

Thanks for taking around! Due to the visa issue, we are not able to present our work in the beautiful state of Hawaii. If you are interested in our work and have any questions, please feel free to contact Zi Lin ☺



Implanting Rational Knowledge into Distributed Representation at Morpheme Level

Zi Lin^{1,2} Yang Liu²

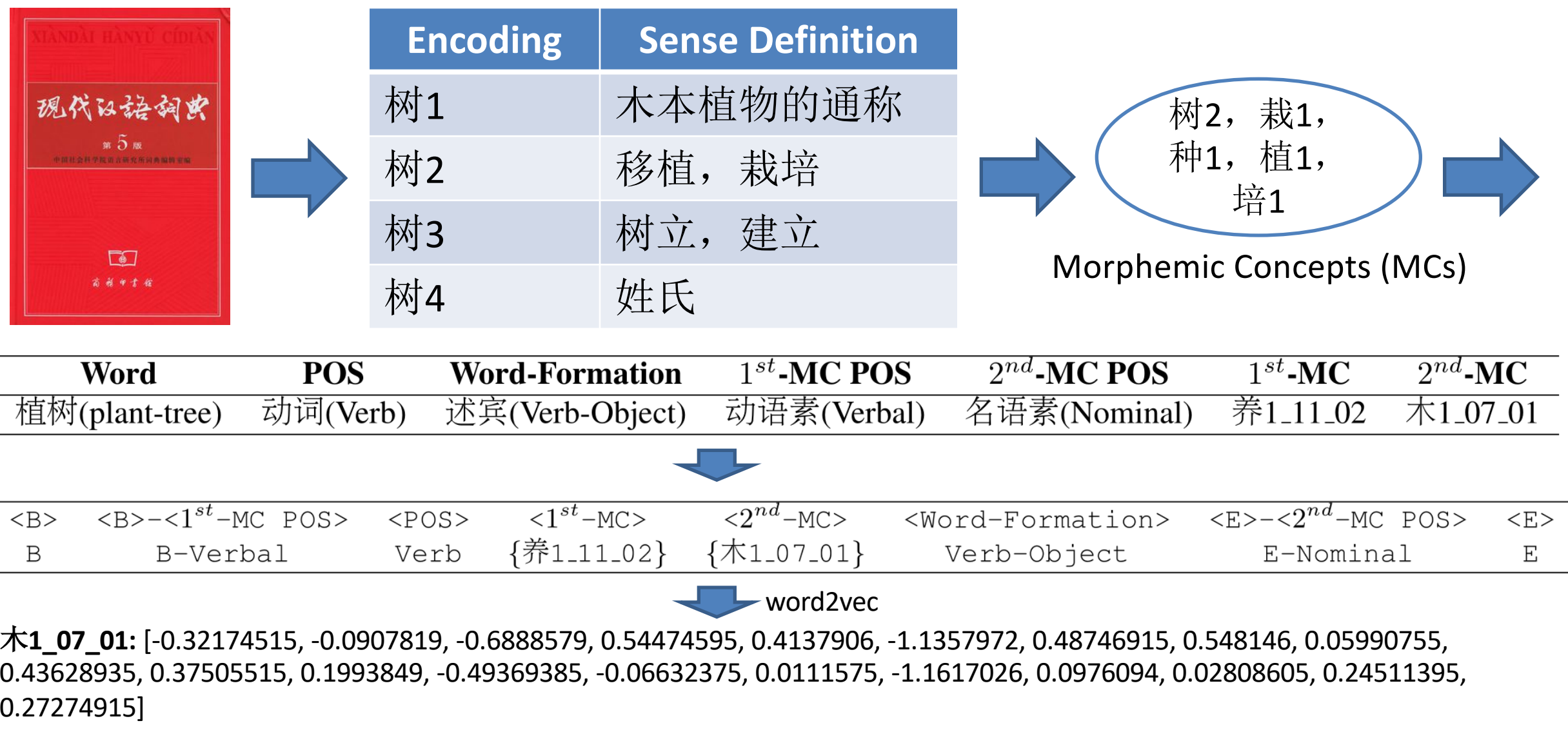
1. Department of Chinese Language and Literature, Peking University
2. Institute of Computational Linguistics, School of EECS, Peking University
`{zi.lin, liuyang}@pku.edu.cn`

Introduction

For parataxis languages like Chinese, word is not a natural unit, and character can provide yet rich **semantic information**. In this work, we:

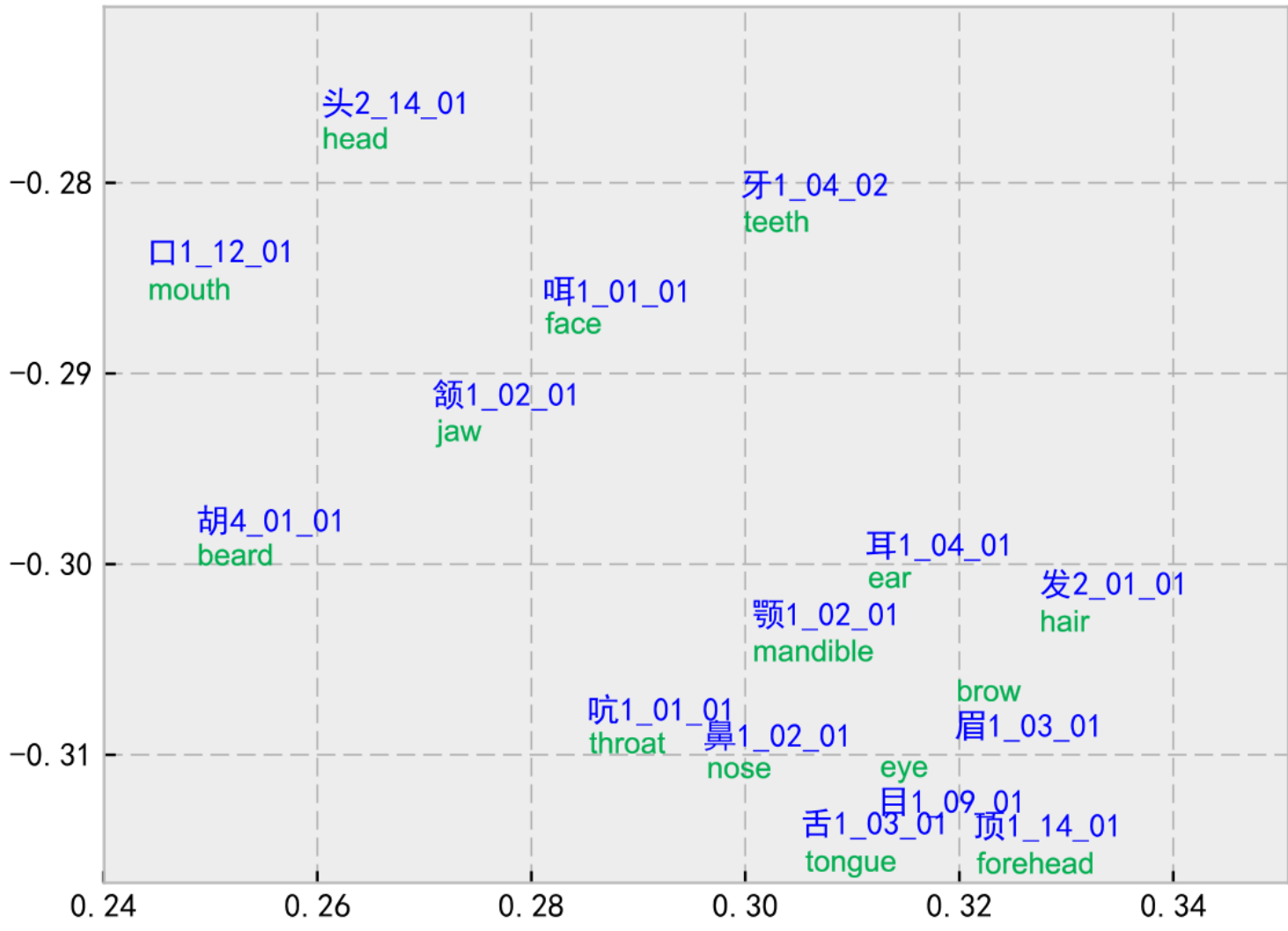
1. Construct the Chinese lexical **ontology** based on semantic word-formation.
2. Propose a novel approach to implanting these pieces of **rational knowledge** into **distributed representation**.

Methods & Results



Intrinsic Evaluation

Paradigmatic Relation



Syntagmatic Relation

MC	Top-Related MCs
{骏1.01.01, 马1.03.01, ..., 驹1.02.02} (horse)	{骏1.01.01, 马1.03.01, ..., 驹1.02.02} (horse) {鞍1.01.01, 鞍1.01.01, ..., 鞍1.01.01} (saddle) {兵1.05.02, 军1.03.01, ..., 卒1.03.01} (soldier)
{鸡1.02.01, 鸭1.01.01, ..., 鹅1.01.01} (fowl)	{仔3.01.01, 子1.13.08, ..., 雏1.02.02} (chick) {野1.07.04} (wild) {坤1.02.02, 母1.06.03, ..., 牝1.01.01} (female) {乳1.05.03, 奶1.03.02} (milk)
{牛1.04.01, 牦1.01.01, ..., 犊1.01.01} (cattle)	{牛1.04.01, 牦1.01.01, ..., 犊1.01.01} (cattle) {土1.07.01, 垆1.01.01, ..., 壤1.03.01} (soil)

