Knowledge Acquisition with Transferable Representation Learning

Muhao Chen
University of Southern California
http://muhaochen.github.io

Understanding Relations Is Prominent In Practice

QA and Semantic Search

Google

mazda car that won 24 Hours of Le Mans

All | Images | News | Shopping | Videos | More | Settings | Tools

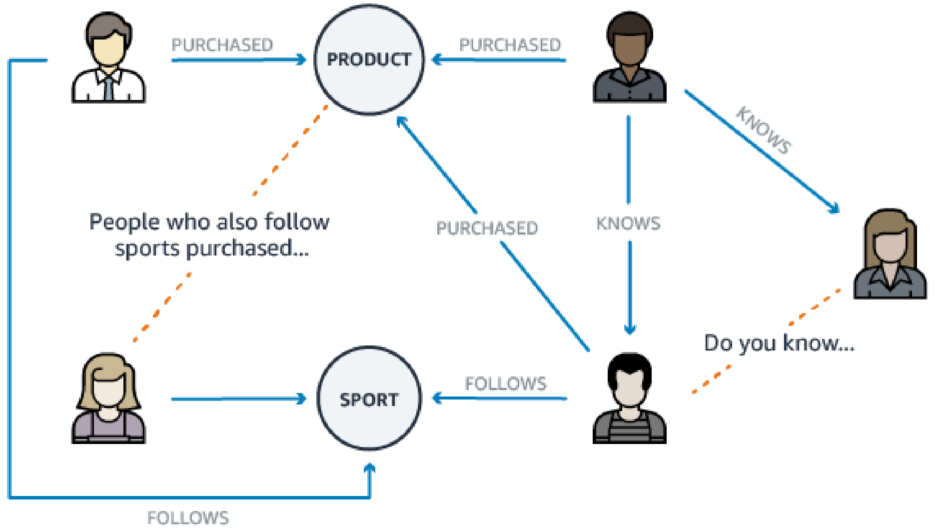
About 34,600,000 results (1.04 seconds)

787B

(?car, ***produced by***, Mazda)
(?car, ***won***, 24 Hours of Le Mans)

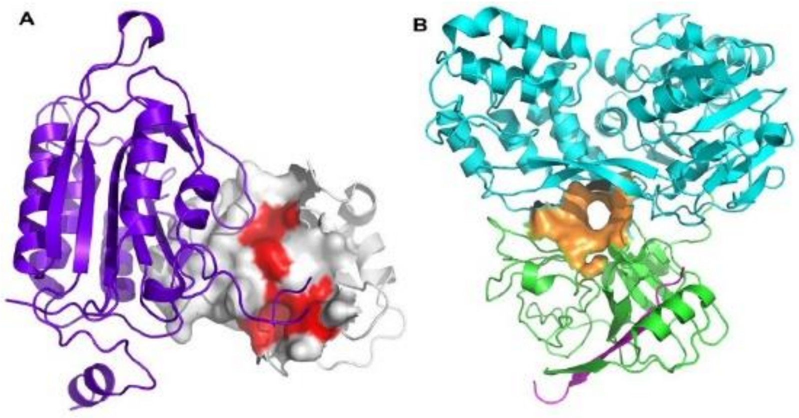
Understanding Relations Is Prominent In Practice

Recommender Systems



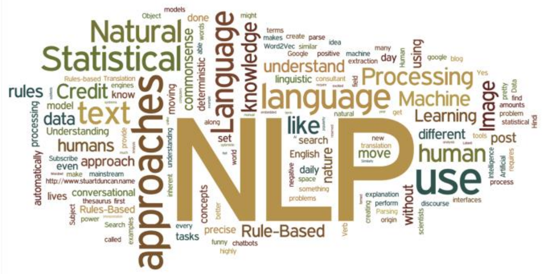
- Co-purchase relations of products
- Social relations of users

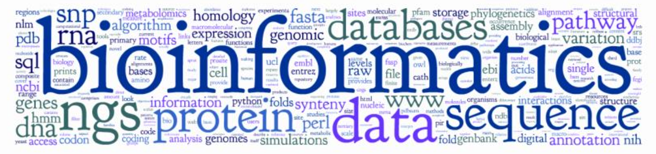
Computational Biology Research



- Interactions of molecules and biomolecules.

Understanding Relations Is Prominent In Practice





- QA
- Discourse relation detection
- Dialogue state tracking
- Event prediction
- Narrative cloze
- Entity/event typing and linking

- Semantic search
- Relational rule mining
- Ontology population
- Ontology matching and knowledge integration

- Interaction prediction of biomolecules
- Mutation effect estimation
- Non-coding RNA alignment
- Drug discovery
- Polypharmacy side effect detection

Multi-relational Data

Knowledge Graphs

Bai du 知识图谱

ConceptNet
An open, multilingual knowledge graph

Google
KNOWLEDGE GRAPH

DBpedia

WIKIDATA

yago
select knowledge

NELL

Freebase

Bio-med Ontologies /Data Banks

MeSH

SKEMPI

neXtprot

STRING

PubMed

GENEONTOLOGY
Unifying Biology

RCSB PDB
PROTEIN DATA BANK

Product & E-Commerce Graphs

product graph

ebay

Alibaba.com

Lexical and Semantic Graphs

WORDNET

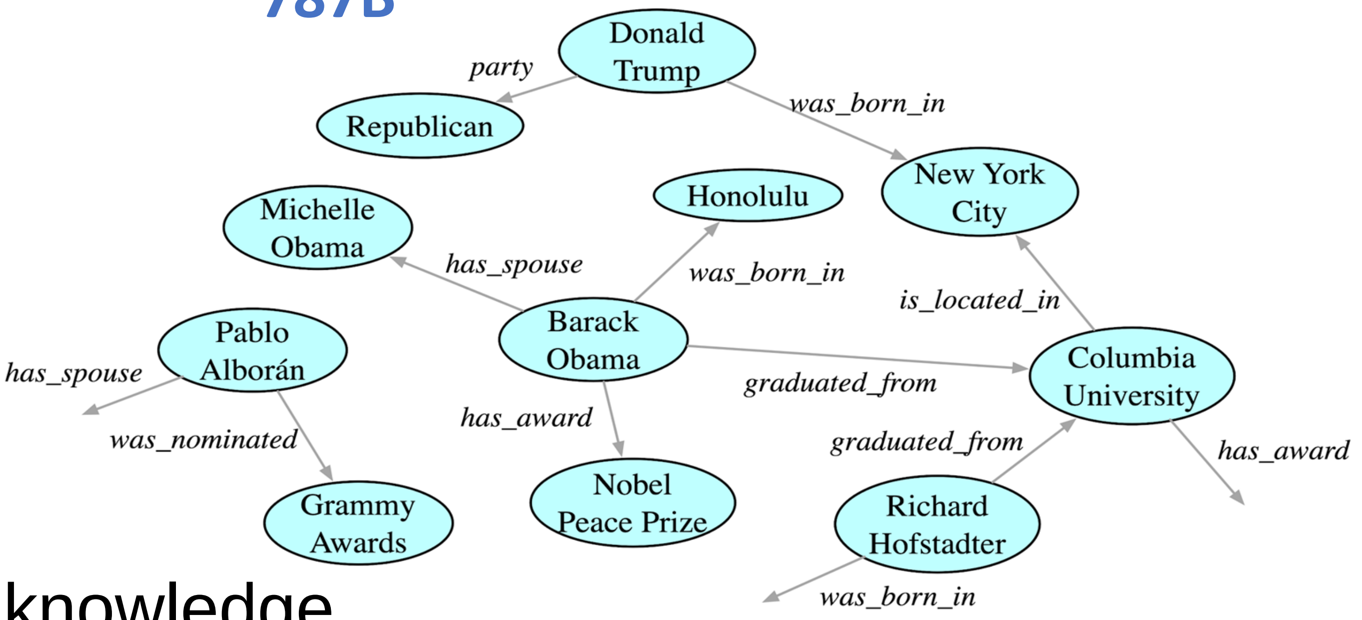
FrameNet

VERBNET

Wiktionary
The free dictionary

Multi-relational Data: Precise But **Expensive** Knowledge Representation

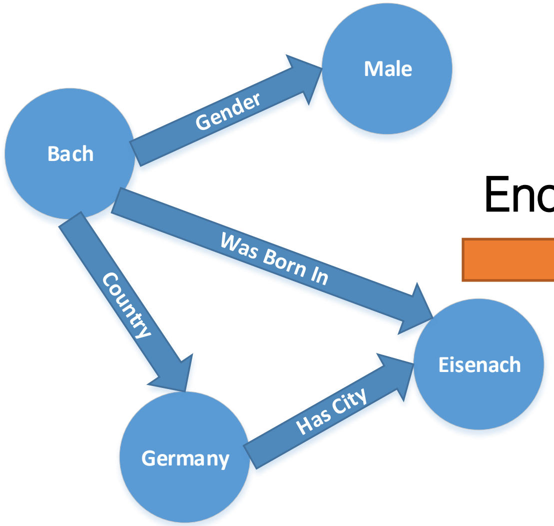
WIKIDATA



Obtaining the structural knowledge

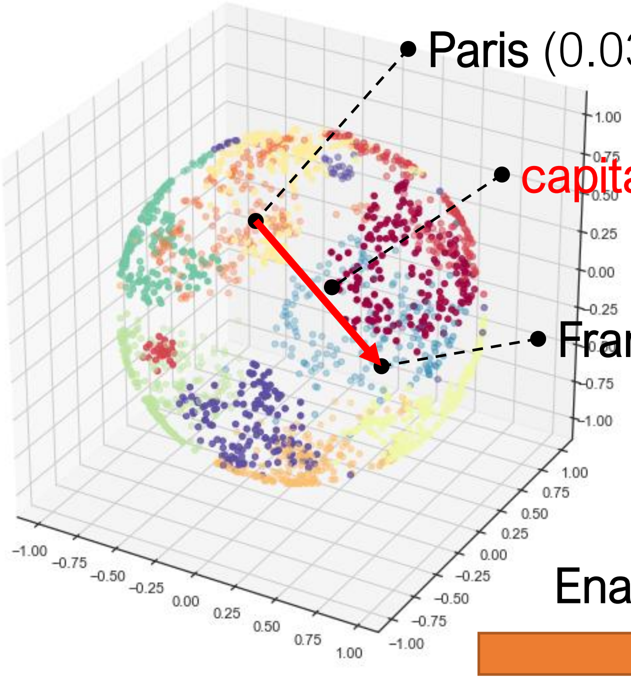
- Is expensive (Avg $5.71 per triple **[Paulheim+, ISWC-18]** in open domain; higher cost in scientific domains).
- Has relied on massive human efforts.
- Has never been close to complete.

Representation Learning: Cheap Knowledge Acquisition from The Embedding Space



Multi-relational data

Encode



Embeddings

Enable

Relational inference as inductive bias (e.g. translational techniques [Bordes+ NIPS-13])

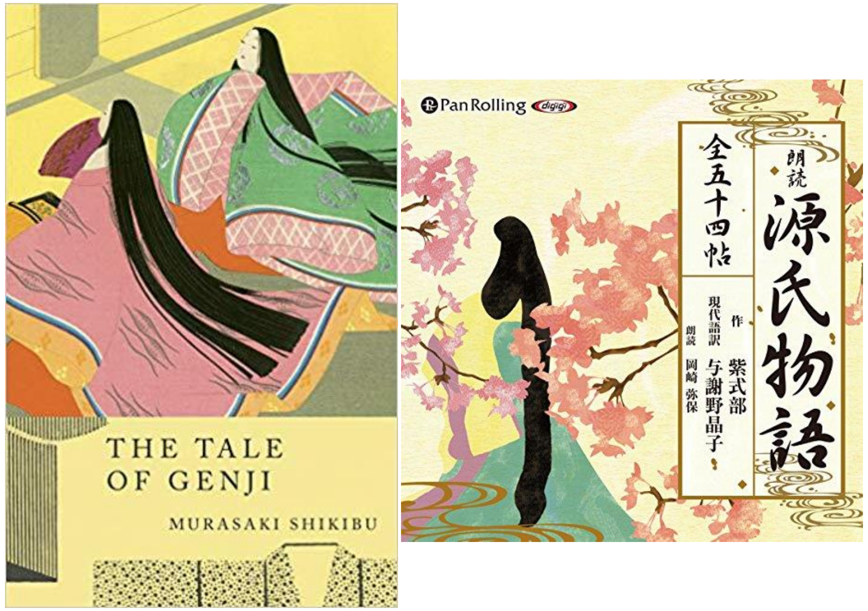
- France – Paris≈ capital
- USD – US ≈ currency
- Bach – German ≈ nationality
- …

Automatically predicting knowledge: 787B + ProducedBy ≈ Mazda

- A much less expensive way for knowledge acquisition
- Yet can still suffer from sparsity and noise of known knowledge

Knowledge Is Not Isolated

Different sources of data can possess complementary knowledge



(The Tale of Genji, *Genre*, **?e**)

DBpedia

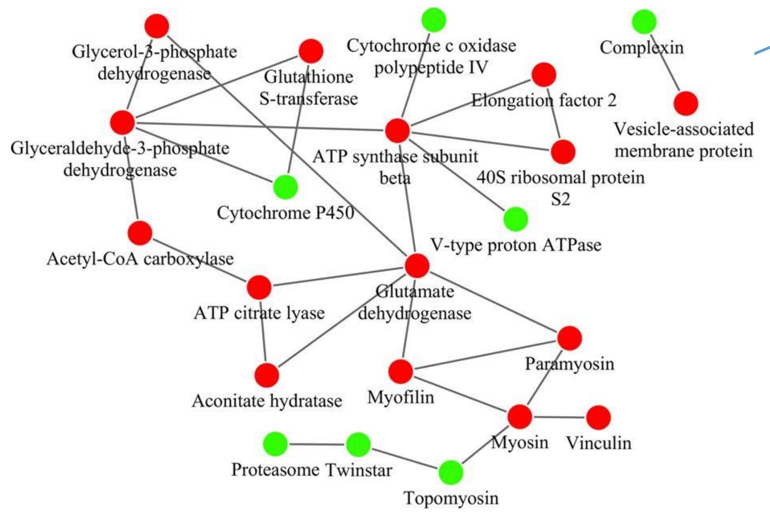
Novel

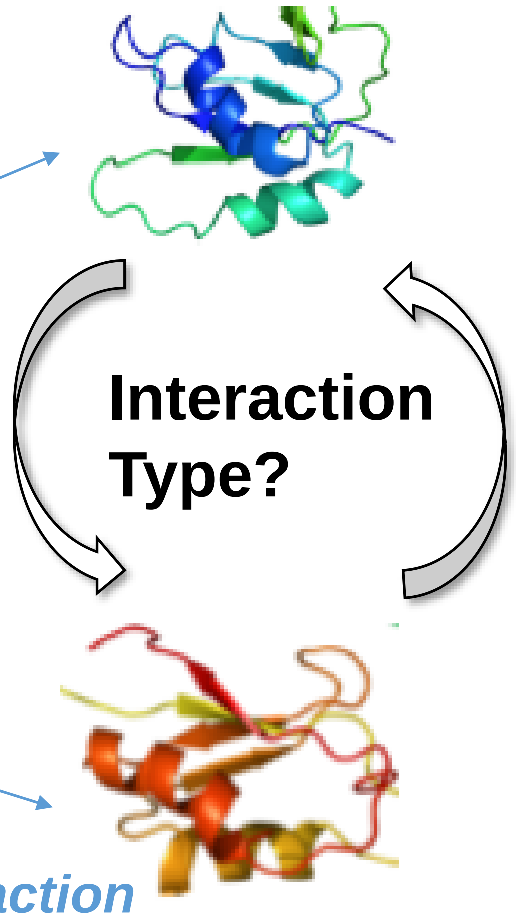
ヂビペディア

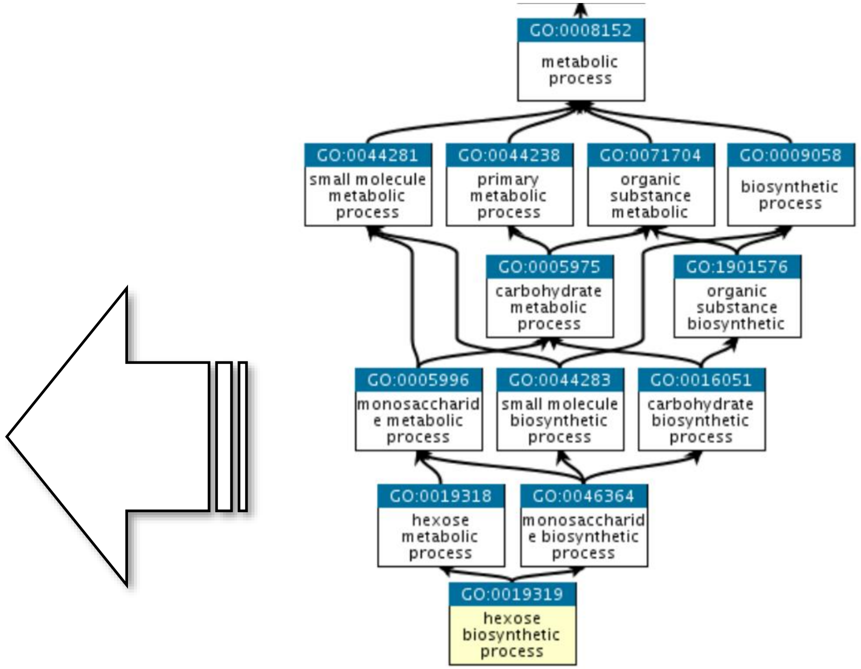
Monogatari (story)
Love story
Royal family story
Realistic novel
Ancient literature

Knowledge Is Not Isolated

Different sources of data can possess complementary knowledge







GENEONTOLOGY
Unifying Biology

Inferring unknown ***protein-protein interaction*** information in **a proteomic knowledge base**

Key Research Questions

Interrelated knowledge in different domains/sources

- Multiple language-specific KGs
- Multiple knowledge bases
- Instance KGs and concept ontologies (different specificity)
- Protein-protein interaction (PPI) data, gene ontologies and cell clusters
- Drug-drug interaction data, disease ontologies and PPI data
- Social networks and product graphs
- …

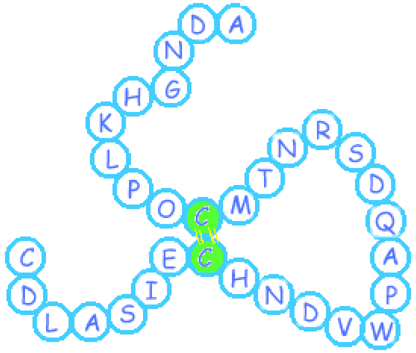
Can we capture the association of knowledge with representation learning?

And use knowledge transfer to populate missing knowledge?

Key Research Questions

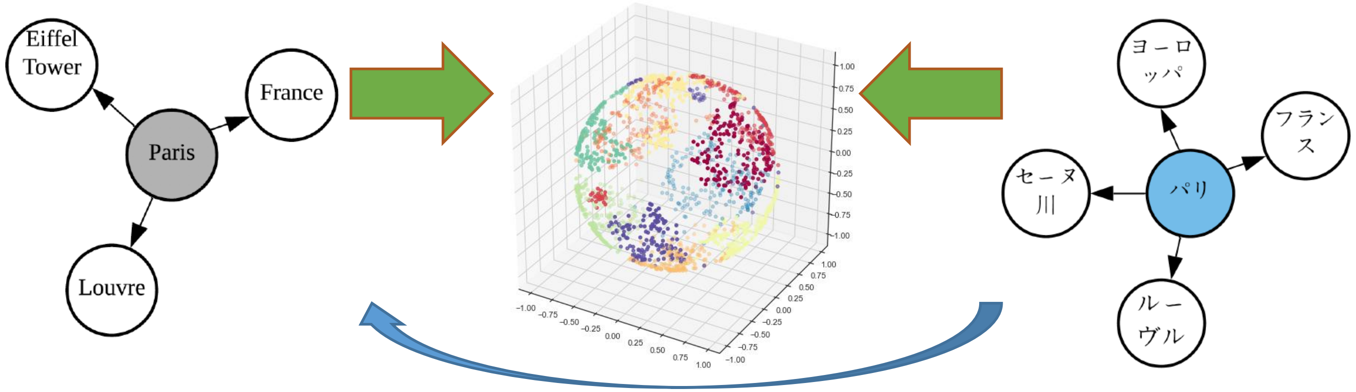
How to acquire structured knowledge from unstructured data?

- Provide globally consistent inference
- Learning to acquire knowledge with limited and indirect supervision
- Acquisition from modalities beyond human languages (molecular and biomolecular sequences, EHR, etc.)



