国际人工智能会议

AAAI 2021论文北京预讲会

UNIVERSITY of WASHINGTON

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

ByteDance AI Lab 字节跳动人工智能实验室

Taxonomy Completion via Triplet Matching Network

Jieyu Zhang[1], Xiangchen Song[2], Ying Zeng[3], Jiaze Chen[3], Jiaming Shen[2], Yuning Mao[2], Lei Li[3]

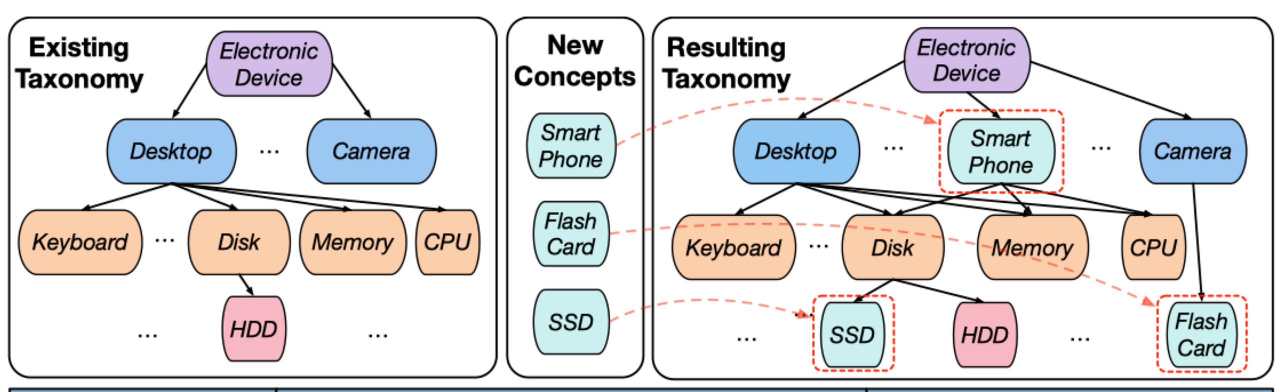
1.Paul G. Allen School of Computer Science & Engineering, University of Washington, 2. Department of Computer Science, University of Illinois Urbana-Champaign, 3. ByteDance AI Lab

INTRODUCTION

Automatically constructing taxonomy finds many applications in e-commerce and web search. Previous approaches focus on the taxonomy expansion, i.e., finding an appropriate hypernym concept from the taxonomy for a new query concept.

In this paper, we formulate a new task, "taxonomy completion", by discovering both the hypernym and hyponym for a query.

We propose Triplet Matching Network (TMN), to find the appropriate <hypernym, hyponym> pairs for a given query concept. Experiments on four real-world large-scale datasets show that **TMN** achieves the best performance on both taxonomy completion task and the previous taxonomy expansion task, outperforming existing methods.



Query	Candidate Position	Label
Smart Phone	< Electronic Device, CPU >	Positive
Smart Phone	< Electronic Device, HDD >	Partially-correct Negative
Smart Phone	< Desktop, CPU>	Partially-correct Negative
Smart Phone	< Desktop, Keyboard>	Negative

METHOD

One-To-Pair Matching

In this work, we seek to learn a model $s: N \times (N \times N) \to R$ with parameter Θ that can measure the relatedness of a query concept n_q and a candidate position, i.e., a pair of concepts $t = \langle n_p, n_c \rangle$, in existing taxonomy $\mathcal{T}^0$. A straightforward instantiation of s is as follows:

$$s(n_q, n_p, n_c) = f(\boldsymbol{x}_q, [\boldsymbol{x}_p, \boldsymbol{x}_c]) = f(\boldsymbol{x}_q, \boldsymbol{x}_t)$$

Auxiliary Learning

We develop multiple auxiliary scorers to capture both coarse- and fine-grained relatedness in one-to-pair matching, and one primal scorer that inputs the internal feature representations of all auxiliary scorers and outputs final matching scores.

$$h(\mathbf{x}_q, \mathbf{x}_t) = \mathbf{x}_q \mathbf{W}^{[1:k]} \mathbf{x}_t + \mathbf{V} \begin{bmatrix} \mathbf{x}_q \\ \mathbf{x}_t \end{bmatrix} + \mathbf{b}$$

$$s_{primal}(n_q, n_p, n_c) = \mathbf{u}_p^T \sigma([\mathbf{h}_1, \dots, \mathbf{h}_l])$$

$$s_1(n_q, n_p) = \mathbf{u}_1^T \sigma(\mathbf{h}_1(\mathbf{x}_q, \mathbf{x}_p))$$

$$s_2(n_q, n_c) = \mathbf{u}_2^T \sigma(\mathbf{h}_2(\mathbf{x}_q, \mathbf{x}_c))$$

$$s_3(n_q, n_p, n_c) = \mathbf{u}_3^T \sigma(\mathbf{h}_3(\mathbf{x}_q, [\mathbf{x}_p, \mathbf{x}_c]))$$

Channel-wise Gating Mechanism

Concepts under the same ancestor are semantically related to each other, which makes it challenging for model to learn the true taxonomic relations based on concept embeddings, especially in bottom-level of a taxonomy.

