UNIVERSITY OF MINNESOTA

Semi-Supervised Convolution Graph Kernels for Relation Extraction

(Intern work with NEC Labs)

Xia Ning
CS&E, University of Minnesota, Twin Cities
Yanjun Qi
NEC Labs America, Inc

NEC
NEC Laboratories America, Inc.
Relentless passion for innovation

Outline

UNIVERSITY OF MINNESOTA

- Background:
  - Related work, problem formulation
- Graph representations for syntactics and semantics
- Convolution Graph Kernels for RE
- Semi-supervised learning framework for RE
- Experimental results

NEC Laboratories America, Inc.
Relentless passion for innovation

Background

University of Minnesota

- Relation Extraction (RE) task in Natural Language Processing (NLP)
  - RE: predict semantic relations between entities from sentences
  - Important for efficient knowledge learning
    - Knowledge identification and acquisition
    - Question answering systems

  Example 1: Malignant paragangliomas have been well described in carriers of mutations of the succinate dehydrogenase B (SDHB) gene.

  $\longrightarrow$ caused (Malignant paragangliomas, SDHB mutations)

  Example 2: Where is Colmar Town?

  $\longrightarrow$ located (Colmar Town, ??) $\rightarrow$ relational database query

- Challenges:
  - Scarce labeled data, abundant unlabeled data
  - Data manipulation and information encoding

NEC Laboratories America, Inc.
Relentless passion for innovation

Background

University of Minnesota

- Problem formulation:
  - Binary classification: $S = w_1 w_2 \ldots e_1 \ldots e_2 \ldots w_{n-1} w_n$

$$\mathcal{F}_R(S) = \begin{cases} +1 & \text{if } e_1 \text{ and } e_2 \text{ are related by relation } R \\ -1 & \text{otherwise} \end{cases}$$

- Related work:
  - Data representations:
    - Words: POS, dictionary indexing, chunk tag, entity type
    - Sentence: string, tree, shortest path
  - Kernels:
    - String kernels [1]: word sequence, linear order
    - All-path graph kernel [2]: random walk, sum of direct product
- Issues with existing methods:
  - Rigid matching between words/word sequences
  - Weak concurrent semantic and syntactic encoding

NEC Laboratories America, Inc.
Relentless passion for innovation

Graph Representation and Comparison

UNIVERSITY OF MINNESOTA

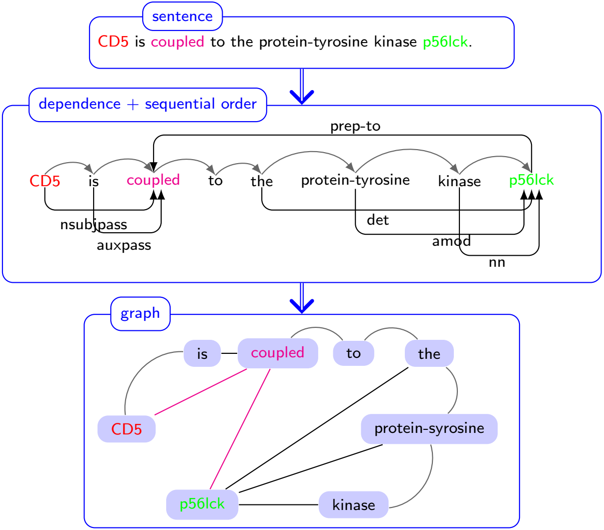
- Motivations:
  - Syntactic: a graph structure naturally exists in a sentence
    - A syntactic "relation" between words:

      Example: cute babies $\rightarrow$ "cute" $\xrightarrow{\text{adjectival modifier}}$ "babies"
    - Enriched graph structures: word $\rightarrow$ vertex, relation $\rightarrow$ edge
  - Semantic: "soft" matching
    - Quantitatively compare words in a semantic-meaningful way

      Example: $sim(\text{"walk"}, \text{"run"}) > sim(\text{"walk"}, \text{"talk"}) > sim(\text{"walk"}, \text{"apple"})$
- Methods:
  - Graph construction:
    - Dependence relations + sequential order
      - Stanford dependence parser
  - Word:
    - Language Model (LM), embedding representations

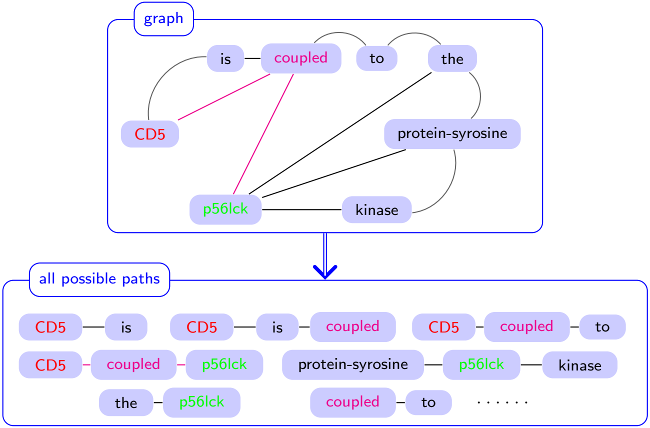
Graph Representation

- Graph Construction



UNIVERSITY OF MINNESOTA

- Graph Representation



NEC Laboratories America, Inc.
Relentless passion for innovation

Convolution Graph Kernel

- Key ideas on Convolution:
  - Decompose big structures into small substructures
  - Kernel on big structures = sum of kernels on small substructures
- Convolution on sentence graphs:
  - Word level kernel + dependence level kernel $\rightarrow$ graph level kernel
  - Multi-level semantic and syntactic information encoding and comparison
  - Multi-level semi-supervision
    - Word embedding: trained from unlabeled sentences
    - Dependence similarity: statics from unlabeled sentences

Convolution Graph Kernel

University of Minnesota

- Kernel $\mathcal{K}_G$ on graphs:

$$\mathcal{K}_G(G, G') = \sum_{p \in P^l(G)} \sum_{p' \in P^l(G')} \mathcal{K}_p(p, p') Pr(p|G) Pr(p'|G')$$

  - Kernel $\mathcal{K}_p$ on single paths:

$$\mathcal{K}_p(p, p') = \begin{cases} \mathcal{K}_w(w_1, w_1') \prod_{i=2,|p|} \mathcal{K}_r(r_{i-1,i}, r'_{i-1,i}) \mathcal{K}_w(w_i, w_i') & \text{if } |p| = |p'| \\ 0 & \text{if } |p| \neq |p'| \end{cases}$$

    - Kernel $\mathcal{K}_w$ on words:
      - Embed each word within a 50-d space: $\Phi(w)$

$$\mathcal{K}_w(w, w') = \exp(-k \times d^2_{Euclidean}(\Phi(w), \Phi(w')))$$

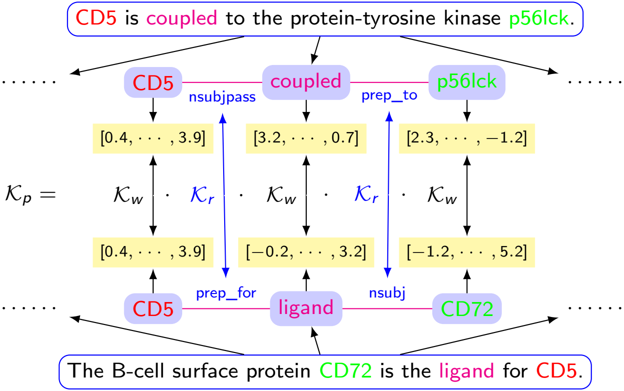
      - Language Model (LM) from semi-supervised deep learning
    - Kernel $\mathcal{K}_r$ on dependencies:
      - Co-occurrence based similarities
      - Co-occurrence when two dependencies related by a common word

NEC Laboratories America, Inc.
Relentless passion for innovation

Convolution Graph Kernel

University of Minnesota

- Kernel on single paths



NEC Laboratories America, Inc.
Relentless passion for innovation

Semi-Supervised Learning Framework

- Semi-supervision on sentence level: self training

1. **NER**
   Name Entity Recognition (NER) to predict potential related entities
2. **entity filter**
   Select sentences which have exactly 2 predicted entities
3. **keyword filter**
   Select sentences which have a relation-related keyword
4. **classification**
   Classify selected sentences using models learned from labeled sentences
5. **rank**
   Select sentences of high prediction scores
6. **incorporation**
   Add selected unlabeled sentences as "pseudo" positive into training set

Experimental Results

University of Minnesota

Table 1: Dataset characteristics for protein-protein interaction

Dataset	# +ppi	# -ppi	dsize	# SD/s	# slen	# ppi/s
AIMED	991	4784	3180	26.9	33.0	5.0
BioInfer	2534	7053	3470	31.5	42.3	8.8
HPRD50	163	270	920	23.4	31.2	3.0
IEPA	335	482	2463	30.0	36.5	1.7
LLL	164	166	537	30.3	37.6	4.3

Table 2: Comparison of SCGK with other methods

Dataset	all-path		ASK		SCGK		SSL-SCGK	
	F	AUC	F	AUC	F	AUC	F	AUC
AIMED	0.564	0.848	0.554	0.824	0.562	0.821	0.572	0.834
BioInfer	0.613	0.819	0.614	0.798	0.606	0.799	0.613	0.806
HPRD50	0.797	0.730	0.727	0.777	0.762	0.819	0.767	0.819
IEPA	0.751	0.851	0.735	0.809	0.737	0.791	0.740	0.797
LLL	0.768	0.834	0.850	0.823	0.849	0.841	0.860	0.847

Experimental Results

UNIVERSITY OF MINNESOTA

Table 3: Effects of path set

	mthd	1		2		3	
		F	AUC	F	AUC	F	AUC
AIMED	upto	0.540	0.795	0.562	0.821	0.561	0.818
	sep			0.560	0.815	0.549	0.800
BioInfer	upto	0.606	0.788	0.606	0.799	0.568	0.753
	sep			0.591	0.776	0.469	0.594
HPRD50	upto	0.755	0.812	0.762	0.819	0.750	0.811
	sep			0.757	0.813	0.738	0.798
IEPA	upto	0.721	0.782	0.737	0.791	0.733	0.794
	sep			0.732	0.796	0.708	0.785
LLL	upto	0.849	0.841	0.833	0.823	0.830	0.805
	sep			0.825	0.754	0.816	0.740

Table 4: Results for SCGK method: BioInfer

k	1		2		3	
	F	AUC	F	AUC	F	AUC
0.005	0.485	0.645	0.498	0.672	0.509	0.685
0.010	0.606	0.788	**0.606**	**0.799**	0.568	0.753
0.020	0.470	0.617	0.484	0.639	0.492	0.648
0.030	0.465	0.589	0.480	0.630	0.489	0.647

NEC Laboratories America, Inc.
Relentless passion for innovation

References

P. Kuksa, Y. Qi, B. Bai, R. Collobert, J. Weston, V. Pavlovic & X. Ning.
Semi-Supervised Abstraction-Augmented String Kernel for Multi-Level Bio-Relation Extraction
ECML PKDD 2010

A. Airola, S. Pyysalo, J. Bjorne, T. Pahikkala, F. Ginter & T. Salakoski
All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning
BMC Bioinformatics, volumn 9, 2008