Roadmap of Research Contributions

Transferable Rep. Learning for Relational Data



How do we capture the **association of knowledge** with **minimal supervision**?

How do we identify and transfer **complementary knowledge**?

Knowledge Acquisition from Unstructured Data

How to provide reliable inference (e.g. ensuring the **logical** or **probabilistic constraints**)?

Generalizable learning, but with **limited supervision**?

Roadmap of Research Contributions

Transferable Rep. Learning for Relational Data

Knowledge Acquisition from Unstructured Data

Method

Minimally supervised knowledge alignment
- Semi-suervised knowledge alignment (first prototype) [IJCAI-17, AKBC-17]◀
- Co-training [IJCAI-18]◀
- Distant supervision [KDD-19]◀
- Visual pivoting [AAAI-21a]◀
- Incidental supervision [EACL-21]◀

Robust embedding learning and knowledge transfer
- Property-aware embedding [SDM-18]
- Hyperbolic embedding [SIGIR-19, EMNLP-20a]◀
- Multi-view learning [IJCAI-19]
- Noise-aware GNN [AAAI-20a]
- Meta-learnable knowledge transfer [EMNLP-20b]◀

Learning with constraints and indirect supervision
- Logical constraints [EMNLP-20c]◀
- Probabilistic soft constraints [AAAI-19]
- Few-shot learning with indirect supervision [CoNLL-20, Best Paper Nomination]◀

Robust and generalizable learning and inference
- Paraphrase-aware retrofitting [EMNLP-19]
- Analogy-aware inference [EMNLP-20d]
- Language modeling for proteins [ISMB-19, *Bioinformatics* [J] 2019, NAR GaB [J] 2020]

Tasks

KB Construction
- KB Completion [AAAI-19, EMNLP-20b]
- Entity alignment [many above]
- Type inference [KDD-19,EMNLP-20a]◀

Natural Language Understanding
- Relation extraction [EMNLP-20c]
- Event prediction [EMNLP-20d]
- Event process typing [CoNLL-20, Best Paper Nomination]◀
- DocRel extraction [ECML-18, Plenary]

Bio/medical Informatics
- Proteomics [ISMB-19,Bioinformatics [J] 2019, NAR GaB [J] 2020]◀
- Diagnostic prediction [AIME-20]
- Disease target prediction [ACM BCB-20, Best Student Paper]◀

Outreach

- Benchmarking and survey paper [PVLDB 2020]
- Transferable Representation Learning Tutorial [AAAI-20b]

- Knowledge Acquisition and Event-centric NLU tutorials [AAAI-21b, ACL-21]

Outline

Transferable Representation Learning of Multi-relational Data

Knowledge Acquisition from Unstructured Data

Future research agenda

Transferable Representation Learning for Multi-relational Data

Capture the knowledge association with minimal supervision in a universal embedding scheme for

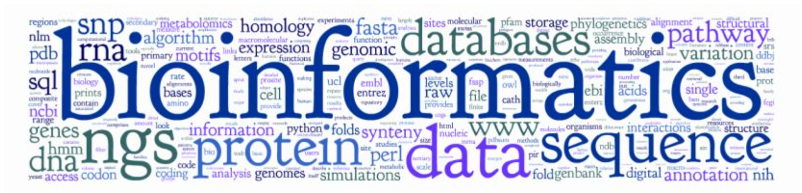
- Multiple language-specific KGs
- Multiple KBs
- Abstract concepts and specific entities
- Proteomic interactions and gene ontologies
- Cells and genomic interaction data
- Molecular data, medical ontologies and drug interaction data
- Social relations and product graphs
- …

Transfer knowledge from some domains to enrich others

A General Methodology to Benefit A Wide Range of Tasks

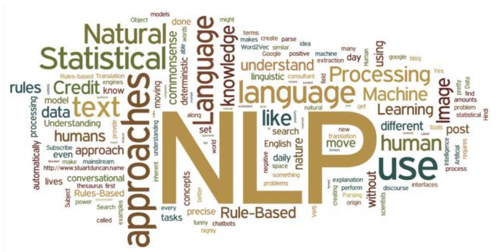
Knowledge Base

Knowledge alignment
KG completion
Ontology population



Protein-protein interaction prediction
Protein binding affinity estimation
Single cell RNA-sequence imputation
Gene Ontology term assignment

Transferable representation learning

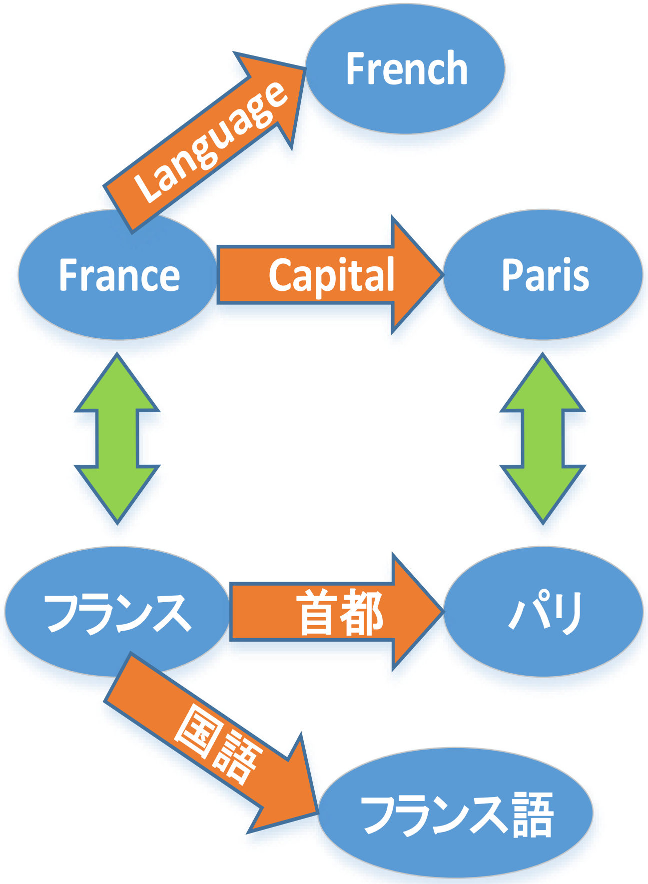


Semantic search
Entity typing
Dialogue state tracking
Paraphrase identification

Medical
INFORMATICS

Polypharmacy side effect detection
Disease and phenotype matching
Clinical event prediction

Scenario 1: KGs in Different Languages



Separately created language-specific KGs

- DBpedia has 125 language-specific versions; Wikidata has 410 of those.

ConceptNet
An open, multilingual knowledge graph

DBpedia

yago
select knowledge

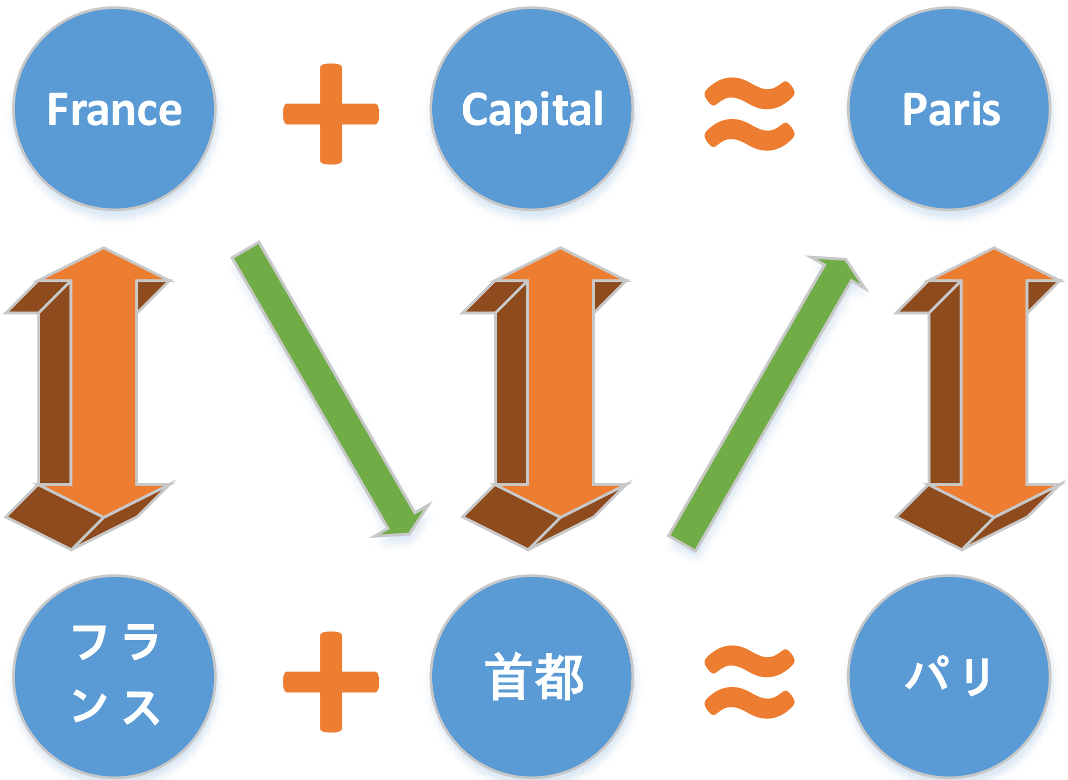
WIKIDATA

Wiktionary
Das freie Wörterbuch

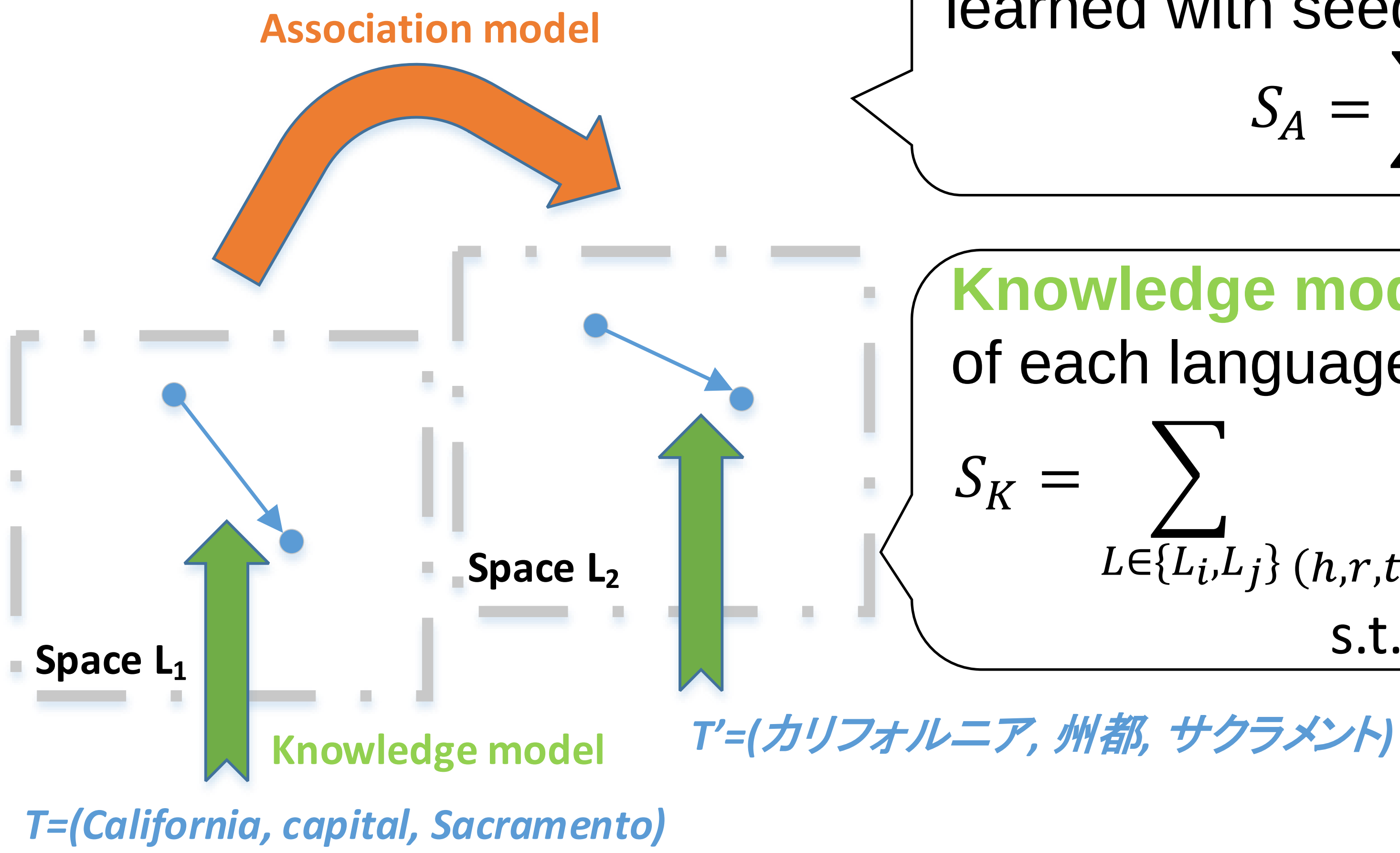
The First Prototype: Simple Translational Model + Supervised Association Learning (MTransE*)

*[IJCAI-17]

- **Training data**: a pair of weakly aligned language-specific KGs
- **Enabling**: cross-lingual semantic transfer + relational inference



Joint Learning of MTransE



Association model: an embedding transformation learned with seed alignment

$$S_A = \sum_{(e,e')\in\delta(L_i,L_j)} \left\| \mathbf{M}_{ij}\mathbf{e} - \mathbf{e}' \right\|$$

Knowledge model: encoding entities and relations of each language as a **translational embedding**

$$S_K = \sum_{L\in\{L_i,L_j\}} \sum_{(h,r,t)\in G_L \wedge (\hat{h},r,\hat{t})\notin G_L} \left[f_r(h,t) - f_r(\hat{h},\hat{t}) + \gamma\right]_+$$

s.t. $f_r(h,t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2$

- **Joint training loss**

$$S_J = S_K + \alpha S_A$$

*[IJCAI-17]

Application: Knowledge Alignment

Table 8: Examples of cross-lingual entity matching.

Entity	Target	Candidates (in ascending order of rank by Euclidean distance)
Barack Obama	French	**Barack Obama**, *George Bush*, *Jimmy Carter*, George Kalkoa
	German	**Barack Obama**, *Bill Clinton*, *George h. w. Bush*, Hamid Karzai
Paris	French	**Paris**, *Amsterdam*, **à Paris**, *Manchester*, De Smet
	German	**Paris**, *Languedoc*, *Constantine*, *Saint-maurice*, *Nancy*
California	French	*San Francisco*, *Los Angeles*, *Santa Monica*, **Californie**
	German	**Kalifornien**, *Los Angeles*, *Palm Springs*, *Santa Monica*

Table 9: Examples of cross-lingual relation matching.

Relation	Target	Candidates (in ascending order of rank by Euclidean distance)
capital	French	**capitale**, *territoire*, pays accrèditant, lieu de veneration
	German	**hauptstadt**, *hauptort*, *gründungsort*, *city*
nationality	French	**nationalié**, **pays de naissance**, *domicile*, *résidence*
	German	**nationalität**, **nation**, letzter start, *sterbeort*
language	French	**langue**, réalisations, lieu deces, *nationalitè*
	German	**sprache**, **originalsprache**, **lang**, *land*

This pilot study got ~30% Hits@1 on DBP15k. But we will introduce lots of improvement to it shortly.

Bold-faced ones are correct answers, *italic* ones are close answers.
Answers do not include those that have pre-existed in training.

Cross-lingual Fact Prediction, e.g.

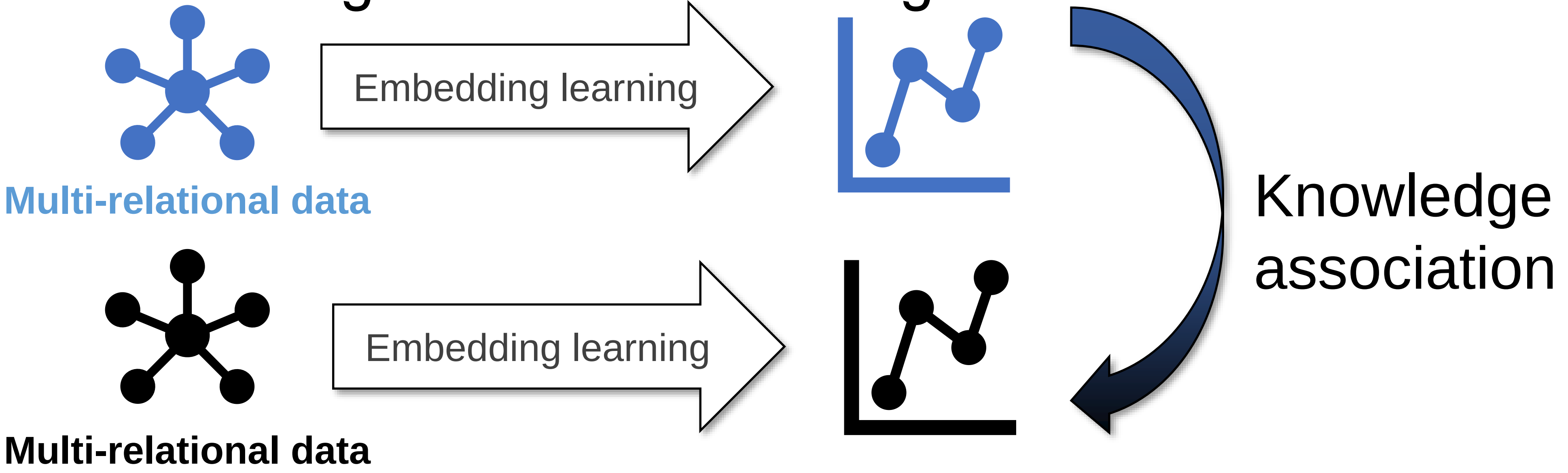
Table 10: Examples of cross-lingual triple completion.

Query	Target	Candidates (in ascending order of rank)
(Adam Lambert, genre, ?t)	French	*musique indèpendante*, **musique alternative**, ode, **glam rock**
	German	**popmusik**, **dance-pop**, no wave, *soul*
(Ronaldinho, position, ?t)	French	**milieu offensif**, **attaquant**, *quarterback*, *latèral gauche*
	German	**stürmer**, *linker flügel*, **angriffsspieler**, *rechter flgel*
(Italy, ?r, Rome)	French	**capitale**, **plus grande ville**, **chef-lieu**, garnison
	German	**hauptstadt**, **hauptort**, verwaltungssitz, stadion
(Barack Obama, ?r, George Bush)	French	*ministre-prèsident*, **prèdècesseur**, *premier ministre*, *prèsident du conseil*
	German	**vorgänger**, **vorgängerin**, besetzung, lied

Bold-faced ones are correct answers, *italic* ones are close answers.
Answers do not include those that have pre-existed in training.

General Framework and Further Improvement

Jointly or iteratively conduct two learning processes: **embedding learning** and **knowledge association learning**



Three directions to improvement

1. More precise embedding alignment requiring less supervision
2. Auxiliary supervision from entity profile information
3. Better embedding learning techniques for inconsistent structures

(1) Semi-supervised Co-training With Entity Descriptions*

*[IJCAI-18]

The alignment information is often limitedly provided to connect KG structures

Iterative co-training of embeddings for KG structures and entity descriptions

DBpedia covers less than 20% entity alignment for En-Fr, and less for other cases.

