IMSM
Institute for Interactive Marketing & Social Media

ANNOUNCEMENT

Bachelor Thesis

KEYWORDS

- Latent Topic Modeling
- Deep Learning
- Music Industry
- Empirical Analysis
- Marketing Analytics

TOPIC: EXPLOITING LATENT TOPICS OF MUSIC SONGS FOR A ROBUST ARTIST CLUSTERING

Entertainment companies (e.g., music labels) face the challenge of finding the right mix of marketing activities for their products (e.g., music albums). The increasing availability of large data sets provides opportunities for companies to better evaluate their marketing effectiveness. Cluster analysis can be applied as a preparatory step to analyze the heterogeneity in marketing effectiveness across different product types and, consequently, to optimize the allocation of marketing expenditures. In addition, cluster analysis can be useful to for targeting purposes (e.g., for recommendation engines on streaming services).

Although music genres are widely used to cluster artists, they are highly subjective and often ambiguous. One reason is that music genres are usually defined subjectively by a person involved in the creation process. To obtain a more objective clustering of artists, machine learning techniques are increasingly used today. However, genre-labels are still often inconsistent because of, for example, different data sources, biased algorithms, and the absence of a standardized set of possible genres. Thus, additional, less ambiguous features are needed to enable a more robust clustering of artists. Assuming that similar artists write songs about similar topics, one possibility would be to exploit the song lyrics to derive the topics of music songs. This approach is promising since most music lyrics are publicity available, texts provide a lot of information and there is a wide range of freely available tools and libraries for the detection of latent topics due to recent progress in text mining techniques (e.g., deep learning and latent topic modeling). Thus, the derived latent topics based on the song lyrics may provide a better foundation for a robust artist clustering than one based on music genres.

As a pre-step for clustering, the aim of this bachelor thesis is to extract latent topics out of music lyrics and test if they can be used to measure artist similarity. To benchmark the results, the open nap music data collection is used (Ellis D.P. et.al., 2002). This dataset contains a similarity measure of 400 different highly-represented artists.

Deep math and data analysis skills are required to write this thesis. Thus, a solid understanding of maths and/or Python/R-programming skills are highly recommended.

LITERATURE:

- **Ellis, D. P., Whitman, B., Berenzweig, A., & Lawrence, S. (2002):** The Quest for Ground Truth in Musical Artist Similarity. In ISMIR.
- **Tian, F., Gao, B., He, D., & Liu, T. Y. (2016):** Sentence Level Recurrent Topic Model: Letting Topics Speak for Themselves. arXiv preprint arXiv:1604.02038.
- **Blei, D. M., & Lafferty, J. D. (2006):** Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (pp. 113-120). ACM.

SUPERVISOR:

- Dr. Nils Wlömert http://www.wu.ac.at/imsm/team/nils-wloemert/en/
- DI Christian Hotz-Behofsits www.wu.ac.at/imsm/jobs/team/christian-hotz-behofsits/

APPLICATION

Applications with CV and transcript of records should be sent to Christian Hotz-Behofsits (christian.hotz-behofsits@wu.ac.at).