Contextualized Word Representations for Multi-Sense Embedding

Kazuki Ashihara† Tomoyuki Kajiwara† Yuki Arase† Satoru Uchida‡
†Osaka University ‡Kyushu University
{ashihara.kazuki, arase}@ist.osaka-u.ac.jp, kajiwara@ids.osaka-u.ac.jp, uchida@flc.kyushu-u.ac.jp

Background

- Distributed word representations are used in many NLP tasks.
- Several word meanings are mixed in a representation.
- There are two approaches to generate multi-sense embeddings in order to distinguish each word senses.
  - To generate representations in advance [1][2]. (Our approach)
  - To generate representations from context [3].
- Previous studies generated multiple word representations using part of speech [1] or topic [2].

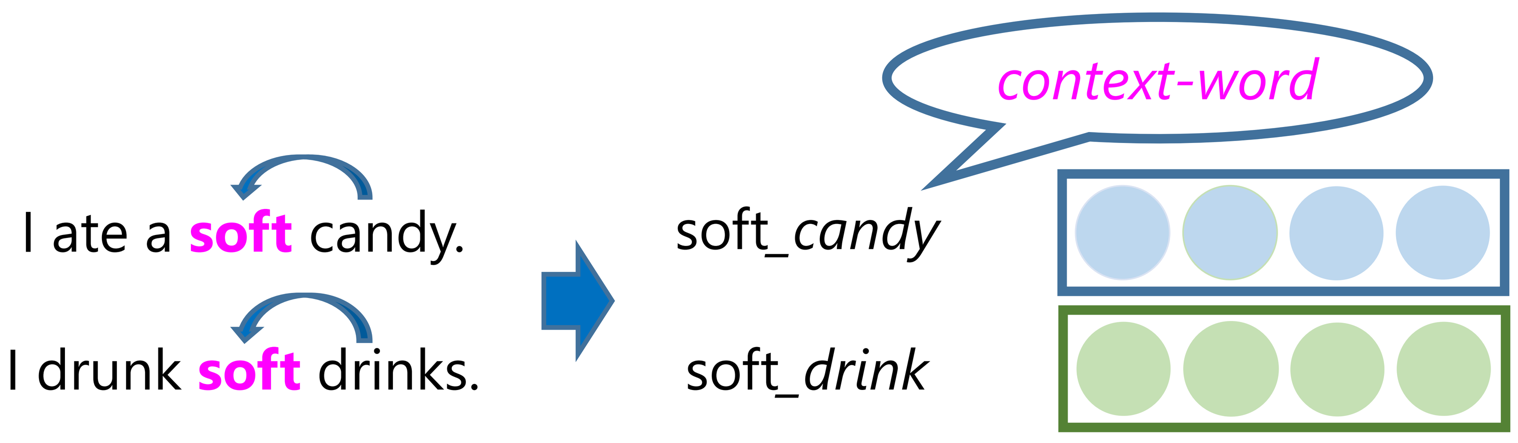
I ate a **soft** candy. → tender

I drank **soft** drinks. → non-alcoholic

