Learning Relations from Social Tagging Data

PRICAI 2018, Nanjing, China, 28-31 August 2018

Hang Dong, Wei Wang, Frans Coenen

Department of Computer Science,

University of Liverpool

Xi'an Jiaotong-Liverpool University

Motivation – Organising social tags semantically

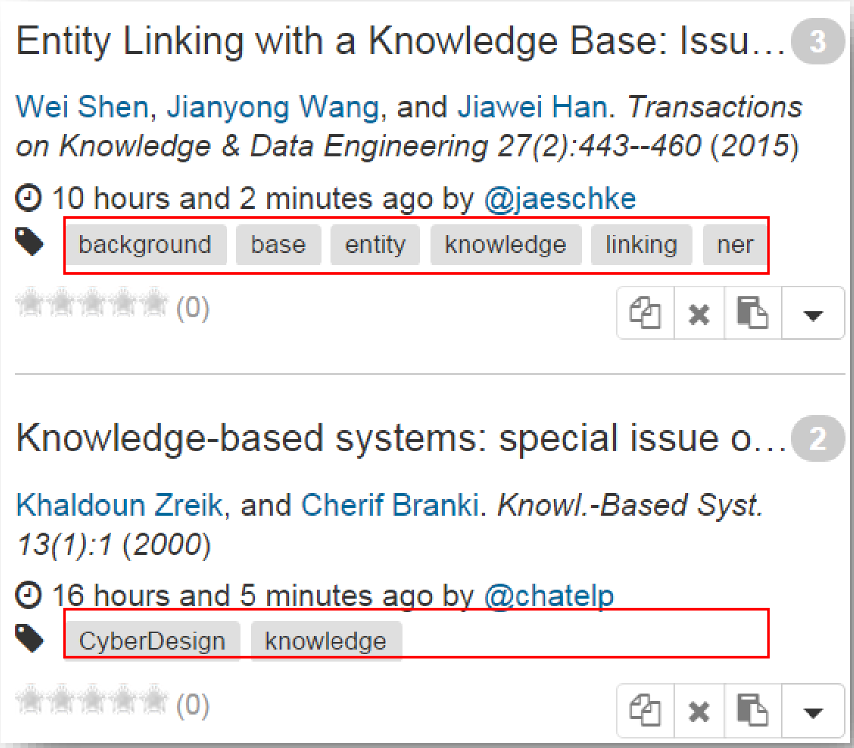
- Social tagging: Users sharing resources – create “keyword” descriptions – terminology of a social group / a domain

- “Folksonomy [social tags] is the result of personal free tagging of pages and objects for one’s own retrieval” (Thomas Vander Wal, 2007)

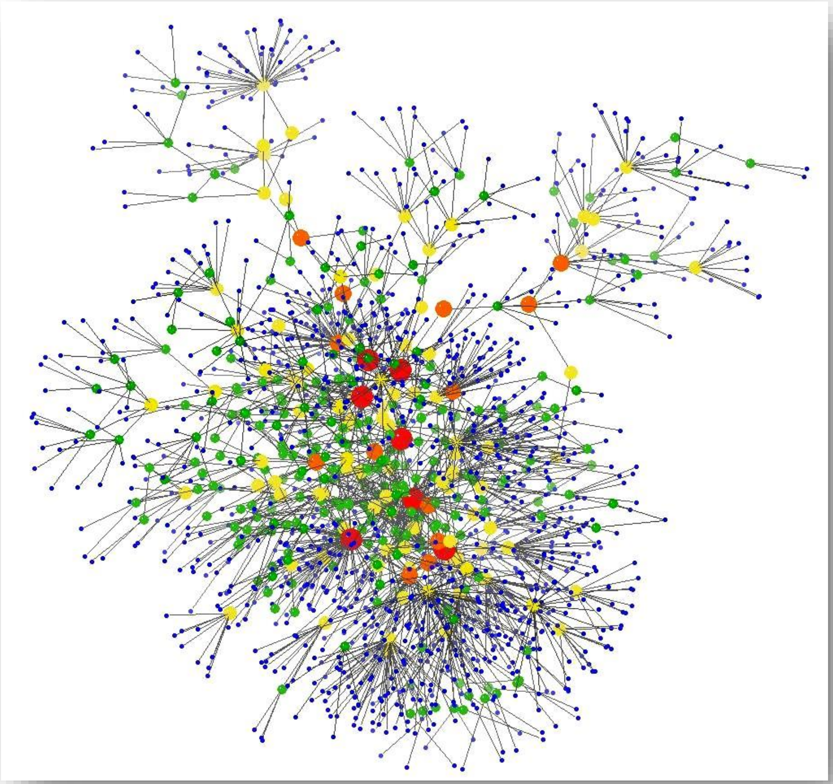
- Issues:
  - Plain (no relations among tags)
  - Noisy and ambiguous, thus not useful to support information retrieval and recommendation.

