# A Family of Fuzzy Orthogonal Projection Models for Monolingual and Cross-lingual Hypernymy Prediction

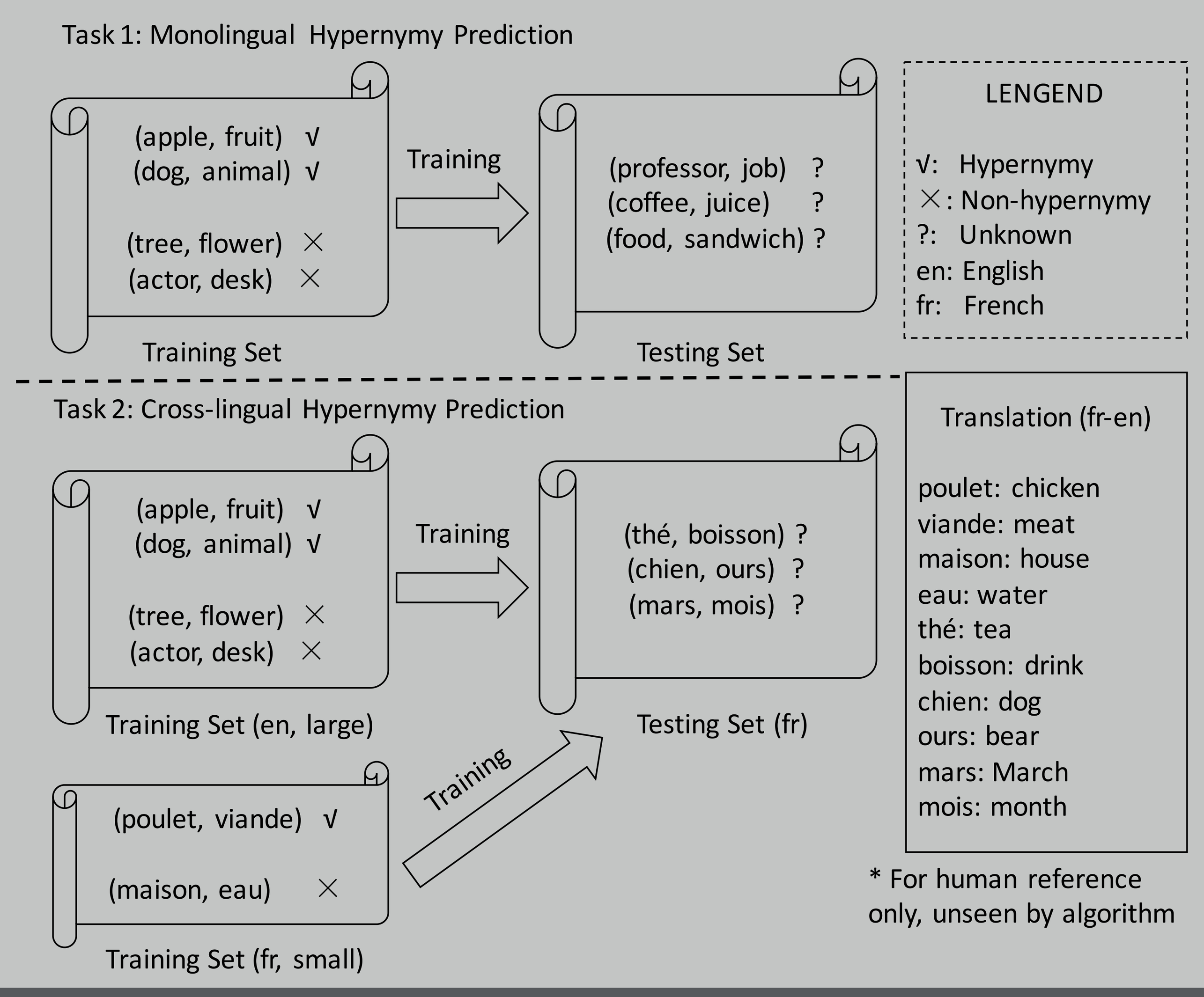

Chengyu Wang[1], Yan Fan[1], Xiaofeng He[1*], Aoying Zhou[2]

[1] School of Computer Science and Software Engineering, East China Normal University
[2] School of Data Science and Engineering, East China Normal University


DaSE Data Science & Engineering

## Introduction

- Learning hypernymy relations is essential for taxonomy construction, fine-grained entity categorization, knowledge base population, etc.
- Previous monolingual distributional models may suffer from "lexical memorization". The sizes of training sets for lower-resourced languages are usually small. Hence, cross-lingual models are necessary for these languages.
- We introduce a family of fuzzy orthogonal projection models for monolingual and cross-lingual hypernymy prediction.




