From Data to Model Programing: Injecting Structured Priors for Knowledge Extraction

Xiang Ren

Department of Computer Science, USC
USC Information Science Institute
USC Machine Learning Center
This **hotel** is my favorite **Hilton property** in **NYC**! It is located right on 42nd street near **Times Square**, it is close to all subways, **Broadways shows**, and next to great restaurants like **Junior’s Cheesecake**, **Virgil’s BBQ** and many others.

--- **TripAdvisor**

**Structured Facts**

1. “Typed” entities
2. “Typed” relationships
This hotel is my favorite Hilton property in NYC! It is located right on 42nd street near Times Square, it is close to all subways, Broadways shows, and next to many great …
Making Machine Learning *Cheaper on Knowledge Extraction*

- Enables *quick* development of applications over various corpora
- Extracts *complex* structures without introducing human errors
Structured Prior Knowledge

Domain Dictionaries

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Canonical Name</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>Donald Trump</td>
<td>Trump, President, Trump, ...</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Ontologies/Knowledge Graphs

- **P1**: (SUBJ-PER, 's children, OBJ-PER) → PER:CHILDREN
- **P2**: (SUBJ-PER, is known as, OBJ-PER) → PER:ALTERNATIVE_NAMES
- **P3**: (SUBJ-ORG, was founded by, OBJ-PER) → ORG:FOUNDED_BY
Challenges of Leveraging Structured Knowledge

• *Noise* in the grounding process

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Canonical Name</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>Wednesday Addams</td>
<td>Wednesday, ...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Today is *Wednesday*. 

*Wednesday Addams* Fictional character
Challenges of Leveraging Structured Knowledge

• *Noise* in the grounding process
• *Incompleteness* of the knowledge sources
Challenges of Leveraging Structured Knowledge

• *Noise* in the grounding process
• *Incompleteness* of the knowledge sources
• *Complex & scalable* reasoning
Previous Work & This Talk

*Learning named entity tagger from domain dictionary* (Shang et al., EMNLP 2018)

*Neural rule grounding* (Zhou et al., 2019)
Previous Work & This Talk

Learning named entity tagger from domain dictionary (Shang et al., EMNLP 2018)

Neural rule grounding (Zhou et al., 2019)
Previous Work & This Talk

Learning named entity tagger from domain dictionary (Shang et al., EMNLP 2018)

Neural rule grounding (Zhou et al., 2019)

KagNet: Learning to Answer Commonsense Questions with Knowledge-aware Graph Networks (Lin et al., 2019)
Learning Named Entity Tagger using *Domain-Specific Dictionary*

EMNLP 2018

*Joint work with Jingbo Shang, Lucas Liu, Xiaotao Gu*
Sequence Tagging: Problem

Every sentence needs to be annotated token by token.

INPUT: Jim bought 300 shares of Acme Corp. in 2006
LABEL: [Jim]:PER bought 300 shares of [Acme Corp.]:ORG in [2006]:Time

Token-level labels by human annotator

BIO: B-PER 0 0 0 0 B-ORG I-ORG 0 B-Time
Challenge: Expensive & Slow on Creating Token-level Training Data

Expensive to adapt to specific domains (e.g., biomedical, business, finance).

Can we generate high-precision, high-recall annotations automatically from domain dictionaries?

Achieved new SoTA on multiple sequence tagging benchmarks with LM-LSTM-CRF architecture (Liu et al., 2018)

(Liu et al., AAAI 2018)
Can We Train Effective Sequence Tagger with Distant Supervision?

No line-by-line annotations, Learn named entity tagger with distant supervision.

Unlabeled corpus + Entity Dictionary → Seq tagging model

“prior knowledge at the input level”
Distant Supervision: Issues with Simple Dictionary Matching

Today is Wednesday.

Name ambiguity & context-agnostic matching $\rightarrow$ false positive

Incomplete dictionary $\rightarrow$ false positive & false negative
AutoNER: Label Filtering & Augmentation

- Removes “irrelevant” entities (and their synonyms) whose canonical names never show up in the corpus

- Introduces out-of-dictionary high-quality phrases* as entities of “unknown” type

---

… Obama Administration Office …

Today is Wednesday.

