Tencent 腾讯 | 腾讯天衍实验室 TENCENT JARVIS Université de Montréal Mila

Fully Exploit Hierarchical Structure for Self-Supervised Taxonomy Expansion

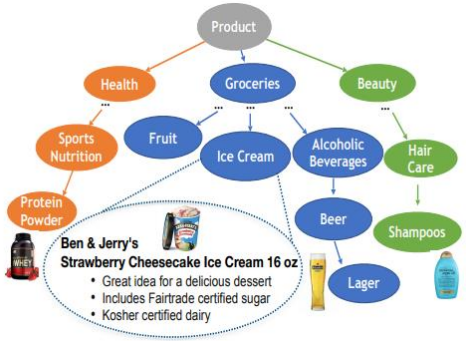
By Suyuchen Wang, Ruihui Zhao, Xi Chen, Yefeng Zheng, and Bang Liu

To Appear in The Web Conference 2021 (WWW'21)

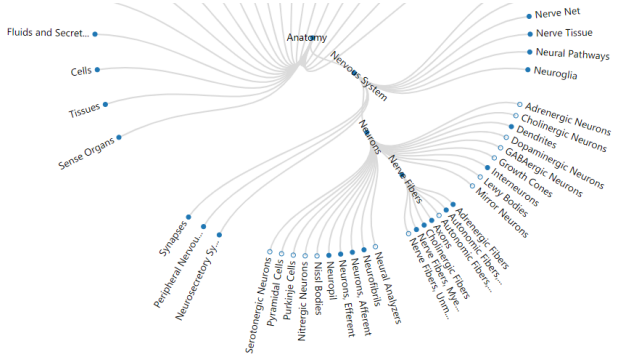
2021.02.27

By 王苏羽晨

Introduction to Taxonomy Expansion



Amazon Product Taxonomy



Medical Subject Headings(MeSH)

Definition

- Taxonomy: A hierarchical structure modeling "is-A" or hypernymy relations

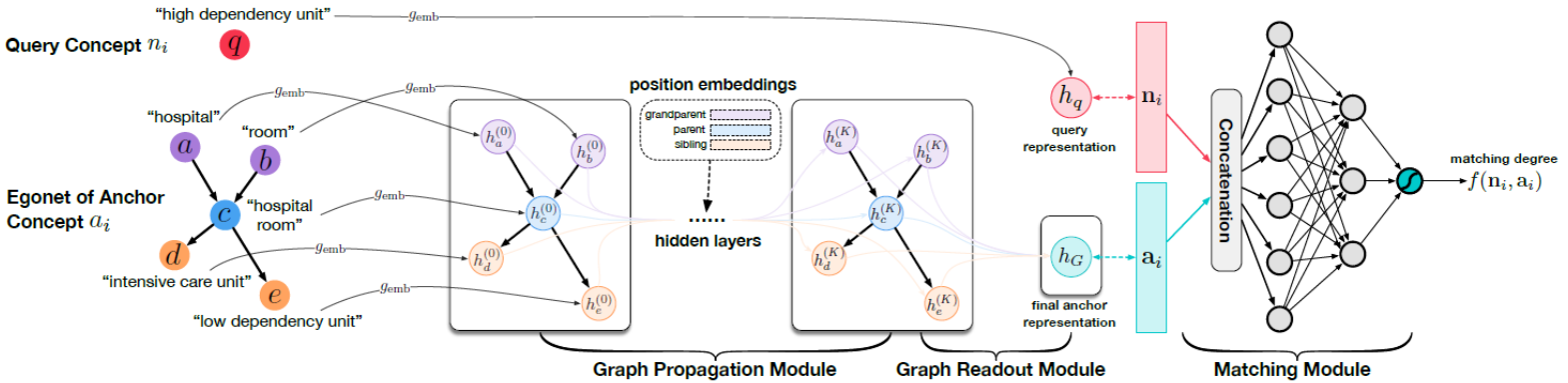
Problem

- Taxonomy curated by domain experts is time-consuming
- Some taxonomies needs to be updated continuously
- Low coverage & inconsistency of taxonomy will hurt the performance of downstream applications

Solution

- Accords to the rules while reducing human effort: Expand a pre-defined seed taxonomy rather than construct from scratch
- Updating taxonomy in a system: Concept mining → taxonomy expansion → item tagging in a row

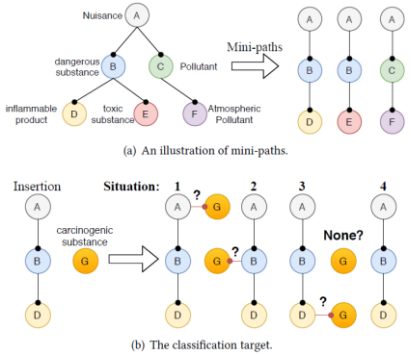
TaxoExpan: Self-supervised Taxonomy Expansion with Position-Enhanced Graph Neural Network (WWW '20)

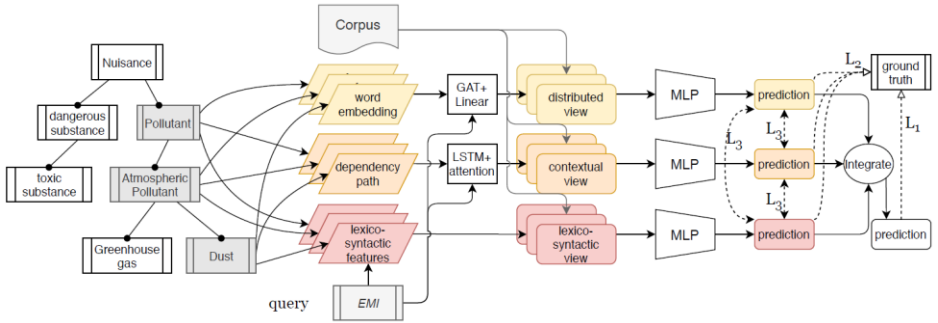


Modeling

- A mere node attaching algorithm for general purpose taxonomy
- Taxonomy modeled as a Bayesian network, do not have to be a tree
- Features of anchor node captured by Positional GAT on "egonet" containing its direct parents and children
- Score by a matching module, whose input is concatenation of query representation (embedding) & anchor representation (GAT results)
- Dataset constructed with self-supervision

STEAM: Self-supervised Taxonomy Expansion with Mini-paths (SIGKDD '20)



