การหา Word Embedding
ด้วย word2vec

	ordered	throne	he	she	killed	poor	…
king	30	20	5	8	10	3	
queen	25	15	3	12	3	2	
slave	5	3	8	6	40	25	
woman	10	5	4	9	5	10	
…							

word embedding

	King	Queen	Slave	Woman
ความรัง	0.90	0.90	0.01	0.20
ความชาย	0.99	0.02	0.50	0.02
อำนาจ	0.8	0.8	0.10	0.40
	⋮	⋮	⋮	⋮

word2vec เป็น algorithm หนึ่งที่ใช้เปลี่ยน word-context matrix ให้เป็น word embedding

word embedding เก็บลักษณะทาง semantic และ syntactic ของแต่ละคำ

= word representation

ใน ฤดูหนาว วิกฤต ฝุ่น จะ ร้ายแรง ขึ้น

context

context

window size = 2

text classification problem

training data :

input	label
วิกฤต	ฝุ่น
วิกฤต	ฤดูหนาว
⋮	⋮

ใน ฤดูหนาว วิกฤต ฝุ่น จะ ร้ายแรง ขึ้น

context context

window size = 2

ใช้ logistic regression

P(ฤดูหนาว | วิกฤต)

P(ฝุ่น | วิกฤต)

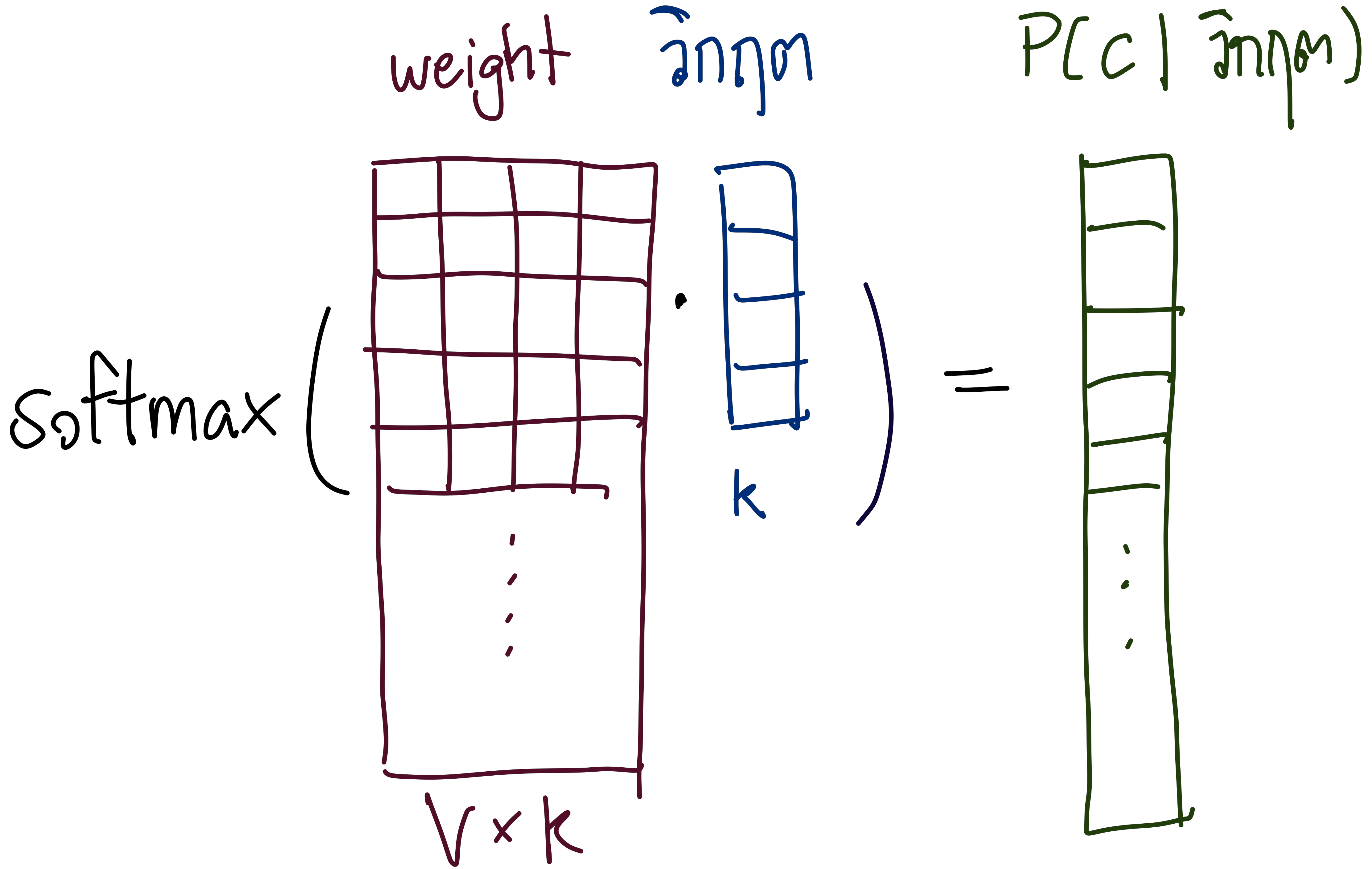
ใน ฤดูหนาว วิกฤต ฝุ่น จะ ร้ายแรง ขึ้น

context

context

window size = 2