… Obama Administration Office …

(Shang et al., EMNLP 2018) *(Shang et al., SIGMOD 2015)
AutoNER: “Tie-or-Break” Schema

- Label the relationship of two consecutive tokens:
  - **Tie**, when the two tokens are matched to the same entity
  - **Unknown**, if at least one of the tokens belongs to an *out-of-dictionary phrase*
  - **Break**, otherwise.

<table>
<thead>
<tr>
<th></th>
<th>Today is <em>Wednesday</em></th>
<th>Today is Wednesday.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BIOES</strong></td>
<td>O   O   S-PER</td>
<td>O   O   O</td>
</tr>
<tr>
<td><strong>“Tie-or-Break”</strong></td>
<td>Break Break</td>
<td>Break Break</td>
</tr>
</tbody>
</table>

(Shang et al., EMNLP 2018)
“Tie-or-Break” Encoding Schema

- Label the relationship of two consecutive tokens:
  - **Tie**, when the two tokens are matched to the same entity
  - **Unknown**, if at least one of the tokens belongs to an *out-of-dictionary phrase*
  - **Break**, otherwise.

<table>
<thead>
<tr>
<th>BIOES</th>
<th><em>Ceramic body and 8GB RAM</em></th>
<th><em>Ceramic body and 8GB RAM</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>B-ASP E-ASP O O O</td>
<td>B-ASP E-ASP O O O</td>
<td></td>
</tr>
<tr>
<td>“Tie-or-Break”</td>
<td>Tie Break Break Break</td>
<td>Tie Break Break Break Unknown</td>
</tr>
</tbody>
</table>

(Shang et al., EMNLP 2018)
AutoNER: Multi-task Prediction of Entity Spans & Types

- char-BiLSTM for learning contextualized representation \( u_i \)

(Shang et al., EMNLP 2018)
AutoNER: Multi-task Prediction of Entity Spans & Types

- char-BiLSTM for learning contextualized representation $\mathbf{u}_i$
- 1st classification layer – “tie” or “break”

$$p(y_i = \text{Break}|\mathbf{u}_i) = \sigma(\mathbf{w}^T \mathbf{u}_i)$$
$$\mathcal{L}_{\text{span}} = \sum_{i|y_i \neq \text{Unknown}} l(y_i, p(y_i = \text{Break}|\mathbf{u}_i))$$

(Shang et al., EMNLP 2018)
AutoNER: Multi-task Prediction of Entity *Spans* & *Types*

- char-BiLSTM for learning contextualized representation
- 1st classification layer – “tie” or “break”
- *candidate entity spans* – merge token(s) between two “break”s

(Shang et al., EMNLP 2018)
AutoNER: Multi-task Prediction of Entity *Spans & Types*

- 2nd classification layer – determine entity types

---

(Shang et al., EMNLP 2018)
Results on Biomedical Domain

- BC5CDR NER dataset: chemical & disease
- Fuzzy-LSTM-CRF: models tokens with “unknown” label
- AutoNER: *close to model trained on clean labeled data*

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictionary Matching (DM)*</td>
<td>93.93</td>
<td>58.35</td>
<td>71.98</td>
</tr>
<tr>
<td>Fuzzy-LSTM-CRF (DM + label cleaning &amp; augmentation)</td>
<td>88.27</td>
<td>76.75</td>
<td>82.11</td>
</tr>
<tr>
<td>AutoNER</td>
<td>88.96</td>
<td>81.00</td>
<td><strong>84.80</strong></td>
</tr>
<tr>
<td>LM-LSTM-CRF on gold-standard</td>
<td>88.84</td>
<td>85.16</td>
<td><strong>86.96</strong></td>
</tr>
</tbody>
</table>

*CTD Chemical and Disease vocabularies: 322,882 Chemical and Disease entity names.*
Results on Tech Review Domain

- LaptopReview NER dataset: **aspect terms**
- Models are harder to generalize
- Still a significant gap to *model trained on clean labeled data*