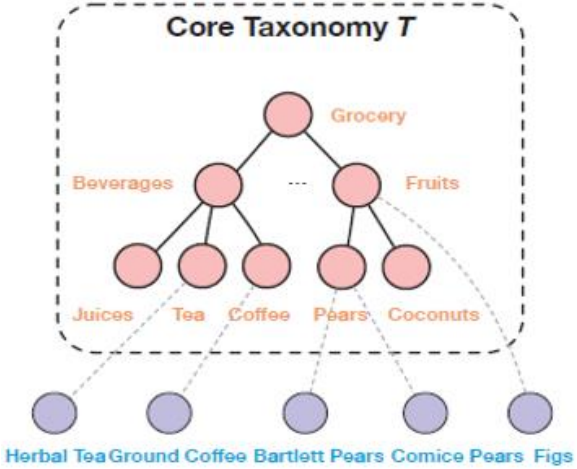


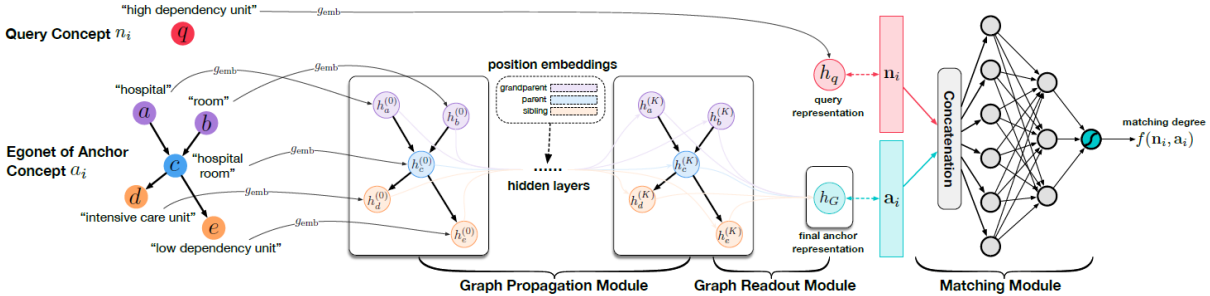
Modeling

- A mere node attaching algorithm for general purpose taxonomy
- Taxonomy modeled as a tree (not a Bayesian network)
- Extract distributed feature by Positional GAT with BERT embedding as input
- Extract contextual feature by weighted-summing the LSTM outputs of the dependency trees of query-anchor co-occurrence in corpus
- Extract lexical-syntactic feature by measuring a series of lexical relationship features
- Loss is the sum of aggregated loss containing 1) classifier loss, 2) single classifier loss and 3) cross-classifier consistency loss
- Dataset constructed with self-supervision

Overall Problems in Previous Solutions

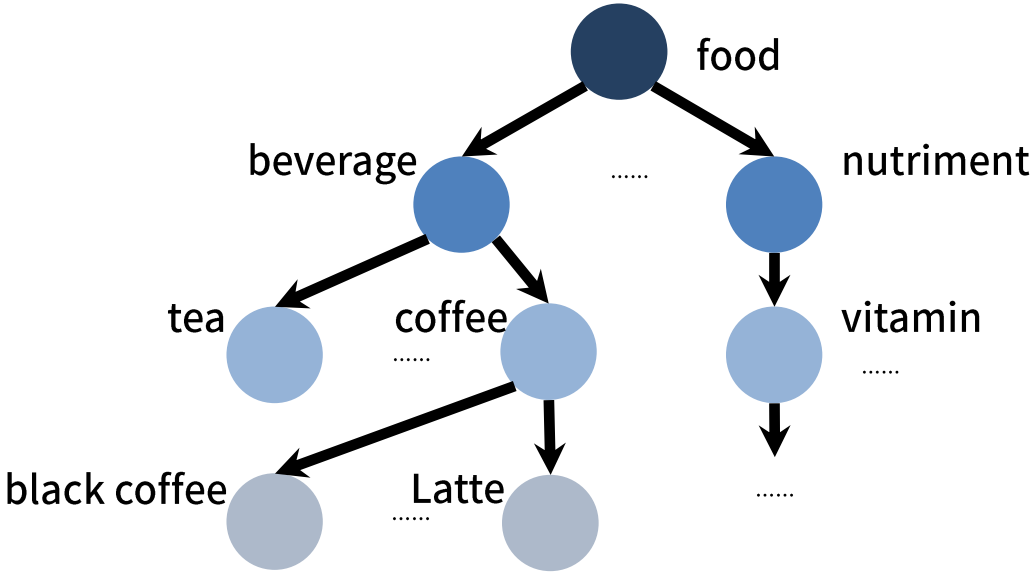
- Coarsely designed lexical and syntactical features: only extract superficial features in node names
- Deficient usage of other text data source: no definition texts are properly used in previous methods
- Insufficient mining of tree structure: learn hypernym detection of a single node, not construction of a hierarchy
- Ignorance of cross-level relationships: the nodes on the right path in taxonomy should share similarity
- High time complexity: need to compare all possible nodes to achieve the result ranking





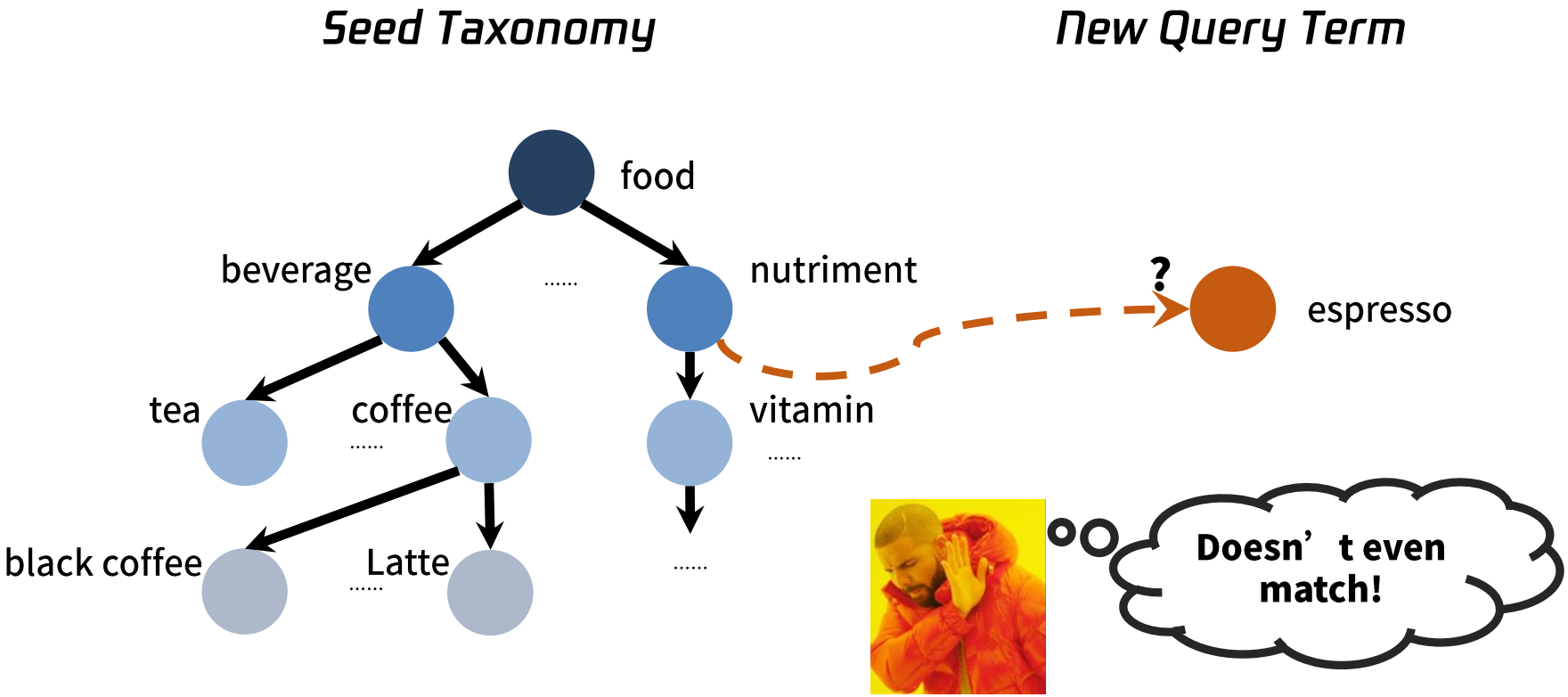
How to Add a New Term to a Taxonomy?