Inter-lingual Link (ILL): (***astronomer***@EN, ***astronome***@FR)

EN triple: (***Ulugh Beg***, *occupation*, ***astronomer***) FR triple: (***Ulugh Beg***, *activité*, ***astronome***)

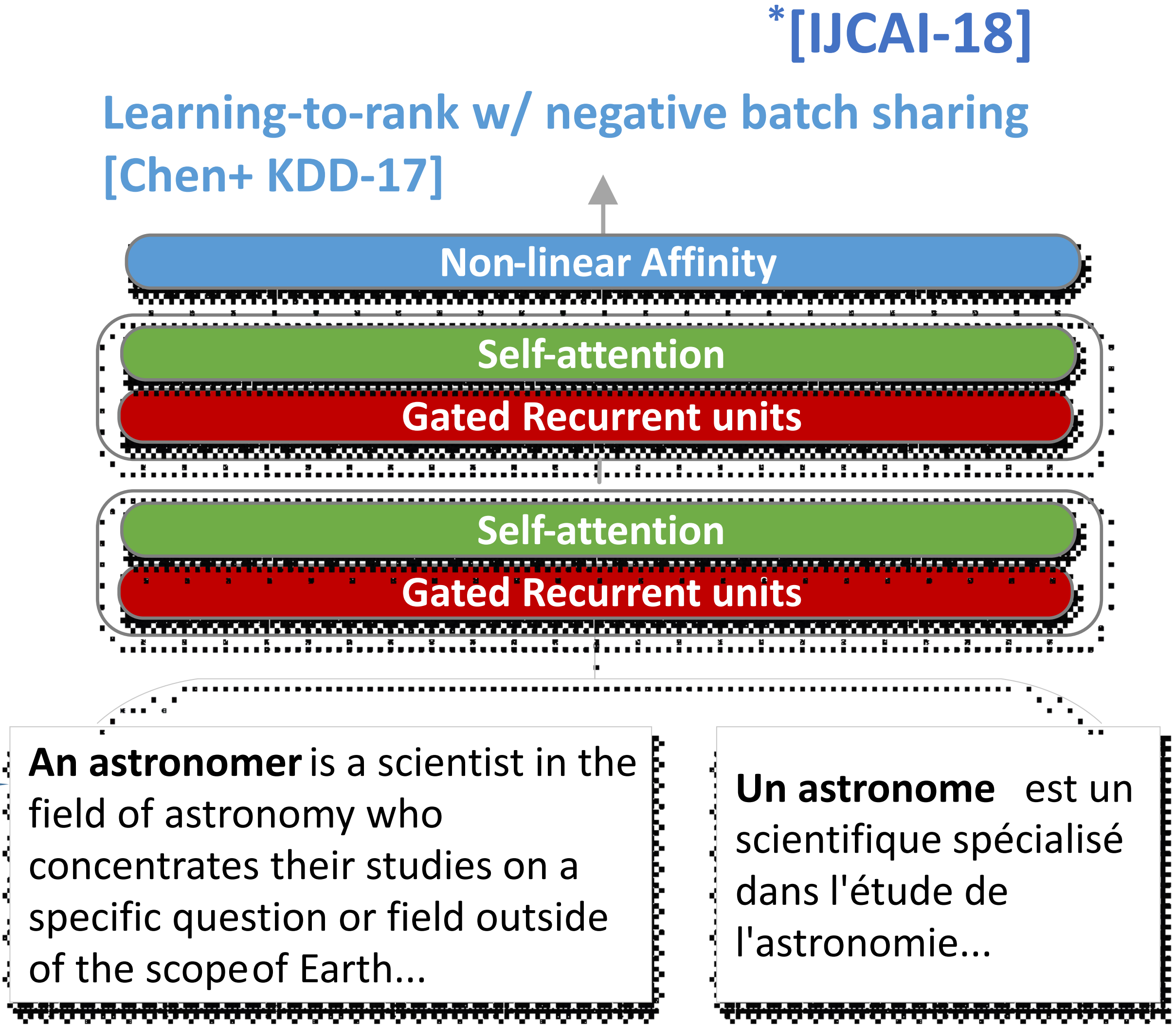
An astronomer is a scientist in the field of astronomy who concentrates their studies on a specific question or field outside of the scope of Earth...

Un astronome est un scientifique spécialisé dans l'étude de l'astronomie...

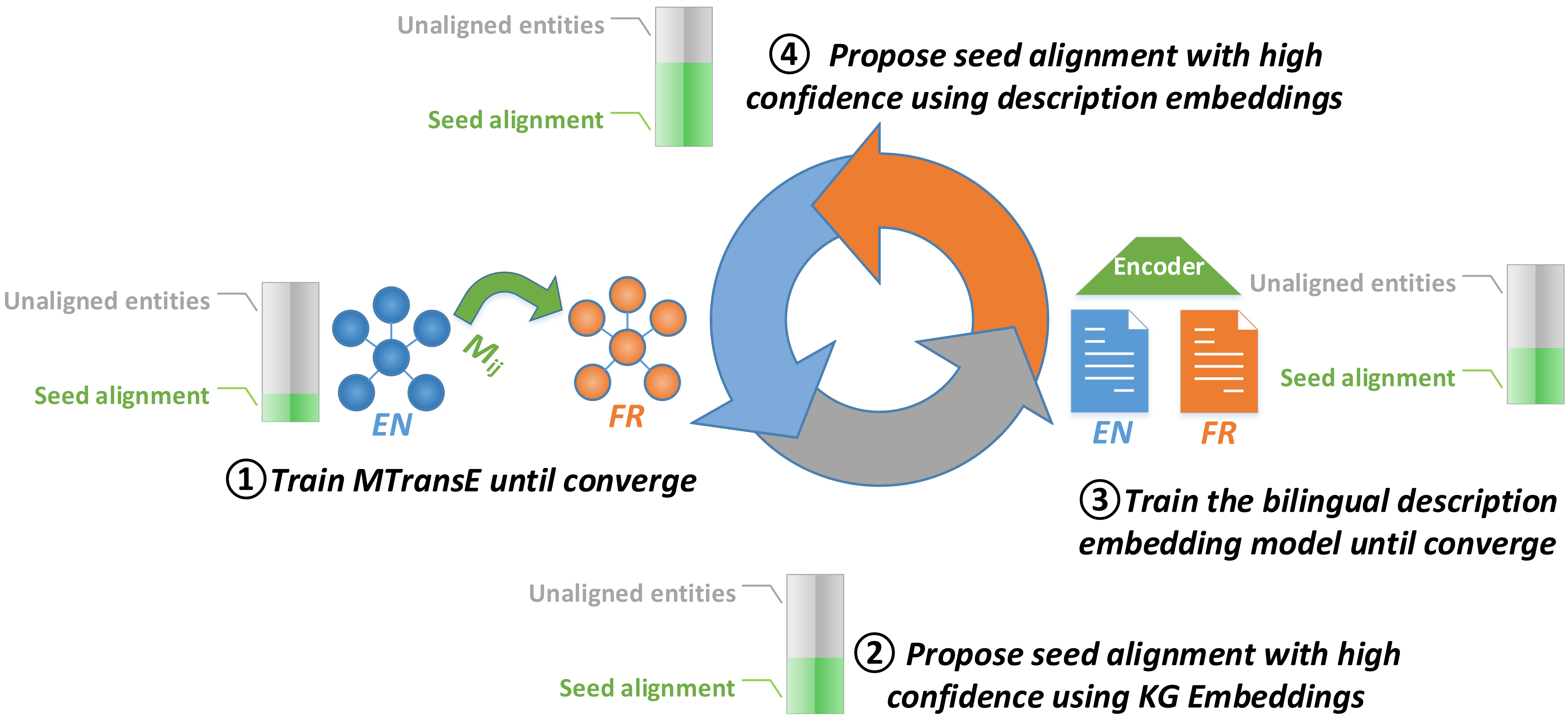
An Entity Description Embedding Model

Siamese document encoder with Self-attention + Pre-trained bilingual word embeddings

To collocate the embeddings of entity description counterparts

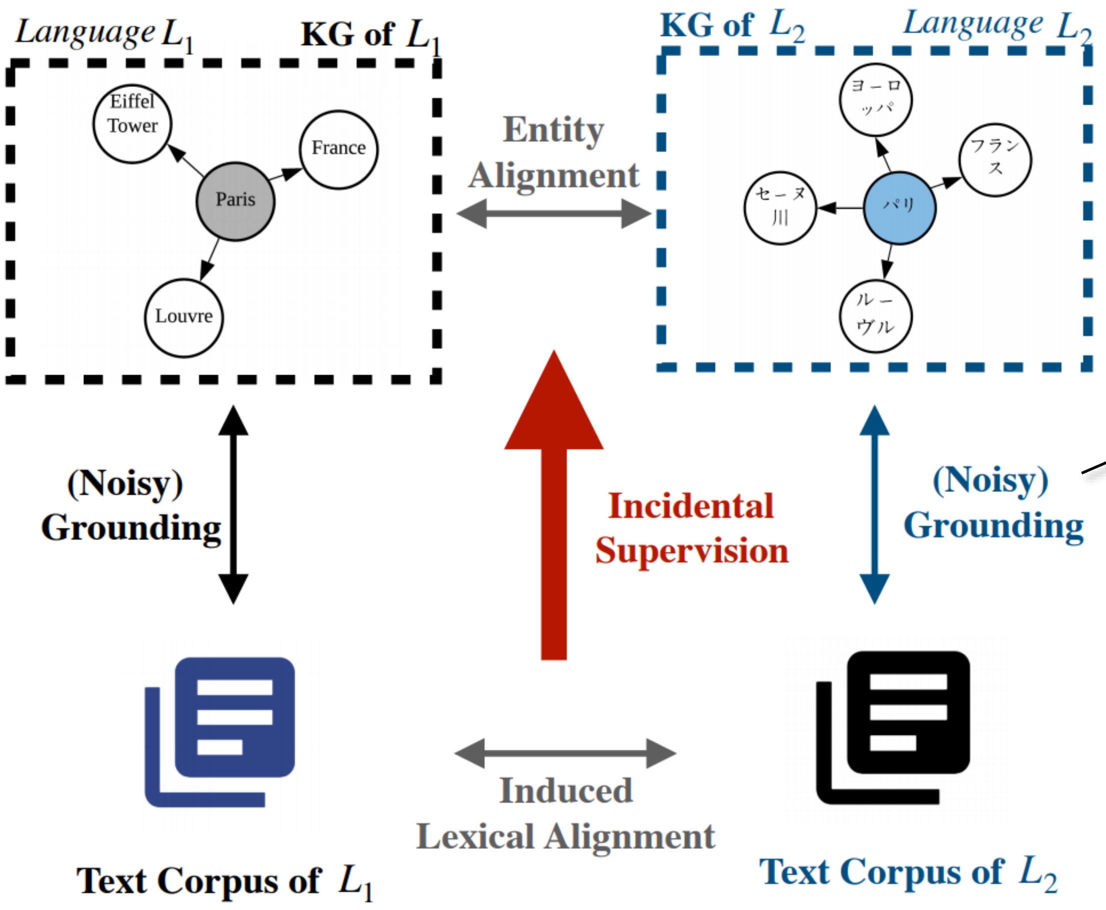


Iterative Co-training Process



(2) Knowledge Alignment Using Incidental Supervision From Free Text*

*[EACL'21 in review]



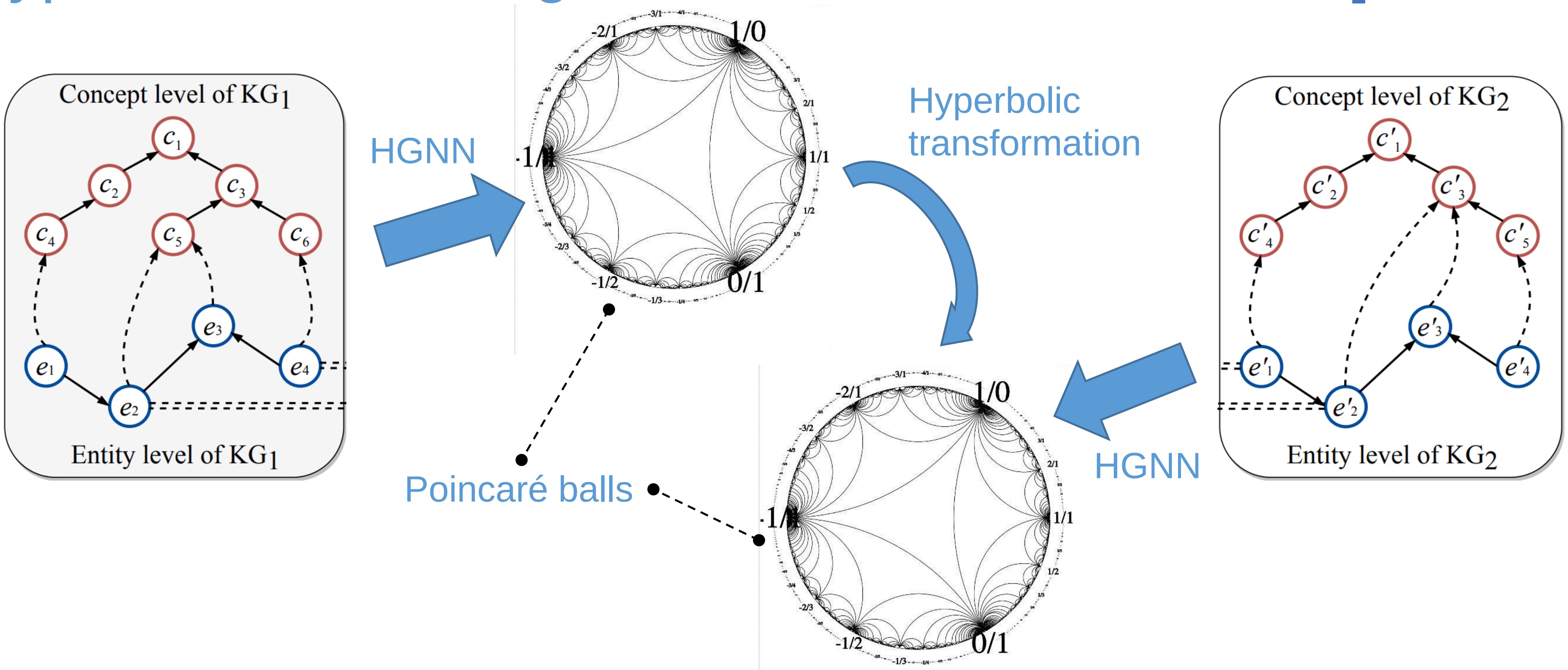
(Noisy) grounding process for text corpora
- Entity discovery and linking (EDL)
- Surface form matching

Three steps

1. ***(Noisy) grounding***: connecting KGs and text corpora
2. ***Embedding learning***: GNN + a neural language model
3. ***Alignment learning***: self-learning for both entity and lexical alignment

(3) Hyperbolic Knowledge Association*

***[EMNLP-20a]**



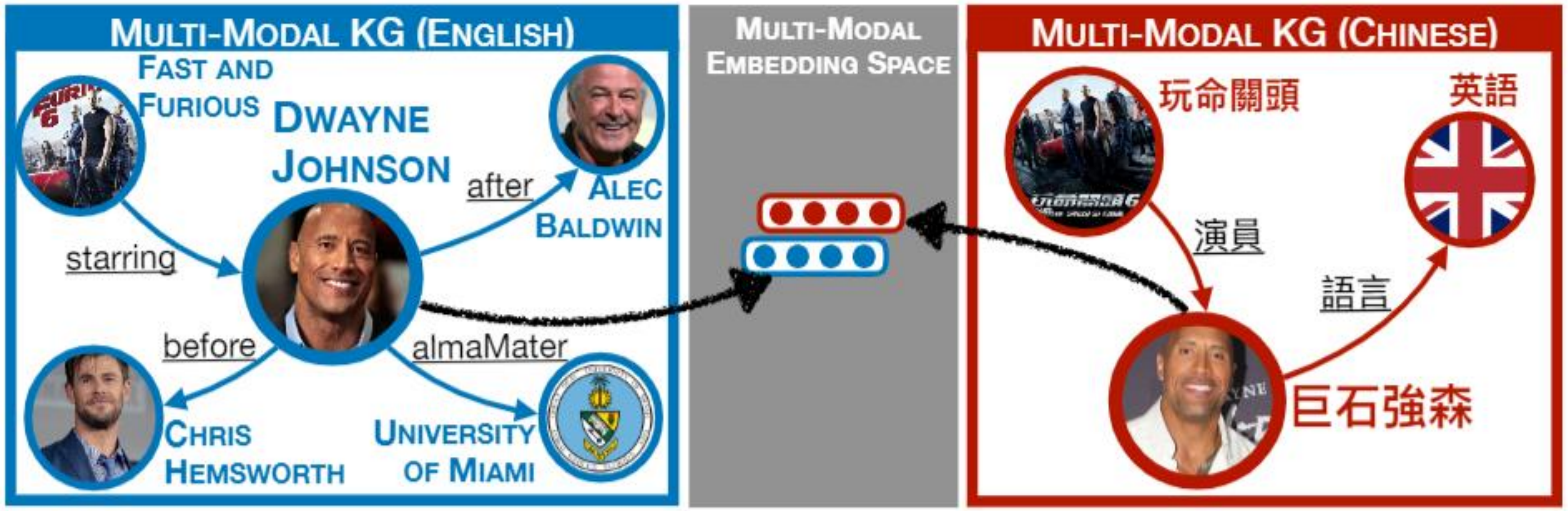
Transferable, ultra low-dimensional **hyperbolic** embeddings **(15-30)**.

- Non-linear distance metric $d_{\mathbb{D}}(\mathbf{u},\mathbf{v}) = \operatorname{arccosh}(1+2\frac{\|\mathbf{u}-\mathbf{v}\|^2}{(1-\|\mathbf{u}\|^2)(1-\|\mathbf{v}\|^2)})$
- Suitable for capturing knowledge association between hierarchical KGs.
- and KGs with significantly different scales (e.g. an instance-graph vs a concept graph).

Also applied to entity type inference.

(4) Multi-modal Entity Alignment*

*[AAAI-21]

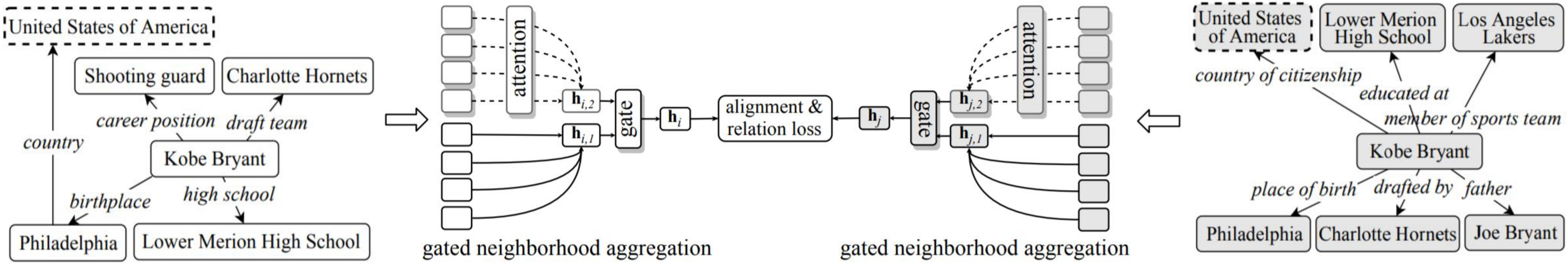


(Unsupervised) visual pivoting by identifying visually similar entities

- ResNet + GCN with bootstrapping
- Particularly benefits long-tail entities

(5) Noise-aware Multi-hop Graph Attention⁺

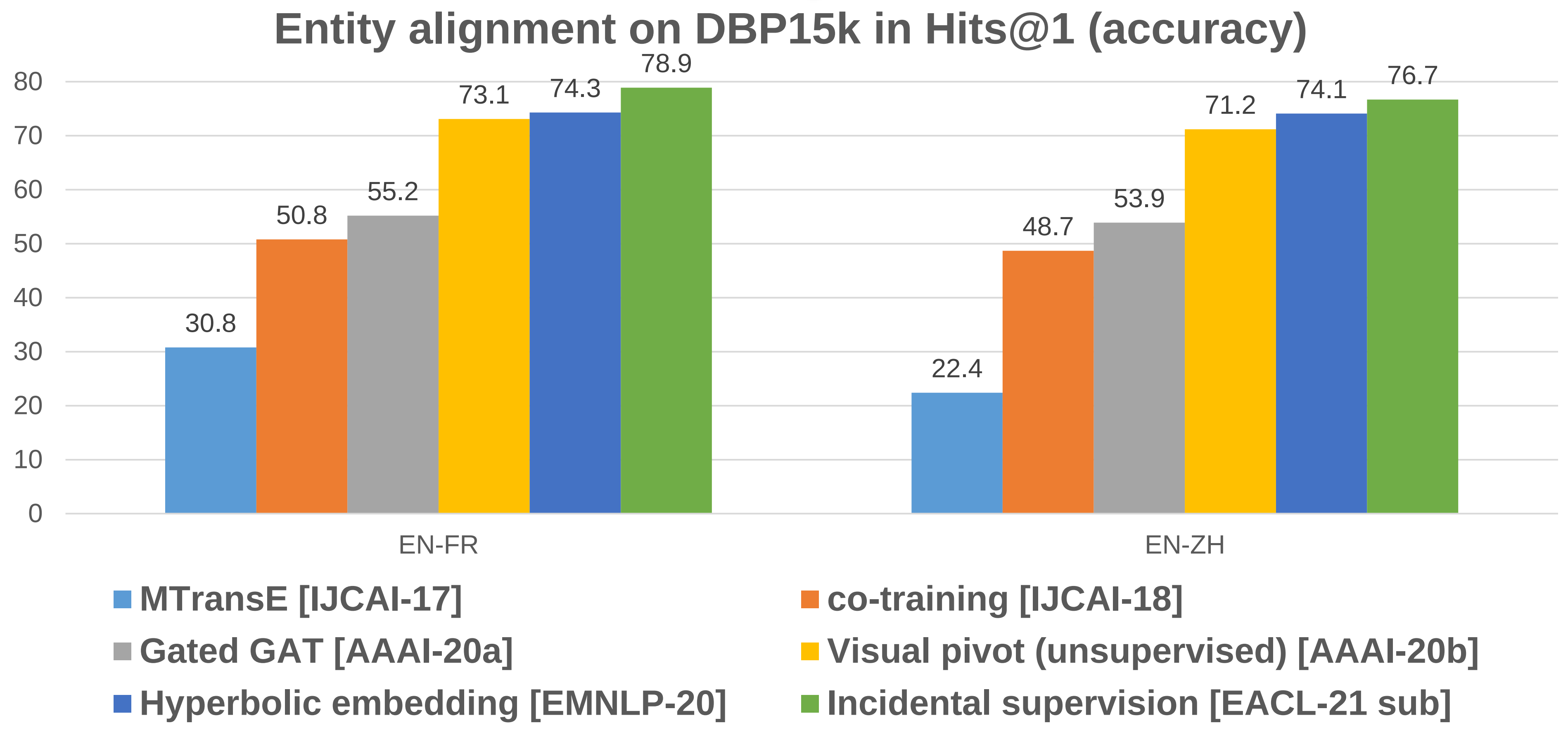
+[AAAI-20]



A robust GNN encoder against the inconsistency of entity neighborhoods in different KGs.