Social tags for movie “Forrest Gump” in MovieLens
https://movielens.org/movies/356

Research aim: from social tagging data to knowledge



http://www.bibsonomy.org/tag/knowledge



http://www.micheltriana.com/blog/2012/01/20/ontology-what

Researcher generated data (user-tag-resource)

Useful knowledge structures, i.e. broader-narrower relations and hierarchies

Challenges

- Not enough contextual information.
  - The effective pattern-based approaches (Hearst, 1992; Wu *et al.*, 2012) are not applicable.

- Sparse: low frequency.
  - The co-occurrence based approaches (see review in García-Silva *et al.*, 2012, Dong *et al.*, 2015, and the features used in Rêgo *et al.*, 2015) are not suitable.

- Tags are ambiguous, noisy.
  - Data cleaning (Dong et al., 2017).
  - Many tags do not match to lexical resources as WordNet or Wikipedia (Andrews & Pane, 2013; Chen, Feng & Liu, 2014).

We need special data representation techniques to characterise the complex meaning of tags.

Types and issues of current methods

- **Heuristics (Co-occurrence) based methods** (set inclusion, graph centrality and association rule) are based on co-occurrence, does not formally define semantic relations (García-Silva *et al.*, 2012).

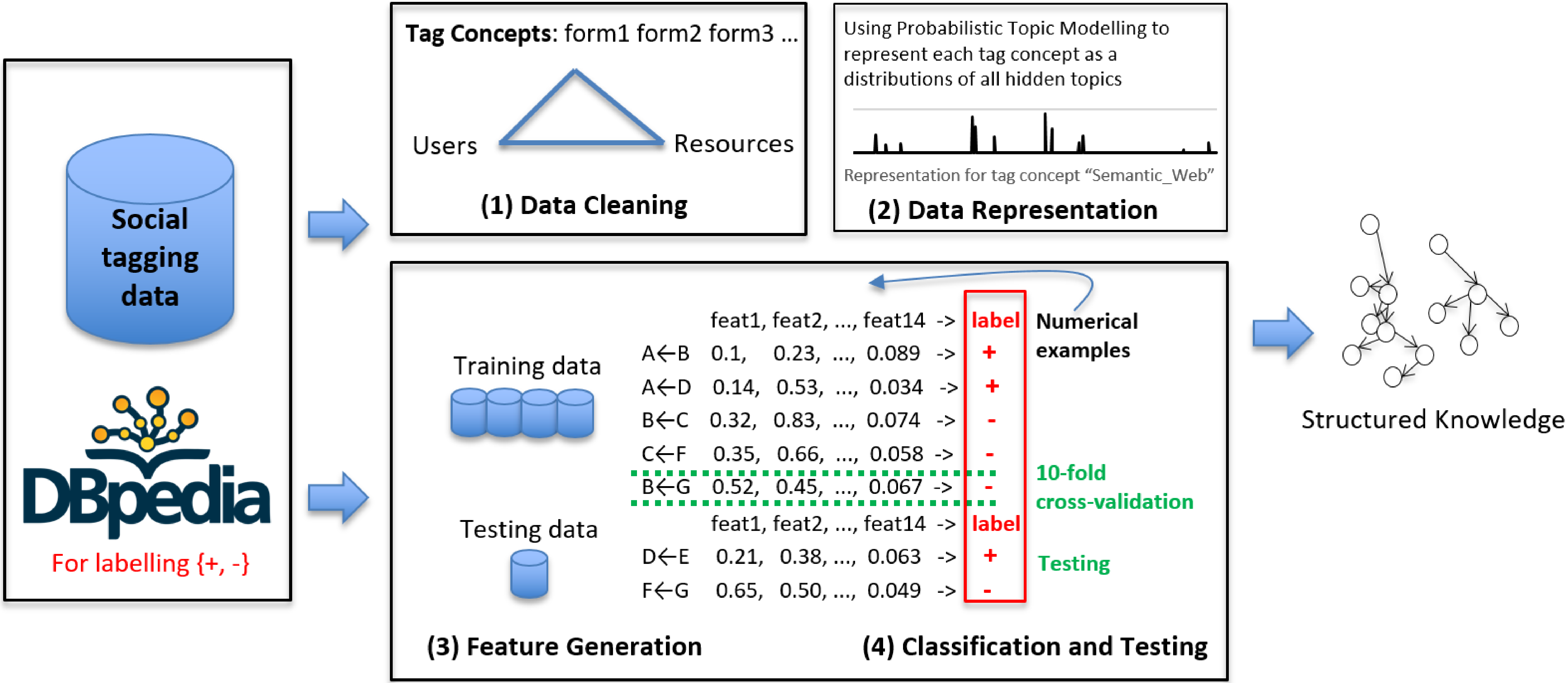
- **Semantic grounding methods** (matching tags to lexical resources) suffer from the low coverage of words and senses in the relatively static lexical resources (Andrews & Pane, 2013; Chen, Feng & Liu, 2014).