Seed Taxonomy

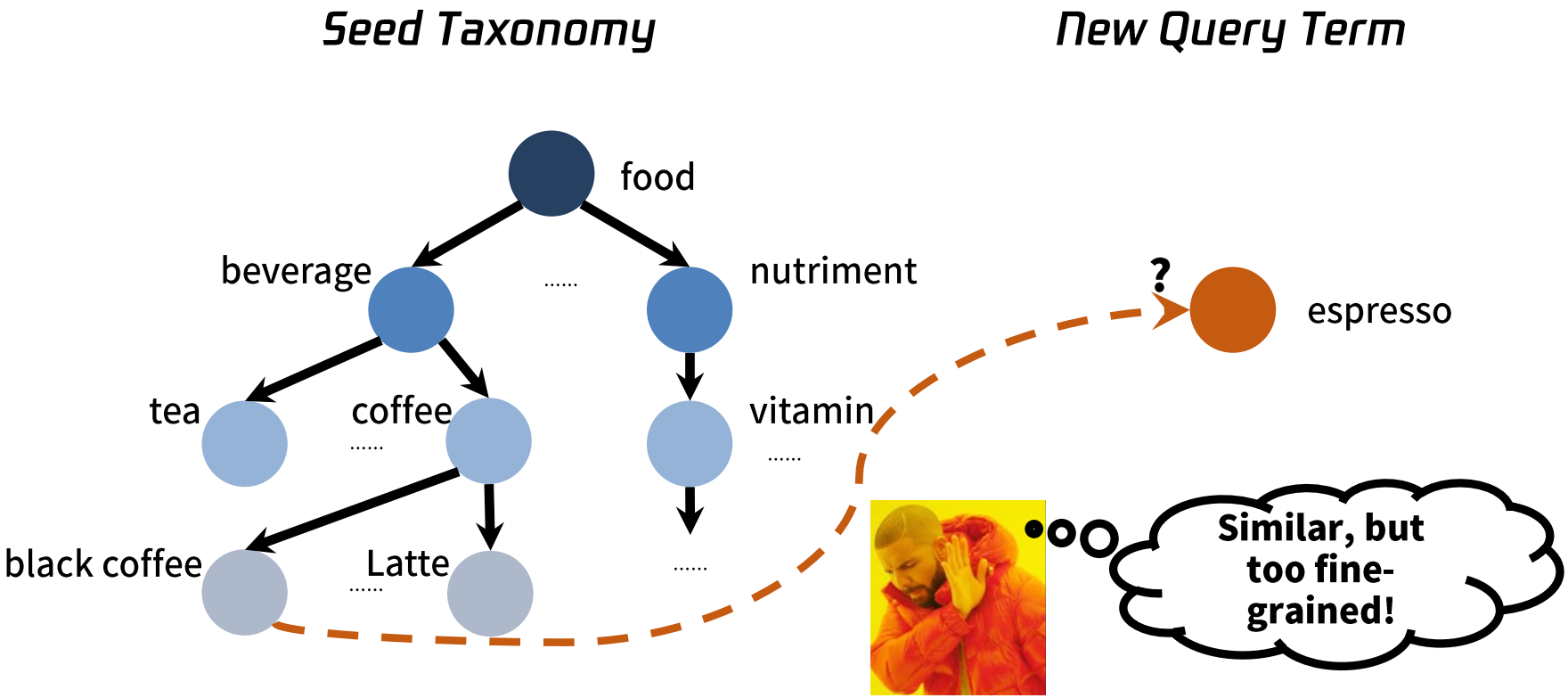


New Query Term

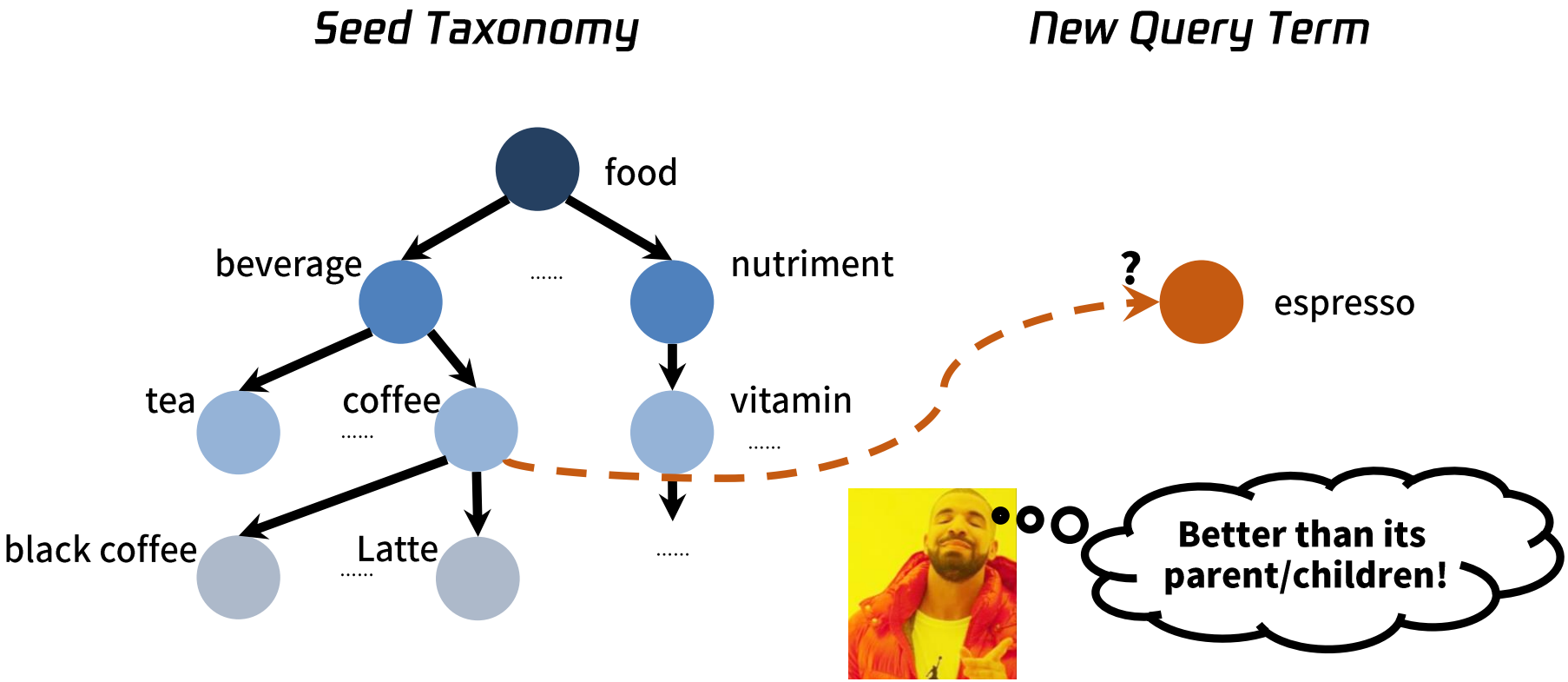
How to Add a New Term to a Taxonomy?



How to Add a New Term to a Taxonomy?



How to Add a New Term to a Taxonomy?



