EUDETECTOR: Leveraging Language Model to Identify EU-Related News

Koustav Rudra, L3S Research Center, Leibniz University Hannover
Danny Tran, Leibniz University Hannover
Miroslav Shaltev, Leibniz Centre for Tropical Marine Research (ZMT), Bremen

The 1st International Workshop on News Recommendation and Intelligence
April 14, Co-located with The Web Conference 2021

EU-News

EU-News: News that is connected/related to EU and has an influence on EU policies

On 9 June Deputy Prime Minister Krasimir Karakachanov expressed concern, saying that North Macedonia was only using Bulgaria to enter the EU and NATO. [...] On 11 June the Bulgarian side of the commission flatly rejected the Macedonian proposal to honour [...] Gotse Delchev, a historical figure shared by the two countries, on 7 October. [...]

News is about film industry
Mention shooting locations

NONEU-News: Not related to EU or has no impact on EU

[...] With many of the major festivals around the world — London, Ireland, Venice, Telluride, and Toronto — happening during September and October, a well-received launch can play a crucial role in a film's quest for sales, distribution, and awards. [...]
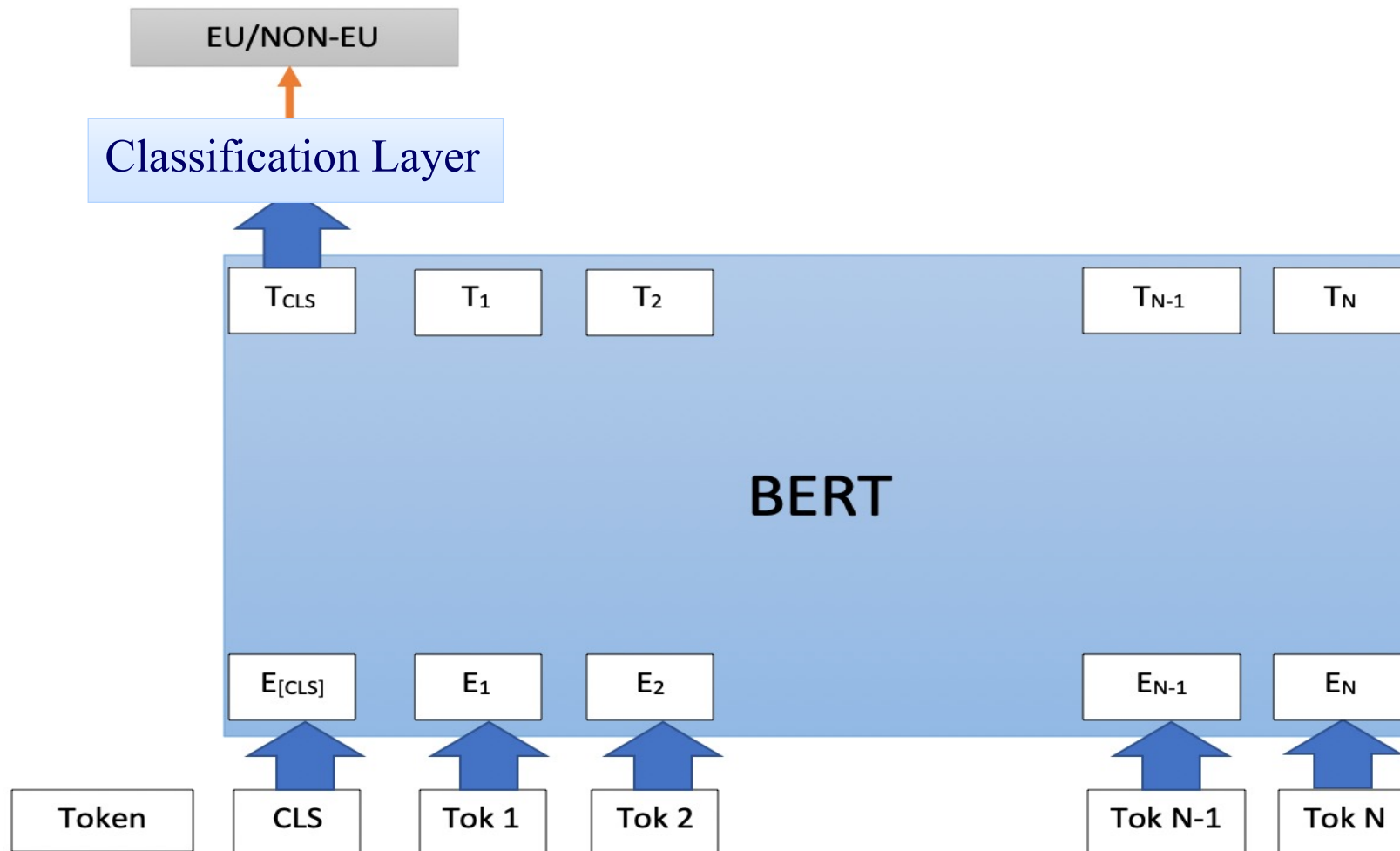
EU-News: Challenges in Detection