- **Machine learning methods**: (i) unsupervised methods could not discriminate among broader-narrower, related and parallel relations (Zhou et al., 2007); (ii) **supervised methods** so far based on data co-occurrence features (Rêgo *et al.*, 2015).

- We proposed a new supervised method, binary classification founded on a set of assumptions using probabilistic topic models.

Supervised learning based on Probabilistic Topic Modeling

Binary classification: input two tag concepts, output whether they have a subsumption (broader-narrower) relation. There are 14 features.



Data Representation

- **Probabilistic Topic Modelling**, Latent Dirichlet Allocation (Blei et al, 2003), to infer the hidden topics in the "Bag-of-Tags". Then we represented each tag as a probability distribution on the hidden topics.
- Input: Bag-of-tags (resources) as documents
- Output: p(word | topic), p(topic | document)

$$v(C_a) = \{p(\mathbf{z}_i|C_a)\}_{i=1}^{\|\mathbf{z}\|} \quad (1)$$

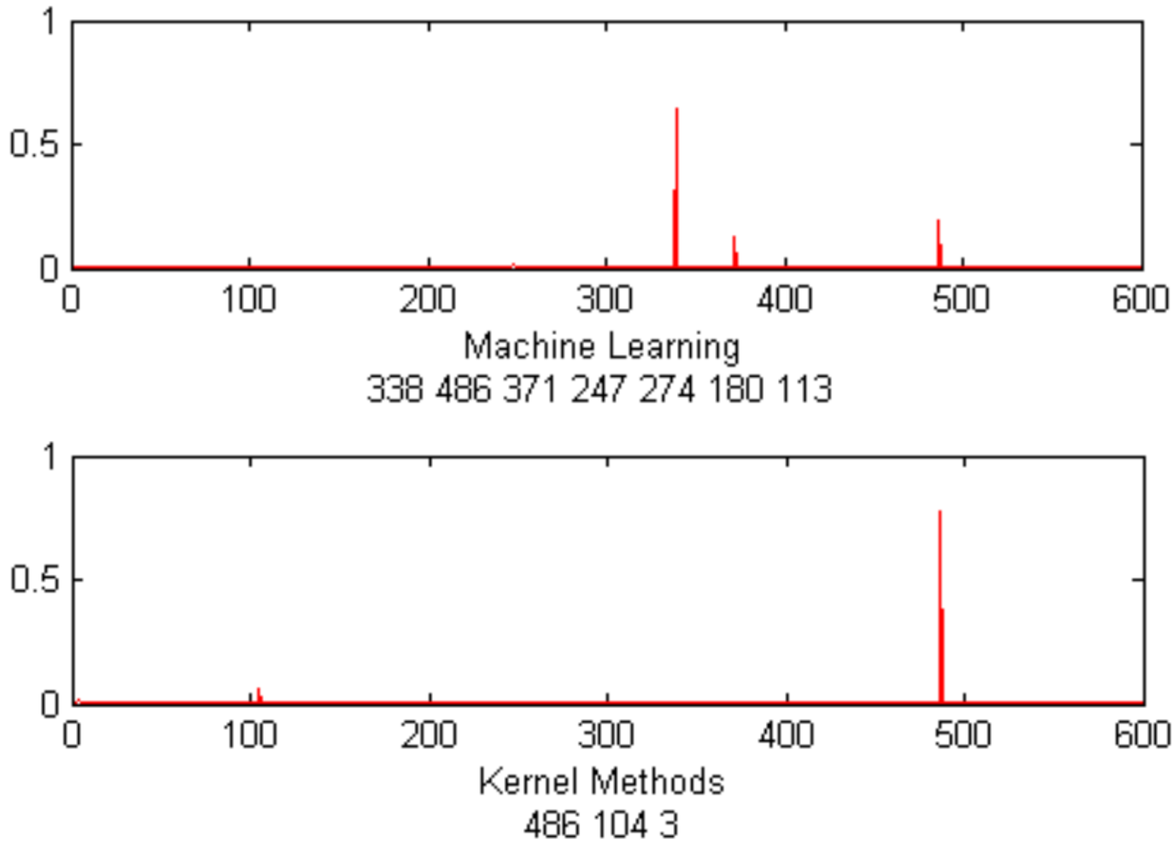
$$p(z|C_a) \propto p(C_a|z) * p(z) \quad (2)$$