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictionary Matching (DM)*</td>
<td>90.68</td>
<td>44.65</td>
<td>59.84</td>
</tr>
<tr>
<td>Fuzzy-LSTM-CRF (DM + label cleaning &amp; augmentation)</td>
<td>85.08</td>
<td>47.09</td>
<td>60.63</td>
</tr>
<tr>
<td><strong>AutoNER</strong></td>
<td>72.27</td>
<td>59.79</td>
<td><strong>65.44</strong></td>
</tr>
<tr>
<td>LM-LSTM-CRF on gold-standard</td>
<td>84.80</td>
<td>66.51</td>
<td><strong>74.55</strong></td>
</tr>
</tbody>
</table>

*13,457 computer terms crawled from a public website.*
AutoNER: Effectiveness on Leveraging Domain Dictionaries

AutoNER ≈ 300 expert annotated articles on BC5CDR dataset
Neural Rule Grounding for Low-Resource Relation Extraction

Joint work with Wenxuan Zhou & Hunter Lin, under submission
Applying Surface Rules for Relation Extraction

SUBJ-PER founded OBJ-ORG

→ founded by

**Matched**
Bill Gates founded Microsoft

**Unmatched**
- Bill Gates launched Microsoft
- Microsoft is founded by Bill Gates
- Bill Gates is born in Seattle

Different words but semantically similar
## Two Types of Methods

<table>
<thead>
<tr>
<th>Deep learning approaches:</th>
<th>Rule-based approaches:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros:</strong></td>
<td><strong>Pros:</strong></td>
</tr>
<tr>
<td>- Latent representation</td>
<td>- Data independent</td>
</tr>
<tr>
<td>- Good generalization</td>
<td>- Easy to interpret</td>
</tr>
<tr>
<td><strong>Cons:</strong></td>
<td>- High precision</td>
</tr>
<tr>
<td>- Data hungry</td>
<td>- Cons:</td>
</tr>
<tr>
<td>- Hard to interpret</td>
<td>- Low recall (Hard to generalize)</td>
</tr>
<tr>
<td></td>
<td>- Missing context information</td>
</tr>
</tbody>
</table>
Learning a DNN with Only Rules & Unlabeled Sentences

\[ R = \{ r_i \} = \{ b_i \rightarrow h_i \} \]

\[ D = \{ x_i \} \]

\[ hard \ matching \]

\[ f \]

\[ r = b \rightarrow h: \ X \ born \ in \ the \ town \ of \ Y \rightarrow (X, \ city\_of\_birth, \ Y) \]
Learning from Patterns/Rules

(A) Bootstrapping

Rules

Finding Matches

Corpus

Extracting Rules

Suffer from error propagation: The errors in model are reinforced and accumulated
Learning from Patterns/Rules

No supervision from either rules or unlabeled data
Learning by Soft Rule Grounding

Proposing a soft rule matcher to match rules on unlabeled sentences
Learning a **Soft Rule Matching** Function

\[ f_s : (S \cup P) \times P \rightarrow [-1, 1] \]

- Perfect matching \( \rightarrow \) score = 1
- Other cases \( \rightarrow \) score = ?

(Zhou et al., 2019)
Sentence Encoding

\[ h_t = \text{BiLSTM}(h_{t-1}, e_t) \]

\[ s_t = v_h^T \tanh(W_h h_t) \]

\[ a_t = \frac{\exp(s_t)}{\sum_{i=1}^{n} \exp(s_i)} \]

\[ c = \sum_{t=1}^{n} a_t h_t \]
Learning a **Soft Rule Matching Function**

\[ l_{\text{sim}} = \max_{p_1 \in P_+} L_+(p, p_1) + \max_{p_2 \in P_-} L_-(p, p_2) \]

\[ L_+ = (\tau_+ - f(p, p_1))^2_+ \]

\[ L_- = (f(p, p_2) - \tau_-)^2_+ \]

\[ f_s(W_1, W_2) = \frac{z_1^T D^T D z_2}{\|z_1 D\| \|z_2 D\|} \]

(Zhou et al., 2019)
Interpretable Soft Rule Matching

Rule body
**REGD**: Soft Rule Matching for Semi-supervised Learning

Assign each unmatched sentence a pseudo label and weight by soft matching.

