A Bayesian approach for comparing hypotheses about sequential data and its applications

Florian Lemmerich[1], Philipp Singer, Martin Becker[2], Lisette Espin-Noboa[1], Dimitar Dimitrov[1], Denis Helic[3], Andreas Hotho[2], and Markus Strohmaier[4]

[1] GESIS - Leibniz Institute for the Social Sciences `firstname.lastname@gesis.org`
[2] University of Würzburg
`lastname@informatik.uni-wuerzburg.de`
[3] Graz University of Technology
`dhelic@tugraz.at`
[4] RWTH Aachen
`markus.strohmaier@humtec.rwth-aachen.de`

Abstract. Sequential data are found in many settings, e.g., as sequences of websites that users visit, as sequences of travel locations, or as sequences of songs a user listened to. To improve the understanding of the mechanisms that underlie the generation of such sequences, we developed a novel approach called *HypTrails*. It utilizes Bayesian hypothesis testing and first-order Markov chain models in order to enable the comparison of a set of *hypotheses* with respect to their plausibility considering some observed data. Each of the hypotheses captures a belief in transitions between the states. It can be derived from theory in the application domain, from other related datasets, or from human intuition. We applied this approach to study several phenomena in the online world, e.g., navigation behavior in Wikipedia or urban mobility data. In this talk, we want to give an introduction to HypTrails [4] and showcase selected real-world applications [3, 1, 2] that utilize it.

References

1. Becker, M., Singer, P., Lemmerich, F., Hotho, A., Helic, D., Strohmaier, M.: Photowalking the city: Comparing hypotheses about urban photo trails on Flickr. In: Social Informatics. pp. 227–244 (2015)
2. Dimitrov, D., Singer, P., Lemmerich, F., Strohmaier, M.: What makes a link successful on Wikipedia? In: International Conference on World Wide Web. pp. 917–926 (2017)
3. Espín-Noboa, L., Lemmerich, F., Singer, P., Strohmaier, M.: Discovering and characterizing mobility patterns in urban spaces: A study of Manhattan taxi data. In: Location and the Web (LocWeb) workshop at WWW (2016)
4. Singer, P., Helic, D., Hotho, A., Strohmaier, M.: Hyptrails: A Bayesian approach for comparing hypotheses about human trails on the web. In: International Conference on World Wide Web. pp. 1003–1013 (2015)