Live [3]
- Online news article virality prediction to countries
- Extension of ELEVATE framework
- Available at: https://elevate.greyc.fr

Semantic Fingerprinting [2]
- Fine-grained entity-level Web content classification
- Concise semantic representation of documents based on their entities

Short Biography

3rd year PhD student
University of Caen, France
Advisor: Prof. Marc Spaniol

Research Interests
- Entity-level analytics
- Data aggregation via LOD
- Deep learning for noisy data analytics

Education
- Master degree in Maths and Computing from Indian Institute of Technology (IIT) Patna (India)
- Bachelor degree in computer science from Guru Jambeshwar University of Science and Technology Hisar (India)

Publications

[1] Govind, Spaniol M. ELEVATE: A Framework for Entity-level Event Diffusion Prediction into Foreign Language Communities. In Proceedings of the 9th ACM on Web Science Conference 2017 Jun 25 (pp. 111-120). ACM.

[2] Govind, Alec C, Spaniol M. Semantic Fingerprinting: A Novel Method for Entity-Level Content Classification. In Proceedings of the 18th International Conference on Web Engineering 2018 Jun 5 (pp. 279-287). Springer.

[3] Govind, Alec C, Spaniol M. ELEVATE-Live: Assessment and Visualization of Online News Virality via Entity-Level Analytics. In Proceedings of the 18th International Conference on Web Engineering 2018 Jun 5 (pp. 482-486). Springer.

Experimental Results

Method	5 days Precision	5 days Recall	5 days F1	5 days #PC	10 days Precision	10 days Recall	10 days F1	10 days #PC	20 days Precision	20 days Recall	20 days F1	20 days #PC
Random	0.07736	0.07691	0.07714	11288	0.07736	0.07691	0.07714	11288	0.07736	0.07691	0.07714	11288
$Inlinks_{CNC}$	0.23571	0.21023	0.22224	5045	0.25685	0.24077	0.24855	5875	0.28541	0.27803	0.28167	6607
$Inlinks_{Adamic}$	0.23553	0.21011	0.22210	5045	0.25685	0.24151	0.24894	5875	0.28470	0.27854	0.28159	6607
$Outlinks_{CNC}$	0.33301	0.42227	0.37236	10568	0.34840	0.45237	0.39363	10845	0.36049	0.46988	0.40798	11049
$Outlinks_{Adamic}$	0.33230	**0.42473**	**0.37287**	10568	0.34928	**0.45498**	**0.39518**	10845	0.36182	**0.47766**	**0.41175**	11049
DirectMapping	0.35574	0.11457	0.17332	1571	0.36171	0.12032	0.18058	1677	0.36565	0.12337	0.18449	1770
$Entities_{DFS}$	0.48845	0.22548	0.30854	4612	0.48580	0.23545	0.31718	4886	0.48444	0.24076	0.32166	5087
$Entities_{BFS}$	**0.53199**	0.19554	0.28597	2842	**0.53082**	0.20154	0.29215	3068	**0.53368**	0.20823	0.29958	3234

Macro-average scores for the adjusted threshold based models (#PC: number of predictions)

Method	5 days Precision	5 days Recall	5 days F1	5 days #PC	10 days Precision	10 days Recall	10 days F1	10 days #PC	20 days Precision	20 days Recall	20 days F1	20 days #PC
$Inlinks_{CNC}$	0.18634	0.20129	0.19353	7055	0.20312	0.22926	0.21540	8336	0.23301	0.25911	0.24537	9220
$Inlinks_{Adamic}$	0.21763	0.18534	0.20020	6028	0.22731	0.20412	0.21509	7011	0.25686	0.23574	0.24585	7687
$Outlinks_{CNC}$	0.27909	**0.38017**	0.32188	14169	0.28725	0.39046	0.33100	14320	0.28945	**0.41396**	0.34069	14533
$Outlinks_{Adamic}$	0.30896	0.37136	0.33730	13315	0.30479	**0.40219**	0.34678	14153	0.30112	0.40503	0.34543	14282
DirectMapping	0.31113	0.14553	0.19830	3538	0.31941	0.14587	0.20028	3529	0.32077	0.14912	0.20359	3755
$Entities_{DFS}$	**0.49333**	0.27747	**0.35517**	7179	**0.47678**	0.28490	**0.35667**	7509	**0.47549**	0.28420	**0.35576**	8210
$Entities_{BFS}$	0.45635	0.23128	0.30698	5413	0.46679	0.22763	0.30603	5408	0.46420	0.23649	0.31334	5702

Macro-average scores for the machine learning approach (#PC: number of predictions)