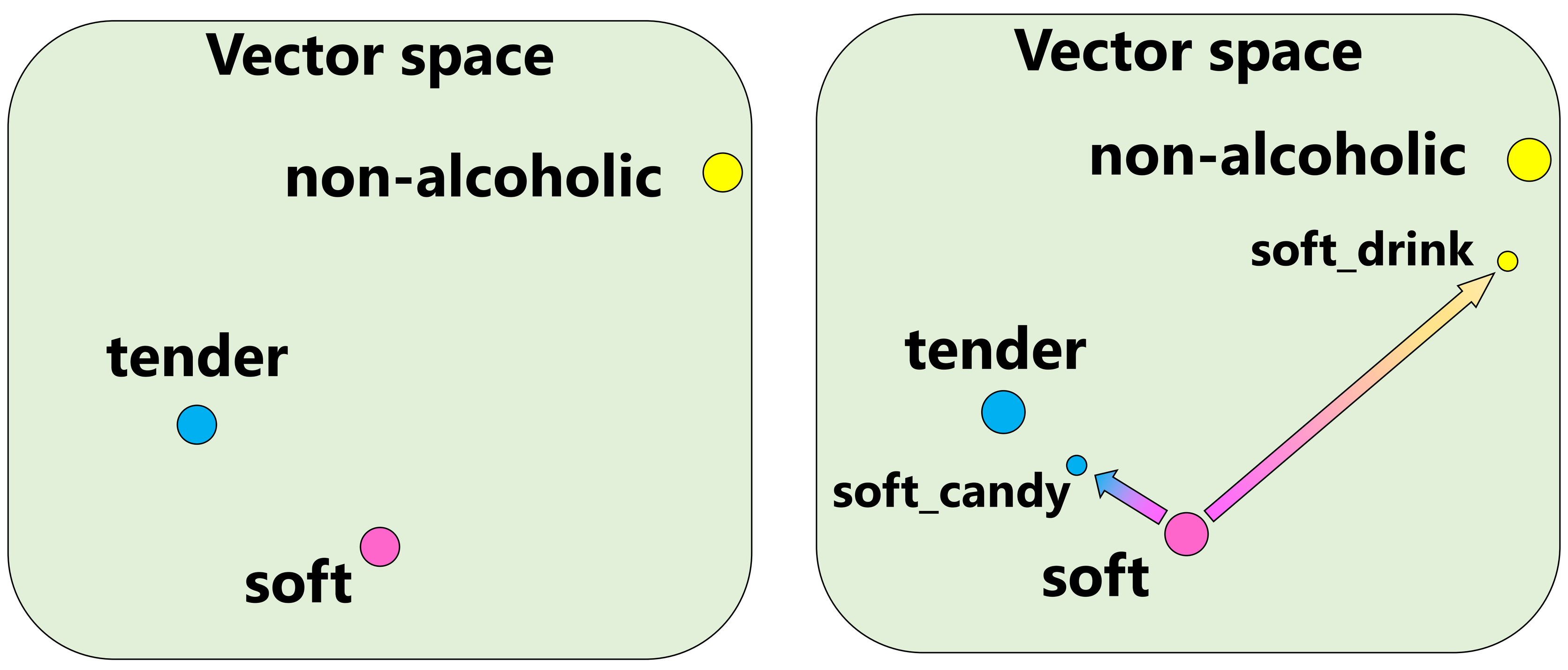
TOPIC : food
POS : adj

Proposed Method

To contextualize a word using **dependency structure relations**



✔ Can distinguish word senses
✘ Data sparseness issue



Pre-training

- Lemmatize
- CBOW algorithm

Post-training

- Initialize using the pre-trained vectors
- Train each word sense

Evaluation Setting

Preprocessing

- 988M sentences from Wikipedia as a training dataset.
- Replace words with frequency of 200 or less to "⟨ unk ⟩" tag
- Parts of speech of the *context-word* were limited to the content word. (i.e. noun, verb, adjective and adverb)

Baseline

model	in-house / ref.	overview
CBOW	in-house	CBOW algorithm
SGNS	ref. [2]	SGNS algorithm
MSSG	ref. [2]	Generates multiple representations by clustering.
POS	in-house	By POS
TOPIC	ref. [2]	By topic
ELMo[3]	in-house	By context

