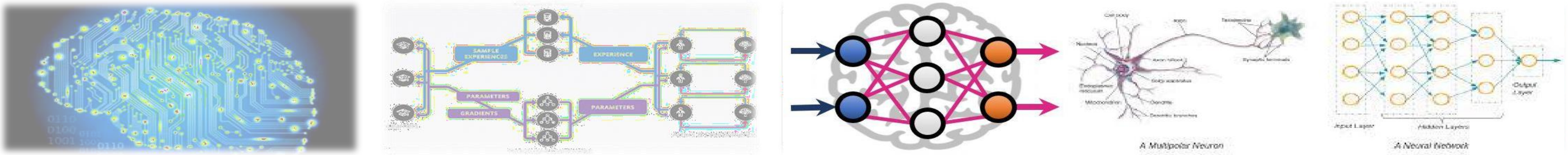


SOTA : Hypernym discovery

state-of-the-art of hypernym discovery

Geonil Yun,
rjsdlf45@gmail.com

2021. 09. 14.

Introduction

- **CRIM – Hybrid Approach**

  - Combination of two approaches

    unsupervised + supervised

    - Unsupervised : Pattern-based approach
    - Supervised : Projection learning approach

  - Dataset – SemEval2018 (general, medical, music)

    - 2 query types, concept and named entity

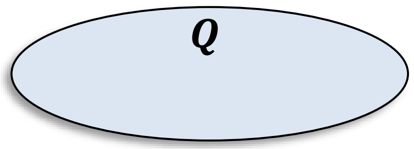
Method

- **CRIM – Hybrid Approach**

  - **Unsupervised** – Pattern-based hypernym discovery

    - Pattern-based approach is known to suffer from **low recall**.
    - Propose algorithm for increasing recall.

① input query $\boldsymbol{q}$, Create empty set $\boldsymbol{Q}$.

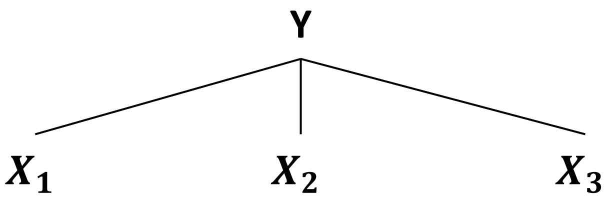


② Find **co-hyponym(q')** by using patterns.

enumeration patterns

“X_1, X_2 and X_3”

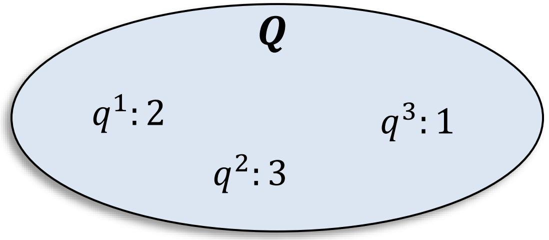
- X_1, X_2, X_3 are co-hyponym.



Method

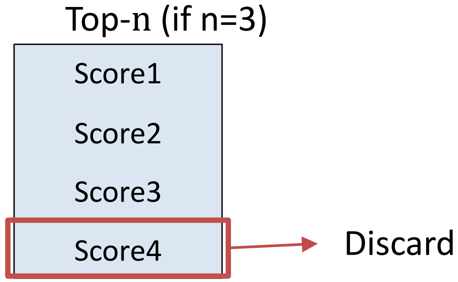
- **CRIM – Hybrid Approach**
  - **Unsupervised** – Pattern-based hypernym discovery

③ Insert q' to Q and record **frequency.**



④ **Score** each co-hyponym $q' \in Q$, and **rank** top-n. (in this paper $n = 5$)

$$\boldsymbol{Score = frequency(q') \cdot similarity(q, q')}$$



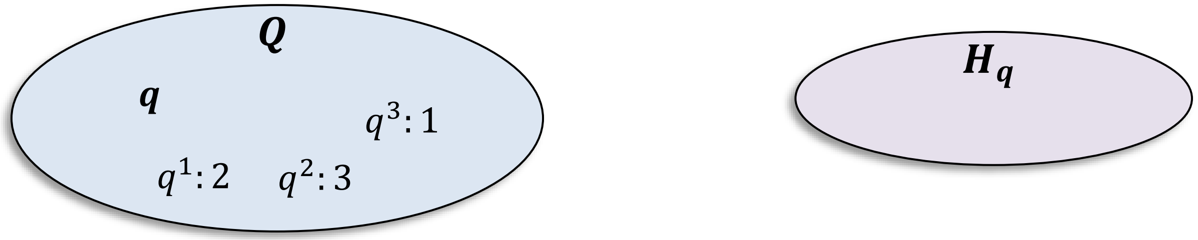
- Similarity is **cosine similarity** of embeddings of q and q'.
- Discard rest of ranking words $\boldsymbol{q'}$ from $\boldsymbol{Q}$.

Method

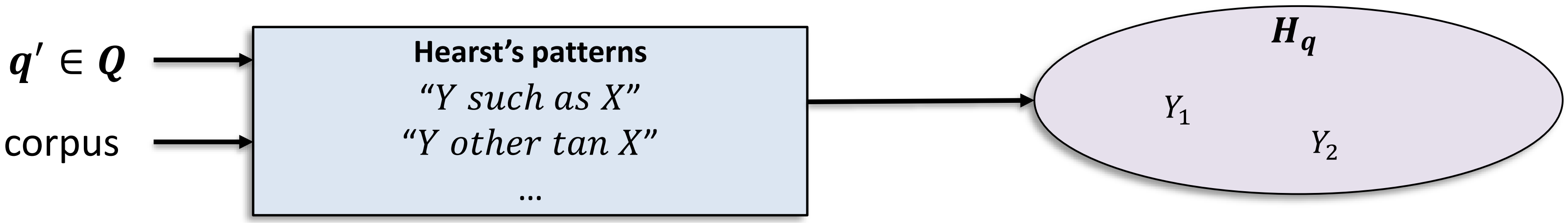
- **CRIM – Hybrid Approach**

  ➢ **Unsupervised** – Pattern-based hypernym discovery

  ⑤ Add query $\boldsymbol{q}$ to $\boldsymbol{Q}$. Create **empty set** of hypernyms $\boldsymbol{H_q}$.



  ⑥ For each $\boldsymbol{q'} \in \boldsymbol{Q}$, search hypernym of $\boldsymbol{q}$' from corpus and add to $\boldsymbol{H_q}$.



  ▪ Find hypernym by using **Hearst's Patterns.**

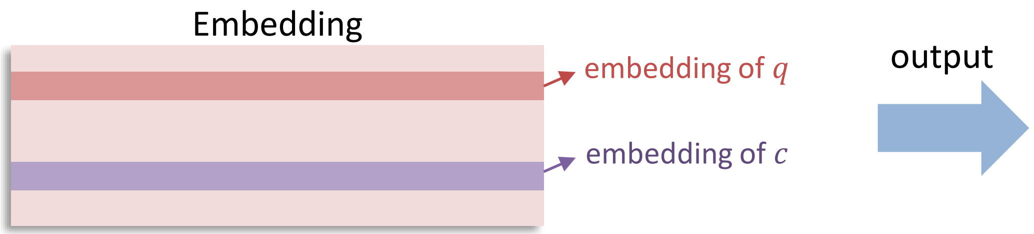
Method

- **CRIM – Hybrid Approach**

  - **Unsupervised** – Pattern-based hypernym discovery

    ⑦ Score each candidate $c \in H_q$, and rank.

$$Score = frequency(c) \cdot similarity(c, q)$$



$$sorted\ [score1, score2, score3\ \ldots]$$

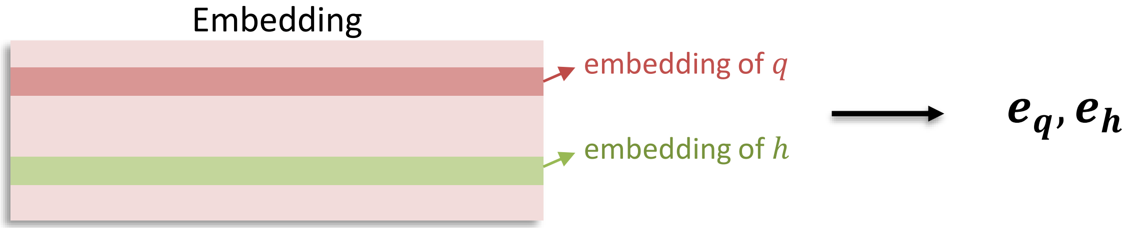
  - Similarity is cosine similarity of embeddings of c and q.
  - Output hypernym candidates according to score.

Method

- **CRIM – Hybrid Approach**
  - **Supervised** – Learning projections for hypernym discovery

  - Apply decision function **to all candidate hypernyms**.
  - Input embeddings of $\boldsymbol{q}$ and $\boldsymbol{h}$.

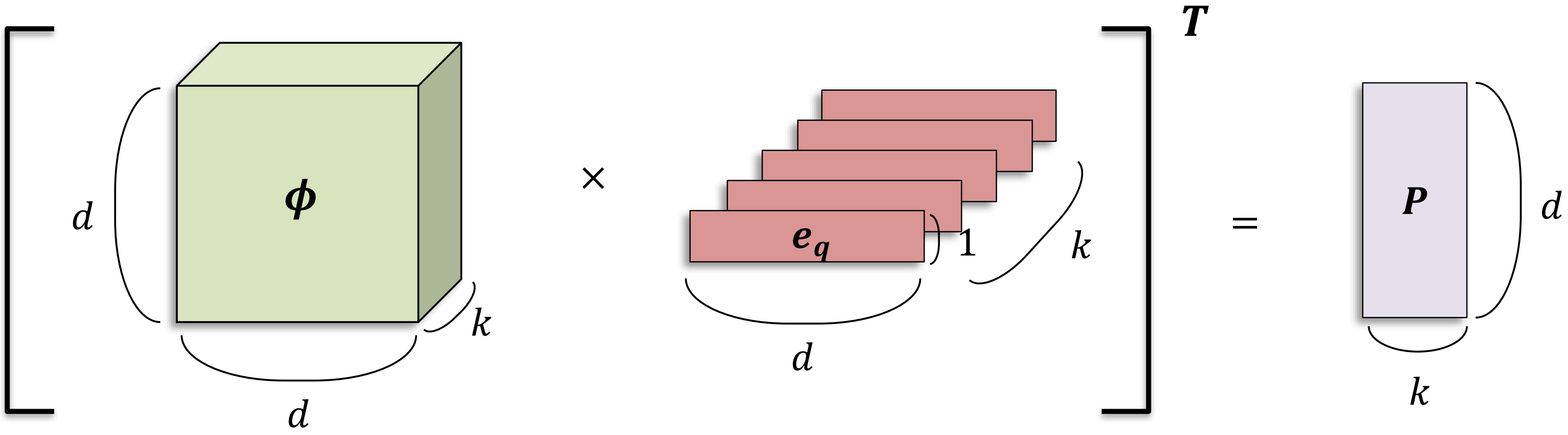
- **Model**



  - For input term q and h, search embedding $\boldsymbol{e_q}, \boldsymbol{e_h} \in \mathbb{R}^{\boldsymbol{d\times 1}}$

Method

- **CRIM – Hybrid Approach**
  - **Supervised** – Learning projections for hypernym discovery
    - **Model**



$$P_i = \left(\phi_i \cdot e_q\right)^T, i \in \{1, \ldots, k\}$$

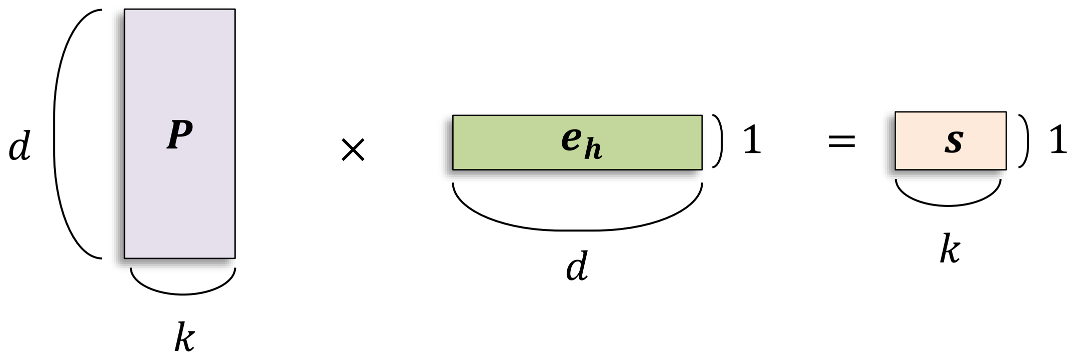
- 3-D projection matrix $\phi_i \in \mathbb{R}^{d \times d}$ for $i \in \{1, \ldots, k\}$, producing matrix $P \in \mathbb{R}^{k \times d}$

Method

- **CRIM – Hybrid Approach**

  - **Supervised** – Learning projections for hypernym discovery

    - **Model**



$$\boldsymbol{s} = \boldsymbol{P} \cdot \boldsymbol{e_h}\ ,\ \boldsymbol{s} \in \mathbb{R}^{\boldsymbol{k \times 1}}$$

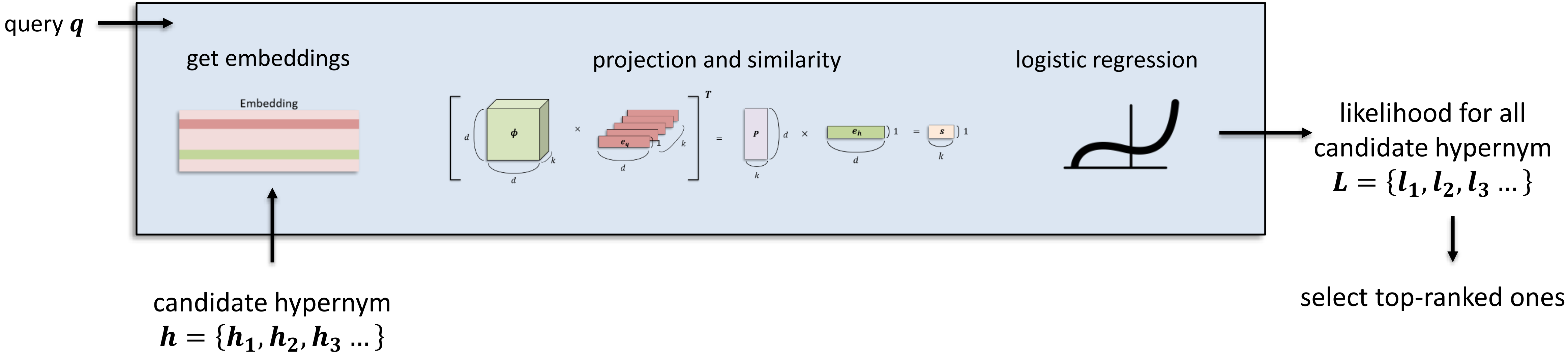
- $\boldsymbol{P}$ is $\boldsymbol{k}$ projections of $\boldsymbol{e_q}$, calculate similarity of $\boldsymbol{P}$ and $\boldsymbol{e_h}$.

Method

- **CRIM – Hybrid Approach**
  - ➢ **Supervised** – Learning projections for hypernym discovery
    - **Model**

$$y = \sigma(W \cdot s + b)$$

- For s, affine transformation and sigmoid activation.
- Logistic regression, likelihood of hypernym.



Method

- **CRIM – Hybrid Approach**

  - **Supervised** – Learning projections for hypernym discovery

    - **Training**

      - Train the model using **negative sampling**.
      - For each pair of (query, hypernym), generate m of negative samples.
      - Loss for **binary cross-entropy**.
      - Minimize loss by using **Adam** optimizer.

$$H(q, h, t) = t \times \log(y) + (1 - t)\log(1 - y)$$

q, h, t indicate query, candidate hypernym, target(1 for positive, 0 for negative) respectively

Method

- **CRIM – Hybrid Approach**
  - **Combination each approach**
    - Take top-100 candidates according to each(pattern-based, learning projection).
    - Normalize scores and sum, rerank the candidates new score

Result

- **Result of CRIM**

	Test set 1A			Test set 2A			Test set 2B		
	MAP	MRR	P@1	MAP	MRR	P@1	MAP	MRR	P@1
Hybrid run 1	19.78	36.10	29.67	34.05	54.64	49.20	40.97	60.93	48.20
Hybrid run 2	19.54	35.94	29.67	31.54	46.19	41.40	40.88	60.18	47.60
Supervised run 1	19.09	34.53	28.33	30.63	44.18	40.00	38.24	53.45	38.20
Supervised run 2	19.11	34.99	28.80	28.51	37.63	34.40	39.95	57.34	43.00
Unsupervised	7.36	15.44	10.73	21.74	47.85	38.60	15.86	34.98	28.60
BaselineSUP	10.60	23.83	19.73	18.84	41.07	35.40	12.99	39.36	33.20
BaselineMFH	8.77	21.39	19.80	28.93	35.80	32.60	33.32	51.48	36.20
Hybrid cross-eval	N/A	N/A	N/A	27.18	49.51	43.20	21.02	37.55	29.20
Supervised cross-eval	N/A	N/A	N/A	22.89	38.30	30.60	16.80	29.28	19.20
BaselineSUP cross-eval	N/A	N/A	N/A	11.66	23.83	17.20	6.31	16.54	9.60

- Run 1 used data augmentation.
- MAP(Mean Average Precision), MRR(Mean reciprocal rank), P@1(Precision at rank1)

Others

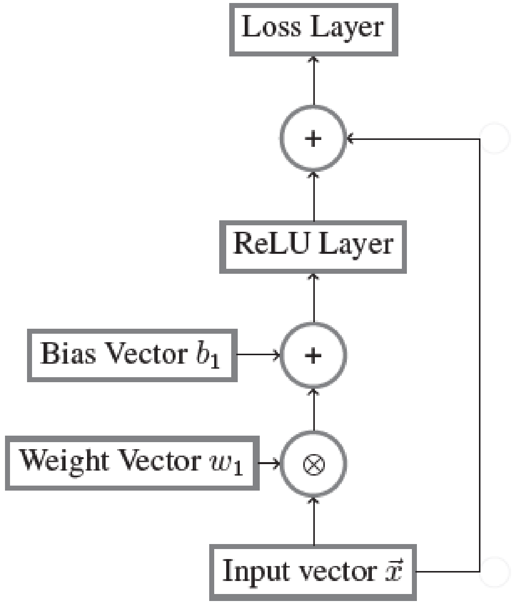
- **Hybrid SVD & NN**

	General			Medical			Music		
Model Variant	MAP	MRR	P@5	MAP	MRR	P@5	MAP	MRR	P@5
Count Hearst Patterns	4.60	11.70	4.10	14.99	43.18	13.70	7.65	26.14	6.76
SVD Hearst Patterns	6.19	15.12	5.65	15.80	47.01	14.53	8.89	29.63	8.05
Hypernyms of Nearest Neighbor	9.85	24.56	8.76	29.57	48.18	34.10	38.65	72.77	38.45
Hybrid of Raw Count & NN	14.82	32.61	13.80	35.29	63.59	38.73	28.22	61.26	30.67
Hybrid of SVD & NN	**15.97**	**34.07**	**15.00**	**37.85**	**64.47**	**40.19**	**54.62**	**77.24**	**55.08**

- Use **Combination of two approaches**(Pattern-based, Distributional).
- Use extended Hearst-pattern(47) and for increasing recall, **SVD to co-occurrence matrix** of Hearst Pattern.
- **Single nearest neighbor** approach, use **cosine-similarity** of vector representations.
- Compare to CRIM, 2(Medical, Music) tasks are better than previous.

Others

- **SPON-Strict Partial Order Network**



$$\boldsymbol{g(\vec{x}) = ReRU(w_1 \otimes \vec{x} + b_1)}$$

$$\boldsymbol{f(\vec{x}) = \vec{x} + g(\vec{x})}$$

$$\boldsymbol{\psi(x, y) = \sum_{i}^{d} max(0, \epsilon + f(\vec{x})_i - \vec{y}_i)}$$

scalar hyper-parameter

- Comprising of **non-negative activations** and **residual connections**.
- Use **negative sampling** for training.
- Augmented SPON is allow **OOV input**.

Others

- **SPON-Strict Partial Order Network**

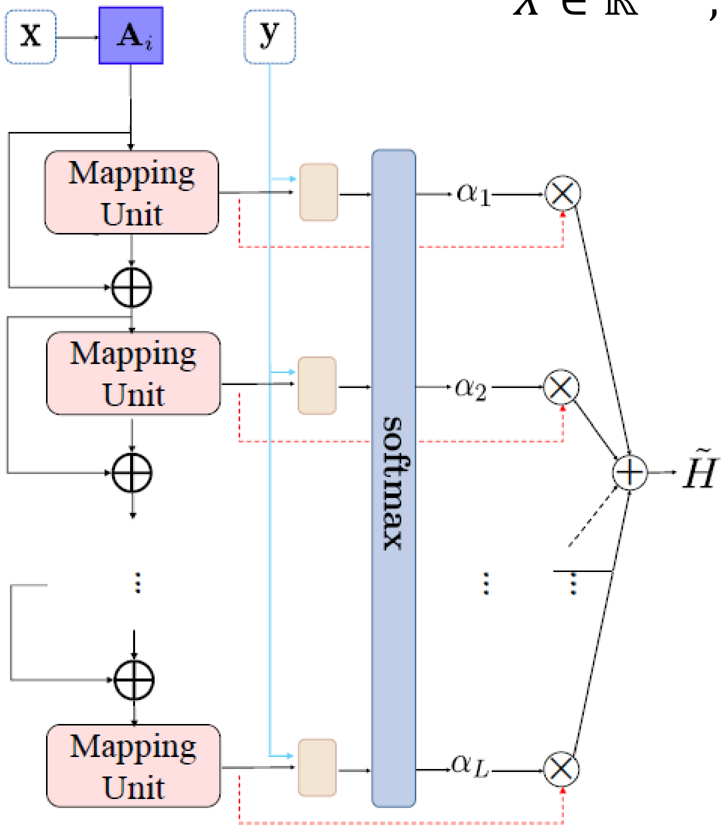
	English			Spanish			Italian		
	MAP	MRR	P@5	MAP	MRR	P@5	MAP	MRR	P@5
CRIM	19.78	36.10	19.03	–	–	–	–	–	–
NLP_HZ	9.37	17.29	9.19	20.04	28.27	20.39	11.37	19.19	11.23
300-sparsans	8.95	19.44	8.63	17.94	37.56	17.06	12.08	25.14	11.73
SPON	**20.20**	**36.95**	**19.40**	**32.64**	**50.48**	**32.76**	**17.88**	**29.80**	**17.95**

	Music		
	MAP	MRR	P@5
CRIM	40.97	60.93	41.31
SPON	**54.70**	**71.20**	**56.30**
	Medical		
	MAP	MRR	P@5
CRIM	**34.05**	**54.64**	**36.77**
SPON	33.50	50.60	35.10

- Compare to CRIM, except 1 task(Medical), other tasks are better than previous.