- Presence of entities is not enough
- Number of EU entities is necessary not a sufficient condition

- Along with the entities, it is necessary to understand their interaction with each other and surrounding context
- Need to identify whether they have any connection/influence on EU

Dataset

- Crawled news articles over a period of one month (November 2019)
- Collected around 93K news articles
- Annotate 5000 news (kappa = 0.77)
- 1662 EU news and 3338 non-EU news

- Interestingly, 67% of news contain atleast one EU entity
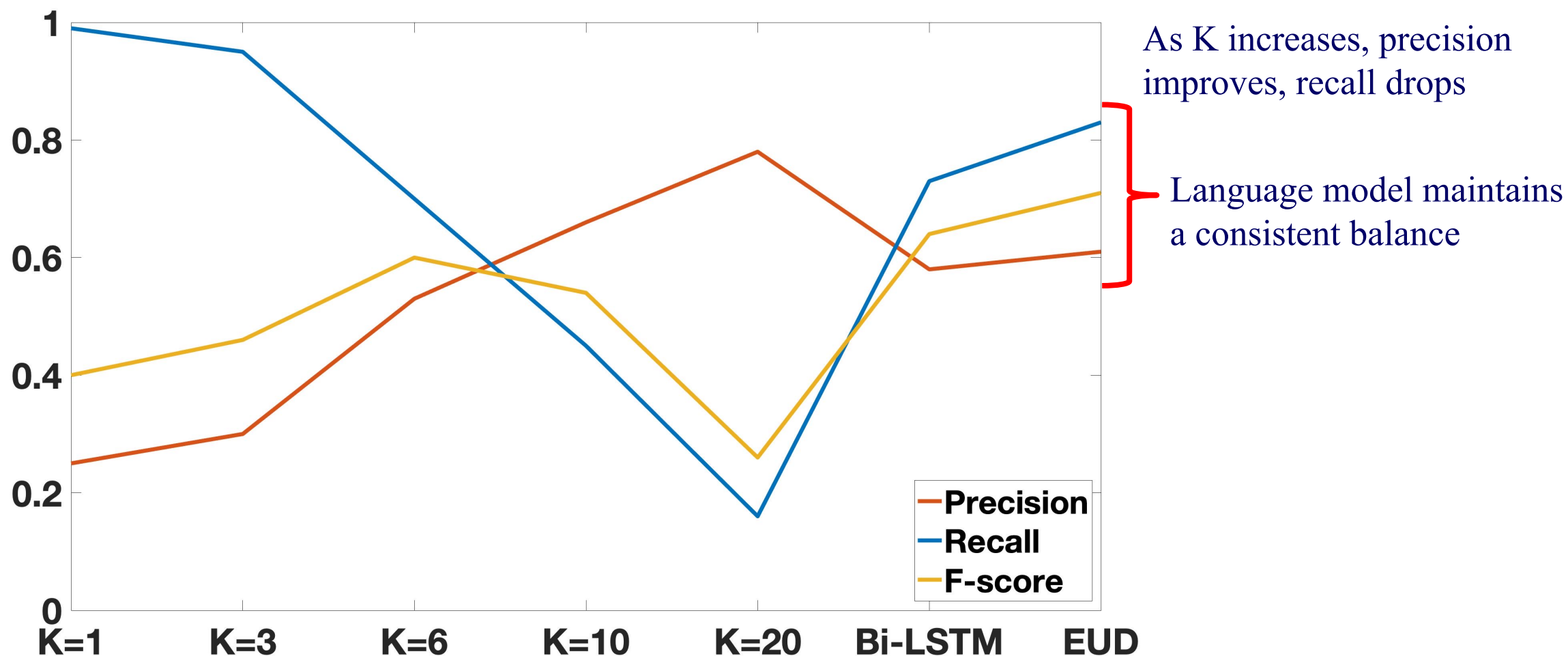
Model: EUDETECTOR