Motivation of Our Model

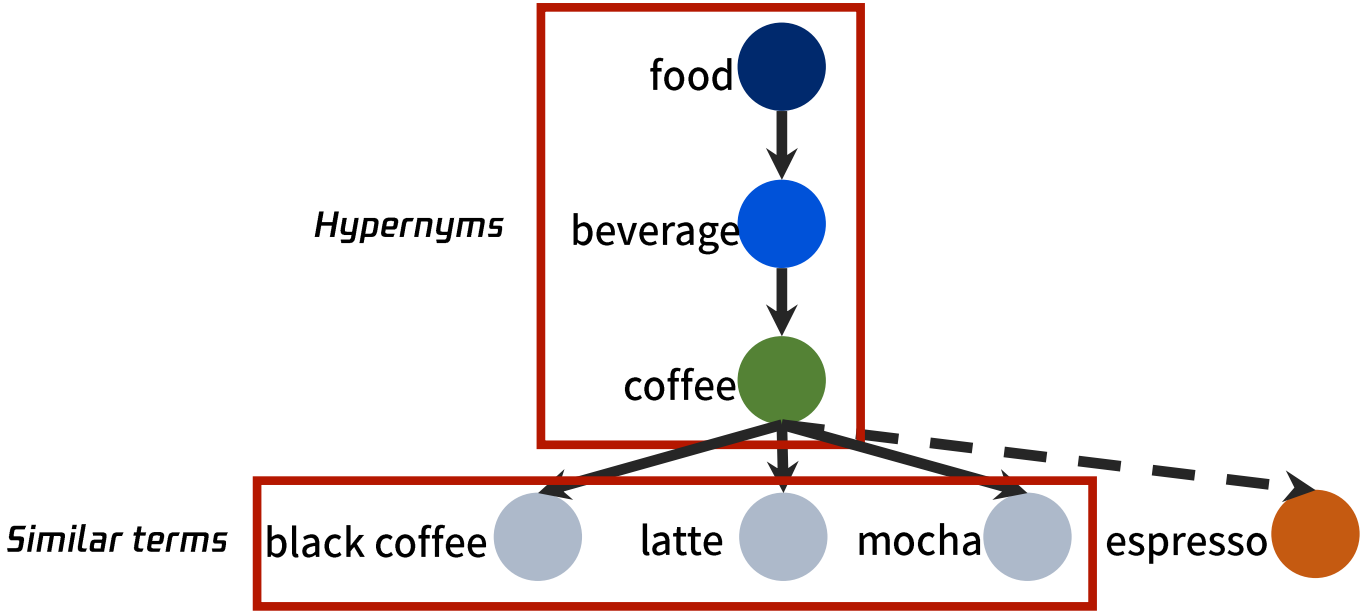
Espresso is a kind of coffee that…

Coffee is a part of espresso…

Term's surface name is insufficient for hypernymy detection

Contextual pattern in large corpora includes noice

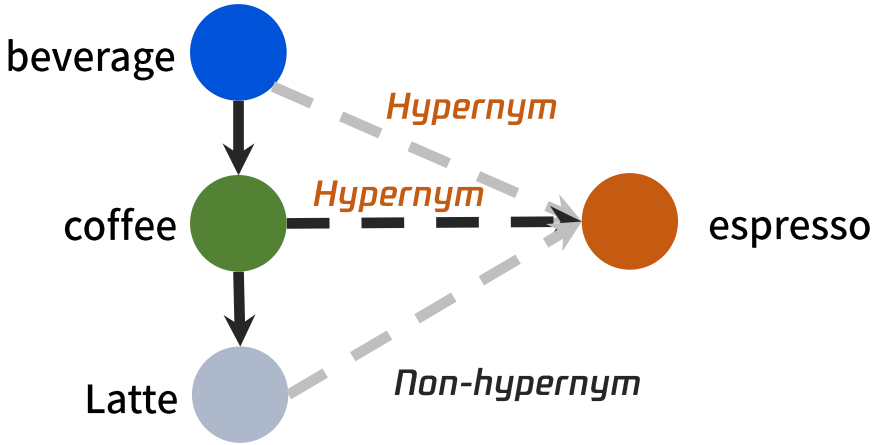
Motivation of Our Model



