Structured Summarisation of News at Scale

Georgiana Ifrim

University College Dublin, Ireland

Abstract

Facilitating news consumption at scale is still quite challenging. Some research effort focused on coming up with useful structures for facilitating news navigation for humans, but benchmarks and objective evaluation of such structures is not common. One area that has progressed recently is news timeline summarisation. In this talk, we present some of our work on long-range large-scale news timeline summarisation. Timelines present the most important events of a topic linearly in chronological order and are commonly used by news editors to organise long-ranging topics for news consumers. Tools for automatic timeline summarisation can address the cost of manual effort and the infeasibility of manually covering many topics, over long time periods and massive news corpora. In this talk, we first compare different high-level approaches to timeline summarisation, identify the modules and features important for this task, and present new state-of-the-art results with a simple new method. We provide several examples of automatic timelines and present both a quantitative and qualitative analysis of these structured news summaries. Most of our tools and datasets are available online on github.

Short Bio

Dr. Georgiana Ifrim is an Associate Professor at the School of Computer Science, UCD, co-lead of the SFI Centre for Research Training in Machine Learning (ML-Labs) and SFI Funded Investigator with the Insight Centre for Data Analytics and VistaMilk SFI Centre. Dr. Ifrim holds a PhD and MSc in Machine Learning, from Max-Planck Institute for Informatics, Germany, and a BSc in Computer Science, from University of Bucharest, Romania. Her research focuses on effective approaches for large-scale sequence learning, time series classification, and text mining. She has published more than 50 peer-reviewed articles in top-ranked international journals and conferences and regularly holds senior positions in the program committees for IJCAI, AAAI, and ECML-PKDD, as well as being a member of the editorial board of the Machine Learning Journal, Springer.