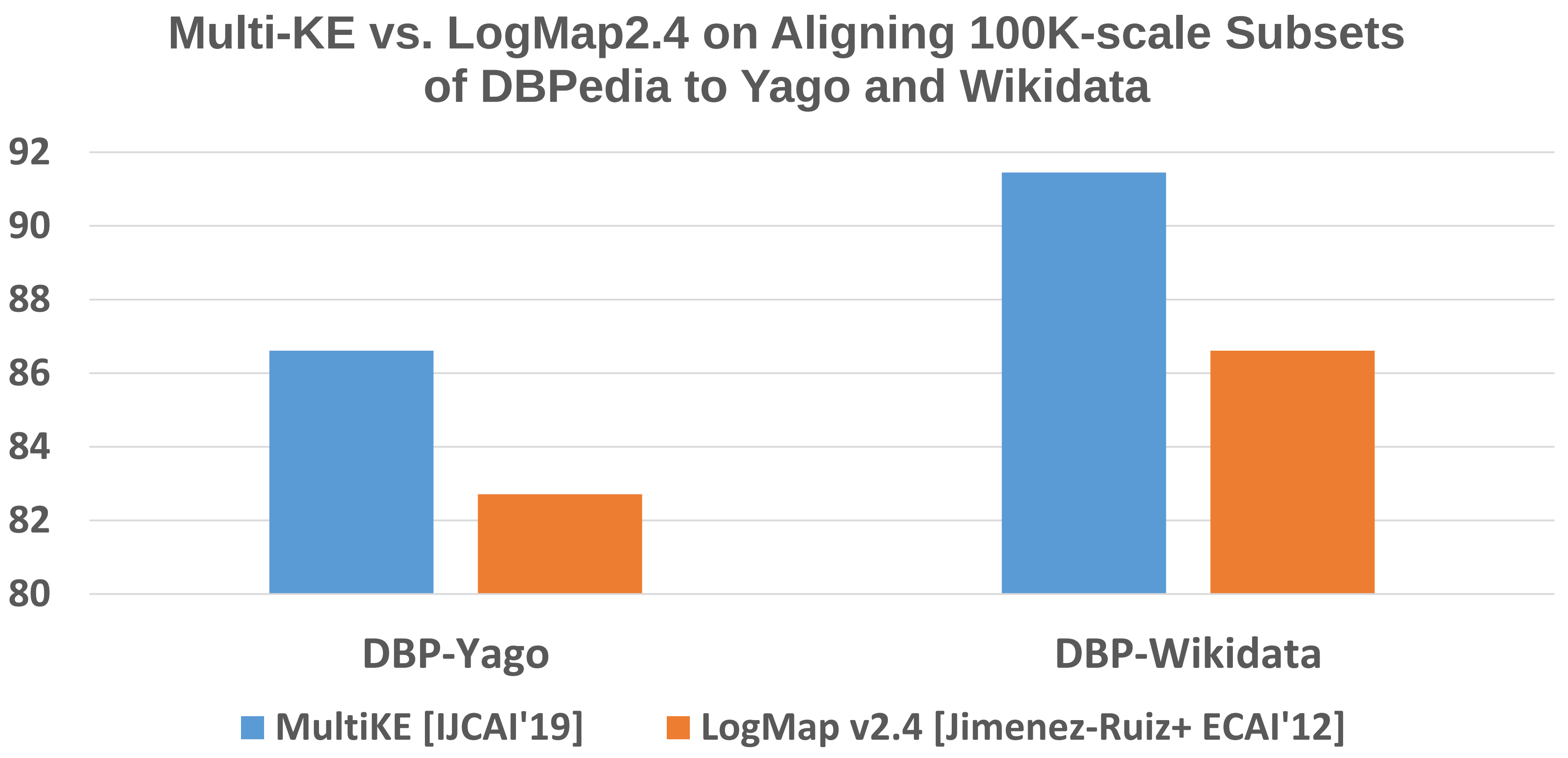
Performance by Our Methods on Semi-supervised Entity Alignment

DBP15k: the benchmark dataset for entity alignment.



***The Candidate space of each test case is 63k~98k entities for each language**

Our Method Outperforms The Well-known Ontology Matching System (LogMap v2.4)



***MultiKE** [IJCAI’19] is a monolingual ontology matching framework with multi-view embeddings of triples, literals, descriptions and attributes.

Recent Advances on Embedding-based Knowledge Alignment

Follow-ups on the same topic

- 2017: IJCAI×2, ISWC×1
- 2018: AAAI×2, COLING×1, ACL×1, EMNLP×1, IJCAI×3
- 2019: AAAI×2, ACL×3, EMNLP×4, ICLR ×1, ICDM×1, ICML×1, IJCAI×6, ISWC×2, KDD×1, WWW×1, WSDM ×1
- 2020: AAAI×3, ACL×1, COLING ×1, CIKM ×1, EMNLP×4, ICDE ×1, ICLR×2, IJCAI×2, ISWC×1, NeurIPS×1, KDD ×2, VLDB×1, WWW×1, WSDM×2

More approaches for embedding learning

- Long-term dependency models, R-GCN, hyperbolic embeddings, holographic embeddings, Gaussian embeddings, etc.

More knowledge association methods

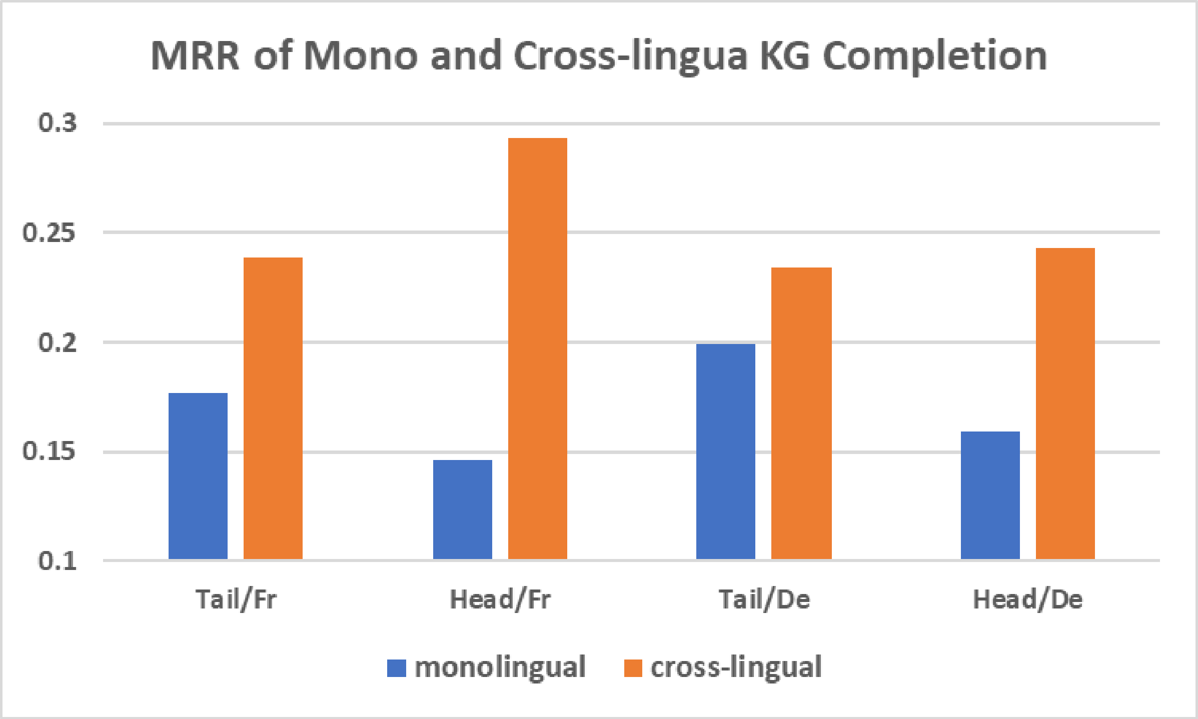
- Adversarial alignment learning, optimal transport, meta learning, noisy supervision, etc.

A systematic summary was given as our AAAI-2020 tutorial, + a benchmarking study and survey in PVLDB vol. 13 (2020).

Relation Inference with Knowledge Transfer

Knowledge transfer to populate a sparser KG (e.g. French)

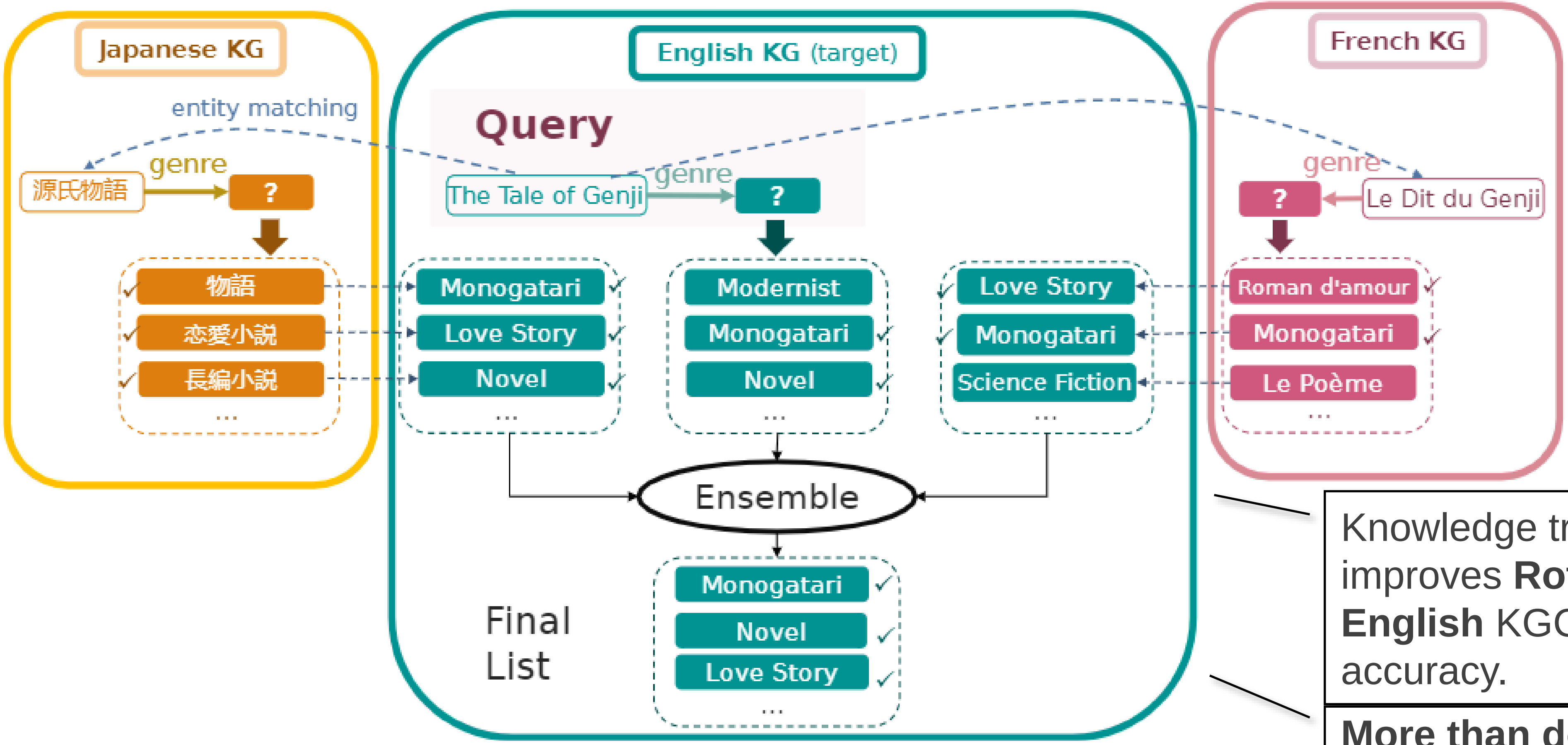
- Obtain the answer of queries *(h, r, ?t)* in the embedding space of a well-populated version (e.g. English) of KG



Cross-lingual knowledge transfer can improve sparse KG completion.

Meta-learnable Knowledge Transfer Among Multiple KGs

[EMNLP-20b]



Knowledge transfer from five langs improves **RotatE**'s [sun+, ICML-19] **English** KGC by **5%** improvement in accuracy.

More than doubling the accuracy in a more low-resource language (**Greek**).

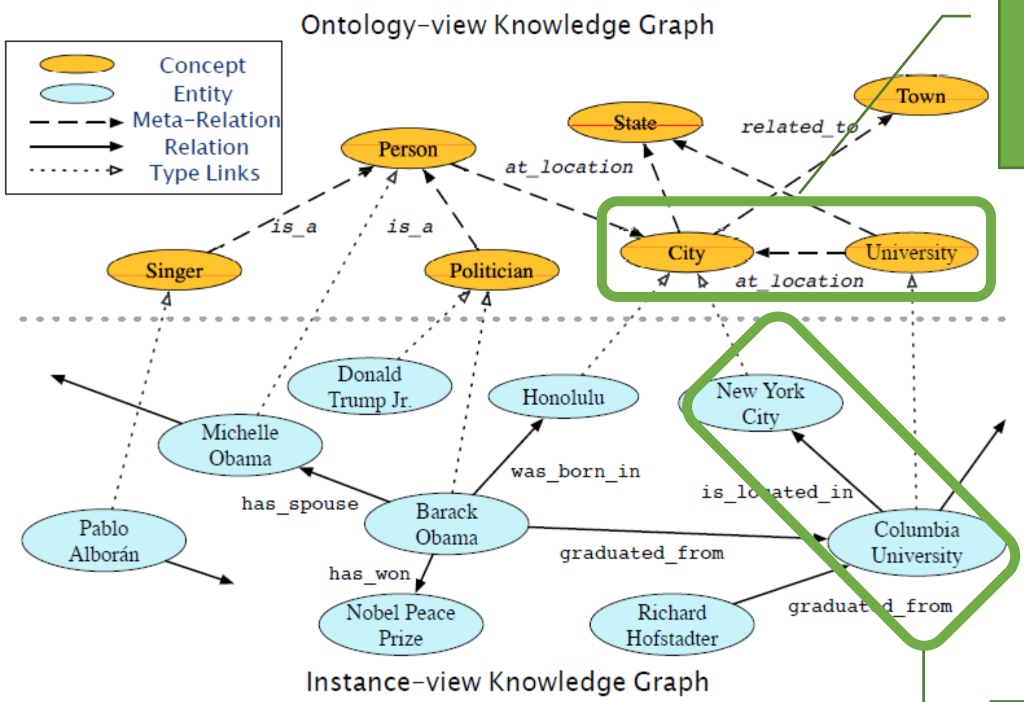
Transferable Embeddings + Meta-learning w/ RankBoost-based Model Weights

Scenario 2.a: Transferable Embedding for Instances and Abstract Concepts*

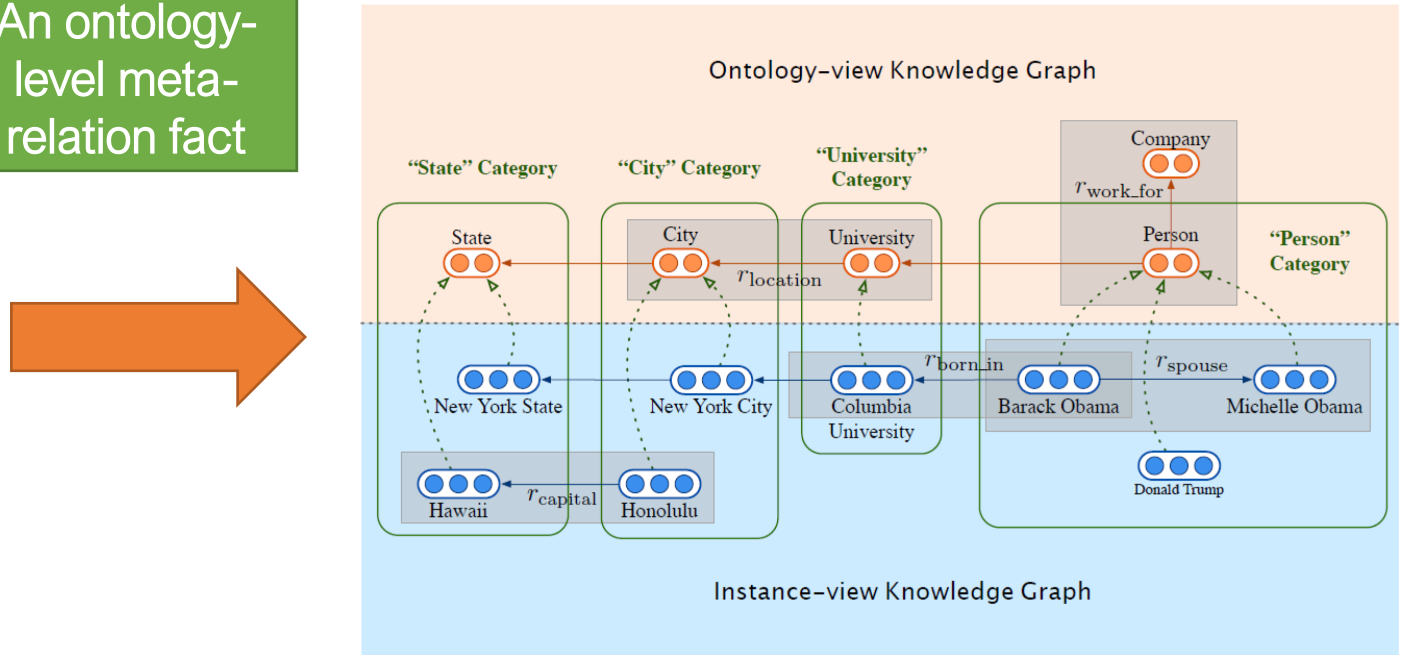
*[KDD-19]

Ontology view: meta-relations of commonsense concepts

Instance view: relations of entities instantiated from concepts



An ontology-level meta-relation fact



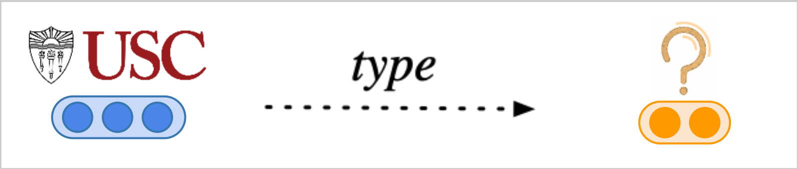
An instance-level relation fact

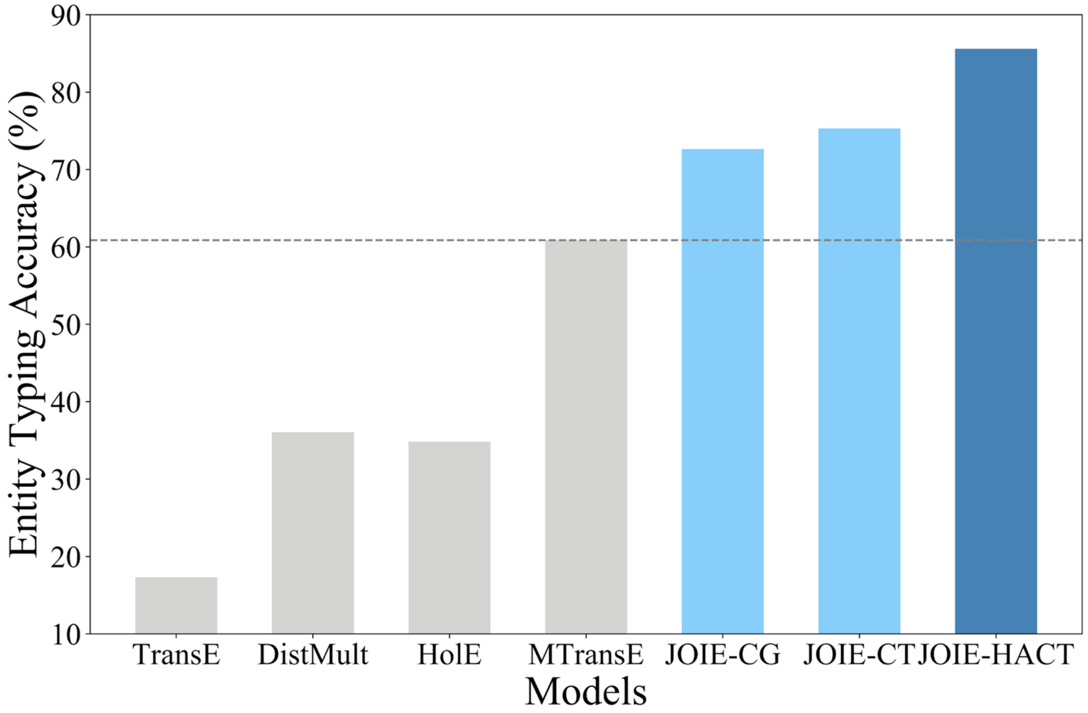
DBpedia yago select knowledge

Application: Entity Typing

*[KDD-19]

Type inference (906 labels) on 40% of >111k entities in YAGO.