Hypernymy relations from parent-child relations

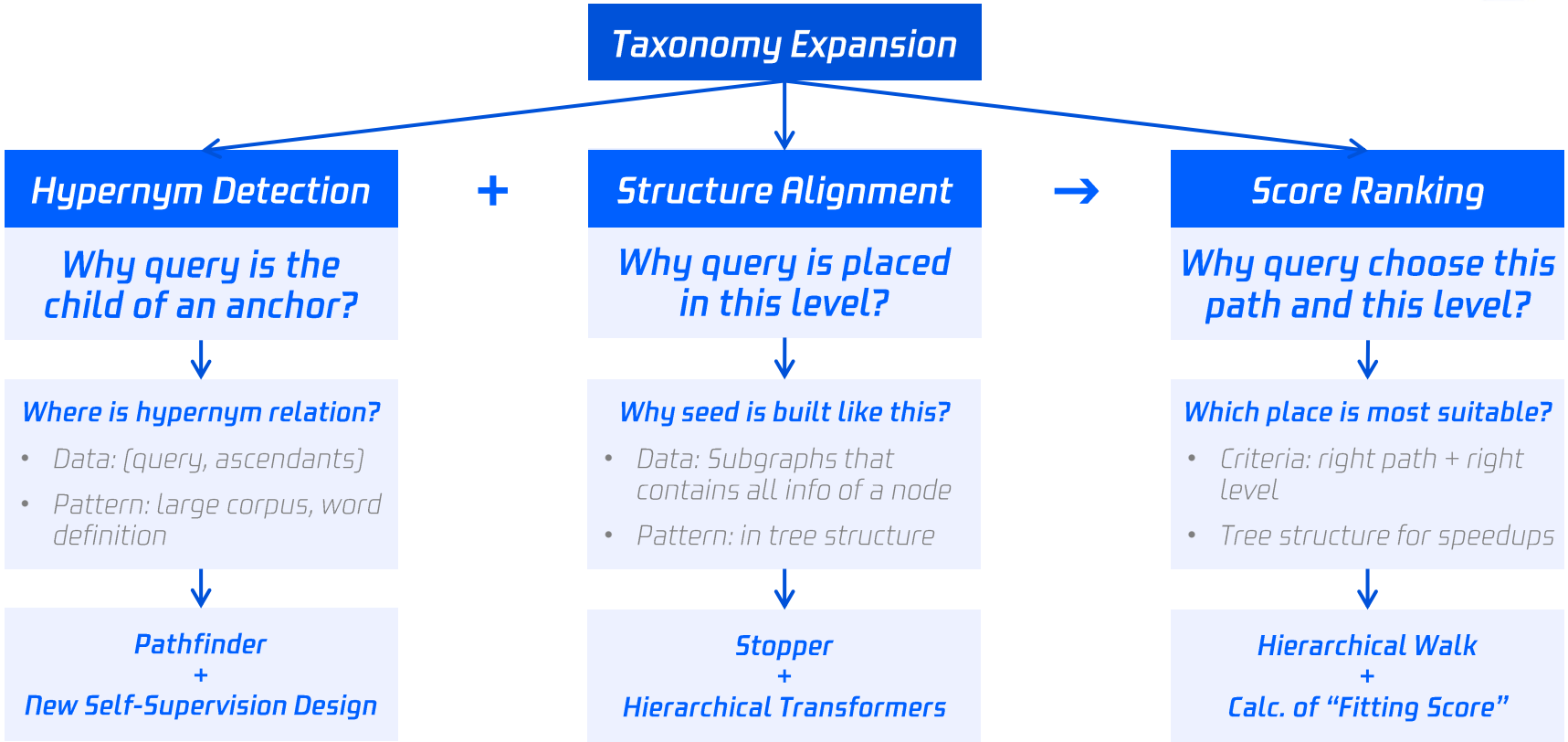
\+ *similarity relations* from sibling relations = *tree coherence*

Motivation of Our Model



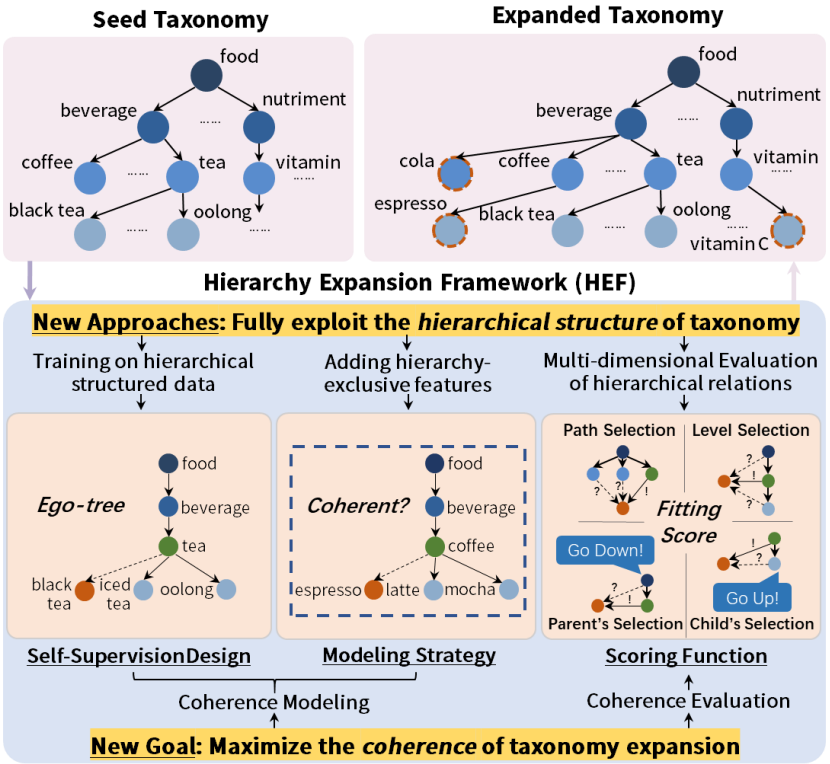
From "is-A" relation to "The most accurate is-A" relation needs comparison with parent/children

