

From Data to Knowledge

Prediction and Analysis of Cultural Heritage with Web Data Augmentation

This topic proposal for a Master's thesis focuses on text analytics and prediction with deep learning.

Problem

- Noisy corpus with inconsistent, missing data entries.
- Patterns and interesting relationships between the different variables in the data are hard to recognize.
- Large volume of data requires smart pre-processing, inferencing as well as automated approaches for mining information.

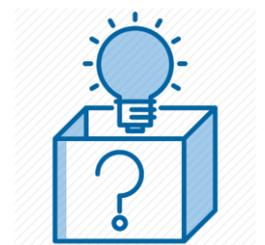
Approach

- Linking and mapping of entities with Web entities (named entity disambiguation).
- Data mining techniques to discover interesting patterns and relationships in the data.
- Predictive analytics with help of Web entities to intelligently fill the missing gaps in data.

About the Project

To extract useful information from unstructured and noisy data, intelligent text analytic techniques are essential. This includes text pre-processing, inferencing, data transformation as well as visualization. Missing or irregular data entries pose further challenges in standard data processing tasks. The data already present on the Semantic Web can prove useful for the mapping and canonicalization of important entities in order to mine information from unstructured data. With the augmentation of meta data obtained from the Web, neural network based prediction techniques can be leveraged for estimating missing entries with a quantifying confidence metric. Furthermore, deep learning based data mining techniques can help discover interesting patterns and relationships in the data – both in semi-supervised and unsupervised fashion.

This Master's thesis is based on the broad theme of data mining and predictive analytics to identify useful information hidden in a large text corpus. Prior experience and interest in deep learning and semantic web techniques would be useful.



Contact

If you are interested in this topic, I would be happy to discuss further, the exact scope for this thesis can be defined as per the expertise and interests of the student. Feel free to stop by at my office F-2.07 or send me an e-mail: Nitisha.Jain@hpi.de.