Examples of long-tail entity typing (Least 15%)

Entity	Model	Top 3 Concept Prediction
Laurence Fishburne	DistMult	football team, club, team
	MTransE	writer, **person**, artist
	JOIE	**person**, artist, philosopher
Warangal City	DistMult	country, village,**city**
	MTransE	administrative region, **city**, settlement
	JOIE	**city**, town, country
Royal Victor -ian Order	DistMult	person, writer, administrative region
	MTransE	election, award, **order**
	JOIE	award, **order**, election

Typing accuracy on long-tail entities (Least 15%)

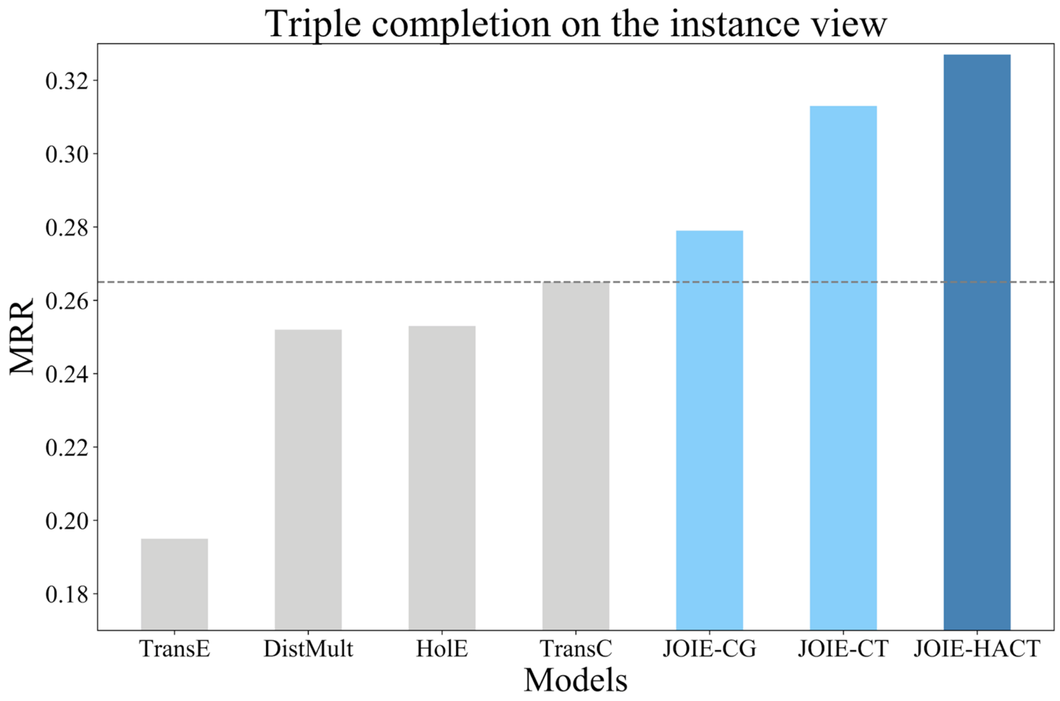
Datasets	YAGO26K-906		
Metrics	MRR	Acc.	Hit@3
DistMult	0.156	10.89	25.33
MTransE	0.526	46.45	67.25
JOIE-TransE-CG	0.708	59.97	79.80
JOIE-TransE-CT	0.737	62.05	82.60
JOIE-HATransE-CT	**0.802**	**69.66**	**87.75**

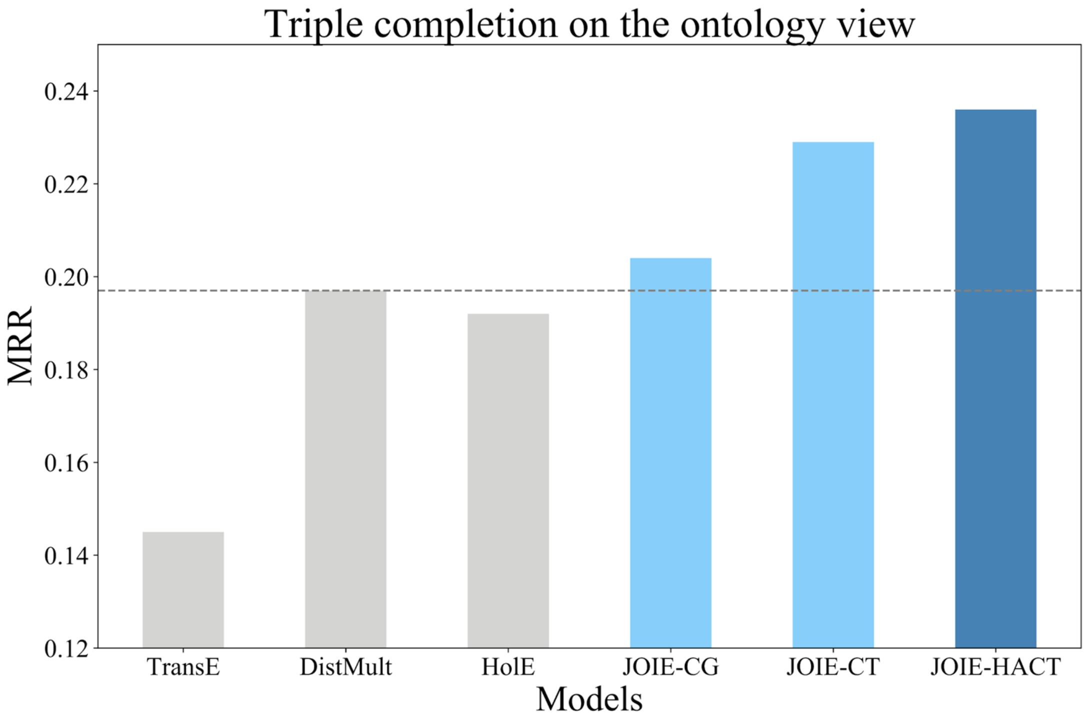
Application: KG Completion

*[KDD-19]

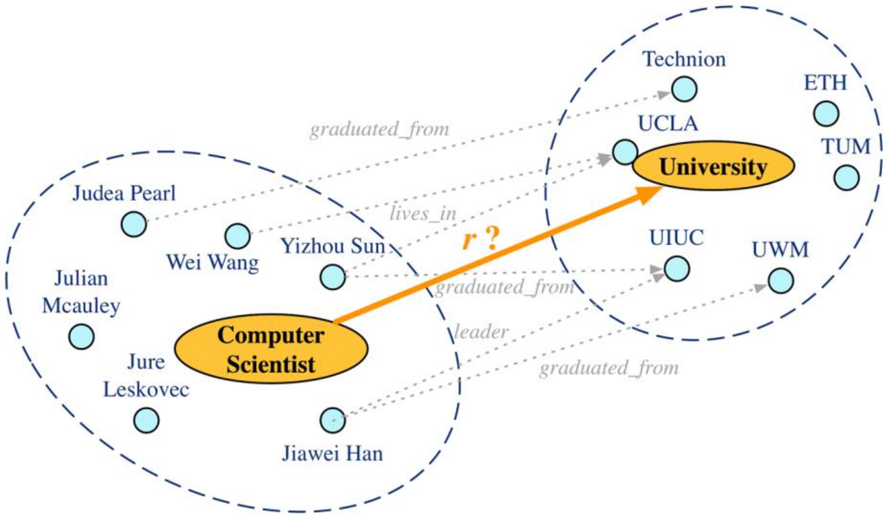
Predicting the 10% held-out relation facts on both views.

Joint representation improves the task on both views.





Transfer Instance-level Knowledge for Ontology Population



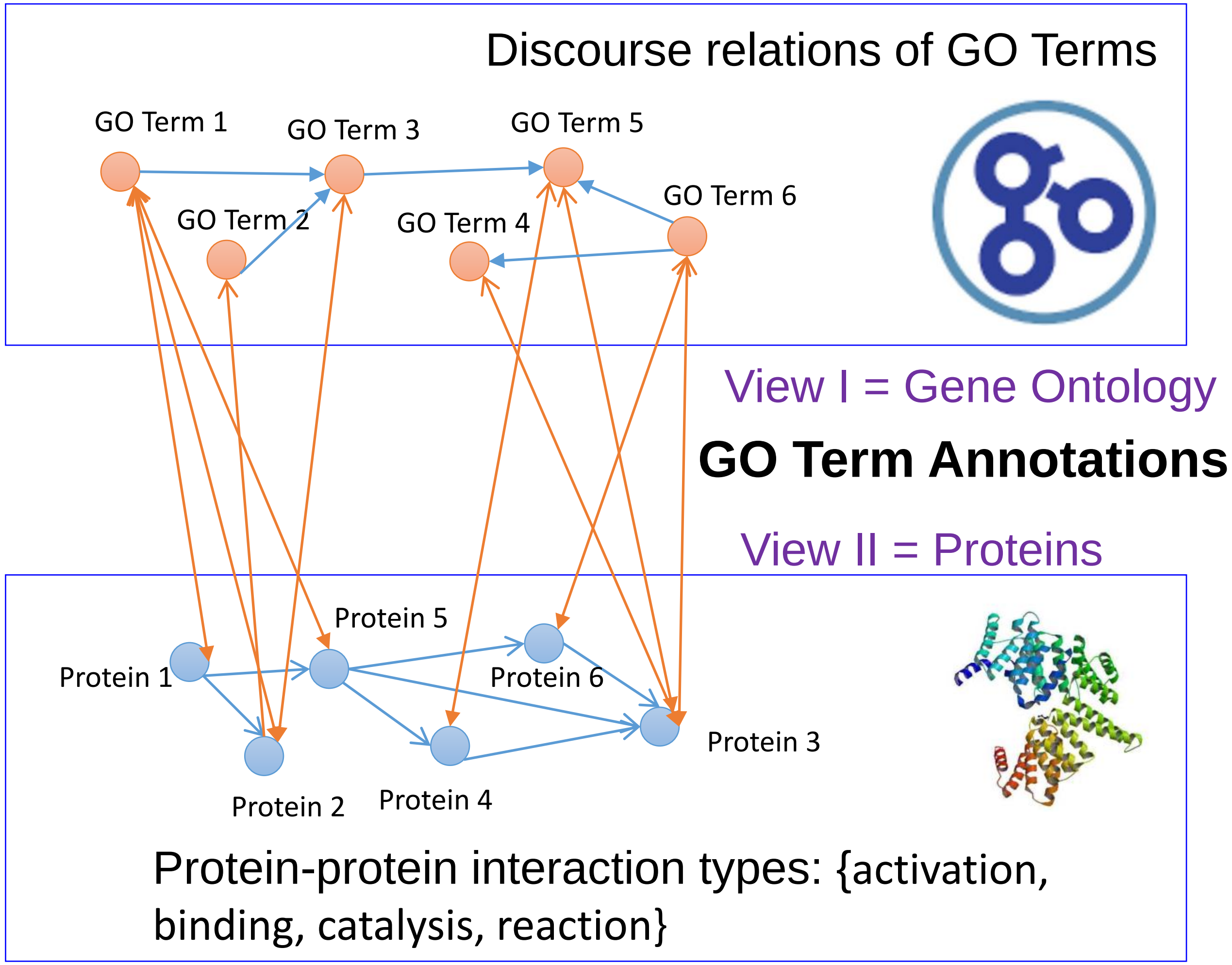
Populating unseen ontological facts by transferring knowledge from instance-view facts.

Examples of ontology population

Query	Top 3 Populated Triples with distances
(scientist, ?r, university)	scientist, ***graduated from***, university (0.499) scientist, ***isLeaderOf***, university (1.082) scientist, *isKnownFor*, university (1.098)
(boxer, ?r, club)	boxer, ***playsFor***, club (1.467) boxer, *isAffiliatedTo*, club (1.474) boxer, ***worksAt***, club (1.479)
(scientist, ?r, scientist)	scientist, ***doctoralAdvisor***, scientist (0.204) scientist, ***doctoralStudent***, scientist (0.221) scientist, ***relative***, scientist (0.228)

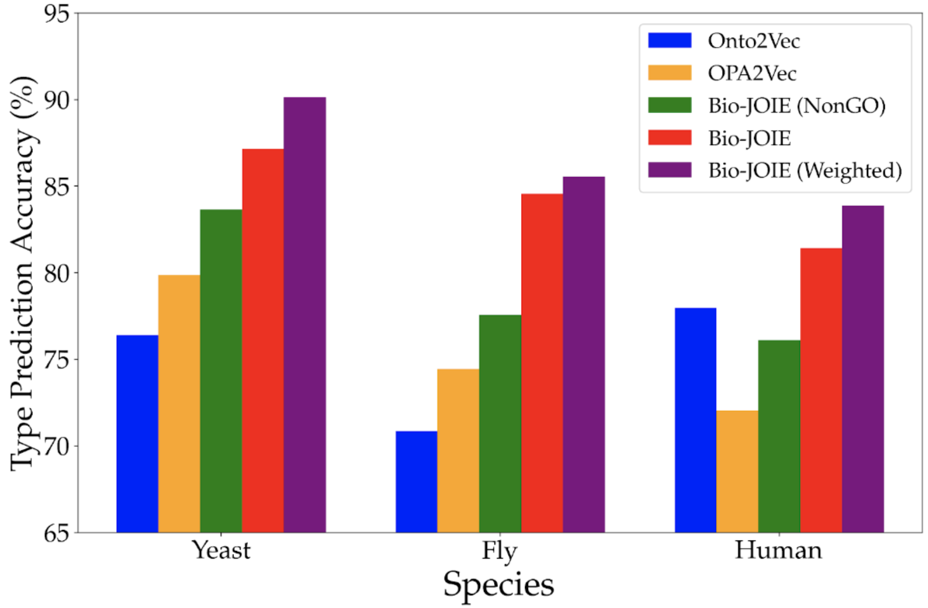
Scenario 3.a: Proteomics and Gene Ontologies

*[ACM BCB-20]
(Best Student Paper)



Transferring knowledge from the gene ontology improves typed protein-protein interaction prediction.

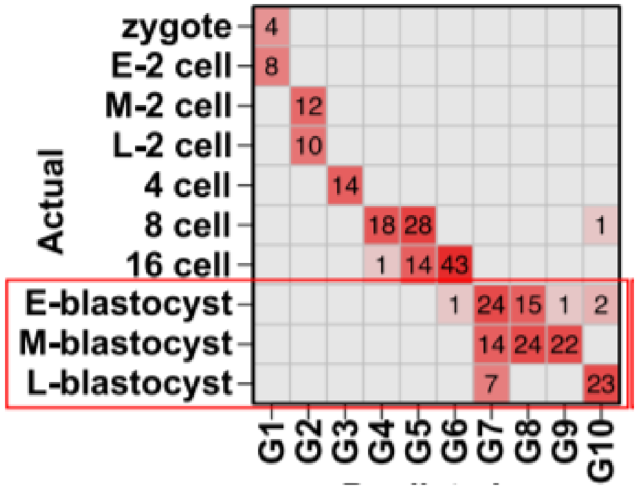
~10% of ACC improvement over SOTA (Opa2Vec, *Bioinformatics* [J] 2019).



and helps disease target prediction for COVID-19 related viral proteins.

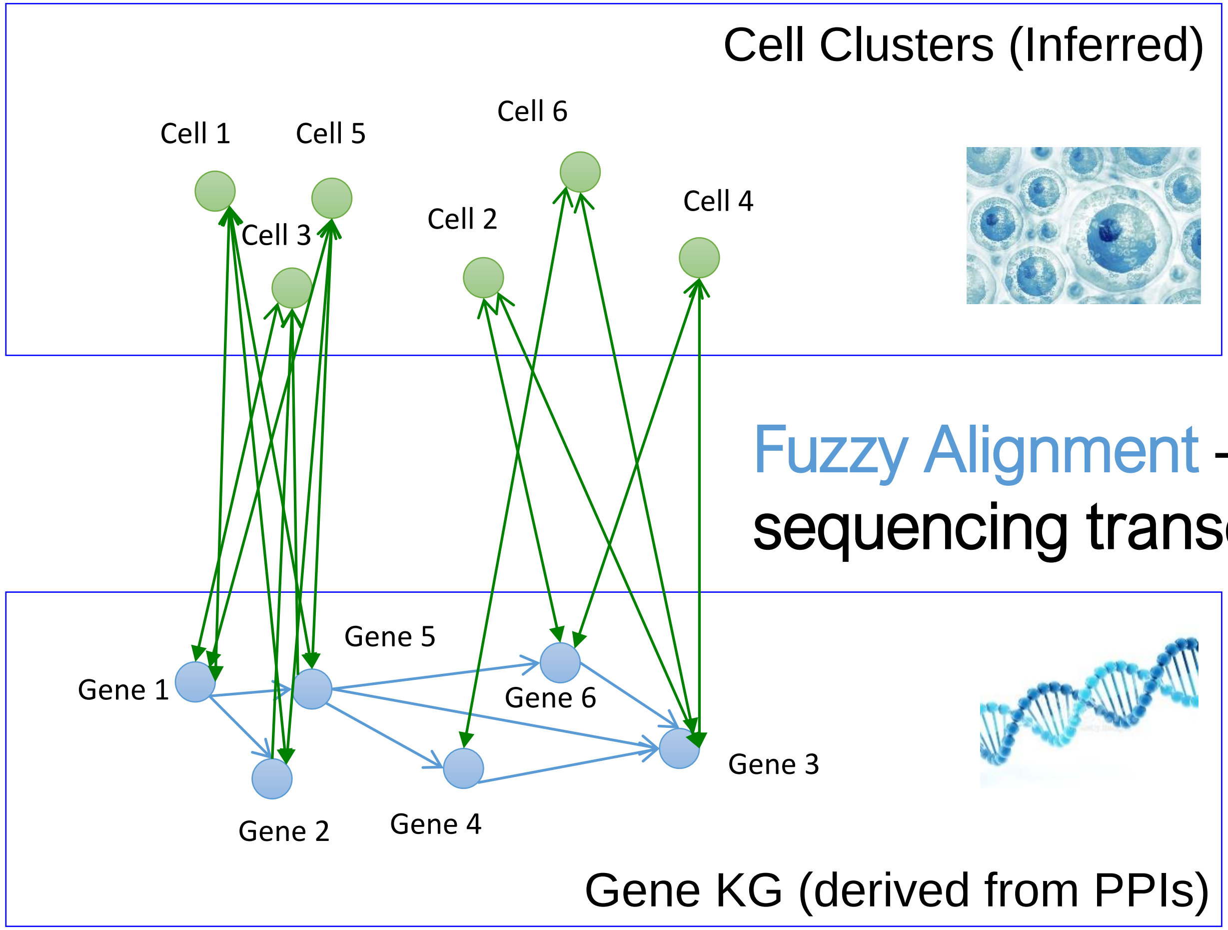
Scenario 3.b: Cell Clustering

At least **10-15%** of ARI improvement over pCMF [Durif+, *Bioinform.* 2019] and others.