Deconstruction of the Taxonomy Expansion Task

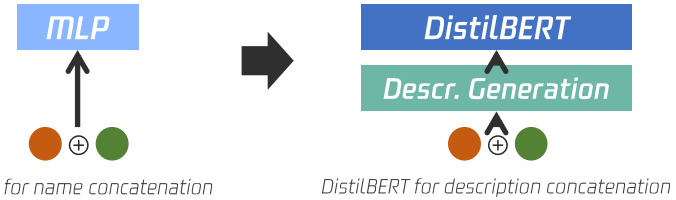


Motivation and Novelties

Major: New approaches and new goal



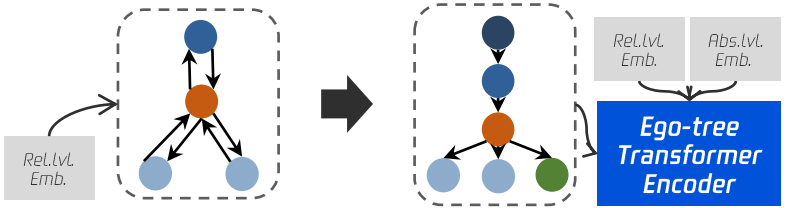
Others: New Implementations



MLP for name concatenation

DistilBERT for description concatenation

Node pair relation detection

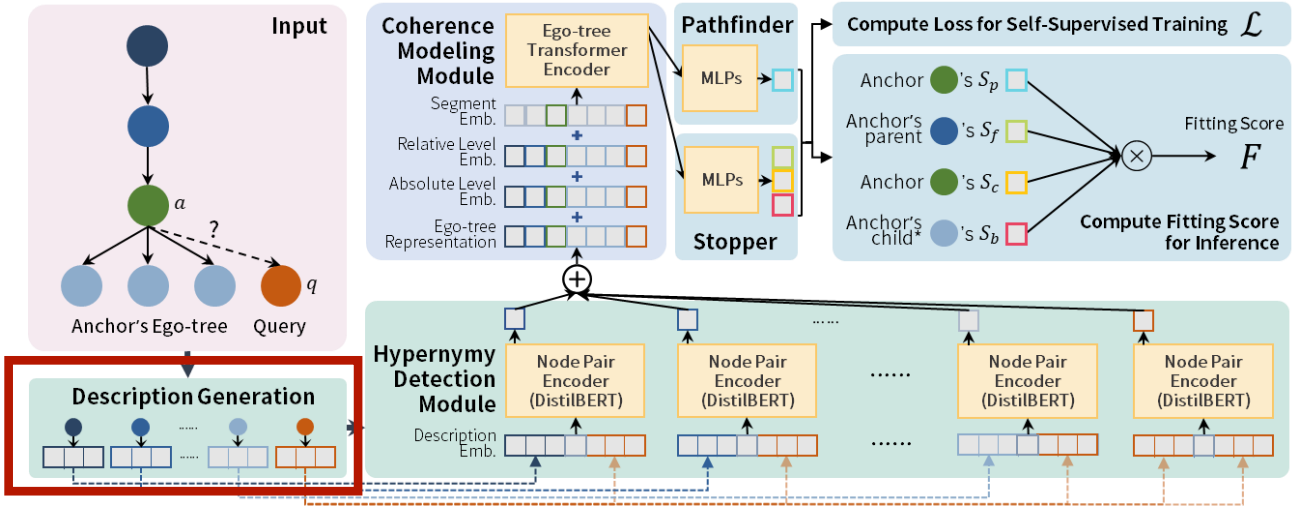


Position-enhanced GAT on egonets

Transformer encoder on ego-trees with query

Structure modeling

Model Design – Description Generation

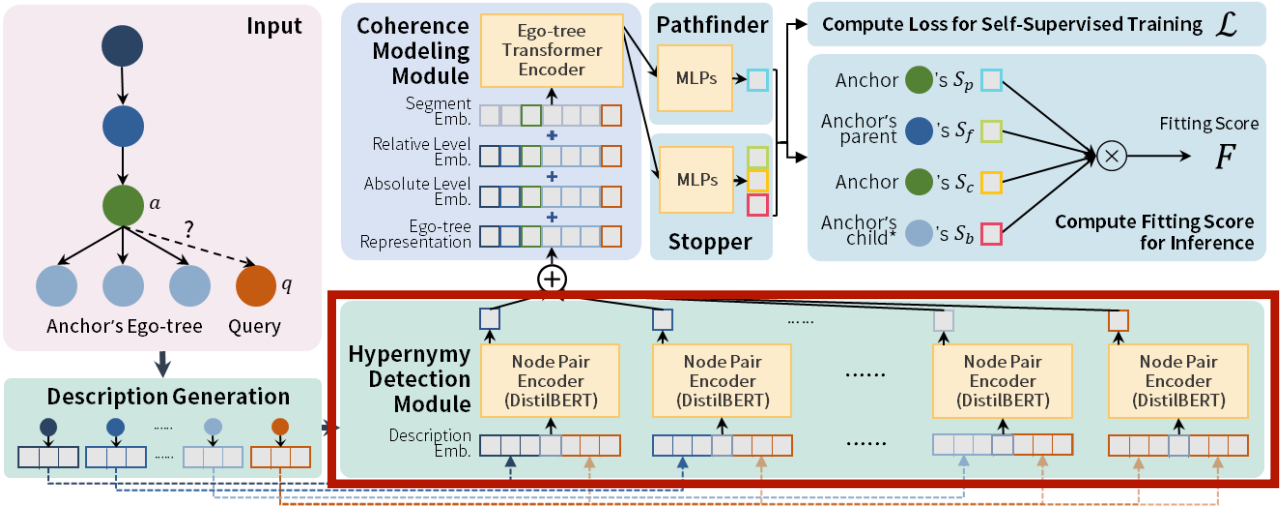


"Adaptation to Climate Change"

↓ algorithm based on DP and WordNet

"adaptation to climate change is the process of adapting to something (such as environmental conditions) to a change in the world's climate"

Model Design – Hypernymy Detection Module

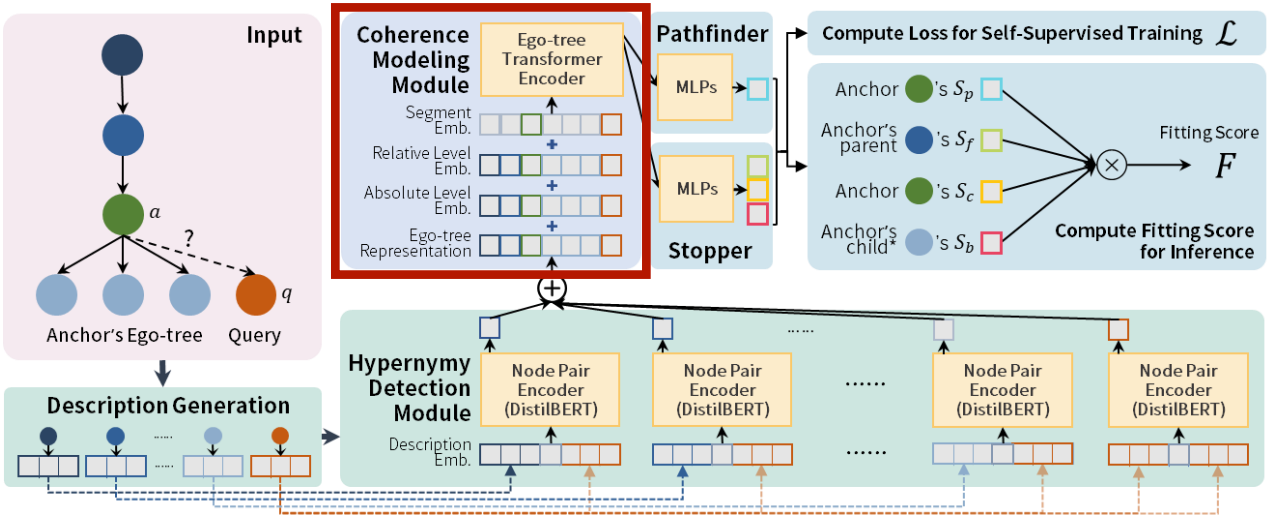


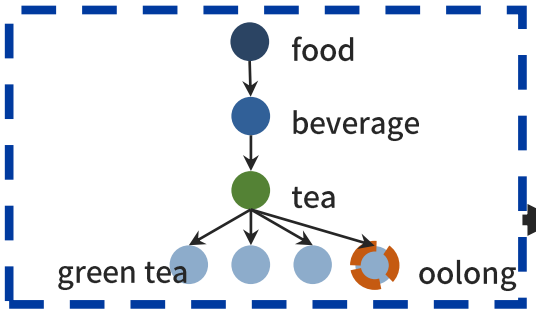
"oolong is Chinese tea leaves that have been partially femented…"

DistilBERT for hypernymy detection

"tea is a tropical evergreen shrub or small tree extensively cultivated…"

Model Design – Coherence Modeling Module

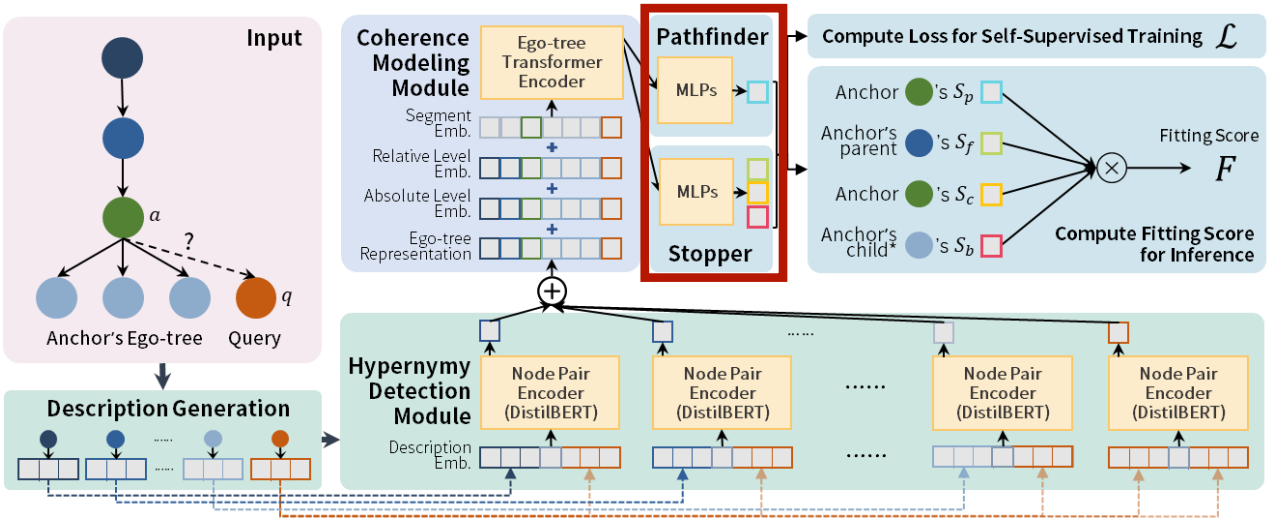




Tree-exclusive features as positional information
(Relative levels w.r.t. query, Absolute level in taxonomy)

Does the extended taxonomy maintains the original design?

Model Design – Pathfinder & Stopper



Tree structure enables dual evaluation:

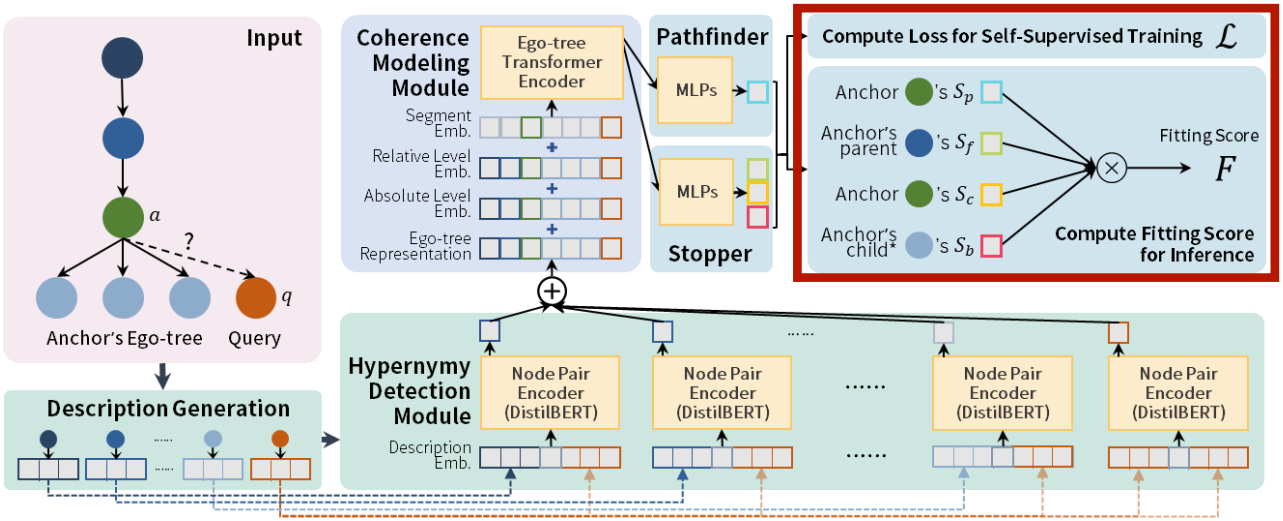
Pathfinder (Path evaluation)

S_p for path evaluation (hypernymy relation?)

