End-to-End Reinforcement Learning for Automatic Taxonomy Induction

Yuning Mao, Xiang Ren, Jiaming Shen, Xiaotao Gu, Jiawei Han

Department of Computer Science, University of Illinois at Urbana-Champaign, USA

Department of Computer Science, University of Southern California, CA, USA

{yuningm2, js2, xiaotao2, hanj}@illinois.edu xiangren@usc.edu

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

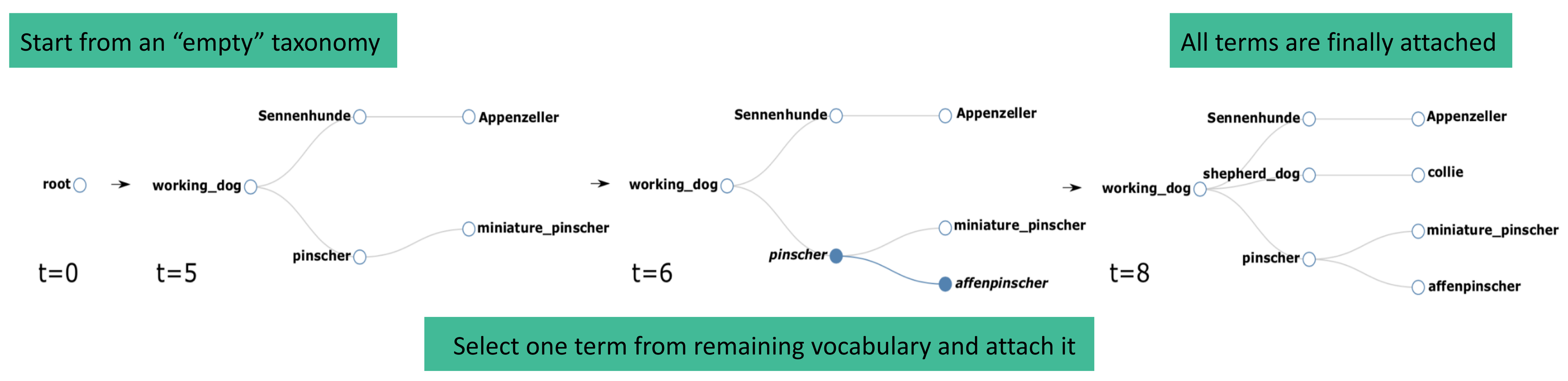
DAIS The Database and Information Systems Laboratory at The University of Illinois at Urbana-Champaign *Large Scale Information Management*

Task

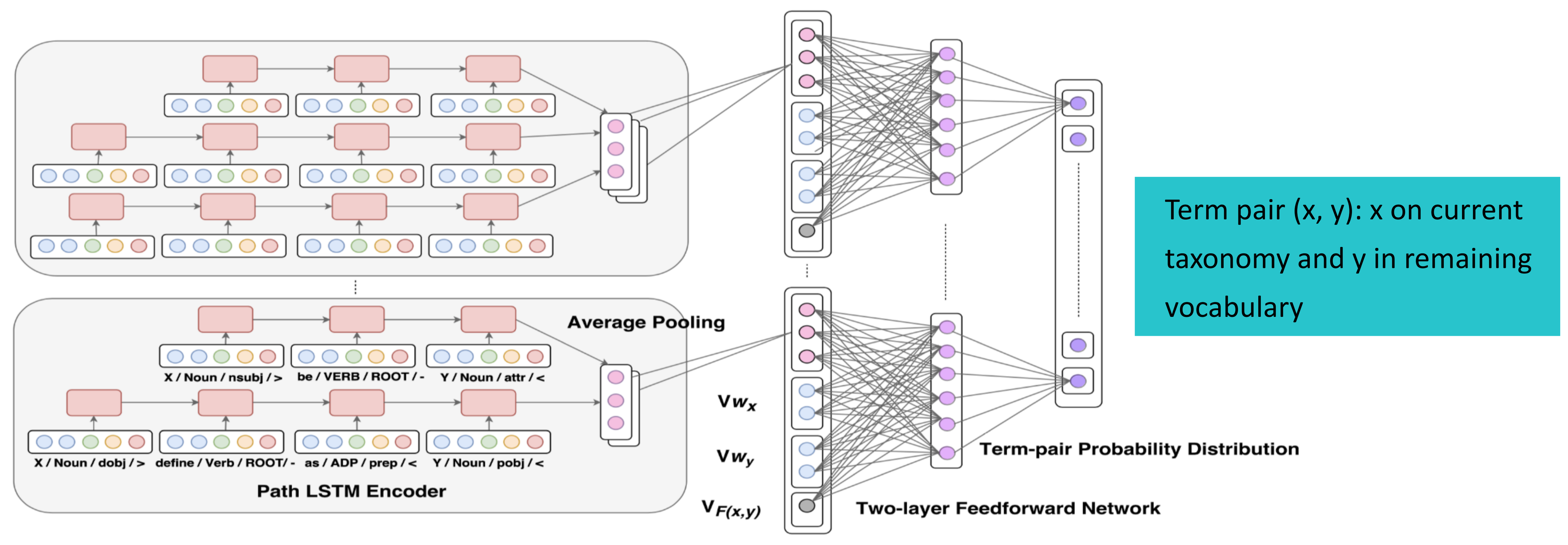
- **Goal: Automatic Taxonomy Induction**
  - Input: 1) a set of training taxonomies
    2) related resources (e.g., background text corpora).
  - Output: given vocabulary V_0, construct a taxonomy T by adding terms from V_0
- **Hypernymy Detection:**
  - Hypernymy pairs (is-a relations) are extracted: (banana, fruit), (panda, mammal), …
  - A noisy hypernym graph is generated
- **Hypernymy Organization**
  - Organize is-a term pairs into a tree-structured hierarchy -> graph pruning
  - maximum spanning tree (MST) (Bansal et al., 2014 [1])
  - minimum-cost flow (MCF) (Gupta et al., 2017 [2])
  - other pruning heuristics (Panchenko et al., 2016 [3])