ใน ฤดูหนาว วิกฤต $\boxed{\text{ฝุ่น}}$ จะ ร้ายแรง ขึ้น


weight
วิกฤต
P(c | วิกฤต)
softmax
k
V × k
=

$$P(\text{ฝุ่น} \mid \text{วิกฤต}) = \frac{\exp\left(u_{\text{ฝุ่น}} \cdot w_{\text{วิกฤต}}\right)}{\sum_{\substack{\text{คำอื่นๆ} \\ v'}} \exp\left(v_{v'} \cdot w_{\text{วิกฤต}}\right)}$$

dot product

$\approx$ similarity

$$= \frac{\exp\left(\sum_k v_{\text{ฝุ่น},k} \cdot \text{วิกฤต}_{\text{วิกฤต},k}\right)}{\sum_{\substack{\text{คำอื่นๆ} \\ v'}} \exp\left(\sum_k v_{v',k} \cdot \text{วิกฤต}_{\text{วิกฤต},k}\right)}$$

Objective function

$$J(\theta) = \sum_{w \text{ in data}} \left(\sum_{c_{left}} -\log P(c_{left} \mid w) + \sum_{c_{right}} -\log P(c_{right} \mid w) \right)$$

- word2vec (Skipgram model) สามารถเปลี่ยน word-context matrix เป็น word embedding ได้อย่างมีประสิทธิภาพ

- Word embedding พวกนี้เก็บลักษณะเฉพาะทาง semantic และ syntactic ไว้สำหรับ โจทย์อื่นๆ ที่ต้องใช้ความเข้าใจของคำ