$$p(z) = \frac{N_z}{N} \quad (3)$$

,where C is a tag concept, z is a topic and N is the occurrences.

TABLE
TAG TOPICS LEARNED USING LATENT DIRICHLET ALLOCATION (LDA) ($T = 600$, ALPHA $= 50/600$, BETA $= 0.01$)

Topic	Most probable 5 tag concepts
62	search web web_search semantic_search social_search
154	cell calcium membrane channel animal
159	language perception speech tone production
231	game game_theory learning theory haifa_games_course
369	child male female cerebral human



Assumptions and Feature Generation

- Assumptions: **Topic similarity**, **Topic distribution**, **Probabilistic association**
- Assumption 1 (Topic Similarity) Tags that have a broader-narrower relations must be similar to each other to some extent (Wang et al., 2009).

TABLE II
SIMILARITY AND DIVERGENCE RELATED FEATURES

Features	Description
Cos_sim	The cosine similarity of two topic distribution vectors
KL_Div1	The Kullback-Leibler Divergence from C_a to C_b
KL_Div2	The Kullback-Leibler Divergence from C_b to C_a
Gen_Jaccard	The generalised Jaccard Index of two topic distribution vectors

For the generalised Jaccard Index,

$$\mathbf{J}(\mathbf{v}(\mathbf{C_a}), \mathbf{v}(\mathbf{C_b})) = \frac{\sum_i \min(v(C_a)_i, v(C_b)_i)}{\sum_i \max(v(C_a)_i, v(C_b)_i))} \qquad (5)$$

- Assumption 2 (Topic Distribution): A broader concept should have a topic distribution spanning over more dimensions; while the narrower concept should span over less dimensions within those of the broader concept.

TABLE III
TOPIC DISTRIBUTION RELATED FEATURES

Features	Description
diff_num_sig	Difference of the number of significant topics
overlapping	Number of overlapping significant topics
diff_max	Difference of the maximum elements in two tag vectors
diff_aver_sig	Difference of the average probability of significant topics

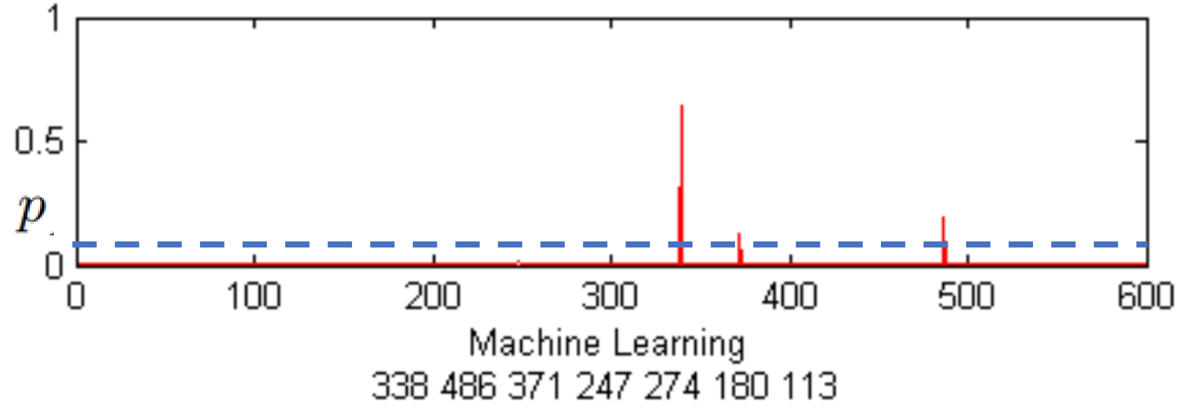
$$\mathbf{z}_a^{sig} = \{z \mid z \in \mathbf{z} \text{ and } p(z|C_a) \geqslant p\} \quad (4)$$