Matched

Bill Gates founded Microsoft

Unmatched

Bill Gates launched Microsoft 0.9
Microsoft is founded by Bill Gates 0.8
Bill Gates is born in Seattle 0.3

\[
\begin{align*}
  u_s &= f(s, p^*) \\
  w_s &= \frac{\exp(\theta u_s)}{\sum_{i=1}^{N_b} \exp(\theta u_i)} \\
  l_u &= -\sum_{i=1}^{n} w_i \cdot \log p(r'_s | s)
\end{align*}
\]
REGD: Soft Rule Matching for Semi-supervised Learning

\[ l = l_a + \alpha \cdot l_{pat} + \beta \cdot l_{sim} + \gamma \cdot l_u \]

(Zhou et al., 2019)
REGD: Soft Rule Matching for Semi-supervised Learning

\[ l = l_a + \alpha \cdot l_{pat} + \beta \cdot l_{sim} + \gamma \cdot l_u \]

supervised learning

trained on \( S_u \): pseudo-labeling

(Zhou et al., 2019)
Performance Comparison

Rules have the highest precision (>80%) but lowest F1
Performance Comparison

Supervised DL models generalize better than rules
Performance Comparison

*Semi-supervised models perform extremely bad since labeled data are scarce*
Performance Comparison

REGD outperforms the competing baselines
Ablation on Components

Base models: PA-LSTM is equivalent to REGD with $l_\alpha$ only; Pseudo-Labeling is similar to adding $l_\mu$ to supervised model.
Predicting on New Relations

- Apply soft rule matching to new relations with unseen rules

<table>
<thead>
<tr>
<th>Method</th>
<th>TACRED</th>
<th>SemEval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Rule (exact match)</td>
<td>100</td>
<td>6.1</td>
</tr>
<tr>
<td>CBOW-GloVe</td>
<td>52.4</td>
<td>86.3</td>
</tr>
<tr>
<td>BERT</td>
<td>66.2</td>
<td>76.8</td>
</tr>
<tr>
<td>REGD</td>
<td>61.4</td>
<td>80.5</td>
</tr>
</tbody>
</table>
KagNet: Learning to Answer Commonsense Questions with Knowledge-aware Graph Networks

Joint work with Bill Lin & Jamin Chen, under submission
What is **Commonsense Reasoning**?

- **Naïve Physics**
  - Humans' *natural understanding of the physical world*
  - The *trophy* would not fit in the brown *suitcase* because it was too **big**. What was too **big**?

- **Folk Psychology**
  - Humans' *innate ability to reason about people's behavior and intentions*
  - *Person A puts his trust in Person B*, because ____? . (A and B are friends.)

- How can we **evaluate** the commonsense reasoning capacity of an NLU model?
  - Recent textual multi-choice QA datasets:
    - **CommonsenseQA** (Talmor et al. NAACL 2019)
    - **CommonsenseNLI** (SWAG & HellaSwag, Zellers et al. 2018, 2019)
    - **SocialIQA** (Sap et al. 2019)
CommonsenseQA dataset  (Talmor et al. 2019 )

Where would I not want a fox?
- hen house,  ❌  england,  ❌  mountains,
-  ❌  english hunt,  ❌  california

Why do people read gossip magazines?
- entertained,  ❌  get information,  ❌  learn,
-  ❌  improve know how,  ❌  lawyer told to

What do all humans want to experience in their own home?
- feel comfortable,  ❌  work hard,  ❌  fall in love,
-  ❌  lay eggs,  ❌  live forever

State-of-the-art Model: Fine-tuning BERT-based classifiers

(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

https://www.tau-nlp.org/commonsenseqa
Our Idea: Imposing External Knowledge

Challenges:

1. How can we find the most relevant paths in KG? *(noisy)*
2. What if the best path is not existent in the KG? *(incomplete)*

Where do *adults* use *glue sticks*?

