Modeling Context-sensitive Selectional Preference with Distributed Representations

Naoya Inoue, Masayuki Ono, Yuichiroh Matsubayashi, Naoaki Okazaki, Kentaro Inui
Graduate School of Information Sciences, Tohoku University

Introduction

- What is Context-sensitive Selectional Preference (CSP)?

CSP = predicate slot preference + contextual fitness

John shot a woman.
Police arrested _______

Arg.	SP	CSP
John	☺	☺
watch	☹	☹
woman	☺	☹

"obj. of arrest should be person"
\+ "someone who is shot is unlikely to be arrested"

Results

- Pseudo-disambiguation test (480k tuples from ClueWeb 2012)

Observed:
(police, arrest, man who shot woman)
versus random negative:
(police, arrest, apple which is declicious)

Van de Cruys 2014	Proposed
0.8635	**0.8947**

- Coreference cluster ranking
(12k pronouns in OntoNotes 5.0 [Hovy et al. 2006])

*In his$_{(i)}$ 40-minute speech$_{(j)}$, Chen$_{(i)}$ declared the determination$_{(k)}$ of the people$_{(l)}$ ... **he**$_{(?)}$ made a statement...*

Van de Cruys 2014	Proposed
0.7420	**0.8265**

- Distributed representations successfully model context-sensitivity!

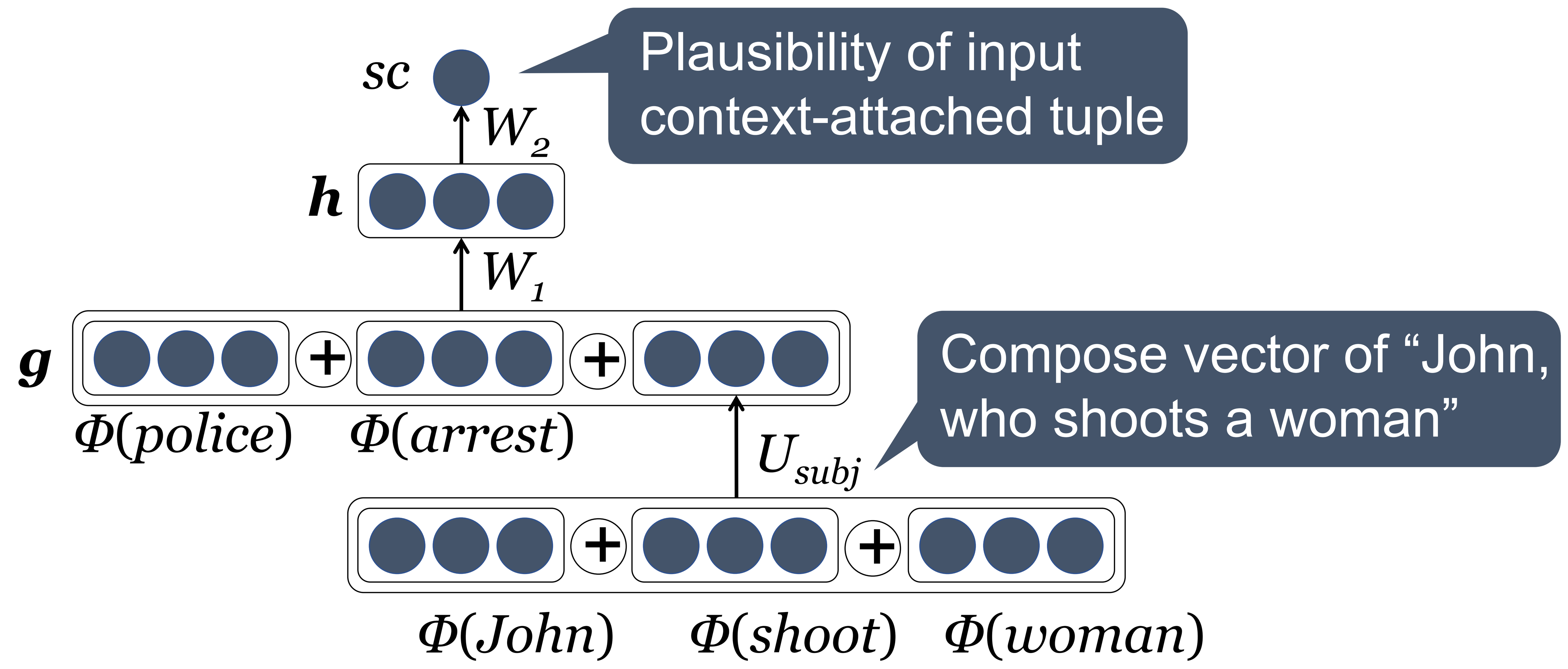
Contacts / Acknowledgement

- Contact author: Naoya Inoue, naoya-i@ecei.tohoku.ac.jp
- This work was partially supported by JSPS KAKENHI Grant Numbers 15H01702, 15H05318, 15K16045, and CREST, JST.

Proposed Model for CSP

Architecture

- Key challenge: how to effectively model context-attached tuples?
- Leverage distributed representations to solve the sparsity issue
- Extend Van de Cruys (2014)'s Neural SP model for CSP



Training

- Generate positive/negative context-attached tuples from ClueWeb12 with coreference resolver (similarly to Chambers & Jurafsky (2008))

- 4,824,394 instances (2,912,624 types) are extracted
- Learn parameters W_*, U_*, Φ via max-margin training:

$$L = max(0, 1 - sc(s^+, v^+, o^+) + sc(s^-, v^+, o^+)) + max(0, 1 - sc(s^+, v^+, o^+) + sc(s^+, v^+, o^-)) + max(0, 1 - sc(s^+, v^+, o^+) + sc(s^-, v^+, o^-))$$

s^-, v^-, o^-: randomly-generated negative example