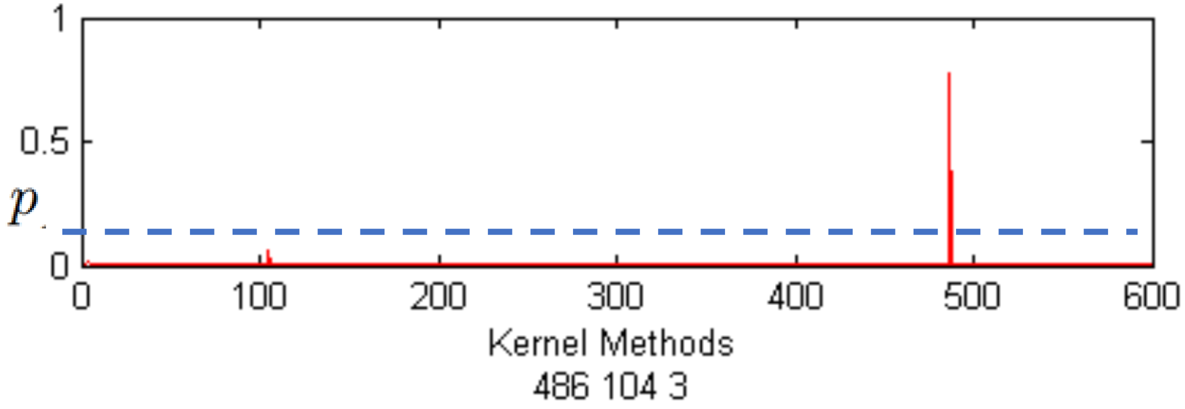
$$\text{diff\_aver\_sig}(C_a, C_b) = \text{Aver}(\mathbf{z}_a^{sig}) - \text{Aver}(\mathbf{z}_b^{sig}) = \frac{\sum(\mathbf{z}_a^{sig})}{\|\mathbf{z}_a^{sig}\|} - \frac{\sum(\mathbf{z}_b^{sig})}{\|\mathbf{z}_b^{sig}\|} \quad (6)$$

$\mathbf{z}_a^{sig}$ is the significant topic set for the concept C_a.

$\mathbf{z}$ is the whole topic set.

p is a probability threshold, 0.1 here.





- Assumption 3 (Probabilistic Association) For two tag concepts having a broader-narrower relation, they should have a strong association with each other, modelled with conditional and joint probability.

TABLE IV
PROBABILISTIC ASSOCIATION FEATURES

Features	Description
$p(C_a \mid C_b)$	The probabilistic association of C_a given C_b
$p(C_b \mid C_a)$	The probabilistic association of C_b given C_a
$p(C_a, C_b)$	The joint probability of C_a and C_b
$p(C_a \mid C_b, R_{a,b})$	The probabilistic association of C_a given C_b and the common root concept $R_{a,b}$
$p(C_b \mid C_a, R_{a,b})$	The probabilistic association of C_b given C_a and the common root concept $R_{a,b}$
$p(C_a, C_b \mid R_{a,b})$	The joint probability of C_a and C_b given the common root concept $R_{a,b}$

$$\begin{aligned} p(C_a|C_b) &= \sum_{z \in \mathbf{z}} p(C_a|z, C_b)p(z|C_b) \\ &= \sum_{z \in \mathbf{z}} p(C_a|z)p(z|C_b) \end{aligned} \quad (7)$$

$$\begin{aligned} p(C_a, C_b) &= p(C_a|C_b)p(C_b) \\ &= p(C_a|C_b) \sum_{z \in \mathbf{z}} p(C_b|z)p(z) \end{aligned} \quad (9)$$

$$\begin{aligned} p(C_a|C_b, R_{a,b}) &= \sum_{z \in \mathbf{z}} p(C_a|z, C_b, R_{a,b})p(z|C_b, R_{a,b}) \\ &= \sum_{z \in \mathbf{z}} p(C_a|z)p(z|C_b, R_{a,b}) \\ &= \sum_{z \in \mathbf{z}} p(C_a|z) \frac{p(C_b, R_{a,b}|z)p(z)}{p(C_b, R_{a,b})} \\ &= \sum_{z \in \mathbf{z}} \frac{p(C_a|z)p(C_b|z)p(R_{a,b}|z)p(z)}{p(C_b, R_{a,b})} \end{aligned} \quad (8)$$

