The University of Nottingham
Malaysia Campus

School of Computer Science

Intelligent Decision Support Systems

Semantic Web and Information Retrieval Research

Automated Learning Terminological Ontologies Problem

- Concept hierarchies are widely (e.g., in digital libraries) for purposes of navigation, browsing, query suggestion and document retrieval.
- They are traditionally designed and maintained manually.
- Manual approach is time-consuming and prone to obsolescence.

Objective

- To learn concept hierarchy and terminological ontology from unstructured text.
- Plausibility of learning is based on the assumption "given sufficient large amount of text in a domain, coverage of knowledge in that domain can be ensured".

Ontology Categorisation

- Formal ontology
- Prototype-based ontology
- Terminological ontology

Existing Methods for Learning Ontologies from Text

- Lexico-syntatic based Approach
- Information Extraction
- Clustering and Classification
- Data Co-occurrence Analysis

Concept Representation

- Concepts are extracted from corpus and represented as documents using words in those documents that are annotated using the concepts.
- Computation of relations between concepts is transformed into computation of relations between representing documents.
- Computation is based on probabilistic topic models.

Dataset

- Crawler and scraper for Web pages collected from digital libraries.
- Tokenising, stopwords removing, POS tagging, stemming and indexing.
- Concept extraction and representation.

Training pLSA and LDA

- pLSA uses Expectation-Maximisation algorithm.
- LDA uses Gibbs sampling.
- 30-90 topics are experimented for training.
- Concepts represented using documents are folded-in learned topic models using same algorithms.

Ontology Learning Tasks

- Terms
- Synonyms
- Concepts
- Concept Hierarchies
- Relations
- Rules

Learning terminological ontology can be decomposed into **learning concepts** and **relations**.

Learning Relations in SKOS Model

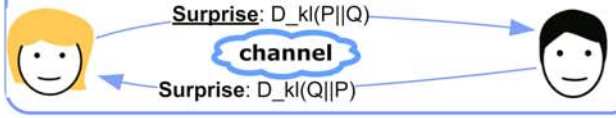
- "Broader" (approximately equivalent to subsumption) and "related".
- "Information Theory Principle for Concept Relationship"

Definition

Information Theory Principle for Concept Relationship: A concept Cp is broader than another concept Cq if the following two conditions hold:

1. (Similarity condition) the similarity measure between them is greater than certain threshold, and
2. (Divergence difference condition) the difference between Kullback-Leibler divergence measures.

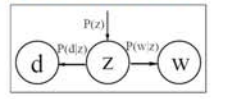
$$D_{KL}(P||Q) - D_{KL}(Q||P) < 0$$



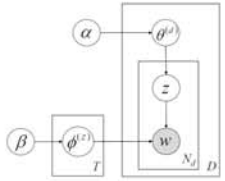
pLSA and LDA Models

- Probabilistic extensions of the Latent Semantic Analysis model.
- Originated from Information Retrieval for document modelling.
- Semantic associations between words and documents are captured by probabilistic topics.

Probabilistic Topic Models



pLSA



LDA

Ontology Learning Algorithms

- Concepts are extracted from corpus and represented as documents using words in those documents that are annotated using the concepts.
- Computation of relations between concepts is transformed into computation of relations between representing documents.
- Computation is based on probabilistic topic models.

Algorithm 1: LSHL

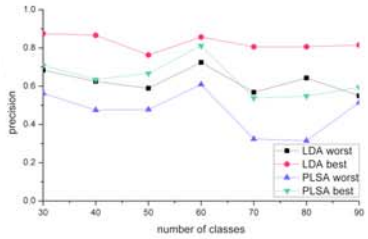
- Local Similarity Hierarchy Learning.
- Learns concept hierarchies.
- Local greedy search.

Algorithm 2: GSHL

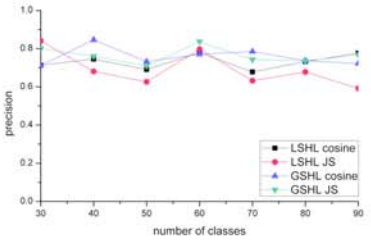
- Global Similarity Hierarchy Learning.
- Able to learn terminological ontologies.
- "broader" and "related" relations.

Experiment and Evaluation

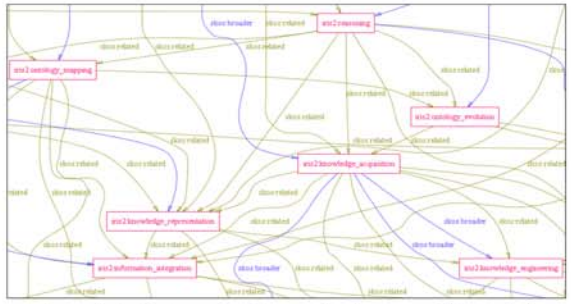
- 672 sets of ontology statements are generated and evaluated by domain experts.
- In almost all of the cases precision of ontology using LDA is better than pLSA.
- The best precision using LDA is 86.6% and the worst is 58%. The best precision using pLSA is 80%, and the worst is 39%.
- The possible reason is the generalisability of LDA to new documents.



Precision comparison



Precision using LDA



A snapshot of the learned ontology centered on concept "Reasoning"

Wang Wei, Payam M. Barnaghi, Andrzej Bargiela
{eyx6ww,payam.barnaghi,Andrzej.Bargiela}@nottingham.edu.my