A Multi-Armed Bandit Approach for Batch Mode Active Learning on Information Networks

• Joint work with Doris, submitted to KDD 2016
• Active learning on heterogeneous information networks via expected entropy reduction
• Centrality by type as primary learners
• Batch mode learning via combinatorial multi-armed bandit
• Adaptive to different classification tasks on the same network structures
• Independent of underlying classification model
• Performance lower-bounded by random
In-progress: Mining Quality Phrases from Massive Text Corpora in Multiple Languages

- Joint work with Jingbo, Jialu, and Xiang
- Automatic Language Detection (47 languages supported)
- Constructing multi-language knowledge base from Wikipedia Entities and Synonyms
- Multi-language punctuations mapping
- Multi-language tokenizer based on Lucene (10 languages supported, can be extended to more: English, French, German, Spanish, Italian, Portuguese, Russian, Arabic, Chinese, Japanese)
Thanks