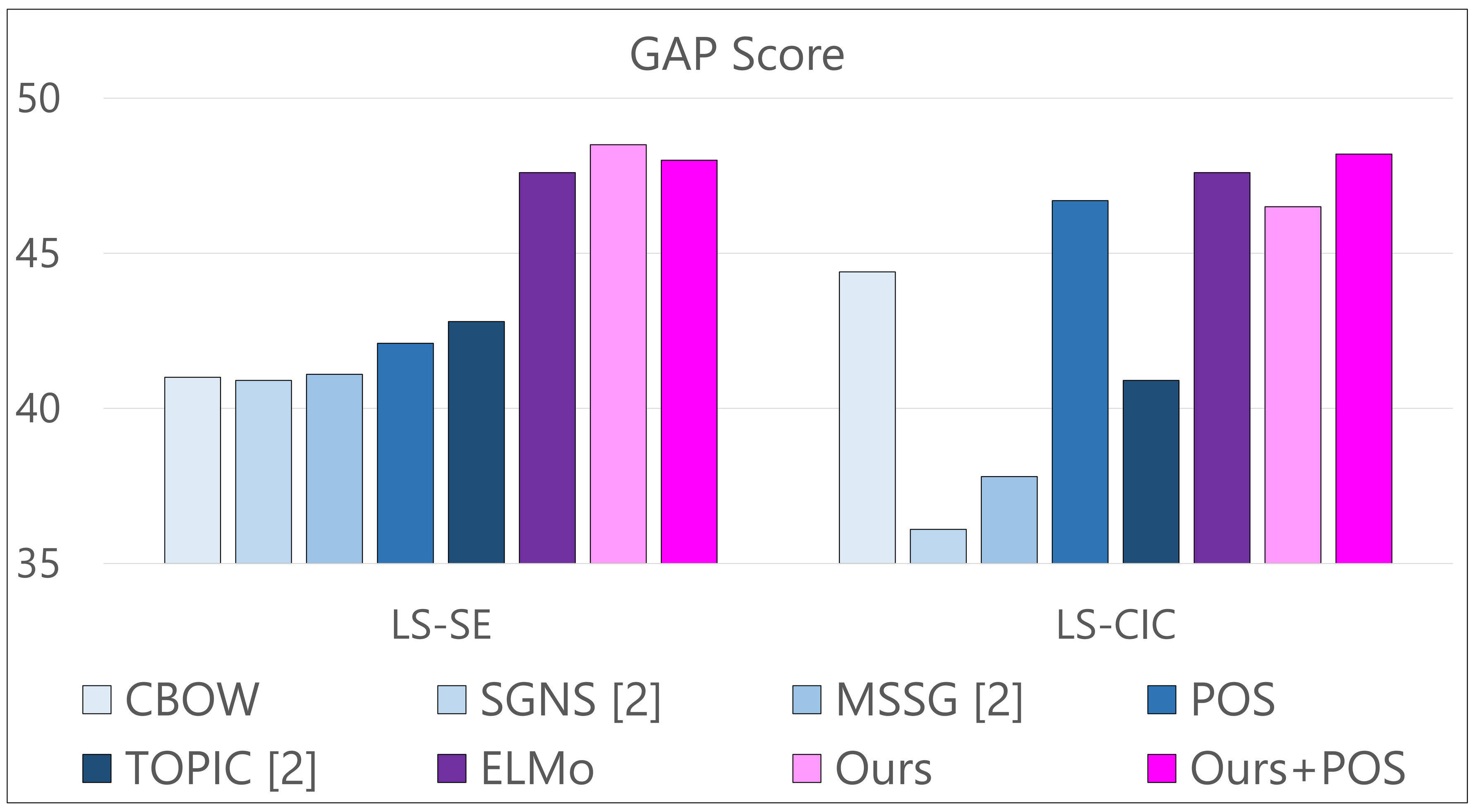
in-house : Our in-house implementation , ref. : Referred to the score from [2]

Lexical Substitution Task

This task requires to …
- rank candidates.
- consider the meaning of the target word in the sentence.
- capture word senses.

TEXT	… you are carrying on two conversations at once and you are required to listen **hard**.
candidates	closely (1), strongly (0), badly (0), carefully (4), …