A: classroom  B: office  C: desk drawer

Structured Commonsense Knowledge (e.g. ConceptNet)
KagNet: Knowledge-Aware Graph Networks
KagNet: Knowledge-Aware Graph Networks

- **Question Answer**
- **Concept Recognition**
- **Graph Construction via Path Finding**

**Schema Graph**
KagNet: Knowledge-Aware Graph Networks

Concept Recognition

Language Encoder (e.g. BERT)

Statement Vector

Graph Vector

GCN-LSTM-HPA

MLP

Plausibility score

Graph Construction via Path Finding

Schema Graph

Question Answer

Concepts

Question Concepts

Answer Concepts

KagNet
The GCN-LSTM-HPA Architecture

1. Encoding Unlabeled Schema Graphs $g$

$C_q$  $P_{i,j}$  $C_a$

GCNs

$P_{i,j}[k]$
The GCN-LSTM-HPA Architecture

1. Encoding Unlabeled Schema Graphs $g$

2. Modeling Relational Paths between $c_i^{(q)}$ and $c_j^{(a)}$
The GCN-LSTM-HPA Architecture

1. Encoding Unlabeled Schema Graphs $g$

2. Modeling Relational Paths between $c^{(q)}_i$ and $c^{(a)}_j$

3. Path-level Attention $W_1$
The GCN-LSTM-HPA Architecture

1. Encoding Unlabeled Schema Graphs $g$

2. Modeling Relational Paths between $c_i(q)$ and $c_j(a)$

3. Path-level Attention $W_1$

4. Concept-Pair-level Attention $W_2$

- $S$: Statement Vector
- $R$, $T$: Relation Graphs
- $P_{i,j}$: Path
- $C_q$, $C_a$: Graphs
- $\alpha(i,j,k)$: Path-level Attention
- $\beta(i,j)$: Concept-Pair-level Attention
- $\beta_{i,j}$: Relation between $R_{i,j}$, $T_{i,j}$
- $\text{LSTM}(P_{i,j}[k])$: LSTM Path Encoder
- $g$: Graph

Modeling Relational Paths between $c_i(q)$ and $c_j(a)$.
# KagNet with Different Base Models & Trained on Varying Amounts of Data

<table>
<thead>
<tr>
<th>Model</th>
<th>10(%) of IHtrain</th>
<th>50(%) of IHtrain</th>
<th>100(%) of IHtrain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IHdev-Acc.(%)</td>
<td>IHtest-Acc.(%)</td>
<td>IHdev-Acc.(%)</td>
</tr>
<tr>
<td>Random guess</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>GPT-FineTUNING</td>
<td>27.55</td>
<td>26.51</td>
<td>32.46</td>
</tr>
<tr>
<td>GPT-KAGNet</td>
<td>28.13</td>
<td><strong>26.98</strong></td>
<td>33.72</td>
</tr>
<tr>
<td>BERT-BASE-FineTUNING</td>
<td>30.11</td>
<td>29.78</td>
<td>38.66</td>
</tr>
<tr>
<td>BERT-BASE-KAGNet</td>
<td>31.05</td>
<td><strong>30.94</strong></td>
<td>40.32</td>
</tr>
<tr>
<td>BERT-LARGE-FineTUNING</td>
<td>35.71</td>
<td>32.88</td>
<td>55.45</td>
</tr>
<tr>
<td>BERT-LARGE-KAGNet</td>
<td>36.82</td>
<td><strong>33.91</strong></td>
<td>58.73</td>
</tr>
<tr>
<td>Human Performance</td>
<td>-</td>
<td>88.9</td>
<td>-</td>
</tr>
</tbody>
</table>
## Result on CommonsenseQA Leaderboard (as of 5/14)

### Version 1.11 Random Split Leaderboard

(12,102 examples with 5 answer choices)