Non-negative Tri-Factorization

$$\underset{\theta}{\operatorname{argmin}} \left\| S - \mathbf{E}_1 \mathbf{U} \mathbf{E}_2^{\mathrm{T}} \right\|$$



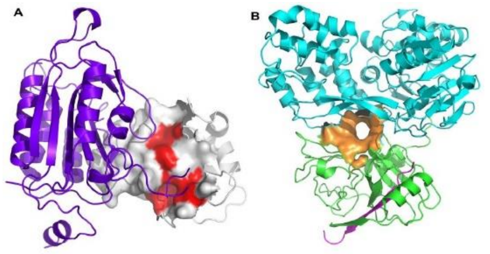
View I = Cells

Fuzzy Alignment – Single-cell RNA sequencing transcripts

View II = Genes

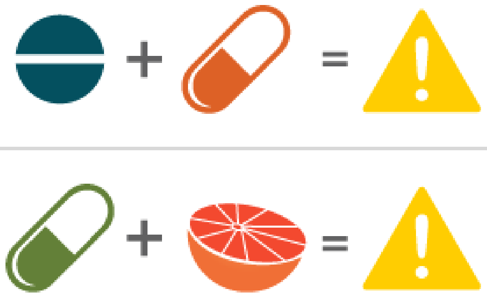
Experiment done on the Zeisel dataset [Zeisel+, *Science* 2015]

More Applications To Be Explored



UMLS

DISEASE ONTOLOGY



Polypharmacy (drug-drug) interaction or drug-target prediction

product graph

Product recommendation

Outline

Transferable Representation Learning of Multi-relational Data

Knowledge Acquisition from Unstructured Data

Future research agenda

Knowledge Acquisition for Events

DARPA & IARPA projects: KAIROS, BETTER, AIDA

Human language always communicates about events.

How to earn a PhD?

Publishing papers

Teaching Apprenticeship

Presenting papers

Doing internships

Oral qualification

Dissertation proposal

Writing a dissertation

Defending a dissertation

Subevents of ***earning a PhD***

What should be the right process?

What contains others?

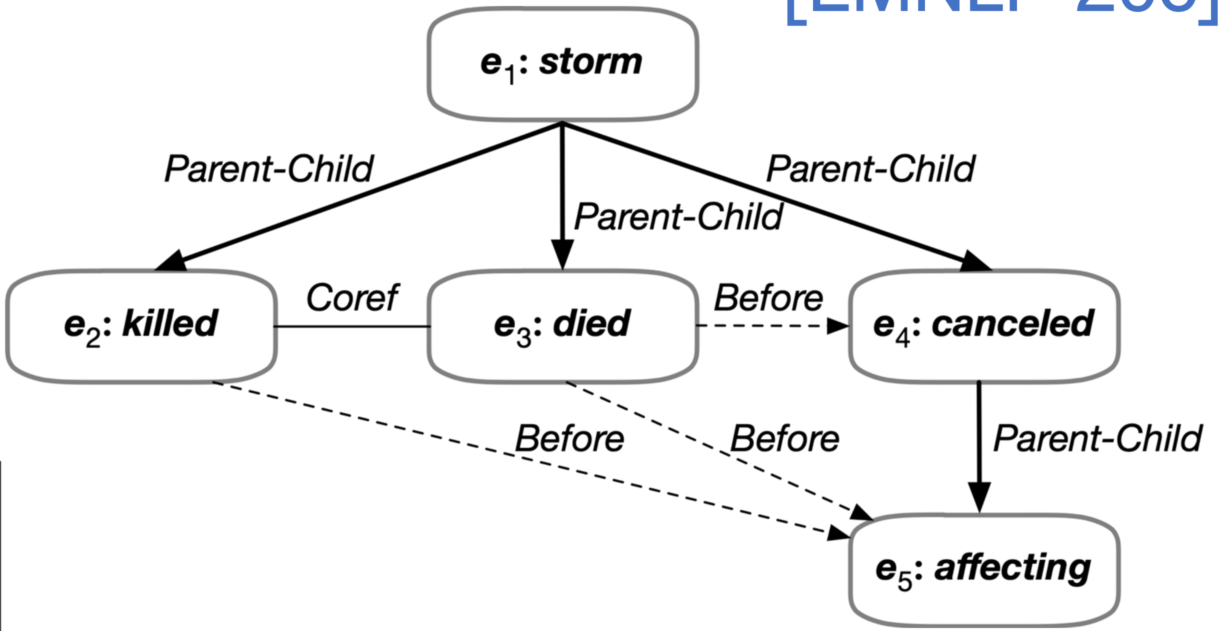
What are depending on others?

What are essential?

Logically Constrained Learning for Event Relation Extraction

*[EMNLP-20c]

- Temporal Relations
- Subevent Relations (Memberships)
- Event Coreference



On Tuesday, there was a typhoon-strength (***e_1:storm***) in Japan. One man got (***e_2:killed***) and thousands of people were left stranded. Police said an 81-year-old man (***e_3:died***) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines (***e_4:canceled***) 230 domestic flights, (***e_5:affecting***) 31,600 passengers.

Goal: inducing the relations of events

A **resource hungry** task with limited labeled data:

- No resource annotates all types of relations
  - TempRel data: TBDense and MATRES
  - Subevent and Coref: HiEve
- Annotations are on ~100 documents

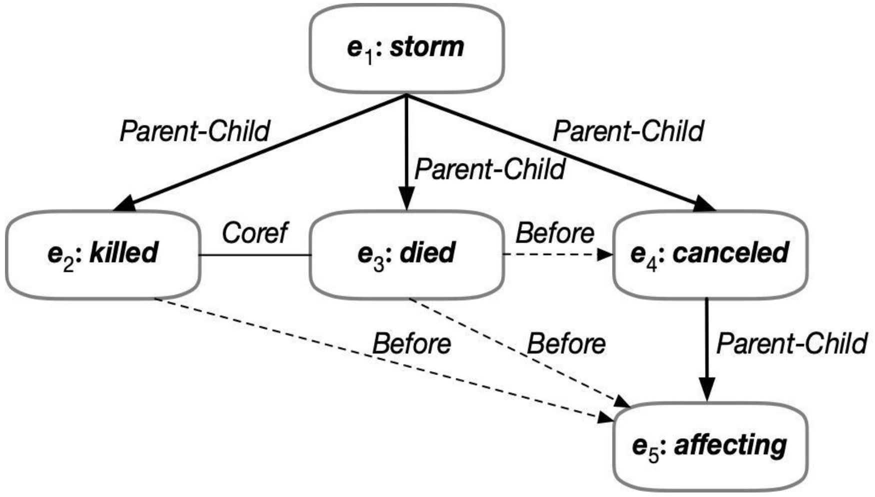
Logical Constraints of Relations

Symmetry
e3:died is BEFORE *e4:canceled*
=> *e4:canceled* is AFTER *e3:died*

Transitivity
e1:storm is PARENT of *e4:canceled*
∧*e4:canceled* is a PARENT of *e5:affecting*
=> *e1:storm* is a PARENT of *e5:affecting*

Conjunction
e3:died is BEFORE *e4:canceled*
∧*e4:canceled* is a PARENT of *e5:affecting*
=> *e3:died* BEFORE *e5:affecting*

(we also consider *Implication* and *Negation*)



Why logical constraints in learning?

- Learning to provide **globally consistent** predictions
- Providing **indirect supervision** across tasks/learning resources

Incorporating Logical Constraints in A Neural Architecture

Symmetry and negation are subsumed within implication loss; Transitivity is subsumed within conjunction loss.

From logical constraints to differentiable functions

- L_A Annotation Loss: $\top \rightarrow r(e_1, e_2)$ $\boxed{\rightarrow}$ $-w_r \log r_{(e_1,e_2)}$
- L_S Implication Loss: $\alpha(e_1, e_2) \leftrightarrow \bar{\alpha}(e_2, e_1)$ $\boxed{\rightarrow}$ $|\log \alpha_{(e_1,e_2)} - \log \bar{\alpha}_{(e_2,e_1)}|$
- L_C Conjunction Loss: $\alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \gamma(e_1, e_3)$ $\boxed{\rightarrow}$ $\log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log \gamma_{(e_1,e_3)}$

  $\alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \neg\delta(e_1, e_3)$ $\boxed{\rightarrow}$ $\log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log(1 - \delta_{(e_1,e_3)})$
- Training Objective: $L = L_A + \lambda_S L_S + \lambda_C L_C$

α \ β	PC	CP	CR	NR	BF	AF	EQ	VG
PC	PC, ¬AF	–	PC, ¬AF	¬CP, ¬CR	BF, ¬CP, ¬CR	–	BF, ¬CP, ¬CR	–
CP	–	CP, ¬BF	CP, ¬BF	¬PC, ¬CR	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	–
CR	PC, ¬AF	CP, ¬BF	CR, EQ	NR	BF, ¬CP, ¬CR	AF, ¬PC, ¬CR	EQ	VG
NR	¬CP, ¬CR	¬PC, ¬CR	NR	–	–	–	–	–
BF	BF, ¬CP, ¬CR	–	BF, ¬CP, ¬CR	–	BF, ¬CP, ¬CR	–	BF, ¬CP, ¬CR	¬AF, ¬EQ
AF	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	–	–	AF, ¬PC, ¬CR	AF, ¬PC, ¬CR	¬BF, ¬EQ
EQ	¬AF	¬BF	EQ	–	BF, ¬CP, ¬CR	AF, ¬PC, ¬CR	EQ	VG, ¬CR
VG	–	–	VG, ¬CR	–	¬AF, ¬EQ	¬BF, ¬EQ	VG	–

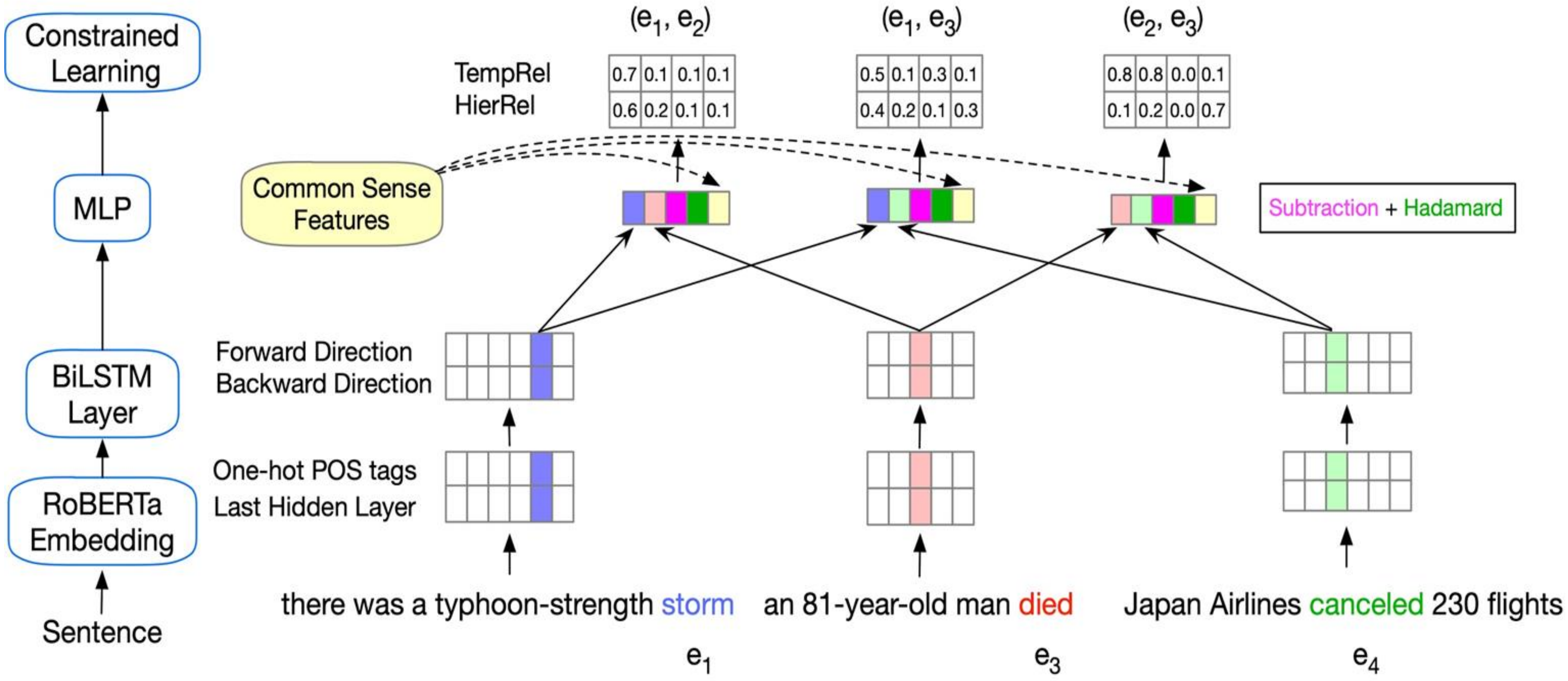
The Joint Constrained Learning Architecture

- Temporal Relations
- Subevent Relations (Memberships)
- Event Coreference

Logical constraints
- Symmetry, transitivity, conjunction, implication.

Converting constraints into differentiable learning objectives

Loss Function: $L = L_A + \lambda_S L_S + \lambda_C L_C$



Logically Constrained Learning for Event Relation Extraction

*[EMNLP-20c]

Constrained learning surpasses SOTA TempRel extraction on MATRES [Ning+, ACL-18] by relatively 3.27% in F_1.

Model	P	R	F_1
CogCompTime (Ning et al., 2018c)	0.616	0.725	0.666
Perceptron (Ning et al., 2018b)	0.660	0.723	0.690
BiLSTM+MAP (Han et al., 2019b)	-	-	0.755
LSTM+CSE+ILP (Ning et al., 2019)	0.713	0.821	0.763
Joint Constrained Learning (ours)	**0.734**	**0.850**	**0.788**

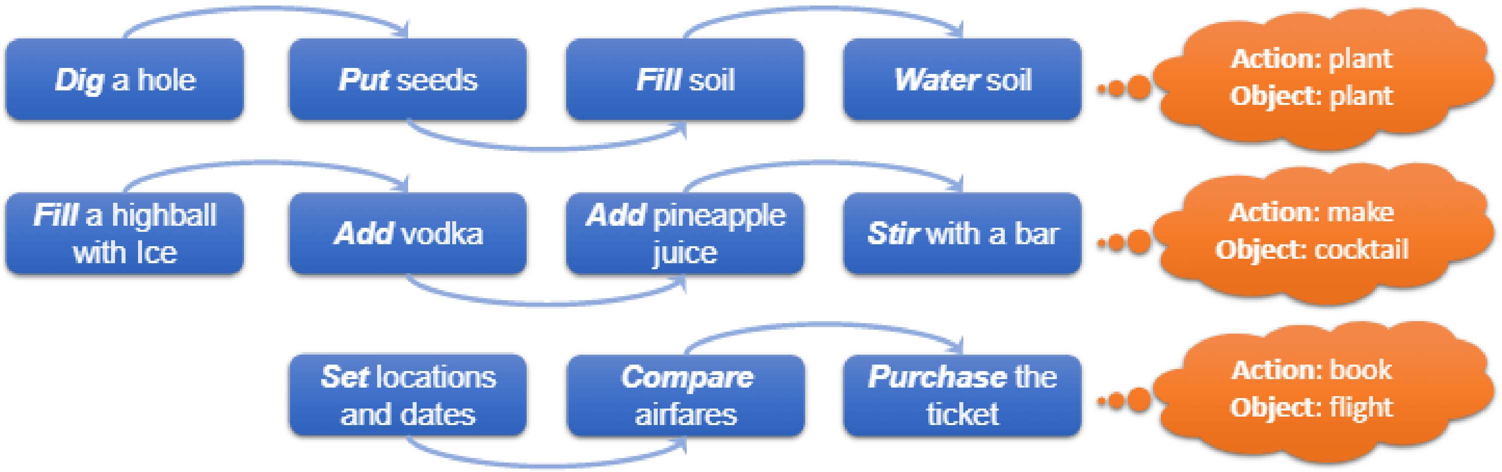
On HiEve [Glavaš+, LREC-14] for subevent extraction, it relatively surpasses previous methods by at least 3.12% in F_1.

	F_1 score		
Model	PC	CP	Avg.
StructLR (Glavaš et al., 2014)	0.522	**0.634**	0.577
TacoLM (Zhou et al., 2020a)	0.485	0.494	0.489
Joint Constrained Learning (ours)	**0.625**	0.564	**0.595**

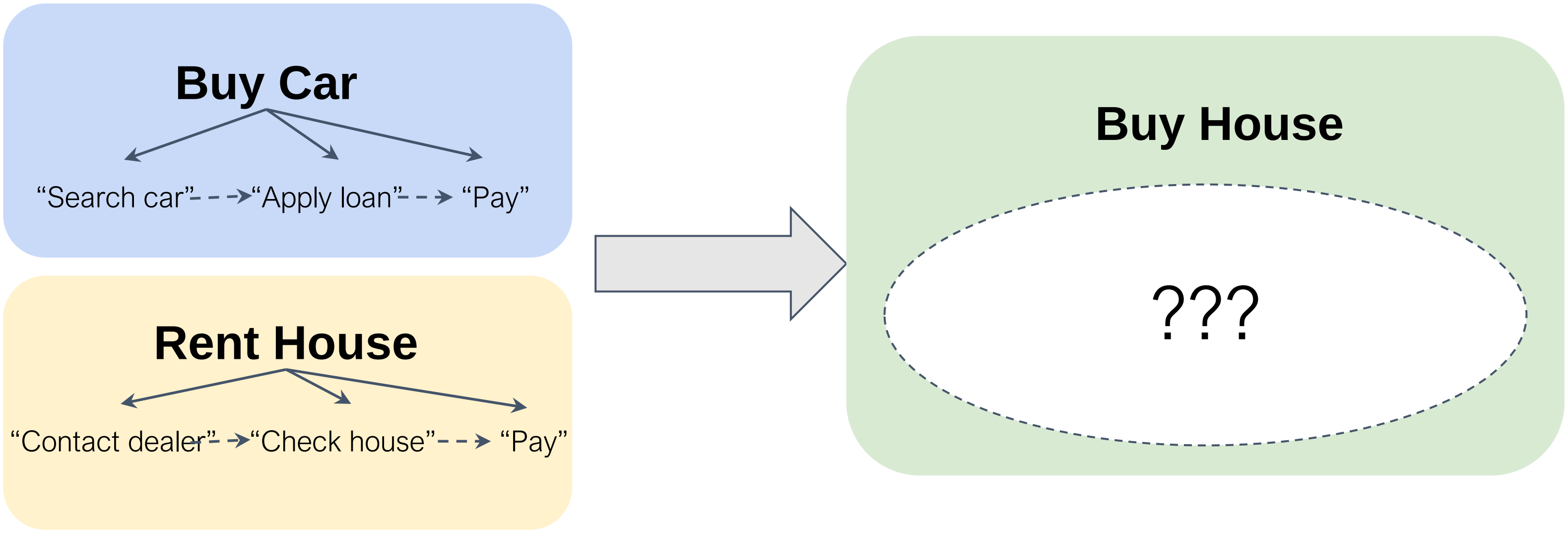
Key Observations

- Constraints are a natural bridge for learning resources with different sets of relations
- Adding constraints in learning is sufficient to enforce logical consistency of outputs, surpassing ILP in inference (w/ constrained learning) by 2.6-12.3% in ACC

More About Eventuality Knowledge Acquisition from Text



Few-shot *intention prediction* for event processes based on indirect supervision from gloss knowledge [CoNLL-20 Best Paper Nomination]



Open-domain event schema induction with analogy-aware inference [EMNLP-20d]

Muhao Chen, Hongming Zhang, Qiang Ning, Heng Ji, Kathleen McKeown, Dan Roth. **Event-centric Natural Language Understanding**. Tutorials in AAAI 2021 and ACL 2021

Probabilistic Constrained Knowledge Acquisition*

*[AAAI-19]