Data preparation

- Social tagging data - Full BibSonomy data from 2005 to July 2015.
  - Concepts extracted through grouping the tag variants together (Dong *et al.*, 2017).
  - Removed resources with less than 3 tag tokens: 7,458 tag concepts and 128,782 publication resources.

- Semantic source - DBpedia "2015-10" for instance labelling.
  - Selected 6 categories,
    - Ml (Category:Machine learning),
    - SW (Category:Semantic Web),
    - Sip (Category:Social information processing),
    - Dm (Category:Data mining),
    - Nlp (Category:Natural language processing),
    - IoT (Category:Internet of Things)
  - 1065 data instances
    - 355 positive instances,
    - 710 negative instances = 355 reversed negative + 355 random negative

Data Cleaning and Concept Extraction

Using user frequency and edited distance to group word forms.

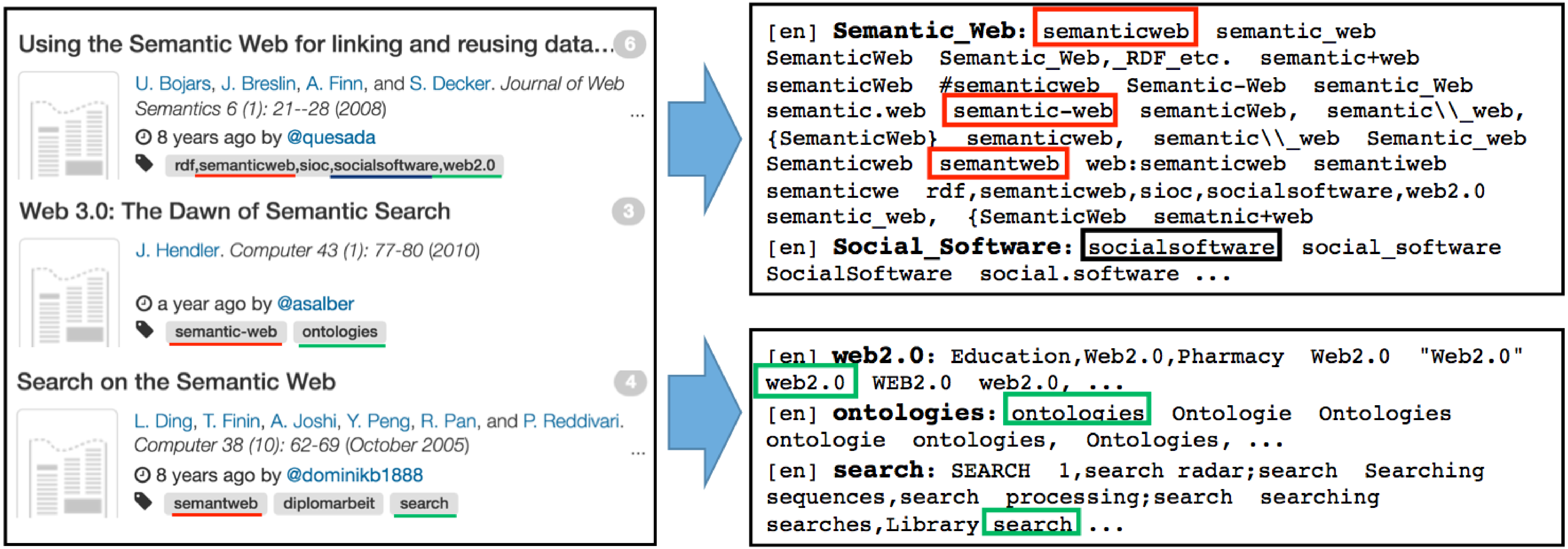


Image in Dong, H., Wang, W., & Coenen, F. (2017). Deriving Dynamic Knowledge from Academic Social Tagging Data: A Novel Research Direction. In iConference 2017 Proceedings (pp. 661-666). https://doi.org/10.9776/17313

