MUSE: Modularizing Unsupervised Sense Embeddings

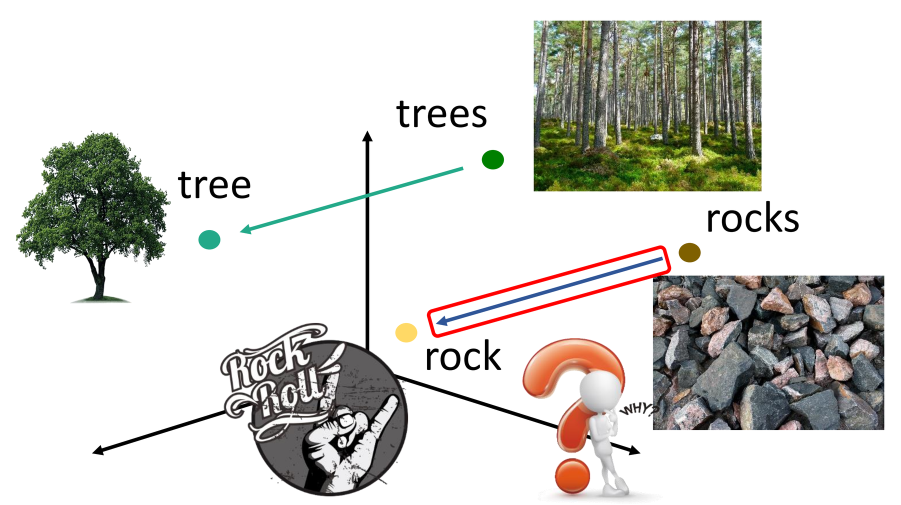
Massachusetts Institute of Technology · CSAIL · National Taiwan University · miu

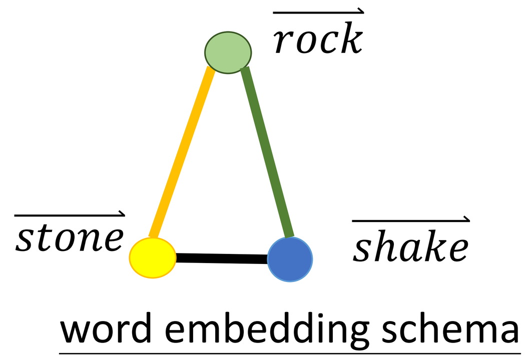
Guang-He Lee and Yun-Nung (Vivian) Chen

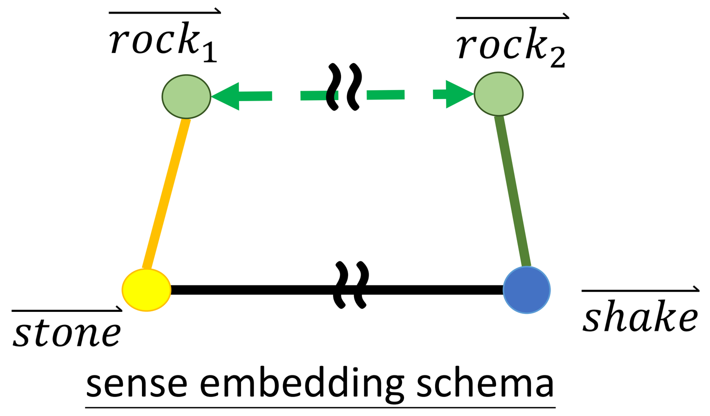
https://github.com/MiuLab/MUSE

Motivation

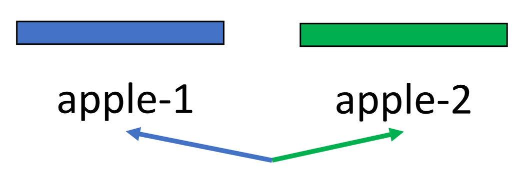
- Why sense embeddings
  - Words are ***polysemous***, but their embeddings are usually not ☹
  - The word embedding is restricted by the triangle inequality
  - Sense embeddings can easily circumvent such constraint







