Generative models of online discussion threads

Pablo Aragón[1,2], **Vicenç Gómez**[1], **Alberto Lumbreras**[3], **and Andreas Kaltenbrunner**[1,4,*]

[1] Universitat Pompeu Fabra
[2] Eurecat, Centre Tecnològic de Catalunya
[3] IRIT, Université de Toulouse, CNRS
[4] NTENT
[*] Corresponding author: kaltenbrunner@gmail.com

ABSTRACT

Online discussion is a core feature of numerous social media platforms and has attracted increasing attention from academia for different and relevant reasons, e.g., the resolution of problems in collaborative editing, question answering and e-learning platforms, the response of online communities to news events, online political and civic participation, etc. Discussions on the Internet commonly occur as a exchange of written messages among two or more participants. These conversations are often represented as threads, which are initiated by a user posting a starting message (a post) and then other users replies to either the post or the earlier replies. Given this sequential posting behavior, online discussion threads follow a tree network structure.

Different modeling approaches have been proposed to identify the governing mechanisms of the network structure of threads. Statistical models of this type are aimed to reproduce the growth of discussion threads through different features, often related to human behavior. This is why they are usually called generative models: they do not only estimate the statistical significance of their corresponding features but also reproduce the temporal arrival patterns of messages that form a discussion thread. The parameters of these models allow to compare different platforms and communities, they even can help to assess the impact of design choices and user interface changes on the way the discussions unfold. Therefore, we aim to provide the participants with state of the art tools and methods for the analysis, diagnosis, management and improvement of online discussion platform and communities.

Outline of the tutorial (3 hours)

- The **theoretical session**[1] will cover the following topics:
  - *Introduction and review of previous work on modelling online discussions.* Historical overview of the different paradigms and platforms for online discussion. Review of previous studies on different online discussion environments like social news, massive open online courses, peer-production, online collaboration, deliberation and citizen participation. We will also review well established results in this topic over the past years including previous work on how social theories like homophily, social influence and emotional contagion help to explain user behaviour in online discussion.
  - *Data-driven modeling of online discussion threads.* Review of state-of-the-art models able to describe the structure and growth process of online discussion threads. This section will emphasize methodological guidelines for correctly addressing statistical questions related to parameter estimation, inference, and model validation as well as limitations of current approaches.
  - *Applications and open research and societal challenges.* A case study of application of generative models for online discussions and presentation and discussion of relevant related open research and societal challenges[1].
- The **practical session**[2] will illustrate the usability of the proposed models on real-world case studies that analyze online discussion threads, e.g., Reddit, Wikipedia, Slashdot, and other relevant platforms. This part will provide the technical skills to understand, use and possibly extend our open software implementation for model learning, inference and validation: https://github.com/alumbreras/discussion-threads

References

1. Aragón, P., Gómez, V., García, D. & Kaltenbrunner, A. Generative models of online discussion threads: state of the art and research challenges. *J. Internet Serv. Appl.* **8**, 15 (2017). DOI 10.1186/s13174-017-0066-z.

2. Lumbreras, A., Jouve, B., Velcin, J. & Guégan, M. Role detection in online forums based on growth models for trees. *Soc. Netw. Analysis Min.* **7**, 49 (2017). DOI 10.1007/s13278-017-0472-z.

This work is supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502).