Toyota competeswith Honda 0.94

Hyundai 0.91

Chrysler 0.76

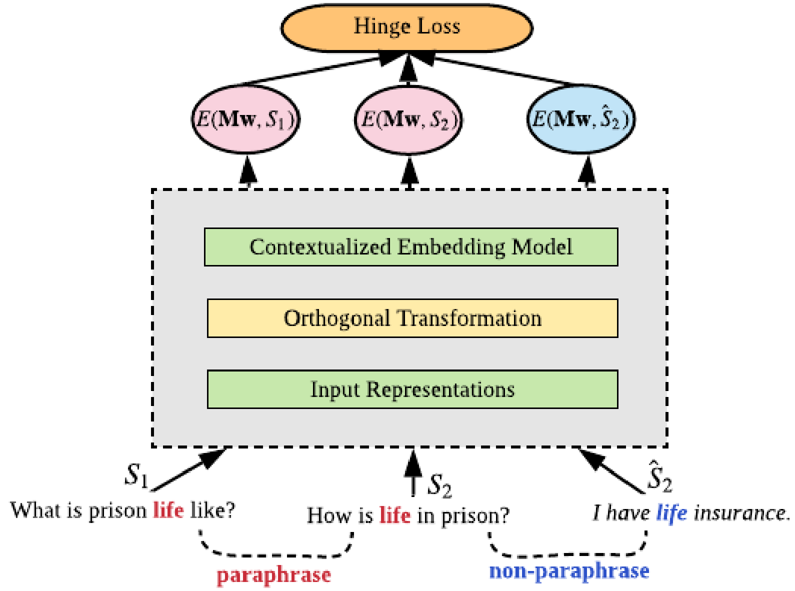
ConceptNet
An open, multilingual knowledge graph

probase

STRING NELL

- Incorporating Probabilistic Soft Logic constraints in learning (w/ Łukasiewicz t-norm)
- Confidence prediction for unseen facts

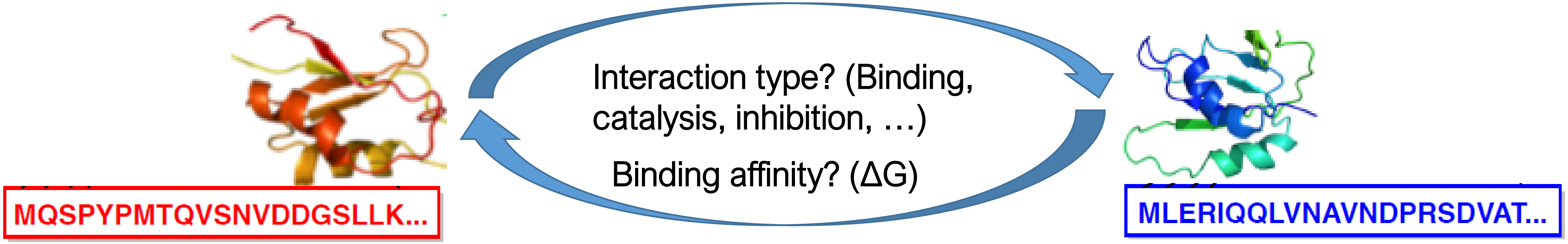
Retrofitting language models for robust discourse relation detection+



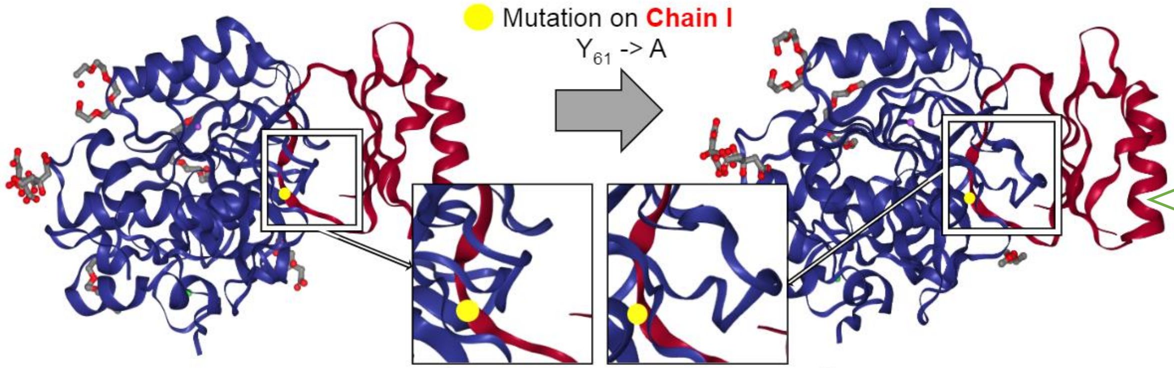
+2.60-3.30% (acc) on textual inference and +3-5% (Pearson's) in textual similarity (SentEval) +5.4% (acc) on Adversarial SQuAD.

*[EMNLP-19]

Knowledge Acquisition Beyond Human Languages



"Entailment model" for Protein-protein interaction prediction [**ISMB'19, *Bioinformatics 2019***].



PDB ID	1TM1 (wild type)	1TO1 (mutant)
Binding Affinity (k_d)	2.24E-12	2.70E-10

ΔΔG estimation on SKEMPIv2 benchmark: ~20% of absolute improvement (0.69->0.88) in Pearson's Corr over SOTA!

Pre-trained language model on wild-type protein sequences helps estimate point mutation effects on proteins [***NAR: Genom. Bioinform.* 2020**].

Outline

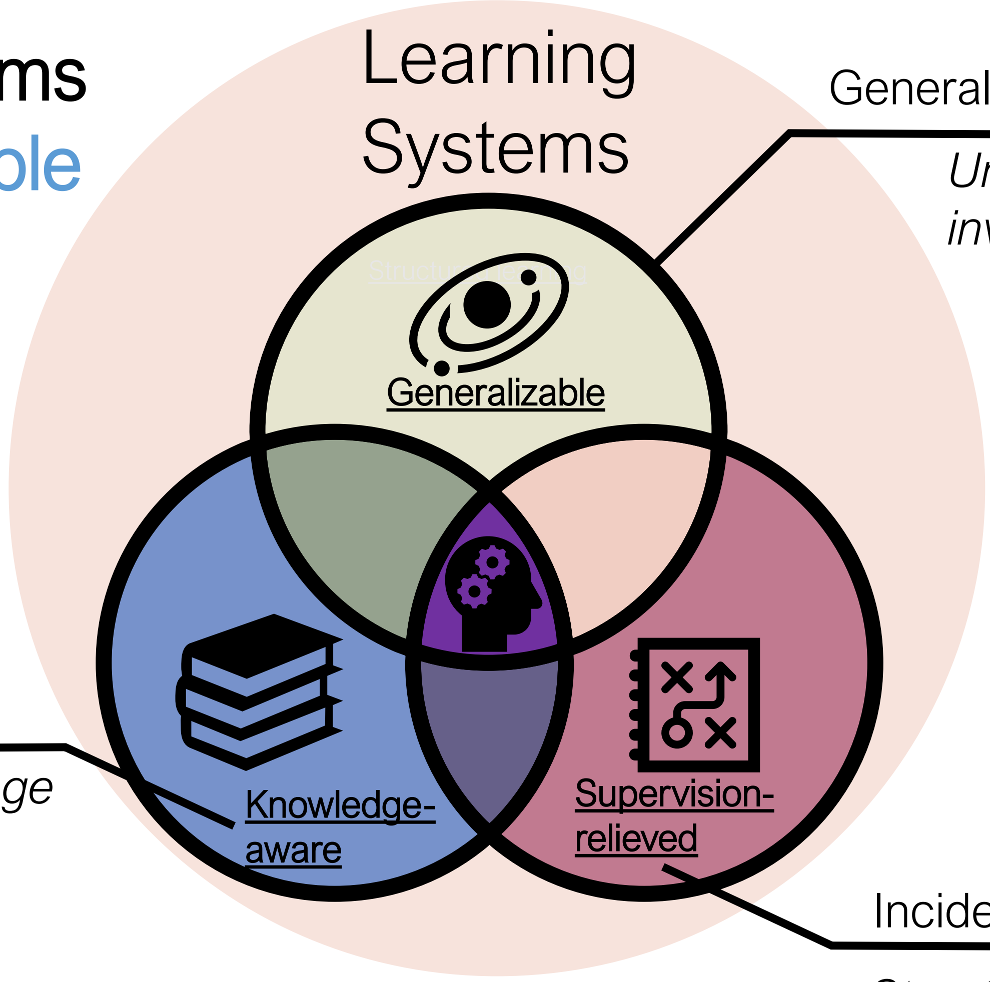
Transferable Representation Learning of Multi-relational Data

Knowledge Acquisition from Unstructured Data

Future research agenda

What's Next

Data-driven learning systems should become more reliable and adaptive.



Generalizable data-driven ML

Universal and domain-invariant feature extraction

Knowledge Integration

Knowledge-aware language modeling, Null prediction

Incidental Supervision

Structured representation of extreme label spaces

Robust Learning Systems with Generalizability

1.2 billion years of evolution distance

0.12 billion years of evolution distance

Yeast

Train

Predict PPI

Arabidopsis

Train

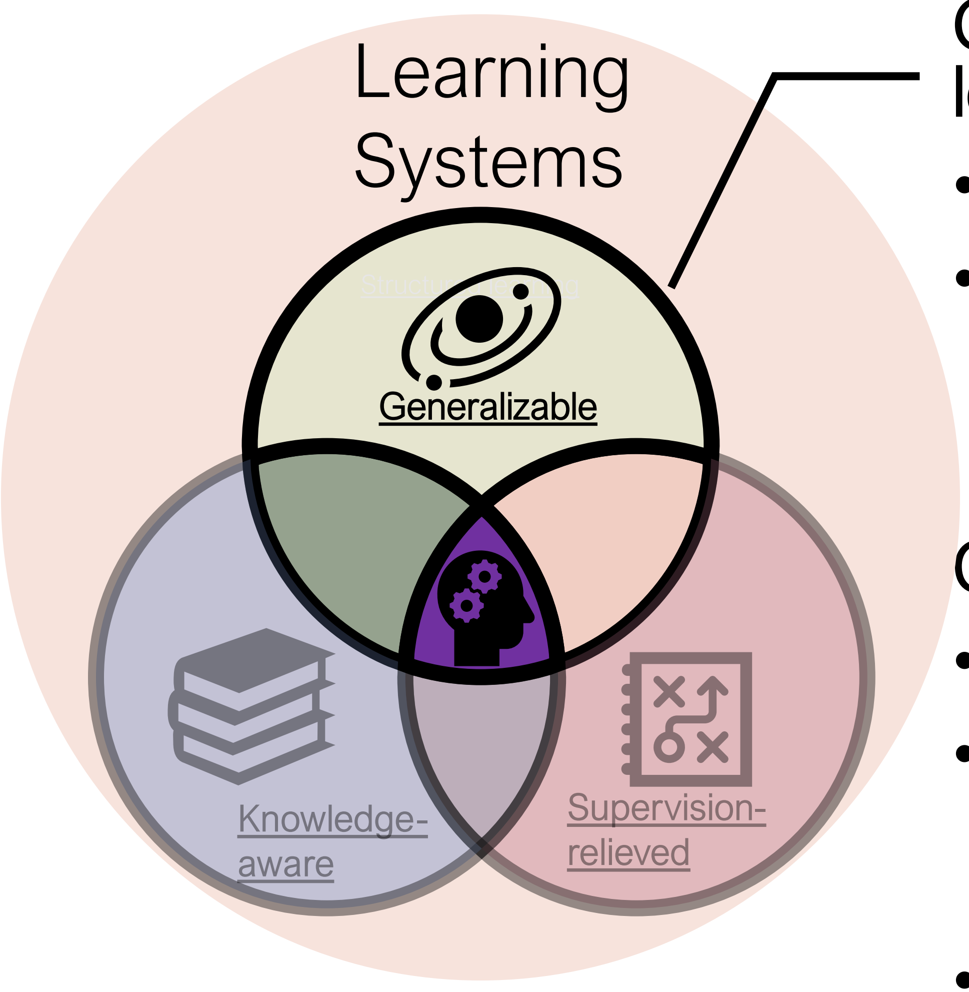
Predict PPI

Tomato

PIPR [ISMB'19]: >97% in F1 scores for intrinsic Protein-protein Interaction (PPI) prediction.

Future direction: using transfer learning to predict PPI for > 1.3 million low-resource species.

Robust Learning Systems with Generalizability



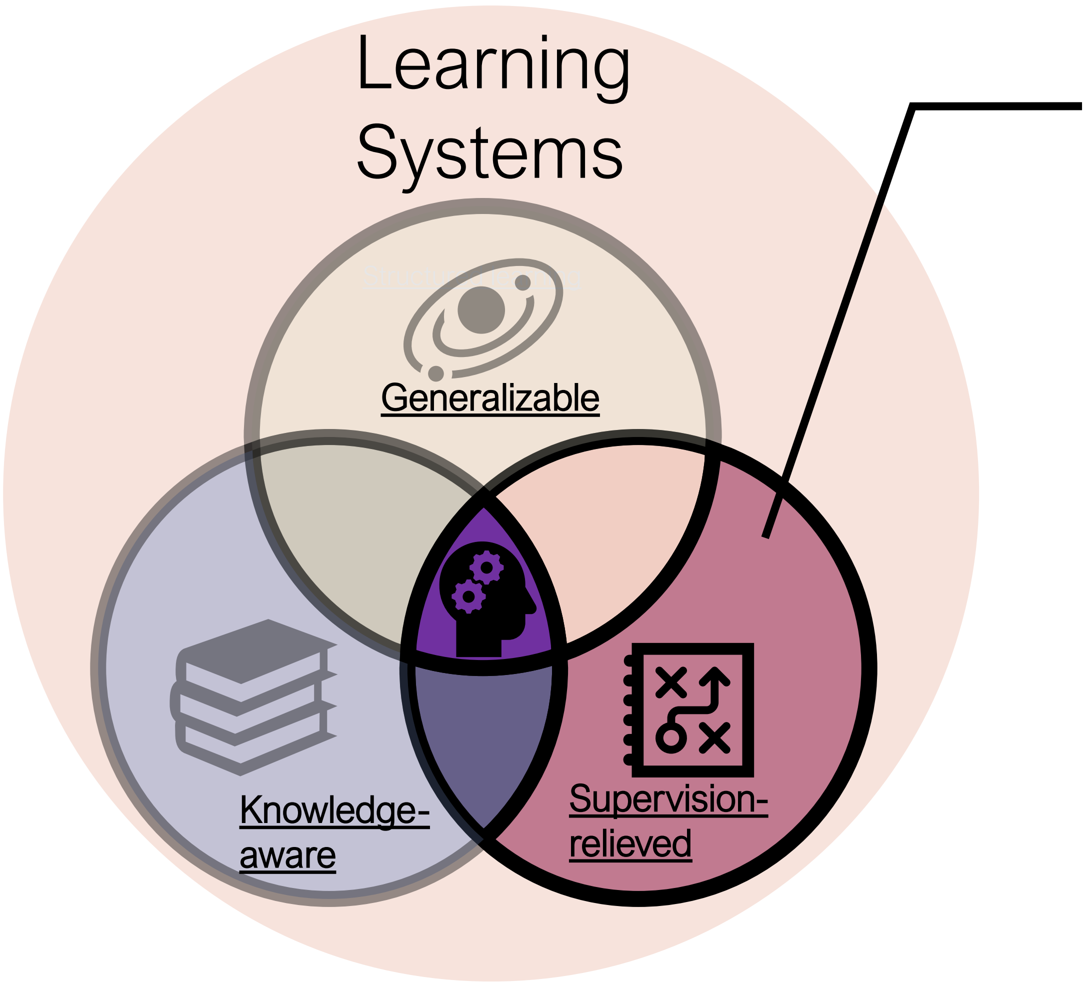
General methodologies for domain-adaptive learning/inference

- Domain-invariant feature extraction
- Massively pre-training
  - Language- and domain-invariant KG embedding (ongoing)
  - Pre-training language models on thousand-species genomic/proteomic data

Cross-domain tasks

- NLP tasks on >6000 low-resource languages
- PPI, folding energy, 3D structure prediction, functional annotation … for > 1.3 million low-resource species
- Clinical data processing [AIME-20] (low-resource due to privacy)

Support Learning With Minimal Supervision



Representing Structures of Feature and Label Spaces

- Non-Euclidean representation learning
- *Set learning* for order-invariant data
  - Concurrent clinical events [AIME-20]

Indirect supervision

- Leverage cheap supervision signals from auxiliary data / tasks
- Learning with noisy labels (ongoing direction)
- Learning/inference with dependency of labels

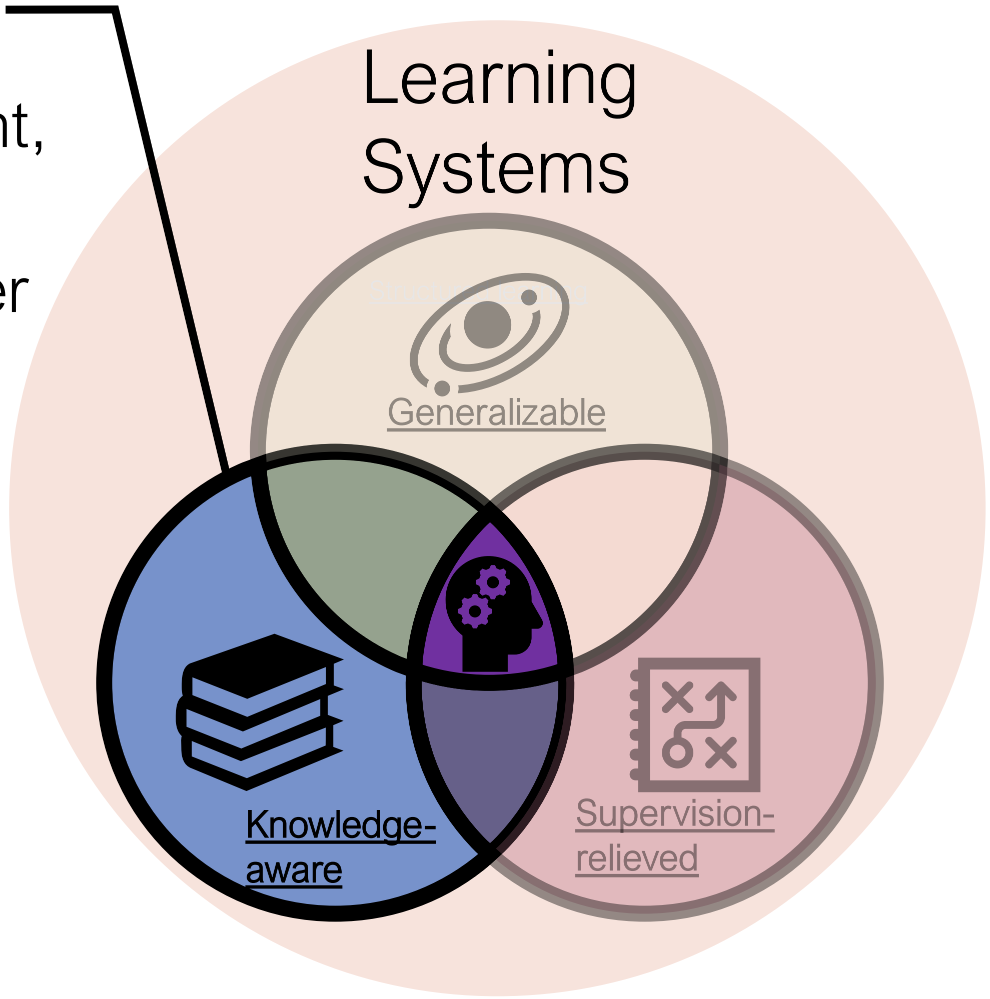
Reliable and Knowledge-aware Learning Systems

Null prediction problem

- Lots of NLU models for entity typing, entity alignment, entity linking, semantic IR, QA, …
- How to let them understand when there is no answer to a query?

Making language models aware of knowledge

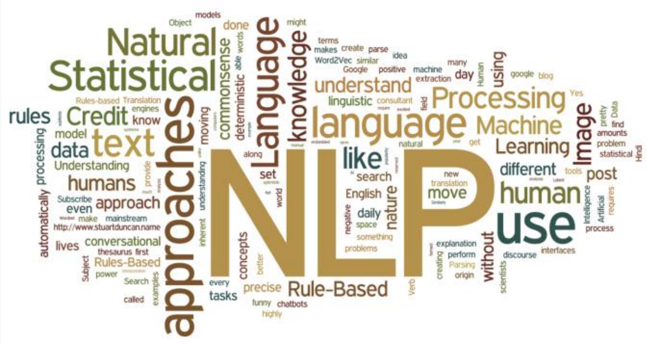
- Eventuality knowledge*
- Temporal knowledge



*Will be discussed in our AAAI-21 and ACL-21 tutorials about **Event-centric NLU**.

Cross-domain and Interdisciplinary Research

A useful technology may benefit multiple research areas and disciplines, and it is important to let it contribute to the Common Good.