## Monolingual Hypernymy Prediction Model: MWP

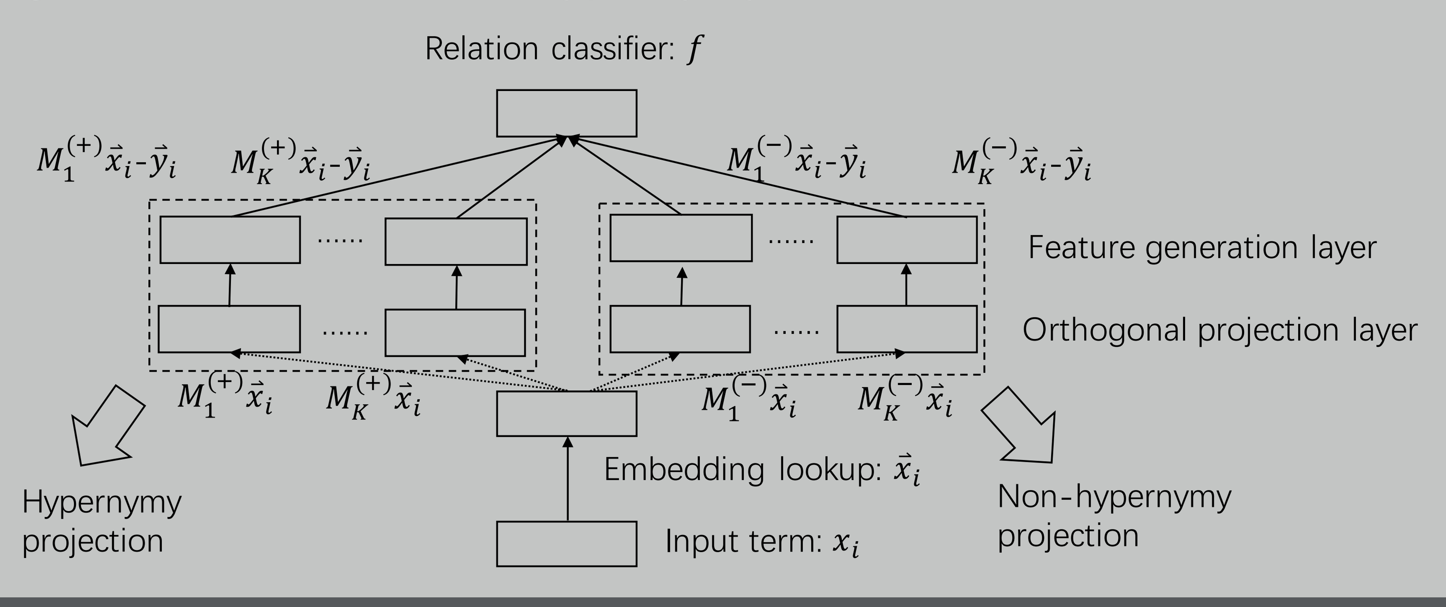
- Hypernymy fuzzy orthogonal projections ($\mathcal{M}^{(+)} = \{\mathbf{M}_1^{(+)}, \cdots, \mathbf{M}_K^{(+)}\}$)
  - Input: a hypernymy relation set $D^{(+)} = \{(x_i, y_i^{(+)})\}$
  - Learning objective ($a_{i,j}^{(+)}$: weight of $(x_i, y_i^{(+)})$ w.r.t. $j$th cluster)

$$\min \tilde{J}(\mathcal{M}^{(+)}) = \frac{1}{2}\sum_{j=1}^{K}\sum_{i=1}^{|D^{(+)}|} a_{i,j}^{(+)}\|\mathbf{M}_j^{(+)}\vec{x}_i - \vec{y}_i^{(+)}\|^2 \text{ s. t. } {\mathbf{M}_j^{(+)}}^T\mathbf{M}_j^{(+)} = \mathbf{I}, \sum_{i=1}^{|D^{(+)}|} a_{i,j}^{(+)} = 1$$