$$\mathbf{g}_p = \theta(\mathbf{W}_1[\mathbf{x}_q, \mathbf{x}_p, \mathbf{x}_c])$$

$$\hat{\mathbf{x}}_p = \mathbf{g}_p \odot \mathbf{x}_p$$

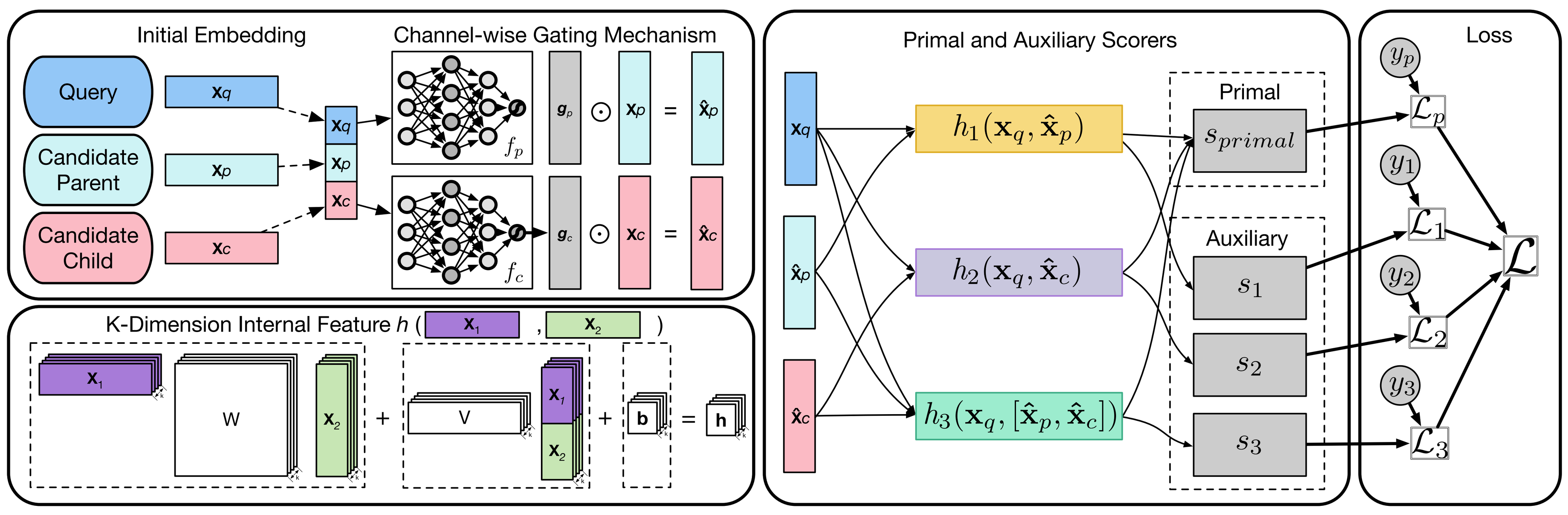
$$\mathbf{g}_c = \theta(\mathbf{W}_2[\mathbf{x}_q, \mathbf{x}_p, \mathbf{x}_c])$$

$$\hat{\mathbf{x}}_c = \mathbf{g}_c \odot \mathbf{x}_c$$

Learning Objective

$$\mathcal{L}(\Theta) = \mathcal{L}_p + \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 + \lambda_3 \mathcal{L}_3$$

TMN FRAMEWORK



SELF-SUPERVISION & LEARNING

Algorithm 1: Self-supervised learning of TMN

Input: A taxonomy $\mathcal{T}^0$; negative size N, batch size B; model $f(\cdot|\Theta)$.

Output: Learned model parameters Θ.

Randomly initialize Θ;

while $\mathcal{L}(\Theta)$ *in Eq. (13) not converge* **do**

  Enumerate nodes in $\mathcal{T}^0$ and sample B nodes without replacement;

  $\mathbb{D} = \{\}$ # current batch of training instances;

  for *each sampled node* n_q **do**

    Select one of its parents n_p and one of its children n_c to construct positive triple $\langle n_q, \langle n_p, n_c \rangle \rangle$

    Generate N negative triples $\{\langle n_q, \langle n_p^l, n_c^l \rangle \rangle |_{l=1}^N\}$;

    $\mathbb{D} \leftarrow \mathbb{D} \cup \{\langle n_q, \langle n_p, n_c \rangle \rangle, \langle n_q, \langle n_p^1, n_c^1 \rangle \rangle, \dots, \langle n_q, \langle n_p^N, n_c^N \rangle \rangle\}$;

  Update Θ based on $\mathbb{D}$.

Return Θ;

Given a node n_p, construct a positive pair $\langle n_p, n_c \rangle$ using one of its parents and one of its children.

Negative samples from valid candidates and label them according to the definition in auxiliary signals.