Experiment

Divide train, validation, test set in 60, 20, 20 ratio
Test over 15 different folds

Baselines:

- Cliff-Calvin System:
  - Identify EU entities using Spacy + Wikidata query
  - Mark it as EU event if #EU entities >= K

- Bi-LSTM model
  - Initiate with glove embeddings

EU-News: Performance

As K increases, precision improves, recall drops

Language model maintains a consistent balance

EU-News: Role of Entities

Data Type	Example
Original Text	Angela Merkel is Germany chancellor. She met US pres... Donald Trump in last year G7 summit
Text-COMP	* * is * chancellor. She met * president * * in last year G7 summit
Text-SUFF	Angela Merkel * Germany * * * US * Donald Trump * * *

Mask entities

Mask context

Entities are more important than context

Comprehensiveness = F_{EU}(Original Text) - F_{EU}(Text-COMP)
Sufficiency = F_{EU}(Original Text) - F_{EU}(Text-SUFF)

Comprehensiveness = 0.214
Sufficiency = 0.116

J. DeYoung, S. Jain, N. F. Rajani, E. Lehman, C. Xiong, R. Socher, B. C. Wallace. ERASER: A Benchmark to Evaluate Rationalized NLP Models. ACL 2020

EUDETECTOR: Application

EUDETECTOR: http://eudetector.l3s.uni-hannover.de
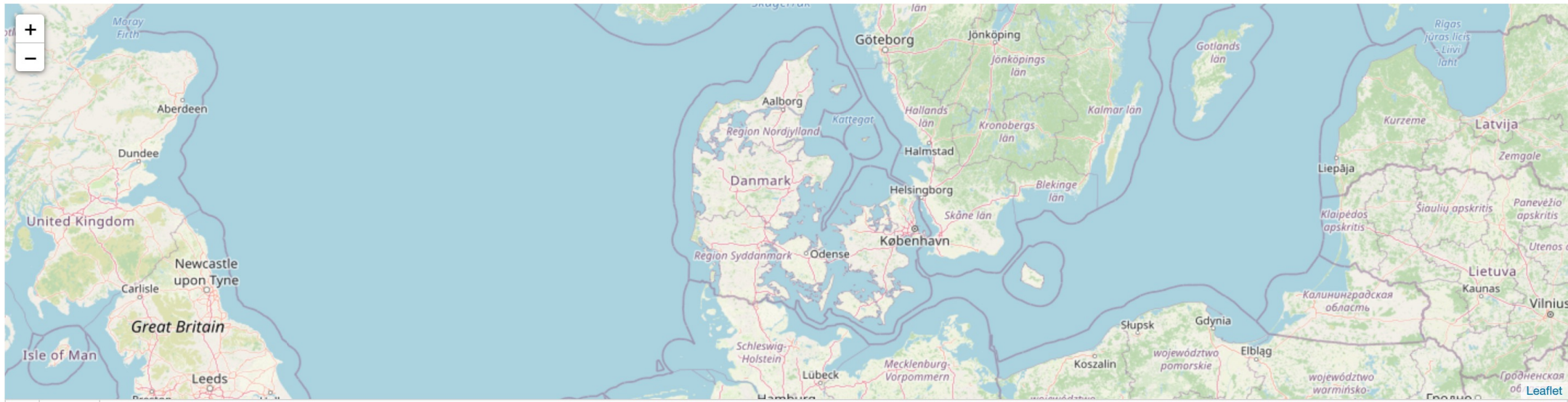
On 11 June the Bulgarian side of the commission flatly rejected the Macedonian proposal to honour Gotse Delchev

or select files

EUDetector   ✕ ▾

Process

Result: /result/?id=7d5ab9a7-fac8-444c-9997-0237219f826f

id	EU	News
0	true	On 11 June the Bulgarian side of the commission flatly rejected the Macedonian proposal to honour Gotse Delchev...

Conclusion

Exploring context is important to identify EU-related news

Mere presence of entities does not ensure EU event

Application may be used as a building block in other systems

KNOWLEDGE · INFORMATION · LEARNING

Thank You

Contact: rudra@l3s.de