- Non-hypernymy fuzzy orthogonal projections ($\mathcal{M}^{(-)} = \{\mathbf{M}_1^{(-)}, \cdots, \mathbf{M}_K^{(-)}\}$)
  - Input: a non-hypernymy relation set $D^{(-)} = \{(x_i, y_i^{(-)})\}$
  - Learning objective ($a_{i,j}^{(-)}$: weight of $(x_i, y_i^{(-)})$ w.r.t. $j$th cluster)

$$\min \tilde{J}(\mathcal{M}^{(-)}) = \frac{1}{2}\sum_{j=1}^{K}\sum_{i=1}^{|D^{(-)}|} a_{i,j}^{(-)}\|\mathbf{M}_j^{(-)}\vec{x}_i - \vec{y}_i^{(-)}\|^2 \text{ s. t. } {\mathbf{M}_j^{(-)}}^T\mathbf{M}_j^{(-)} = \mathbf{I}, \sum_{i=1}^{|D^{(-)}|} a_{i,j}^{(-)} = 1$$

- Training neural network-based hypernymy classifier




## Cross-lingual Hypernymy Prediction Models: TMWP and ITMWP

- Task input
  - Hypernymy sets of source and target languages ($D_S^{(+)}$ and $D_T^{(+)}$)
  - Non-hypernymy sets of source and target languages ($D_S^{(-)}$ and $D_T^{(-)}$)
  - Unlabeled relations $U_T = \{(x_i, y_i)\}$ of the target language
- Transfer MWP Model (TMWP)
  - Learning fuzzy orthogonal projections for hypernymy relations
  - $\beta$: balance factor of two languages, $\mathbf{S}$: translation matrix (via bilingual lexicon induction), $\gamma_i^{(+)}$: additional weight of $(x_i, y_i^{(+)}) \in D_S^{(+)} \cup D_T^{(+)}$

$$\min \tilde{J}(\mathcal{M}^{(+)}) = \frac{\beta}{2}\sum_{j=1}^{K}\sum_{i=1}^{|D_S^{(+)}|} a_{i,j}^{(+)}\gamma_i^{(+)}\|\mathbf{M}_j^{(+)}\mathbf{S}\vec{x}_i - \mathbf{S}\vec{y}_i^{(+)}\|^2 + \frac{1-\beta}{2}\sum_{j=1}^{K}\sum_{i=1}^{|D_T^{(+)}|} a_{i,j}^{(+)}$$
$$\|\mathbf{M}_j^{(+)}\vec{x}_i - \vec{y}_i^{(+)}\|^2 \text{ s. t. } {\mathbf{M}_j^{(+)}}^T\mathbf{M}_j^{(+)} = \mathbf{I}, \sum_{i=1}^{|D_S^{(+)}|} a_{i,j}^{(+)}\gamma_i^{(+)} = 1, \sum_{i=1}^{|D_T^{(+)}|} a_{i,j}^{(+)} = 1$$

## Cross-lingual Hypernymy Prediction Models: TMWP and ITMWP

- Transfer MWP Model (TMWP)
  - Learning fuzzy orthogonal projections for non-hypernymy relations
  - $\gamma_i^{(-)}$: additional weight of $(x_i, y_i^{(-)}) \in D_S^{(-)} \cup D_T^{(-)}$

$$\min \tilde{J}(\mathcal{M}^{(-)}) = \frac{\beta}{2}\sum_{j=1}^{K}\sum_{i=1}^{|D_S^{(-)}|} a_{i,j}^{(-)}\gamma_i^{(-)}\|\mathbf{M}_j^{(-)}\mathbf{S}\vec{x}_i - \mathbf{S}\vec{y}_i^{(-)}\|^2 + \frac{1-\beta}{2}\sum_{j=1}^{K}\sum_{i=1}^{|D_T^{(-)}|} a_{i,j}^{(-)}$$
$$\|\mathbf{M}_j^{(-)}\vec{x}_i - \vec{y}_i^{(-)}\|^2 \text{ s. t. } {\mathbf{M}_j^{(-)}}^T\mathbf{M}_j^{(-)} = \mathbf{I}, \sum_{i=1}^{|D_S^{(-)}|} a_{i,j}^{(-)}\gamma_i^{(-)} = 1, \sum_{i=1}^{|D_T^{(-)}|} a_{i,j}^{(-)} = 1$$

  - Training neural network-based hypernymy classifier
- Iterative Transfer MWP Model (ITMWP)
  - Employing semi-supervised training data augmentation techniques over the TMWP model to improve the performance