EXPERIMENT RESULTS

Dataset

Dataset	$\lvert\mathcal{N}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{D}\rvert$
MAG-CS	24,754	42,329	6
MAG-Psychology	23,187	30,041	6
WordNet-Verb	13,936	13,408	13
WordNet-Noun	83,073	76,812	20

Ablation Study

Method	MAG-Psychology			
	MR	MRR	Recall@10	Prec@10
TMN w/o CG	265.729	0.385	0.298	0.061
TMN w/o s_1&s_2	258.382	0.458	0.368	0.075
TMN w/o s_1	269.058	0.471	0.382	0.123
TMN w/o s_2	229.306	0.474	0.381	0.078
TMN w/o s_3	342.021	0.326	0.213	0.043
TMN	212.298	0.471	0.377	0.077

Overall Performance

Method	MAG-Psychology							
	MR	MRR	Recall@1	Recall@5	Recall@10	Prec@1	Prec@5	Prec@10
Closest-Position	5201.604	0.168	0.030	0.072	0.107	0.062	0.029	0.022
Single Layer Model	435.548 ± 4.057	0.350 ± 0.002	0.090 ± 0.003	0.209 ± 0.002	0.274 ± 0.003	0.183 ± 0.006	0.085 ± 0.001	0.056 ± 0.001
Multiple Layer Model	**297.644** ± 8.097	**0.413** ± 0.002	**0.110** ± 0.001	**0.265** ± 0.004	**0.334** ± 0.002	**0.224** ± 0.003	**0.108** ± 0.001	**0.068** ± 0.000
Bilinear Model	2113.024 ± 9.231	0.032 ± 0.000	0.000 ± 0.000	0.001 ± 0.000	0.003 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.001 ± 0.000
Neural Tensor Network	299.004 ± 6.294	0.380 ± 0.004	0.066 ± 0.004	0.207 ± 0.002	0.291 ± 0.004	0.134 ± 0.008	0.084 ± 0.001	0.059 ± 0.001
TaxoExpan	728.725 ± 2.096	0.253 ± 0.001	0.015 ± 0.001	0.092 ± 0.001	0.163 ± 0.001	0.031 ± 0.001	0.038 ± 0.000	0.033 ± 0.000
ARBORIST	547.723 ± 20.165	0.344 ± 0.012	0.062 ± 0.009	0.185 ± 0.011	0.256 ± 0.013	0.126 ± 0.018	0.076 ± 0.004	0.052 ± 0.003
TMN	**212.298** ± 3.051	**0.471** ± 0.001	**0.141** ± 0.001	**0.305** ± 0.004	**0.377** ± 0.002	**0.287** ± 0.001	**0.124** ± 0.001	**0.077** ± 0.000

Method	WordNet-Verb							
	MR	MRR	Recall@1	Recall@5	Recall@10	Prec@1	Prec@5	Prec@10
Closest-Position	34778.772	0.144	0.011	0.045	0.075	0.020	0.016	0.013
Single Layer Model	2798.243 ± 61.384	0.140 ± 0.009	0.029 ± 0.005	0.065 ± 0.006	0.093 ± 0.008	0.044 ± 0.007	0.019 ± 0.002	0.014 ± 0.001
Multiple Layer Model	2039.213 ± 240.577	0.227 ± 0.020	0.050 ± 0.006	0.120 ± 0.009	0.160 ± 0.015	0.075 ± 0.009	0.036 ± 0.003	0.024 ± 0.002
Bilinear Model	1863.915 ± 5.685	0.175 ± 0.001	0.012 ± 0.001	0.054 ± 0.000	0.096 ± 0.001	0.017 ± 0.001	0.016 ± 0.000	0.015 ± 0.000
Neural Tensor Network	**1599.196** ± 18.409	**0.255** ± 0.003	**0.051** ± 0.002	**0.125** ± 0.006	**0.176** ± 0.005	**0.076** ± 0.003	**0.038** ± 0.002	**0.027** ± 0.001
TaxoExpan	1799.939 ± 4.511	0.227 ± 0.002	0.024 ± 0.001	0.095 ± 0.001	0.140 ± 0.002	0.036 ± 0.002	0.029 ± 0.000	0.021 ± 0.000
ARBORIST	1637.025 ± 4.950	0.206 ± 0.011	0.016 ± 0.004	0.073 ± 0.011	0.116 ± 0.011	0.024 ± 0.006	0.022 ± 0.003	0.018 ± 0.002
TMN	**1445.801** ± 27.209	**0.304** ± 0.005	**0.072** ± 0.003	**0.163** ± 0.005	**0.215** ± 0.001	**0.108** ± 0.005	**0.049** ± 0.002	**0.032** ± 0.000

TMN can be easily **reduced to taxonomy expansion**. *TaxoExpan* and *ARBORIST* are directly trained on taxonomy expansion task while **TMN** is trained on completion task. **TMN** outperforms the others in both dataset with a large margin.

CONCLUSION & FUTURE WORK

We study taxonomy completion without manually labeled supervised data. We propose a novel **TMN** framework to solve the one-to-pair matching problem in taxonomy completion, which can be applied on other applications where one-to-pair matching problem exists.

Interesting future work includes leveraging current method to cleaning the existing taxonomy, and incorporating feedback from downstream to generate more diverse (auxiliary) supervision signals for taxonomy completion.

主办方：中国中文信息学会青年工作委员会

承办方：智源社区