- Word embedding เป็นพื้นฐานของ NLP + Deep learning เกือบทั้งหมดในตอนนี้

ประเมินประสิทธิภาพของ Word Embedding

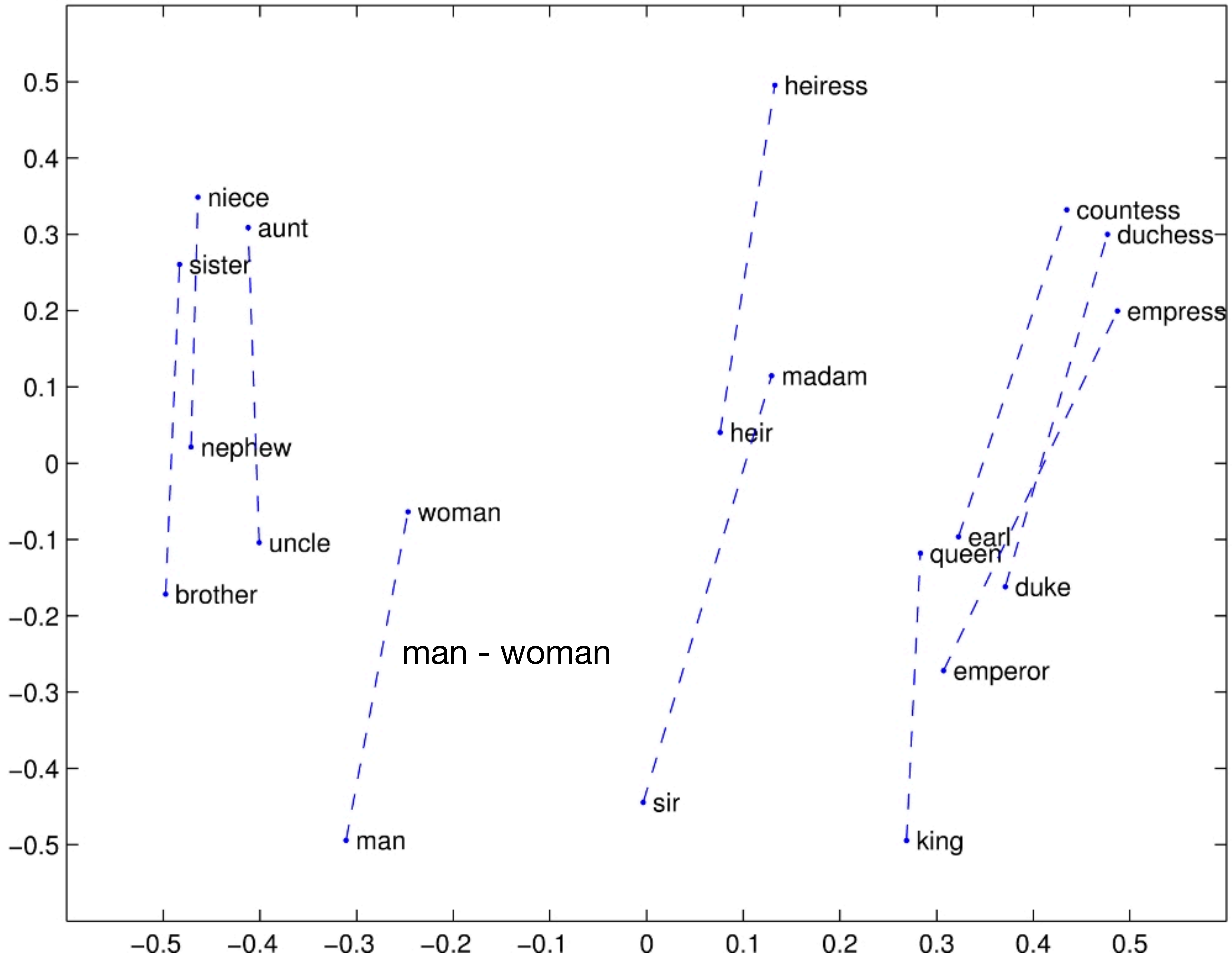
Evaluate Word Embeddings

- การประเมินผลเฉพาะตัว (Intrinsic evaluation)
  - Semantic analogy test
  - Syntactic/morphological analogy test
  - Word similarity test
- การประเมินผลจากโจทย์อื่นๆ (Extrinsic evaluation)
  - นำไปใช้เป็น feature ของ text classification

Semantic Analogy Test

- Bangkok:Thailand = Paris:France
- Mexico:peso = Korea:won
- uncle:aunt = king:queen
- boy:girl = grandpa:grandma

Mikolov et al., (2013) Distributed Representations of Words and Phrases and their Compositionality