Tag grounding for instance labelling

About: Machine learning

An Entity of Type : Concept, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

is skos:broader of	dbc:Artificial_neural_networks dbc:Classification_algorithms dbc:Data_mining_and_machine_learning_software dbc:Evolutionary_algorithms dbc:Machine_learning_researchers dbc:Kernel_methods_for_machine_learning dbc:Artificial_intelligence_conferences dbc:Ensemble_learning dbc:Log-linear_models
is dct:subject of	dbr:Darkforest dbr:Supervised_learning dbr:Mixture_model dbr:Rademacher_complexity dbr:Kernel_embedding_of_distributions dbr:Product_of_experts dbr:Deeplearning4j dbr:Google_DeepMind dbr:Adaptive_projected_subgradient_method

DBpedia concept pairs

http://dbpedia.org/page/Category:Machine_learning

```
learning_to_rank <- machine_learning
chromosome <- genetic_algorithms
schema <- genetic_algorithms
pattern_recognition <- machine_learning
formal_concept_analysis <- machine_learning
semantic_analysis <- machine_learning
deep_learning <- machine_learning
unsupervised_learning <- machine_learning
mixture_model <- cluster_analyse
margin <- support_vector_machines
inheritance <- genetic_algorithms
selection <- genetic_algorithms
support_vector_machines <- machine_learning
evolutionary_algorithms <- machine_learning
cluster_analyse <- machine_learning
bayesian_networks <- machine_learning
speciation <- evolutionary_algorithms
evolvability <- machine_learning
stability <- machine_learning
schema <- genetic_programming
generative_model <- machine_learning
mixture_model <- machine_learning
```

Matched tag concept pairs (positive data)

Classification and Testing

- Baselines:
  - **Topic similarity with "Information Theory Principle for Concept Relationship"** in the study by Wang *et al.* (2009), equivalent to the feature set **S1**.
  - **Co-occurrence based features** (support, confidence, cosine similarity, inclusion and generalisation degree, mutual overlapping and taxonomy search) in the study by Rêgo *et al.* (2015), denoted as the feature set **S4**.

- Training 80%, testing 20%.
- Parameters C and γ for SVM Gaussian Kernel (Hsu, 2003) were tuned with 10-fold cross-validation on the training data.
- For using Latent Dirichlet Allocation to generate features, we set
  - the topic-word hyperparameter α as 50/|z|,
  - the document-topic hyperparameter β as 0.01,
  - the number of topic |z| as 600, empirically selected based on the perplexity measure.

Classification Results

Classifiers: Logistic Regression (LR), SVM Gaussian Kernels (SVM), and Weighted-SVM (C+/C- = 2) (Veropoulos et al., 1999).

S1: Topic Similarity Features (Wang *et al*., 2009)

S2: Topic Distribution Based Features

S3: Probabilistic Association Features

S4: Co-occurrence Features (Rêgo *et al.*, 2015)

Table 2. Classification results using different feature set combinations

		Recall	Precision	F_1 score	Accuracy	AUC
S1 + S2 + S3 (Full features in our approach)	LR	54.9%	60.0%	57.4%	72.8%	0.808
	SVM	**73.2%**	65.0%	68.9%	77.9%	0.814
	weighed-SVM	**100.0%**	42.0%	59.2%	54.0%	0.792
Wang et al. [28](**S1**)	LR	12.7%	47.4%	20.0%	66.2%	0.585
	SVM	38.0%	58.7%	46.2%	70.4%	0.648
Rêgo et al. [22] (**S4**)	LR	16.9%	63.2%	26.7%	69.0%	0.657
	SVM	22.5%	57.1%	32.3%	68.6%	0.563
S1 + S2 + S3 + S4	LR	56.3%	62.5%	59.3%	74.2%	0.808
	SVM	71.8%	64.6%	68.0%	77.5%	0.818
S2	LR	22.5%	59.3%	32.7%	69.0%	0.752
	SVM	59.2%	55.3%	57.1%	70.4%	0.688
S3	LR	4.2%	37.5%	7.6%	65.7%	0.769
	SVM	5.6%	50.0%	10.1%	66.7%	0.794
S1 + S2	LR	42.3%	61.2%	50.0%	71.8%	0.761
	SVM	63.4%	57.0%	60.0%	71.8%	0.699
S1 + S3	LR	32.4%	54.8%	40.7%	68.5%	0.700
	SVM	62.0%	64.7%	63.3%	76.1%	0.776
S2 + S3	LR	33.8%	60.0%	43.2%	70.4%	0.787
	SVM	59.2%	60.9%	60.0%	73.7%	0.743

* **S1** denotes Similarity and Divergence Based Features; **S2**, Topic distribution Based Features; **S3**, Probabilistic Association Features; **S4**, the baseline feature set in [22] including support, confidence, cosine similarity, inclusion and generalisation degree, mutual overlapping and taxonomy search

Examples of the learned relations

Table 3. Examples of learned relations from Bibsonomy tags using the three feature sets **S1+S2+S3** with SVM

narrower → broader concept	narrower → broader concept
social_graphs → social_networks	semantic_analysis → machine_learning
mixture_model → data_mining	unsupervised_learning → machine_learning
folksonomy → collective_intelligence	latent_variables → bayesian_networks
semantic_search → semantic_web	sentiment_analysis → natural_language_processing
delicious →social_bookmarking	word_sense_disambiguation → natural_language_processing

Conclusion and Future Studies

- Relation extraction from social tags as a supervised learning problem.
- A novel method to derive domain independent features to learn broad-narrower relation. Three assumptions, including Topic similarity, Topic distribution, Probabilistic association, help capture tag relations based on human cognitive processing of information.

- Future studies:
  - **Heterogeneous Knowledge Bases** for tag grounding and instance labelling.
  - **Knowledge Base Enrichment**: identify new relations to enrich KBs.
  - **Deep learning** approaches: neural network architectures for relation extraction.

Key References

- P. Andrews, and J. Pane, “Sense induction in folksonomies: a review,” *Artificial Intelligence Review*, vol. 40, no. 2, pp. 147-174, 2013.
- D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- J. Chen, S. Feng, and J. Liu, “Topic sense induction from social tags based on non-negative matrix factorization,” *Information Sciences, vol. 280, pp. 16-25, 2014.*
- H. Dong, W. Wang and H.-N. Liang, "Learning Structured Knowledge from Social Tagging Data: A Critical Review of Methods and Techniques," *8th IEEE International Conference on Social Computing and Networking (IEEE SocialCom 2015),* Chengdu, 2015, pp. 307-314.
- H. Dong, W. Wang, F. Coenen. Deriving Dynamic Knowledge from Academic Social Tagging Data: A Novel Research Direction, *iConference 2017*, Wuhan, P.R. China, pp. 661-666.
- A. García-Silva, O. Corcho, H. Alani, and A. Gómez-Pérez, “Review of the state of the art: Discovering and associating semantics to tags in folksonomies,” *The Knowledge Engineering Review*, vol. 27, no. 01, pp. 57-85, Mar. 2012. T. V. Wal, “Folksonomy,” http://vanderwal.net/folksonomy.html, 2007.
- M. A. Hearst (1992, August). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2* (pp. 539-545). Association for Computational Linguistics.
- C. W. Hsu, C.C. Chang, C.J. Lin, *A practical guide to support vector classification*. Technical report, Department of Computer Science, National Taiwan University, 2003.
- A. S. C. Rego, L. B. Marinho, and C. E. S. Pires, “A supervised learning approach to detect subsumption relations between tags in folksonomies,” in *Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC ’15)*. ACM, 2015, pp. 409–415.
- R. R. Souza, D. Tudhope, and M. B. Almeida, “Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems,”2012.
- K. Veropoulos, C. Campbell, N. Cristianini, et al.: Controlling the sensitivity of support vector machines. In: Proceedings of the IJCAI, pp. 55–60 (1999)
- T. Vander Wal. Folksonomy (2007). http://vanderwal.net/folksonomy.html. Accessed 08 August 2018
- W. Wang, P.M. Barnaghi, A. Bargiela. Probabilistic topic models for learning terminological ontologies. *IEEE Trans. Knowl. Data Eng.* 22(7), 1028–1040, 2010.
- W. Wu, H. Li, H. Wang, & K. Q. Zhu (2012, May). Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 481-492). ACM.
- M. Zhou, S. Bao, X. Wu, and Y. Yu, An unsupervised model for exploring hierarchical semantics from social annotations*. In: Aberer K. et al. (eds) The Semantic Web. ISWC 2007, ASWC 2007. Lecture Notes in Computer Science, vol 4825.* Springer, Berlin, Heidelberg.

Thank you for your attention!

Hang.Dong@liverpool.ac.uk

Home page: http://cgi.csc.liv.ac.uk/~hang/