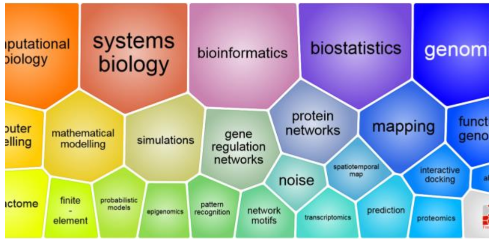
Natural Language Processing

QA & Dialogue systems

Knowledge Representation

ebay
amazon
NETFLIX

Recommendation Systems



Computational Biology

Acknowledgement

My 72 coauthors from 15 institutes. We wrote 50 papers together in the past 6 years.

- Carlo Zaniolo, Kai-Wei Chang, Wei Wang, Alex Bui, Yizhou Sun, Mario Gerla, Eleazar Eskin, Jessica Li, Demetr Terzopoulos, Chelsea Ju, Dat Duong, Guangyu Zhou, Tianran Zhang, Jyun-Yu Jiang, Junheng Hao, Wenchao Yu, Weijia Shi, Shirley Chen, Jieyu Zhao, Kuan-hao Huang, Jiaqi Gu, Qi Zhao, Pengyuan Du, Tuan Le, Seunghyun Yoo, Zijun Xue, Pei Zhou, Tao Zhou, Zheng Wang, James Zhang, Ankith Uppunda (UCLA)
- Dan Roth, Ben Zhou, Haoyu Wang, Hongming Zhang (UPenn)
- Nigel Collier, Fangyu Liu (Cambridge)
- Heng Ji, Manling Li (UIUC)
- Andrew McCallum, Michael Boratko (UMass)
- Kathleen McKeown (Columbia)
- Bryan Perozzi, Gang Huang, Mohan Yang, Shi Gao, Jie Mao (Google)
- Chris Quirk (Microsoft)
- Changping Meng, Jennifer Neville (Purdue)
- Steven Skiena, Yingtao Tian, Haochen Chen, Xiaofei Sun, Syed Fahad Sultan (SUNY SBU)
- X. Sean Wang, Jingheng Zhou (Fudan)
- Zequn Sun, Jiacheng Huang, Wei Hu, Yuzhong Qu, Chengming Wang, Lingbing Guo, Qingheng Zhang, Jiacheng Huang (NJU)
- Qiang Ning (AI2)
- Changjun Fan, Li Zeng, Zhong Liu (NUDT)
- Chengkai Li, Farahnaz Akrami (U. Texas)

Thank You

Sequence-based Protein-Protein Interaction (PPI) Prediction*

UCLA

*[*ISMB'19, Bioinformatics 2019*]

Amino acid sequence 1

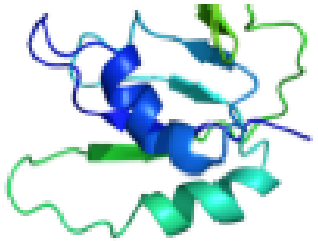
MQSPYPMTQVSNVDDGSLLK...

Interact or not?

Interaction type? (Binding, catalysis, inhibition, …)

Binding affinity? (ΔG)

Amino acid sequence 2



MLERIQQLVNAVNDPRSDVAT...

PIPR: Multifaceted Protein-Protein Based on Only Sequences

*[*ISMB'19, Bioinformatics 2019*]

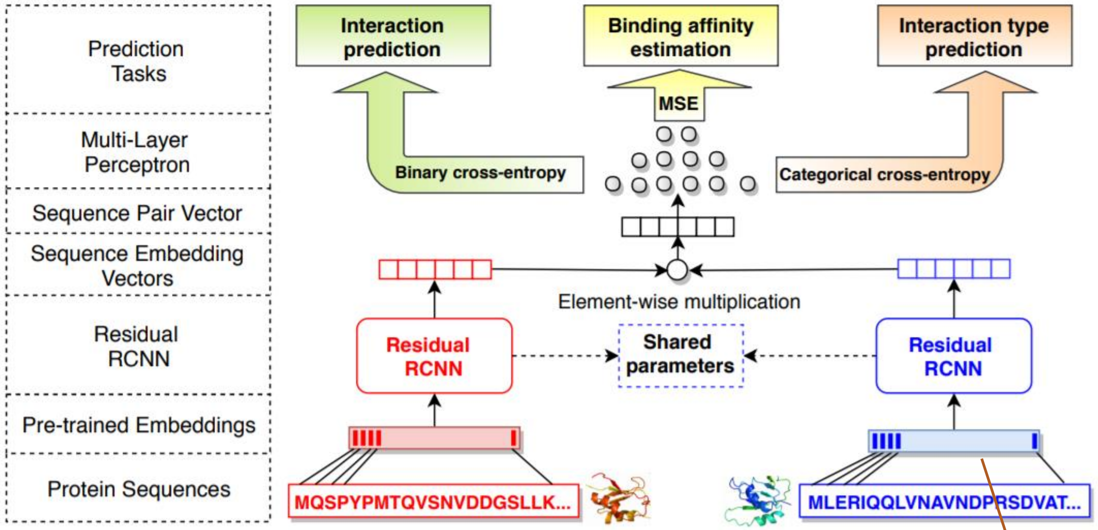
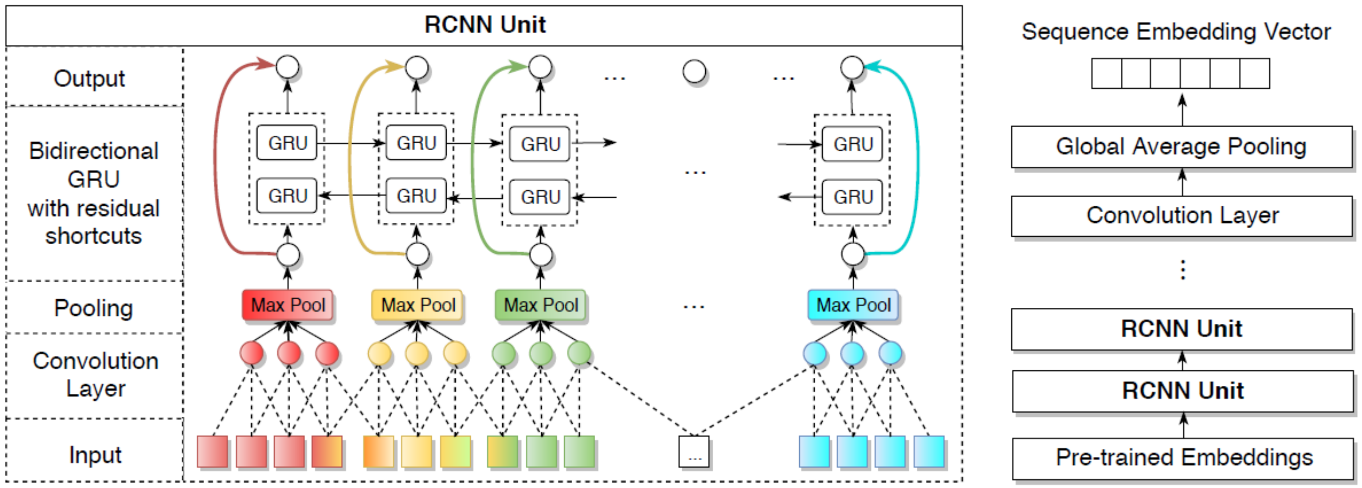


Fig. 2: The overall learning architecture of our framework.



Siamese architecture for capturing multi-faceted PPI information.

Physicochemical property-aware embeddings of amino acids.

Residual RCNN for multi-granular feature aggregation.

Multi-faceted PPI Prediction

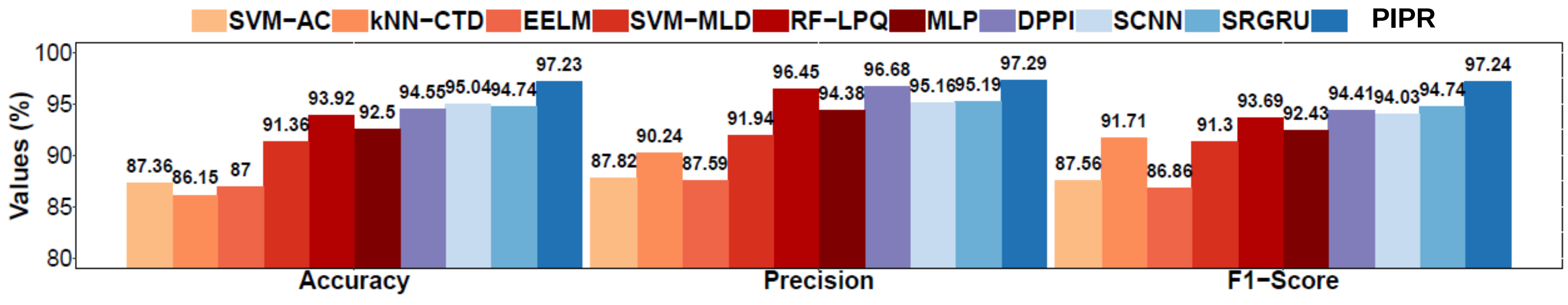


Fig. 3: Evaluation of binary PPI prediction on the Yeast dataset.

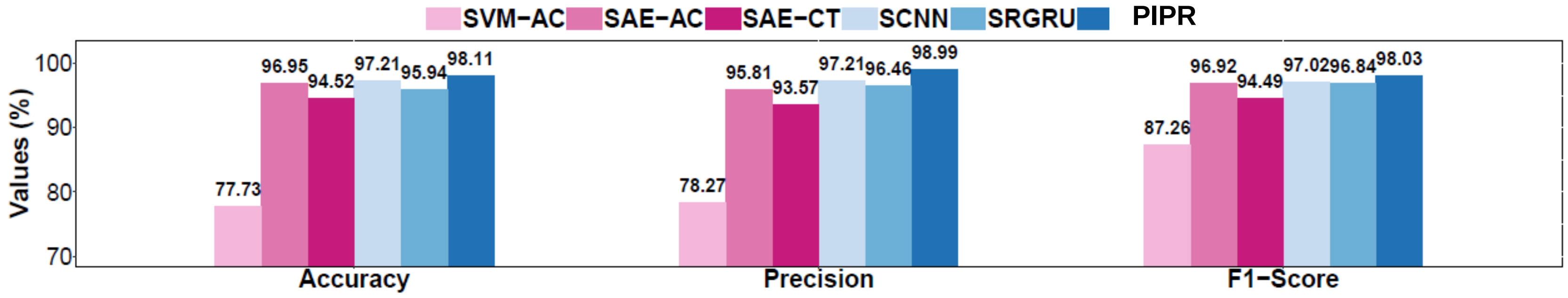


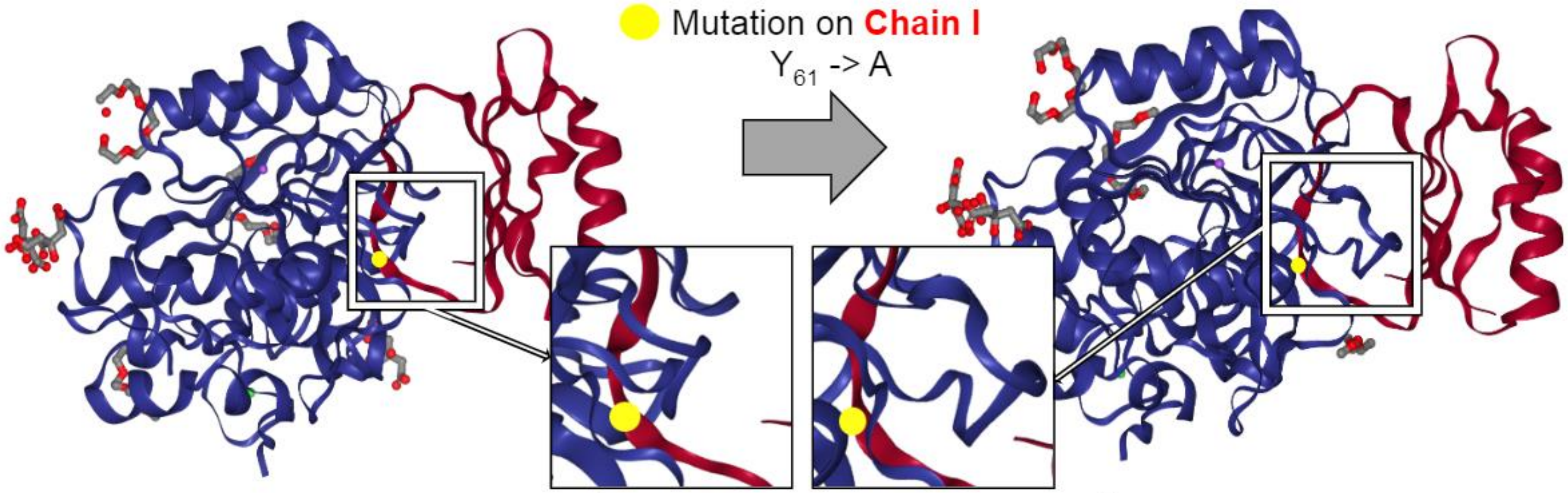
Fig. 4: Evaluation of binary PPI prediction on the Human dataset.

>97% in F1 scores for PPI prediction on Yeast and Human.

- Also reaches SotA performance on PPI type prediction and binding affinity estimation on three other benchmark datasets

*[*ISMB'19, Bioinformatics 2019*]

Point Mutation Effect Estimation



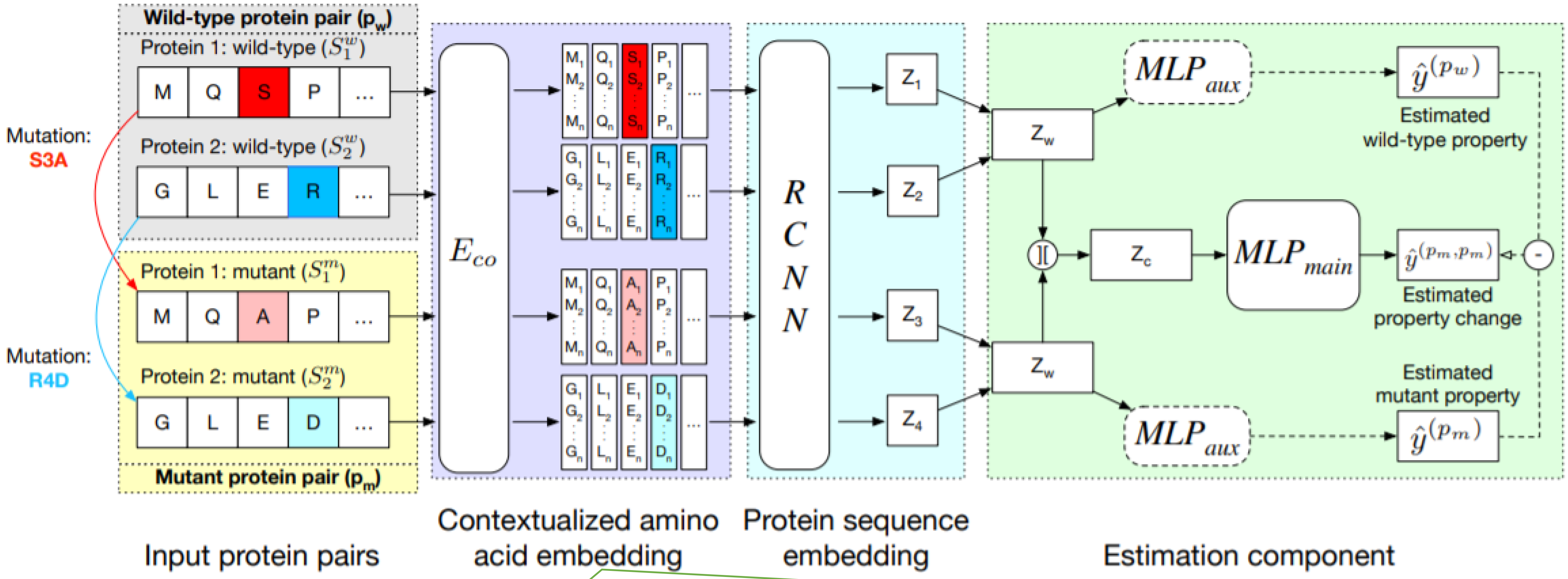
PDB ID	1TM1 (wild type)	1TO1 (mutant)
Binding Affinity (k_d)	2.24E-12	2.70E-10

1 point mutation leads to ~100× change of binding affinity

- Mutations (very **slight changes**) are very difficult to be captured
- **1 or 2 point mutations** may cause a **significant change** to a PPI property

MuPIPR: Pre-trained Amino Acid Language Model + Multi-task Learning

UCLA



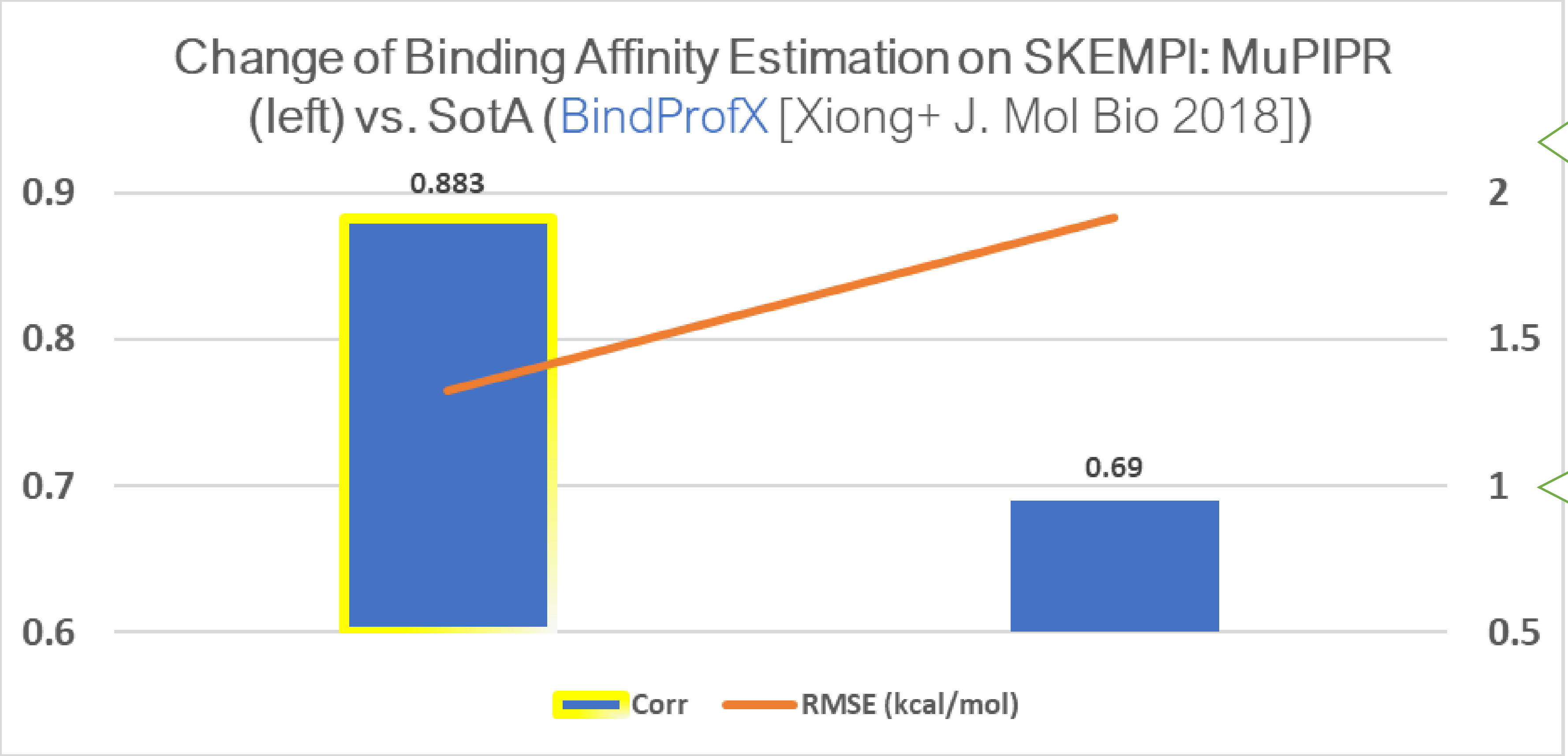
Pre-trained BiLSTM language model on wild-type proteins help propagating point mutation effects

*[*NAR Genom. Bioinform. 2020*]

UCLA

Estimation of PPI Property Changes

*[*NAR Genom. Bioinform. 2020*]



On SKEMPI: ~20% of **absolute improvement** in Pearson's Corr over SotA.

Also offering strong performance on *de novo* prediction.

Also significantly better in estimating change of buried surface area (ΔBSA)

- MuPIPR 0.695 vs. SotA 0.329 in Corr