Methodology

- **Taxo-RL:**
  - An end-to-end reinforcement learning (RL) model that combines hypernymy detection and organization
  - Determines **which** term to select and **where** to place it on the taxonomy via a policy network
  - Global taxonomy structure is captured
  - Edges are assessed based on contributions



- **RL Component - States:**
  - The state at time t comprises:
    - the current taxonomy T_t (terms & structure)
    - the remaining vocabulary V_t
  - Update deterministically
- **RL Component - Actions:**
  1. select a term x_1 from the remaining vocabulary V_t
  2. remove x_1 from V_t
  3. attach x_1 as a hyponym of one term x_2 that is already on the current taxonomy T_t
  - $|V_t| = |V_{t-1}| - 1$, $|T_t| = |T_{t-1}| + 1$
  - Action Space: $|V_t| \times |T_t|$
  - Episode Length: $|V_0|$
- **RL Component - Rewards:**
  - Evaluation Metrics: $P_a = \frac{|\text{is-a}_{\text{sys}} \wedge \text{is-a}_{\text{gold}}|}{|\text{is-a}_{\text{sys}}|}, R_a = \frac{|\text{is-a}_{\text{sys}} \wedge \text{is-a}_{\text{gold}}|}{|\text{is-a}_{\text{gold}}|}$
    - Ancestor-F1
    - Edge-F1
  - Reward Shaping: R_t = Edge-F1(t) – Edge-F1(t-1)
- **Action (term-pair) Representation**
  - Dependency Paths between x, y
  - W_x: Word Embedding of x
  - W_y: Word Embedding of y
  - F(x, y): **S**urface (Ends with, Contains, etc.), **F**requency (pattern-based co-occur info), and **G**enerality (edge not too general or narrow) Features



Policy Network

Experimental Results

- **Compared methods:**
  - TAXI [3]: pattern-based method that ranked 1st in the SemEval-2016 Task 13 competition
  - HypeNET [4]: state-of-the-art hypernymy detection method
  - HypeNET + MST (maximum spanning tree): post-processing of HypeNET to prune the hypernym graph into a tree
  - Bansal et al. (2014) [1]: state-of-the-art taxonomy induction method
  - SubSeq [2]: state-of-the-art results on the SemEval-2016 Task 13
  - Taxo-RL (**RE**, with virtual root embedding), Taxo-RL (**NR**, with new root addition), Taxo-RL (NR) + **FG** (with frequency and generality features)
  - Taxo-RL (**partial**, allows partial taxonomy), Taxo-RL (**full**, has to use all terms in the vocabulary)
- **Performance Study on End-to-End Taxonomy Induction:**
  - WordNet (533/144/144 taxonomies for training, validation, and test set, size (10, 50], depth=4, animals, daily necessities, etc.)

Model	P_a	R_a	$F1_a$	P_e	R_e	$F1_e$
TAXI	66.1	13.9	23.0	54.8	18.0	27.1
HypeNET	32.8	26.7	29.4	26.1	17.2	20.7
HypeNET+MST	33.7	41.1	37.0	29.2	29.2	29.2
Taxo-RL (RE)	35.8	47.4	40.8	35.4	35.4	35.4
Taxo-RL (NR)	41.3	49.2	**44.9**	35.6	35.6	**35.6**
Bansal et al. (2014)	48.0	55.2	51.4	-	-	-
Taxo-RL (NR) + FG	52.9	58.6	**55.6**	43.8	43.8	**43.8**

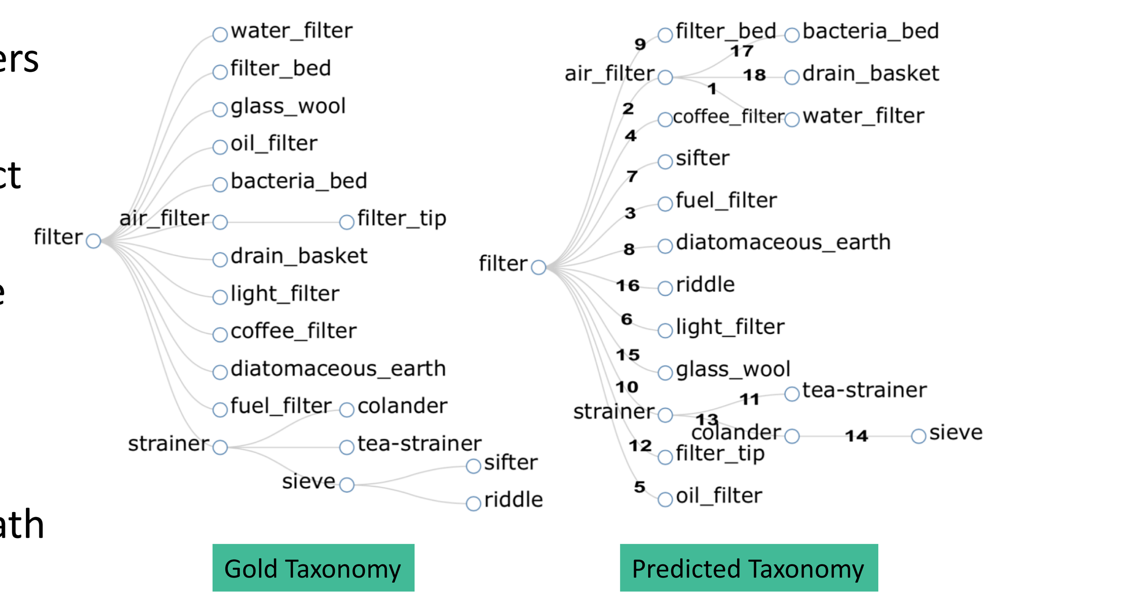
- **Testing on Hypernymy Organization:**
  - SemEval-2016 Task 13 (test set only, hundreds of terms, environment, science domain)

	Model	P_a	R_a	$F1_a$	P_e	R_e	$F1_e$
Env	TAXI (DAG)	50.1	32.7	39.6	33.8	26.8	29.9
	TAXI (tree)	**67.5**	30.8	42.3	**41.1**	23.1	29.6
	SubSeq	-	-	-	-	-	22.4
	Taxo-RL (Partial)	51.6	36.4	42.7	37.5	24.2	29.4
	Taxo-RL (Full)	47.2	**54.6**	**50.6**	32.3	**32.3**	**32.3**
Sci	TAXI (DAG)	61.6	41.7	49.7	38.8	34.8	36.7
	TAXI (tree)	76.8	38.3	51.1	44.8	28.8	35.1
	SubSeq	-	-	-	-	-	39.9
	Taxo-RL (Partial)	**84.6**	34.4	48.9	**56.9**	33.0	**41.8**
	Taxo-RL (Full)	68.3	**52.9**	**59.6**	37.9	**37.9**	37.9

- **Ablation Study:**
  - Multiple sources of information are complementary to each other

Model	P_a	R_a	$F1_a$	$F1_e$
Distributional Info	27.1	24.3	25.6	13.8
Path-based Info	27.8	48.5	33.7	27.4
D + P	36.6	39.4	37.9	28.3
D + P + Surface Features	41.3	49.2	44.9	35.6
D + P + S + FG	**52.9**	**58.6**	**55.6**	**43.8**

- **Case Studies:**
  - Numbers indicate the orders of term pair selections
  - (air filter, filter, 2) -> correct root
  - (fuel filter, filter, 3), (coffee filter, filter, 4) -> substring inclusion
  - -> (colander, strainer, 13), (glass wool, filter, 16) -> path and distributional info



Conclusion and References

- **Conclusion**:
  - Learns the representations of term pairs by optimizing a holistic tree metric
  - Reduces error propagation between two phases
  - Achieves new state-of-the-art results
- **References**:

[1] Mohit Bansal, David Burkett, Gerard De Melo, and Dan Klein. ACL 2014. Structured learning for taxonomy induction with belief propagation.

[2] Amit Gupta, Rémi Lebret, Hamza Harkous, and Karl Aberer. CIKM 2017. Taxonomy induction using hypernym subsequences.

[3] Alexander Panchenko, Stefano Faralli, Eugen Ruppert, Steffen Remus, Huber Naets, Cedrick Fairon, Simone Paolo Ponzetto, and Chris Biemann. SemEval 2016. Taxi: a taxonomy induction method based on lexico-syntactic patterns, substrings and focused crawling.

[4] Vered Shwartz, Yoav Goldberg, and Ido Dagan. ACL 2016. Improving hypernymy detection with an integrated path-based and distributional method

ARL NETWORK SCIENCE CTA ILLINOIS