The task of Lexical Substitution

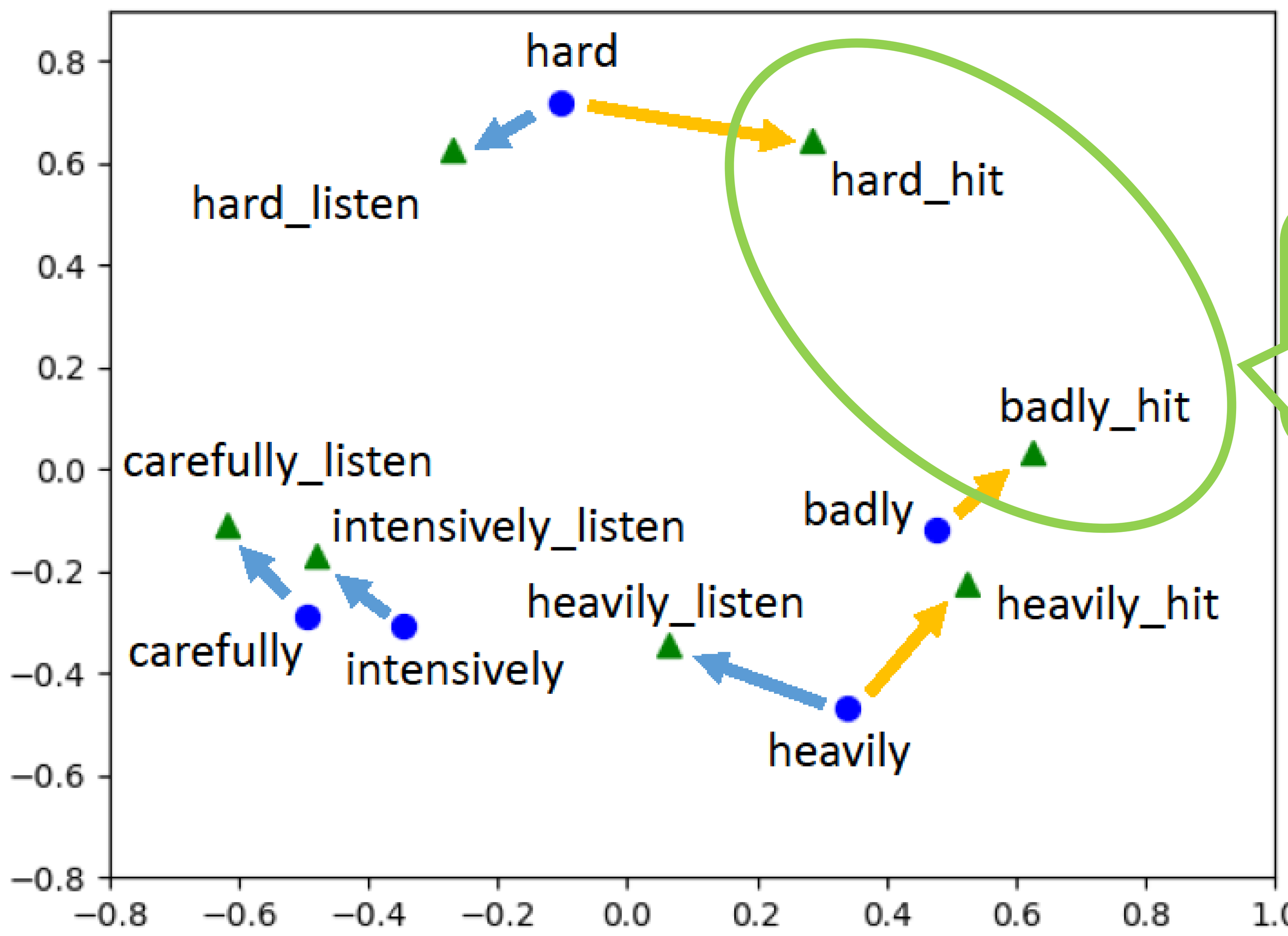


GAP calculates ranking accuracy by considering the weight of candidate.

Target	hard
TEXT1	… you are carrying on two conversations at once and you are required to <u>listen</u> **hard**.
output	carefully (4), intensively (0), closely (1), intently (1), …
TEXT2	One event in particular <u>hits</u> the platoon **hard** : the death of its platoon leader, …
output	badly (3), heavily (0), strongly (0), severely (1), firmly (0), …

Example outputs in a Lexical Substitution Task. The target words in the input are presented in **bold** and *context-words* are <u>underlined</u>.

- Output is different for each context.
- Ours ranked correct candidate words at the top.



Move closer after post-training

2D representation of trained vectors

Summary

- We proposed a method that generates word representations of different senses using *context-words* as clues to distinguish word senses.
- The extensive evaluations confirmed the effectiveness of our method.

[Reference]
[1] Gustavo Henrique Paetzold and Lucia Specia. 2016. "Unsupervised Lexical Simplification for Non-Native Speakers," In Proc. of AAAI, pages 3761-3767.
[2] Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2017. "Learning Topic-Sensitive Word Representations," In Proc. of ACL, pages 441-447.
[3] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. "Deep Contextualized Word Representations," In Proc. of NAACL, pages 2227-2237.