https://nlp.stanford.edu/projects/glove/

Semantic Analogy Test

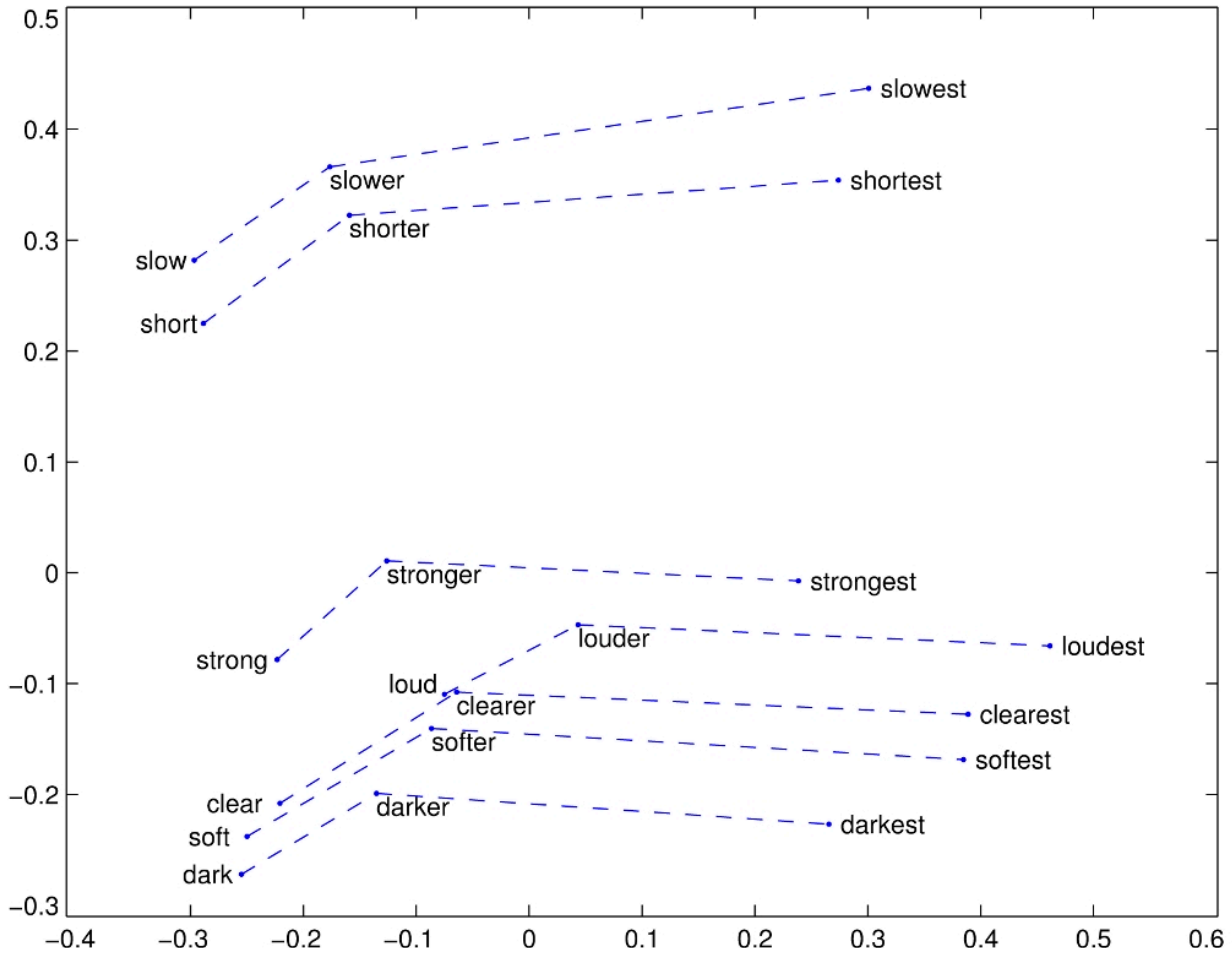
- Bangkok:Thailand = Paris:France
  - Bangkok - Thailand = Paris - France
- Mexico:peso = Korea:won
- uncle:aunt = king:queen
- boy:girl = grandpa:grandma

Mikolov et al., (2013) Distributed Representations of Words and Phrases and their Compositionality

Morphological Analogy Test

- 9 types of English morphology e.g.
  - amazing:amazingly = possible:possibly
  - clear:unclear = known:unknown
  - bad:worse = big:bigger
  - dancing:danced = sleeping:slept