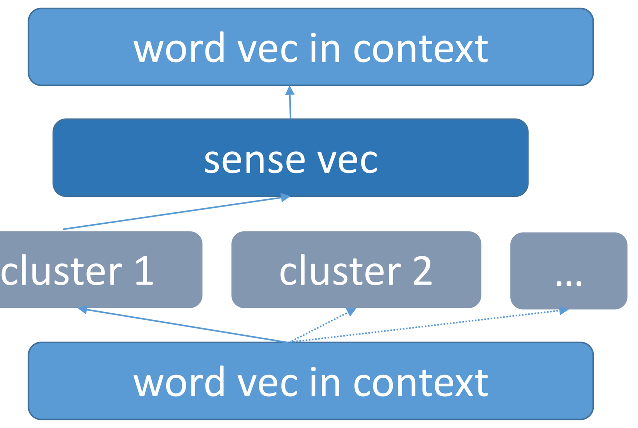
- Key mechanisms



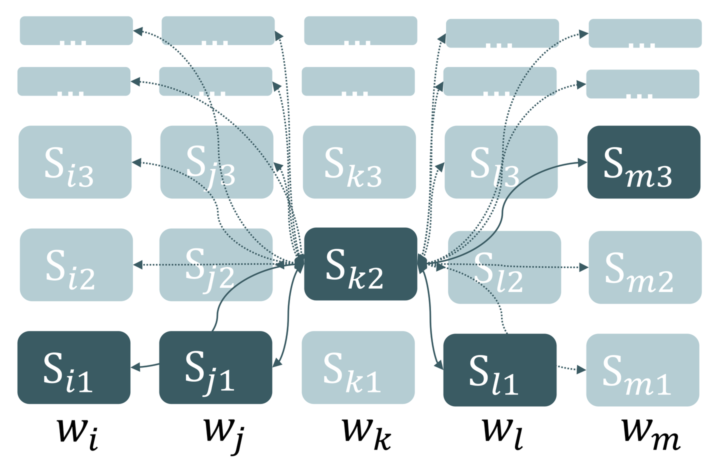
Smartphone companies including ***apple*** blackberry, and sony will be invited.

Traditional Frameworks

- Clustering based on contexts as sense ID
  - ◯ ***Efficient*** sense selection
  - ✗ Embedding learning involves word tokens