<table>
<thead>
<tr>
<th>Model</th>
<th>Affiliation</th>
<th>Date</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td></td>
<td>03/10/2019</td>
<td>88.9</td>
</tr>
<tr>
<td>KagNet</td>
<td>Anonymous</td>
<td>05/14/2019</td>
<td>58.9</td>
</tr>
<tr>
<td>CoSE</td>
<td>Anonymous</td>
<td>04/12/2019</td>
<td>58.2</td>
</tr>
<tr>
<td>SGN-lite</td>
<td>Anonymous</td>
<td>04/20/2019</td>
<td>57.1</td>
</tr>
<tr>
<td>BERTLarge</td>
<td>Tel-Aviv University</td>
<td>03/10/2019</td>
<td>56.7</td>
</tr>
<tr>
<td>BERTBase</td>
<td>University College London</td>
<td>03/13/2019</td>
<td>53.0</td>
</tr>
<tr>
<td>BERTBase</td>
<td>University of Melbourne</td>
<td>04/22/2019</td>
<td>52.6</td>
</tr>
<tr>
<td>GPT</td>
<td>Tel-Aviv University</td>
<td>03/10/2019</td>
<td>45.5</td>
</tr>
<tr>
<td>ESIM+GLOVE</td>
<td>Tel-Aviv University</td>
<td>03/10/2019</td>
<td>34.1</td>
</tr>
<tr>
<td>ESIM+ELMO</td>
<td>Tel-Aviv University</td>
<td>03/10/2019</td>
<td>32.8</td>
</tr>
</tbody>
</table>

[https://www.tau-nlp.org/csqa-leaderboard](https://www.tau-nlp.org/csqa-leaderboard)
Knowledge-Injection Baseline Methods

Table 3: Comparisons with knowledge-aware baseline methods using the [in-house split](both easy and hard mode) on top of BLSTM as the sentence encoder.

Table 4: Ablation study on the KAGNet framework.
Transferability

BERT-KerNet  BERT-FineTune

CSQA  SWAG  WSC

59.01%  53.51%
56.53%  51.23%

No Training!

Interpretability

What do you fill with ink to write on an A4 paper?
A: fountain pen ✓ (KagNet); B: printer (BERT);
C: squid  D: pencil case (GPT); E: newspaper

1. select concept pairs of high att. scores
ink PartOf fountain_pen
ink RelatedTo container IsA fountain_pen
fill HasSubEvent ink AtLocation fountain_pen
fill RelatedTo container IsA fountain_pen
write UsedFor pen
write UsedFor pen IsA fountain_pen
paper RelatedTo write UsedFor fountain_pen

2. Ranking via path-level attn.
Summary

• Learnings
  • Where to solicit complex rules?
  • Coverage of KG grounding; completeness of KG
  • Scalability

• Some open problems
  • Inducing transferrable, latent structures from pre-trained models
  • Modular network for modeling compositional rules
  • Modeling “human efforts” in the objective
Community

• Deep Learning for Low-resource NLP (DeepLo): ACL 2018, EMNLP 2019
• Learning on Limited Data (LLD) Workshop: NeurIPS 2018, ICLR 2019
• Automated Knowledge Base Construction (AKBC)

• Open-source tools
  • AutoNER toolkit: multiple training options (distant training, LM-augmentation, etc.) for building sequence taggers [https://github.com/shangjingbo1226/AutoNER](https://github.com/shangjingbo1226/AutoNER)

• PubMed literature search powered by an auto-constructed, open knowledge graph [http://usc.edu/life-inet](http://usc.edu/life-inet)
Students

Bill Lin
Priya Irukulapati
Woojeong Jin
Wenxuan Zhou

Collaborators

Jure Leskovec, Computer Science, Stanford University
Dan MacFarland, Sociology, Stanford University
Dan Jurafsky, Computer Science, Stanford University
Jiawei Han, Computer Science, UIUC
Kennth Yates, Clinical Education, USC
Craig Knoblock, USC ISI
Curt Langlotz, Bioinformatics, Stanford University
Heng Ji, Computer Science, UIUC
Kuansan Wang, Microsoft Academic
Xiaolin Shi, Snapchat
Mark Musen, Bioinformatics, Stanford University

Research Partnerships

Funding

Microsoft Academic
SPARK AT STANFORD
bioRxiv
Semantic Scholar
ByteDance
BioPortal
J.P. Morgan
Google
Schmidt Family Foundation
Amazon
Adobe
Thank You!

• Injecting structured prior knowledge into various knowledge extraction tasks: input level vs. model level
• Aim to lower the reliance on traditional human-annotated data
• Learnings:
  • Where to solicit complex rules?
  • Coverage of KG grounding; completeness of KG
  • Scalability of computational models

• Technology Transfer: