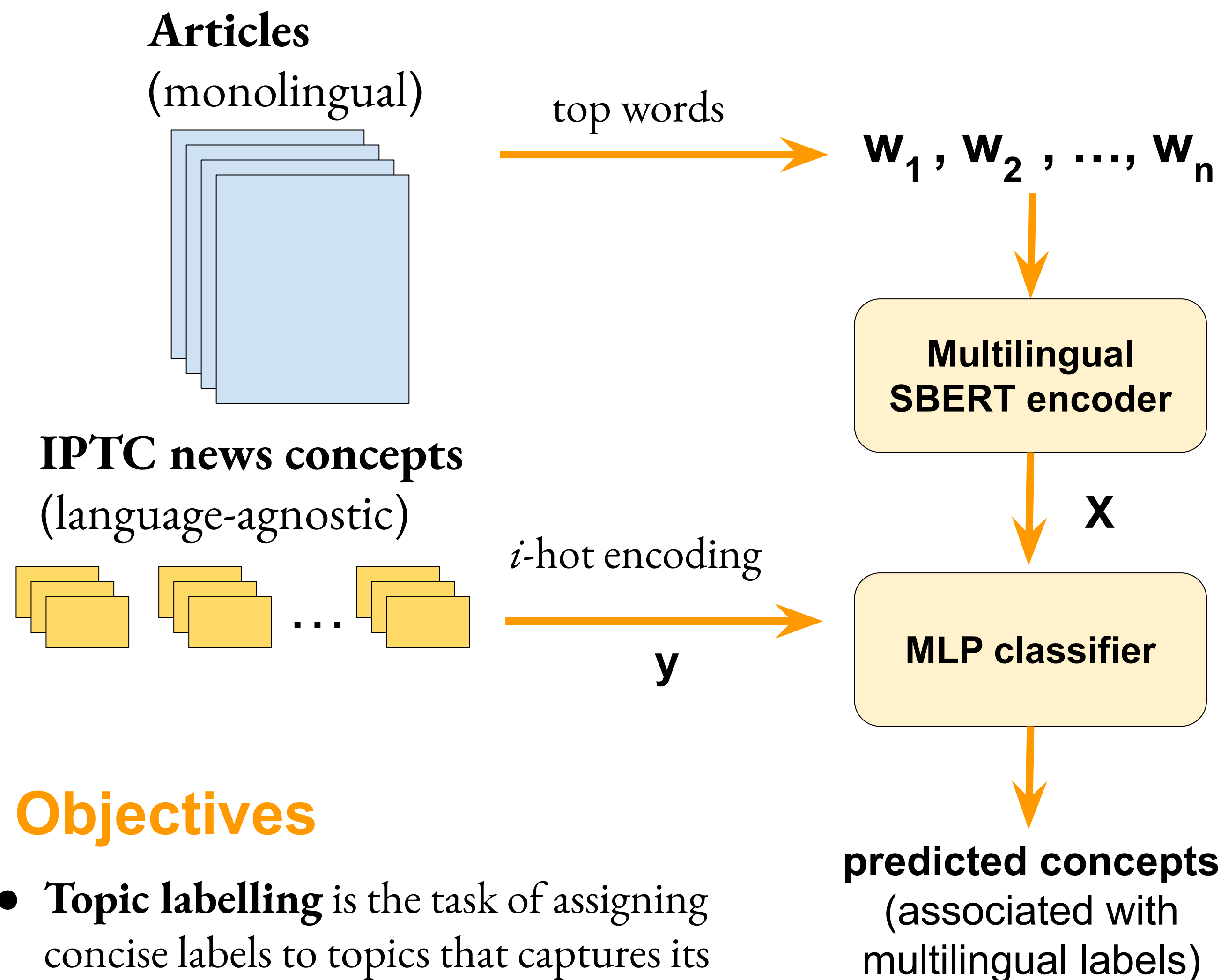




Multilingual Topic Labelling of News Topics using Ontological Mapping

Elaine Zosa, Lidia Pivovarova, Michele Boggia, and Sardana Ivanova
University of Helsinki
{firstname.lastname}@helsinki.fi

Training



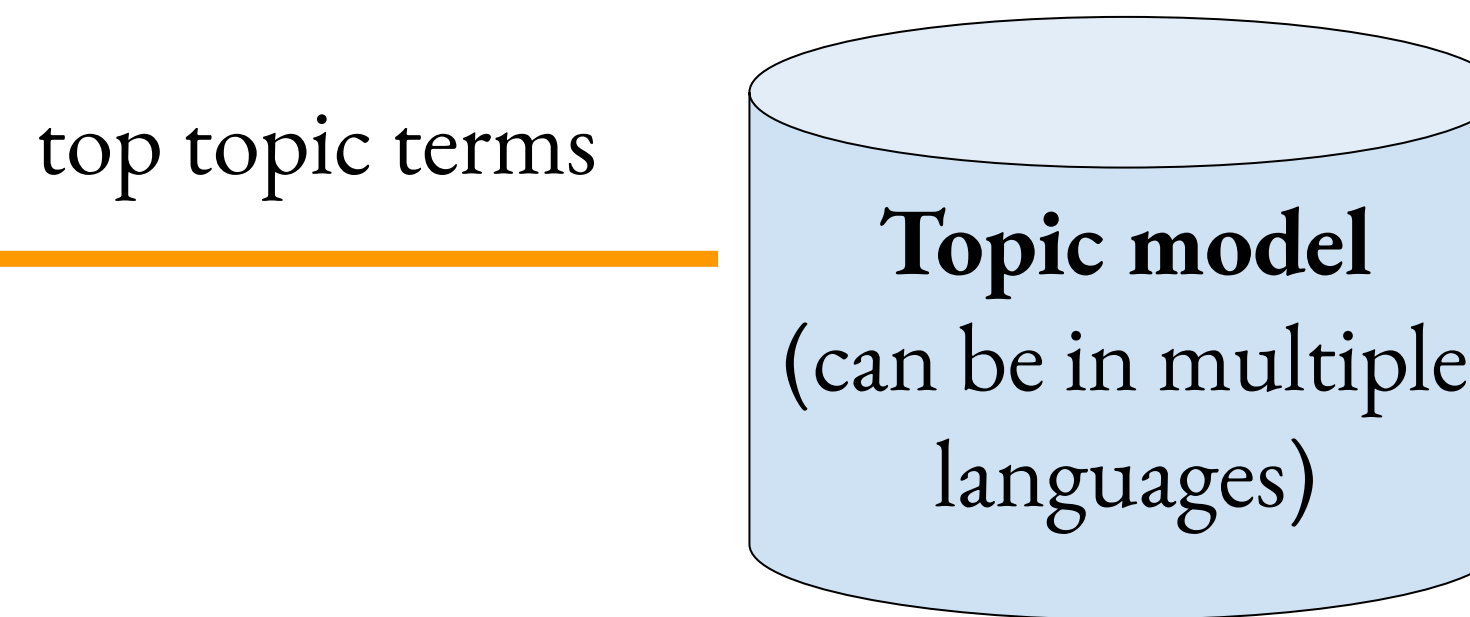
Objectives

- **Topic labelling** is the task of assigning concise labels to topics that captures its semantic content
- **Topic labels** aid topic interpretation and navigation of a corpus
- Multilingual news collections need to be more accessible to a broad range of users
- We need **multilingual** topic labels
- We need topic labelling methods that works on **topics** in **multiple languages**

Our method

- We cast topic labelling as a classification problem
- **Distant supervision**: top article words used as *pseudo topics* for training (Alokaili, et al. 2020)
- **Zero-shot inference**: produces labels for topics in languages not seen during training
- Produces **multilingual labels**
- Labels are on par with state-of-the-art label generation methods (see paper)

Testing/inference



Sample outputs

Topic (Finnish): räikkönen, bottas, ajaa (*to drive*), hamilton, mercedes, ferrari

Predicted labels:

EN: motor racing/FI: autourheilu
EN: formula one/ FI: formula 1

Topic (English): film, movie star, director, hollywood, actor, minute

Predicted labels:

EN: film festival/FI: elokuvajuhlat
EN: cinema industry/FI: elokuvateollisuus

More info!
Code & link to paper



<https://github.com/ezosa/topic-labelling>

References

- Alokaili, A., Aletras, N., Stevenson, M.: Automatic generation of topic labels. (2020)
- Popa, C., Rebedea, T.: BART-TL: Weakly-supervised topic label generation. (2021)
- Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using siamese BERT-networks. (2019)
- Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y.: BERTScore: Evaluating Text Generation with BERT. (2019)

Acknowledgements

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grants 770299 (NewsEye) and 825153 (Embeddia).