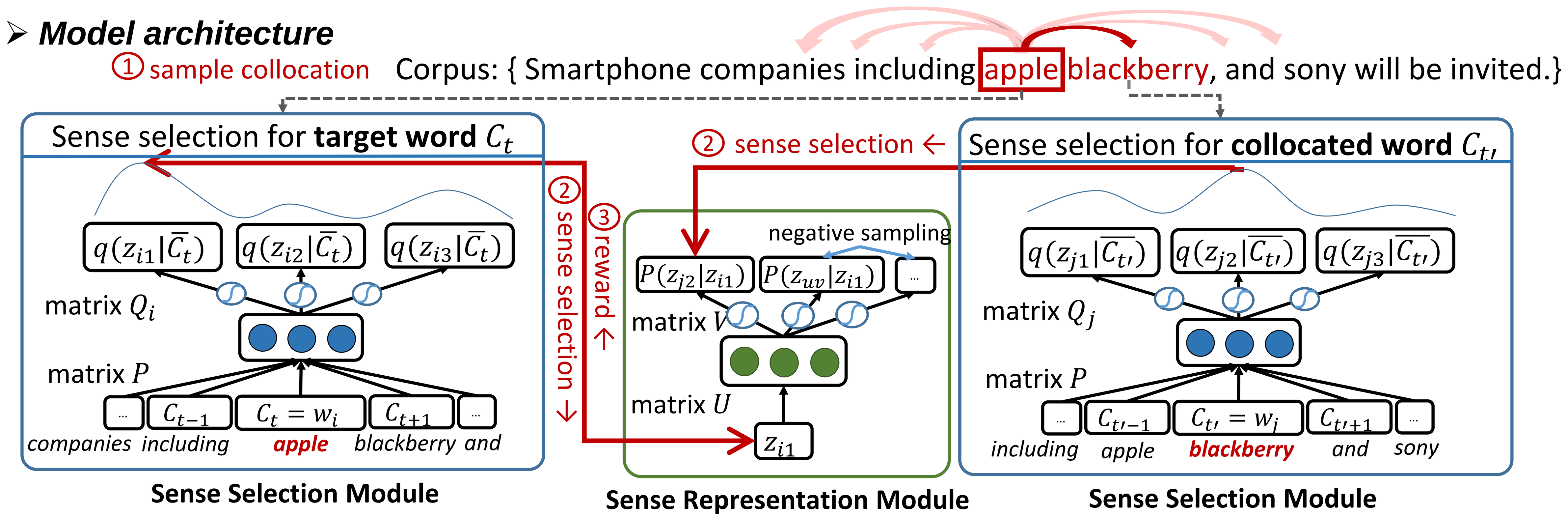


- Sense selection by a distribution
  - ◯ No word tokens are involved
  - ✗ ***Inefficient*** EM algorithms

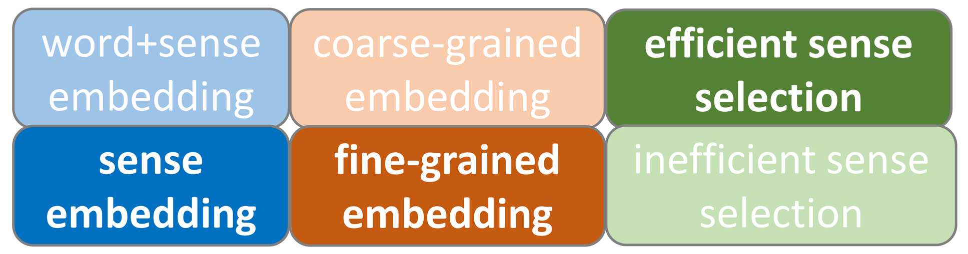


The Proposed Approach: MUSE

- ***Model architecture***



- Advantages from both worlds



- *Issue 1: impossible for a single model*
- Solution: modular framework
  - Distinct modules for specific mechanisms

- *Issue 2: how to formulate single objective?*
- Solution: Markov Decision Process (MDP)
  - sense selection → state/action in MDP
  - sense representation → reward in MDP

❖ **sense selection: linear neural network**

- Policy-based: $\pi(z_{ik} \mid \bar{C}_t) = \frac{\exp(Q_{ik}^T \sum_{j \in \bar{C}_t} P_j)}{\sum_{k' \in Z_i} \exp(Q_{ik'}^T \sum_{j \in \bar{C}_t} P_j)}$
- Value-based: $q(z_{ik} \mid \bar{C}_t) = \sigma(Q_{ik}^T \sum_{j \in \bar{C}_t} P_j)$

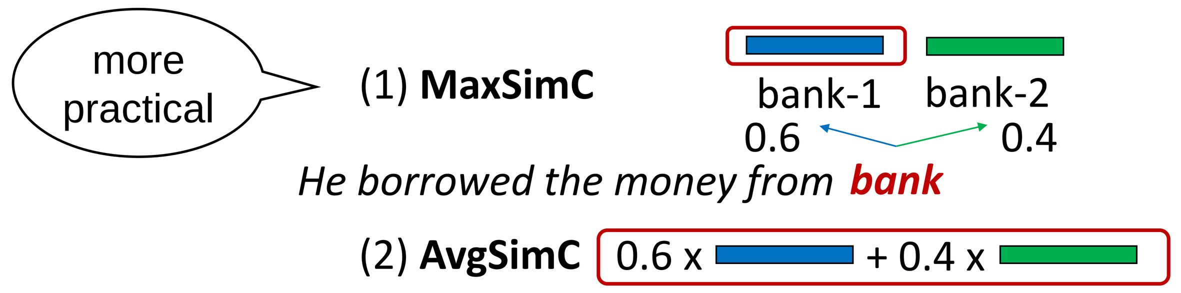
❖ **sense representation: skip-gram**

- $\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) = \log \sigma(U_{z_{ik}}^T V_{z_{jl}}) + \sum_{v=1}^{M} \mathbb{E}_{z_{uv} \sim p_{neg}(z)}[\log \sigma(-U_{z_{ik}}^T V_{z_{uv}})]$