## Monolingual Experiments

- Task 1: Supervised Hypernymy Detection
  - BLESS and ENTAILMENT: General-domain datasets
  - ANIMAL, PLANT and VEHICLE: Domain-specific datasets

| Method | BLESS | ENTAILMENT | ANIMAL | PLANT | VEHICLE |
|---|---|---|---|---|---|
| Mikolov et al. (2013) | 0.84 | 0.83 | 0.80 | 0.81 | 0.82 |
| Yu et al. (2015) | 0.90 | 0.87 | 0.67 | 0.65 | 0.70 |
| Tuan et al. (2016) | 0.93 | 0.91 | 0.89 | 0.92 | 0.89 |
| Nguyen et al. (2017) | 0.94 | 0.91 | 0.83 | 0.91 | 0.83 |
| MWP (Non-orthogonal) | 0.95 | 0.90 | 0.90 | 0.92 | 0.87 |
| MWP | 0.97 | 0.92 | 0.92 | 0.94 | 0.90 |

- Task 2: Unsupervised Hypernymy Relation Classification
  - The unsupervised measure: $\tilde{s}(x_i, y_i) = \|\mathcal{F}^{(-)}(\vec{x}_i, \vec{y}_i)\|_2 - \|\mathcal{F}^{(+)}(\vec{x}_i, \vec{y}_i)\|_2$
  - BLESS: Hypernymy vs. Reverse-hypernymy
  - WBLESS: Hypernymy vs. Other relations

| Method | BLESS | WBLESS |
|---|---|---|
| Santus et al. (2014) | 0.87 | - |
| Weeds et al. (2014) | - | 0.75 |
| Kiela et al. (2015) | 0.88 | 0.75 |
| Nguyen et al. (2017) | 0.92 | 0.87 |
| Roller et al. (2018) | 0.96 | 0.87 |
| MWP (Non-orthogonal) | 0.95 | 0.89 |
| MWP | 0.97 | 0.92 |

## Cross-lingual Experiments

- Experimental Settings
  - Source language: 17,394 hypernymy relations and 67,930 non-hypernymy relations in English
  - Target languages: labeled word pairs of seven languages derived from the Open Multilingual Wordnet project
  - Learning settings: "few-shot learning" for target languaes
- Experimental Results

| Method | French | Chinese | Japanese | Italian | Thai | Finnish | Greek |
|---|---|---|---|---|---|---|---|
| Task: cross-lingual hypernymy direction classification | | | | | | | |
| Santus et al. (2014) | 0.65 | 0.65 | 0.68 | 0.61 | 0.63 | 0.70 | 0.62 |
| Weeds et al. (2014) | 0.76 | 0.71 | 0.77 | 0.76 | 0.72 | 0.77 | 0.70 |
| Kiela et al. (2015) | 0.67 | 0.65 | 0.71 | 0.68 | 0.65 | 0.70 | 0.62 |
| Shwartz et al. (2016) | 0.79 | 0.67 | 0.71 | 0.72 | 0.66 | 0.75 | 0.66 |
| TMWP (N) | 0.78 | 0.71 | 0.75 | 0.76 | 0.73 | 0.76 | 0.71 |
| TMWP | 0.80 | 0.72 | 0.76 | 0.78 | 0.75 | 0.78 | 0.73 |
| ITMWP (N) | 0.82 | 0.72 | 0.76 | 0.78 | 0.75 | 0.81 | 0.72 |
| ITMWP | 0.81 | 0.74 | 0.78 | 0.81 | 0.78 | 0.81 | 0.75 |
| Task: cross-lingual hypernymy detection | | | | | | | |
| Santus et al. (2014) | 0.67 | 0.63 | 0.67 | 0.62 | 0.64 | 0.62 | 0.64 |
| Weeds et al. (2014) | 0.74 | 0.66 | 0.68 | 0.71 | 0.62 | 0.68 | 0.69 |
| Kiela et al. (2015) | 0.70 | 0.61 | 0.65 | 0.68 | 0.57 | 0.61 | 0.67 |
| Shwartz et al. (2016) | 0.72 | 0.66 | 0.69 | 0.64 | 0.66 | 0.69 | 0.70 |
| TMWP (N) | 0.72 | 0.67 | 0.70 | 0.70 | 0.68 | 0.71 | 0.70 |
| TMWP | 0.75 | 0.71 | 0.76 | 0.72 | 0.69 | 0.72 | 0.71 |
| ITMWP (N) | 0.72 | 0.74 | 0.77 | 0.74 | 0.67 | 0.71 | 0.72 |
| ITMWP | 0.76 | 0.73 | 0.78 | 0.74 | 0.72 | 0.73 | 0.73 |