

Abstract for GR-TR Conference on Statistical Mechanics and Dynamical Systems

Topic: Other

Preference: Poster

A Probabilistic Latent Semantic Approach to Classification of Literary Texts

Mehmet Can Yavuz*, Muhittin Mungan†

Boğaziçi University, Dpt. of Physics, Bebek 34342 Istanbul, Turkey

* Electronic Address: mcanyavuz@gmail.com

† Electronic Address: mmungan@boun.edu.tr

Probabilistic Latent Semantic Analysis (pLSA)[1][2] is an unsupervised learning algorithm to mine keywords or main concepts in large data sets. The algorithm returns mixture probabilities based on a latent class model which can be used for classifications.[3]

In this poster, I present a pLSA approach to the classification and analysis of literary texts in terms of differentiations and similarities by genres, by eras and by authors.

-
- [1] Thomas Hoffman, *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval (SIGIR-99)*, (1999).
 - [2] Scott Deerwester, Susan Dumais, George Furnas, Richard Harshman, Thomas Landauer, Karen Lochbaum, Lynn Streeter *Journal of American Society for Information Science(1986-1998)*, **41,6**, *ABI/INFORM Global* pg. 391 (1990).
 - [3] Thomas Hoffman, *Advances in Neural Information Processing Systems*, **12**, pp-914-920, *MIT Press*, (2000).