Stopper (Level evaluation + comparison)

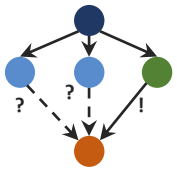
S_f for suggestion as parent (child may be better)
S_c for level evaluation (suitable level?)
S_b for suggestion as children (parent may be better)

Model Design – Pathfinder & Stopper

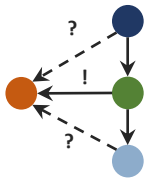


Fitting score integrating Path + Level, Parent + Children Selection

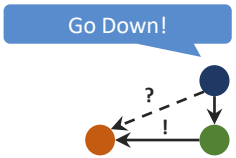
Path Selection



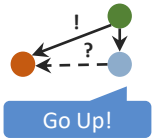
Level Selection



Parent's Selection



Child's Selection



Results of Main Experiments and Ablation Studies

	SemEval-16 Task 13 - Environment			SemEval-16 Task 13 - Science			SemEval-16 Task 13 – Food		
Methods	**Acc**	**MRR**	**Mean Wu-Palmer**	**Acc**	**MRR**	**Mean Wu-Palmer**	**Acc**	**MRR**	**Mean Wu-Palmer**
BERT+MLP	11.1	21.5	47.9	11.5	15.7	43.6	10.5	14.9	47.0
TAXI (ACL '16)	16.7	-	44.7	13.0	-	32.9	18.2	-	39.2
HypeNet (ACL '16)	16.7	23.7	55.8	15.4	22.6	50.7	20.5	27.3	63.2
TaxoExpan (WWW '20)	11.1	32.3	54.8	27.8	44.8	57.6	27.6	40.5	54.2
STEAM (SIGKDD '20)	36.1	46.9	69.6	36.5	48.3	68.2	34.2	43.4	67.0
HEF	**55.3**	**65.3**	**71.4**	**53.6**	**62.7**	**75.6**	**47.9**	**55.5**	**73.5**

SemEval-16 Task 13 Environment	Acc	MRR	Mean Wu-Palmer
HEF	**55.3**	**65.3**	**71.4**
- WordNet Descr.	41.5	55.3	62.6
- Ego-tree + Egonet	45.3	60.6	69.9
- Relative Level Emb.	49.1	59.2	60.9
- Absolute Level Emb.	49.1	60.6	68.4
Stopper Only	52.8	62.5	68.7
Pathfinder + Current Only	50.9	62.1	66.8
Current Only	41.5	54.7	58.6

	Environment	Science	Food
Nodes	261	429	1486
Edges	260	428	1485
Depth	6	8	8

Acc = Hit@1 = $\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_i = \hat{y}_i)$

MRR (Mean Reciprocal Rank) = $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\{rank(y_i)\}}$

Mean Wu-Palmer Similarity = $\frac{1}{n}\sum_{i=1}^{n}\frac{2\times depth(LCA(y_i,\hat{y_i}))}{depth(y_i)+depth(\hat{y_i})}$

Case Studies and Future Works

Query (q)	Ground Truth ($\hat{p}$)	Scores	Prediction (p)	Scores
q: **paddy** is rice in the husk either gathered or still in the field	$\hat{p}$: **rice** is grains used as food either unpolished or more often polished	$S_p = 0.9997$ $S_c = 0.4599$	p: **rice** is grains used as food either unpolished or more often polished	$S_p = 0.9997$ $S_c = 0.4599$
	parent($\hat{p}$): **starches** is a commercial preparation of starch that is used to stiffen textile fabrics in laundering	$S_f = 0.9755$	parent(p): **starches** is a commercial preparation of starch that is used to stiffen textile fabrics in laundering	$S_f = 0.9755$
$F(\hat{p}, q) = 0.4483$ $\hat{p}$'s Ranking: 1	$c^*_{\hat{p}}$: **white rice** is having husk or outer brown layers removed	$S_b = 0.9995$	c^*_p: **white rice** is having husk or outer brown layers removed	$S_b = 0.9995$
q: **fish meal** is ground dried fish used as fertilizer and as feed for domestic livestock	$\hat{p}$: **feed** is food for domestic livestock	$S_p = 0.9993$ $S_c = 0.3169$	p: **feed** is food for domestic livestock	$S_p = 0.9993$ $S_c = 0.3169$
	parent($\hat{p}$): **food** is any substance that can be metabolized by an animal to give energy and build tissue	$S_f = 0.9984$	parent(p): **food** is any substance that can be metabolized by an animal to give energy and build tissue	$S_f = 0.9984$
$F(\hat{p}, q) = 0.3158$ $\hat{p}$'s Ranking: 1	$c^*_{\hat{p}}$: **mash** is mixture of ground animal feeds	$S_b = 0.9988$	c^*_p: **mash** is mixture of ground animal feeds	$S_b = 0.9988$
q: **bourguignon** is reduced red wine with onions and parsley and thyme and butter	$\hat{p}$: **sauce** is flavorful relish or dressing or topping served as an accompaniment$\cdots$	$S_p = 0.0002$ $S_c = 0.0001$	p: **wine** is a red as dark as red wine	$S_p = 0.9997$ $S_c = 0.1399$
	parent($\hat{p}$): **condiment** is a preparation (a sauce or relish or spice) to enhance flavor or enjoyment	$S_f = 0.0004$	parent(p): **alcohol** is any of a series of volatile hydroxyl compounds that are made from hydrocarbons by distillation	$S_f = 0.9812$
$F(\hat{p}, q) = 1e-11$ $\hat{p}$'s Ranking: 328	$c^*_{\hat{p}}$: **bercy** is butter creamed with white wine and shallots and parsley	$S_b = 0.9997$	c^*_p: **red wine** is wine having a red color derived from skins of dark-colored grapes	$S_b = 0.8784$
q: **hot fudge sauce** is hot thick chocolate sauce served hot	$\hat{p}$: **chocolate sauce** is sauce made with unsweetened chocolate or cocoa$\cdots$	$S_p = 0.9471$ $S_c = 9e-5$	p: **sauce** is flavorful relish or dressing or topping served as an accompaniment$\cdots$	$S_p = 0.9995$ $S_c = 0.0172$
	parent($\hat{p}$): **sauce** is flavorful relish or dressing or topping served as an accompaniment$\cdots$	$S_f = 0.9617$	parent(p): **condiment** is a preparation (a sauce or relish or spice) to enhance flavor or enjoyment	$S_f = 0.9888$
$F(\hat{p}, q) = 6e-6$ $\hat{p}$'s Ranking: 20	$c^*_{\hat{p}}$: **None**	$S_b = 0.0700$	c^*_p: **lyonnaise sauce** is brown sauce with sauteed chopped onions and parsley$\cdots$	$S_b = 0.9995$

3rd: Bourguignon is also a kind of sauce for cooking meat. The description is incorrect.

4th: Leaf nodes' backward scores are assigned to be a constant. The backward score assignment is incorrect.

What's Next?

No WordNet descriptions for Chinese?: Discover approaches for generating useful descriptions suggesting hypernymy relations for Chinese concepts

New ways for integrating larger corpus: Find ways to use larger corpus or query logs to boost the performance of hypernymy detection / coherence modeling

···and to fix the bad cases:

Better description generation algorithm: Develop a DP+ML algorithm to generate domain-related descriptions for terms

Backward scores for leaf node's inexistent child: New approaches for obtaining the backward scores

Some useful links

Our paper: https://arxiv.org/abs/2101.11268

An awesome list for taxonomy-related research:
https://github.com/mickeystroller/awesome-taxonomy

Thanks