Others

- **Recurrent Mapping Model**



$X \in \mathbb{R}^{d\times 1}, A_i \in \mathbb{R}^{d\times d}$

$t = A_i X$

$h_l \in \mathbb{R}^{d\times 1}, l^{\text{th}}$ level

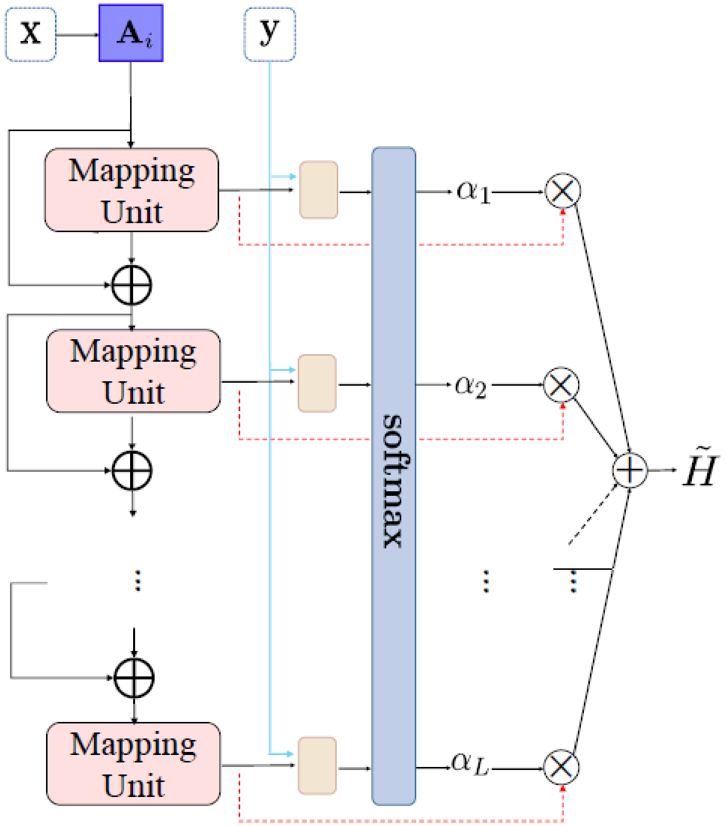
$\hat{h}_{l+1} = W_\phi h_l, \qquad W_\phi \in \mathbb{R}^{d\times d}, h_0 = t$

$h_{l+1} = \hat{h}_{l+1} + h_l$

- Assume hypernym terms may come from different concept levels.
- Comprising with recurrent structure and residual connections.
- W_ϕ is learnable projection matrix, and shared during each Mapping unit.

Others

- **Recurrent Mapping Model**



$$\alpha_l = \frac{\exp(h_l \cdot y)}{\sum_{k=1}^{L} \exp(h_k \cdot y)}$$

$$\hat{H} = \sum_{l=1}^{L} \alpha_l \cdot h_l$$

$$s_y = \hat{H} \cdot y$$

- Weight vector can get by calculating h_l and candidate hypernym.
- Attention weight of level is viewed as prob of lying in level.

Others

- **Recurrent Mapping Model**

Model	1A			2A			2B		
	MAP	MRR	P@5	MAP	MRR	P@5	MAP	MRR	P@5
MFH	8.77	21.39	7.81	28.93	35.80	34.20	33.32	51.48	35.76
vTE	10.60	23.83	9.91	18.84	41.07	20.71	12.99	39.36	12.41
Sparsans	8.95	19.44	8.63	17.94	37.56	17.06	12.08	25.14	11.73
NLP-HZ	9.37	17.29	9.19	20.04	28.27	20.39	11.37	19.19	11.23
Hybrid	15.97	34.07	15.00	37.85	**64.47**	**40.19**	54.62	**77.24**	55.08
CRIM	19.78	36.10	19.03	34.05	54.64	36.77	40.97	60.93	41.31
SPON	20.20	36.95	19.40	33.50	50.60	35.10	54.70	71.20	56.30
RMM	**27.12** ±0.12	**39.07** ±0.42	**23.41** ±0.22	**38.56** ±0.34	54.89 ±0.63	37.17 ±0.33	**63.86** ±0.06	74.75 ±0.52	**61.61** ±0.25

- Compare to others, except 3 metrics(Medical-MRR, P@5 and Music-MRR) ,others are better than previous.

References

1) Bernier-Colborne, G., & Barriere, C. (2018, June). Crim at semeval-2018 task 9: A hybrid approach to hypernym discovery. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 725-731).
2) Held, W., & Habash, N. (2019, July). The effectiveness of simple hybrid systems for hypernym discovery. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3362-3367).
3) Dash, S., Chowdhury, M. F. M., Gliozzo, A., Mihindukulasooriya, N., & Fauceglia, N. R. (2020, April). Hypernym detection using strict partial order networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 7626-7633).
4) Bai, Y., Zhang, R., Kong, F., Chen, J., & Mao, Y. Hypernym Discovery via a Recurrent Mapping Model.

Thank you