09_11_2022

Selection process (Inclusion/Exclusion criteria) (Further refined during the selection process)

In order to be selected for this review, a paper should propose or actively apply, either with a primary or secondary focus a topic labeling technique.
Papers appearing in the selected research do not necessarily need to describe the implementation of a novel labeling approach, but it is important that they do not meet any of the following **exclusion criteria**:

- The paper does not actively apply any topic labeling techniques
- The labels do not posses descriptive properties with regards to the specifics of the topics content
  - Tang et al., 2019, associates (binary) sentiment labels to topics
  - Bahrainian et al., 2018, extracts topics from a corpus of 28 years of scholarly articles divided into one year time slices. A binary label (continued/not continued) is assigned to each topic to indicate if it's covered in the subsequent time slice.
  - Figueiredo & Jorge, 2019, use a SVM classifier to assign binary labels to LDA topics in order to classify tweets as either "relevant" or "irrelevant".
- All the described labeling approaches are taken from existing work and re-proposed as-is (on the same corpus and set of topics)
- The paper and/or the analysed corpus do not match the imposed language restrictions
- The paper is a systematic review (secondary/tertiary study)

Selection process

Paper selection (in progress)

Decision Support Systems

- A Text Analytics Approach for Online Retailing Service Improvement: Evidence from Twitter
- Data-driven decision-making in credit risk management: The information value of analyst reports
- Sourcing product innovation intelligence from online reviews
- How do consumers in the sharing economy value sharing? Evidence from online reviews

Expert Systems with Applications

- Document-based topic coherence measures for news media text
- Providing recommendations for communities of learners in MOOCs ecosystems
- Recent trends in mathematical expressions recognition: An LDA-based analysis
- Topic2Labels: A framework to annotate and classify the social media data through LDA topics and deep learning models for crisis response
- Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints
- Social media analysis by innovative hybrid algorithms with label propagation
- The climate change Twitter dataset
- Supporting digital content marketing and messaging through topic modelling and decision trees
- Large scale analysis of open MOOC reviews to support learners' course selection
- Criteria determination of analytic hierarchy process using a topic model
- Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews
- W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis

Information Sciences

- A weakly-supervised graph-based joint sentiment topic model for multi-topic sentiment analysis
- Identifying impact of intrinsic factors on topic preferences in online social media: A nonparametric hierarchical Bayesian approach
- Author topic model for co-occurring normal documents and short texts to explore individual user preferences

Journal of Infometrics

- Application of machine learning techniques to assess the trends and alignment of the funded research output
- Developing a topic-driven method for interdisciplinarity analysis
- Improving fitness: Mapping research priorities against societal needs on obesity
- Is it all bafflegab? – Linguistic and meta characteristics of research articles in prestigious economics journals
- Topic-linked innovation paths in science and technology
- Does deep learning help topic extraction? A kernel k-means clustering method with word embedding
- Exploring scientific trajectories of a large-scale dataset using topic-integrated path extraction

Knowledge-Based Systems

- Experimental explorations on short text topic mining between LDA and NMF based

Schemes
- Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016
- Identifying topical influencers on twitter based on user behavior and network topology
- Influence Factorization for identifying authorities in Twitter
- Learning document representation via topic-enhanced LSTM model

ACL

- PhraseCTM: Correlated Topic Modeling on Phrases within Markov Random Fields
- Spatial Aggregation Facilitates Discovery of Spatial Topics
- Generating Summaries with Topic Templates and Structured Convolutional Decoders (*)
- Neural Mixed Counting Models for Dispersed Topic Discovery
- An Unsupervised Neural Attention Model for Aspect Extraction (*)

CIKM

- One Rating to Rule Them All? Evidence of Multidimensionality in Human Assessment of Topic Labeling Quality
- ConCET: Entity-Aware Topic Classification for Open-Domain Conversational Agents

COLING

- Twitter Topic Classification (*)
- Community Topic: Topic model inference by consecutive word community discovery
- Model-Free Context-Aware Word Composition
- Mining Crowdsourcing Problems from Discussion Forums of Workers

EACL

- Multimodal Topic Labelling
- BART-TL: Weakly-Supervised Topic Label Generation
- Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews

ECIR

- Multilingual Topic Labelling of News Topics Using Ontological Mapping
- Labeling Topics with Images Using a Neural Network (*)

ECML PKDD

- Survival Factorization on Diffusion Networks

KDD

- TaxoGen: Unsupervised Topic Taxonomy Construction by Adaptive Term Embedding and Clustering (*)
- Hierarchical Topic Mining via Joint Spherical Tree and Text Embedding (*)

EMNLP

- Neural Topic Modeling with Cycle-Consistent Adversarial Training
- Condolence and Empathy in Online Communities
- Exophoric Pronoun Resolution in Dialogues with Topic Regularization (*)
- Phrase-BERT: Improved Phrase Embeddings from BERT with an Application to Corpus Exploration (*)

Cases requiring further evaluation

Pre-existing labels (supervised topic model)

The set of available labels is provided by the dataset used to build the model. Generally, the topic model learns to generate the topics together with an associated label.
In this context, selection of papers making use of supervised topic modelling techniques will require an analysis on the expressiveness of the labels.

- A topic-sensitive trust evaluation approach for users in online communities (Knowledge-based systems)
  - Stack Exchange Q&A dataset with 174 distinct topics
- A Semi-discriminative Approach for Sub-sentence Level Topic Classification on a Small Dataset
  - Dataset of 3076 reviews. Each sentence is labeled with one out of 17 predefined tags
- Automatic Phenotyping by a Seed-guided Topic Model
  - 3.9 million clinical documents across 19 age groups. labels for 11 phenotypes
- Neural Models for Documents with Metadata
  - Datasets: 20 newsgroups, IMDB corpus of 50,000 movie reviews, and the UIUC Yahoo answers dataset with 150,000 documents in 15 categories
- SCDV : Sparse Composite Document Vectors using soft clustering over distributional representations
  - multi-class experiments on 20NewsGroup dataset and multi-label classification experiments on Reuters-21578 dataset
- Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training (*)
  - 9 classes for product reviews and 12 classed for restaurant reviews

- Adapting Topic Models using Lexical Associations with Tree Priors (*)
  - Amazon reviews and 20NewsGroups
- Topically Driven Neural Language Model
  - 20NEWS dataset
- Emotional Contagion-Based Social Sentiment Mining in Social Networks by Introducing Network Communities
  - Reddit dataset containing the threads of four sub-forums (i.e. Movies, Politics, Science and Olympic)
- Analyzing Bayesian Crosslingual Transfer in Topic Models
  - Only three labels (culture, technology, and education)
- Multi-source Neural Topic Modeling in Multi-view Embedding Spaces
  - Four short-text corpora (20NSshort, TMNtitle, R21578title and Ohsumedtitle), one small corpus (20NSsmall) and two large corpora (TMN and Ohsumed)
- Seed-Guided Topic Discovery with Out-of-Vocabulary Seeds
  - SciDocs (11 category names), (10 categories and 100K reviews), Twitter (9 categories)
- Inductive Topic Variational Graph Auto-Encoder for Text Classification
  - 20NewsGroups, Ohsumed, R52 and R8, and MR

Clustering methods

- Improving Deep Embedded Clustering via Learning Cluster-level Representations

Hierarchical methods

- TaxoGen: Unsupervised Topic Taxonomy Construction by Adaptive Term Embedding and Clustering
- Hierarchical Topic Mining via Joint Spherical Tree and Text Embedding
- CoRel: Seed-Guided Topical Taxonomy Construction by Concept Learning and Relation Transferring

Image labels

- Labeling Topics with Images Using a Neural Network

Descriptive topic labels applied directly to corpus content

- CTM - A Model for Large-Scale Multi-View Tweet Topic Classification
  - Individual labels applied to tweets

Next steps

- Finish the paper selection procedure
  - Evaluate the edge cases (starting from papers making use of supervised topic modelling techniques)
- Extract references from the selected papers and use the results to:
  - Perform snowballing
  - Build a graph using Pajek
- Depending on the size of the final selection (after snowballing)
  - Increase/Decrease the time-frame / selected venues

#Thesis/Temporary notes#