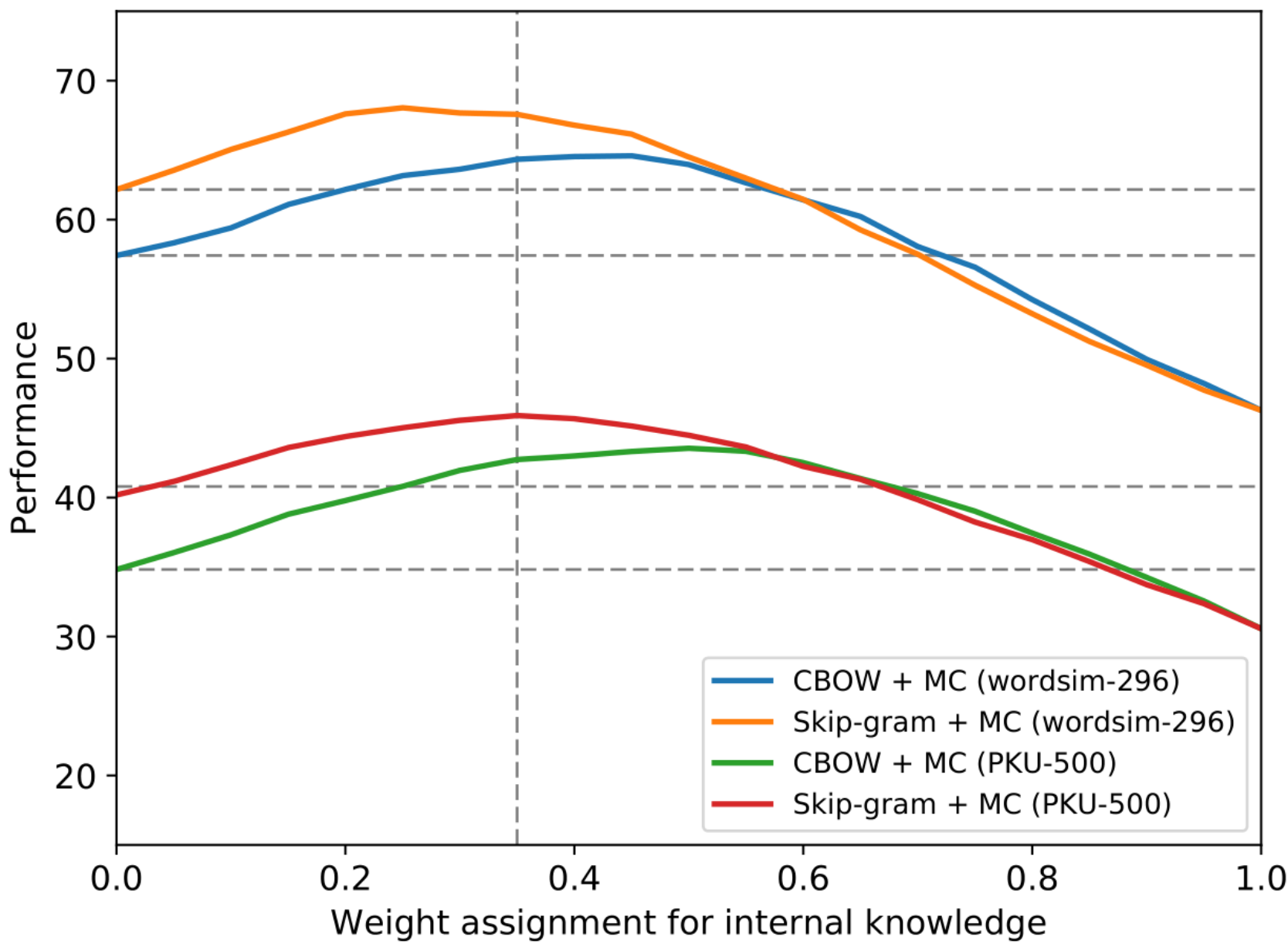
Extrinsic Evaluation

Word Similarity Measurement

Model	wordsim-296	PKU-500
CBOW	57.43	34.82
Skip-gram	62.17	40.19
CWE	58.60	39.25
SE-WRL	61.40	40.89
MC	46.28	30.57
CBOW+MC	64.35	42.74
Skip-gram+MC	67.58	45.91

Table 10: Evaluation results on wordsim-296 and PKU-500 ($\rho \times 100$)

Different Weight of Internal Knowledge



Contribution

1. Taking advantage of such lexical and semantic knowledge representation, the constructed ontology may meet a variety of needs in **humanities** and **NLP**.
2. We, for the first time, put forward an approach to implanting the structured rational knowledge into distributed representation by merely using the lexicon.
3. It is a convenient way by obtaining unambiguous morpheme embeddings simply based on the descriptions in the lexicon, which naturally avoids heavy disambiguation in the corpus as before.

Reference

- [1] Yang Liu, Zi Lin, and Sichen Kang. 2018. Towards a Description of Chinese Morphemic Concepts and Semantic Word-Formation. Journal of Chinese Information Processing, 32(2):11–20.
- [2] Yuanhe Tian and Yang Liu. 2016. Lexical Knowledge Representation and Sense Prediction of Chinese Unknown Words. Journal of Chinese Information Processing, 30(6):26–34.
- [3] Yuanhe Tian and Yang Liu. 2016. Semantic Word-formation Based Chinese Word Similarity Computing. Journal of Chinese Information Processing, 31(1), 94-101.