- *Issue 3: how to optimize modules?*
- Solution: Reinforcement Learning (RL)
  - Policy-based: maximizes the ***expected rewards***
  - Value-based: ***estimates*** the ***rewards*** directly
- *Issue 4: how to conduct sense selection?*
- Solution: Exploration
  - Policy gradient: sampling
  - Value-based: greedy, ε-Greedy, Boltzmann sampling

Experiment 1: Contextual Word Similarities

- Setup
  - ***Dataset***: April 2010 Wikipedia dump
  - ***Context window***: 5
  - ***Embedding dim***: 300
- Evaluation: similarity on *contextual* word pairs
  - *... east **bank** of the Des Moines River ...*
  - *... basis of all **money** laundering ...*



- Compared target: human-judged similarity

Approach	MaxSimC	AvgSimC	
Huang et al., 2012	26.1	65.7	clustering
Neelakantan et al., 2014	60.1	<u>69.3</u>	clustering
Tian et al., 2014	63.6	65.4	probabilistic
Li & Jurafsky, 2015	<u>66.6</u>	66.8	probabilistic
Bartunov et al., 2016	53.8	61.2	probabilistic
Qiu et al., 2016	64.9	66.1	probabilistic
MUSE-Policy	66.1	67.4	
MUSE-Greedy	66.3	68.3	
MUSE-ε-Greedy	67.4[+]	68.6	
MUSE-Boltzmann	**67.9**[+]	68.7	

➢ MUSE achieves the state-of-the-art on MaxSimC

Context	k-NN Senses
braves finish the season in **tie** with the los angeles dodgers	scoreless otl shootout 6-6 hingis 3-3 7-7 0-0
his later years proudly wore **tie** with the chinese characters for	pants trousers shirt juventus blazer socks anfield
of the mulberry or the **blackberry** and minos sent him to	cranberries maple vaccinium apricot apple
of the large number of **blackberry** users in the us federal	smartphones sap microsoft ipv6 smartphone
shells and/or high explosive squash **head** hesh and/or anti-tank	venter thorax neck spear millimeters fusiform
head was shaven to prevent **head** lice serious threat back then	shaved thatcher loki thorax mao luthor chest
appoint john pope republican as **head** of the new army of	multi-party appoints unicameral beria appointed

➢ MUSE can effectively separate different senses in an unsupervised way.

Experiment 2: Synonym Selection

- Task: Q: yield (A) submit (B) challenge (C) boast (D) scorn

	Approach	ESL-50	RD-300	TOEFL-80
word embeddings	Global Context	47.73	45.07	60.87
word embeddings	SkipGram	52.08	55.66	66.67
word sense disambiguation	IMS+SkipGram	41.67	53.77	66.67
unsupervised sense embeddings	EM	27.08	33.96	40.00
unsupervised sense embeddings	MSSG (Neelakantan et al., '14)	<u>57.14</u>	<u>58.93</u>	78.26
unsupervised sense embeddings	CRP (Li & Jurafsky, '15)	50.00	55.36	<u>82.61</u>
unsupervised sense embeddings	MUSE-Policy	52.38	51.79	79.71
unsupervised sense embeddings	MUSE-Greedy	57.14	58.93	79.71
unsupervised sense embeddings	MUSE-ε-Greedy	61.90[+]	62.50[+]	84.06[+]
unsupervised sense embeddings	MUSE-Boltzmann	**64.29**[+]	66.07[+]	**88.41**[+]
supervised sense embeddings	Retro-GlobalContext	63.64	66.20	71.01
supervised sense embeddings	Retro-SkipGram	56.25	65.09	73.33

➢ MUSE with exploration outperforms all baselines.

➢ MUSE beat some supervised systems w/o any supervision.

Conclusion

- **Efficiency**: *purely sense-level* representation learning with *linear-time* sense decoding
- **Modeling**: single objective for modular unsupervised sense embedding learning
- **Learning**: leverage *RL* to model the sense selection process
- **Exploration**: introduce various exploration mechanisms for the sense selection for *robustness*
- **Experiment**: *state-of-the-art* performance

guanghe@mit.edu
guanghelee.github.io

y.v.chen@ieee.org
vivianchen.idv.tw

Code Available: http://github.com/MiuLab/MUSE