Mikolov et al., (2013) Distributed Representations of Words and Phrases and their Compositionality



https://nlp.stanford.edu/projects/glove/

Word Similarity Test

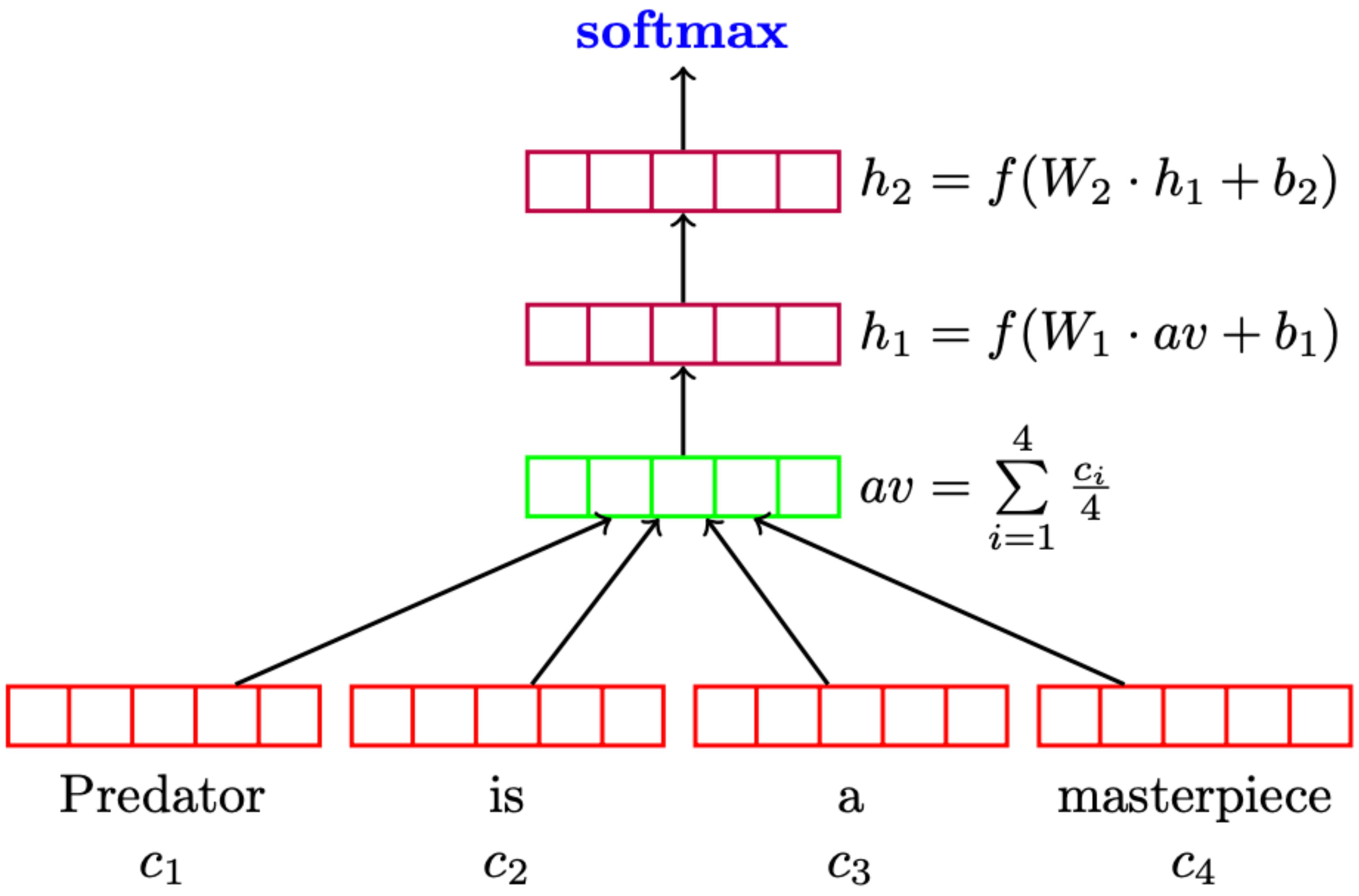
Word_1	Word_2	Similarity score [0,10]
love	sex	6.77
stock	jaguar	0.92
money	cash	9.15
development	issue	3.97
lad	brother	4.46

$$\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2} \sqrt{\sum\limits_{i=1}^{n} B_i^2}},$$

http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/

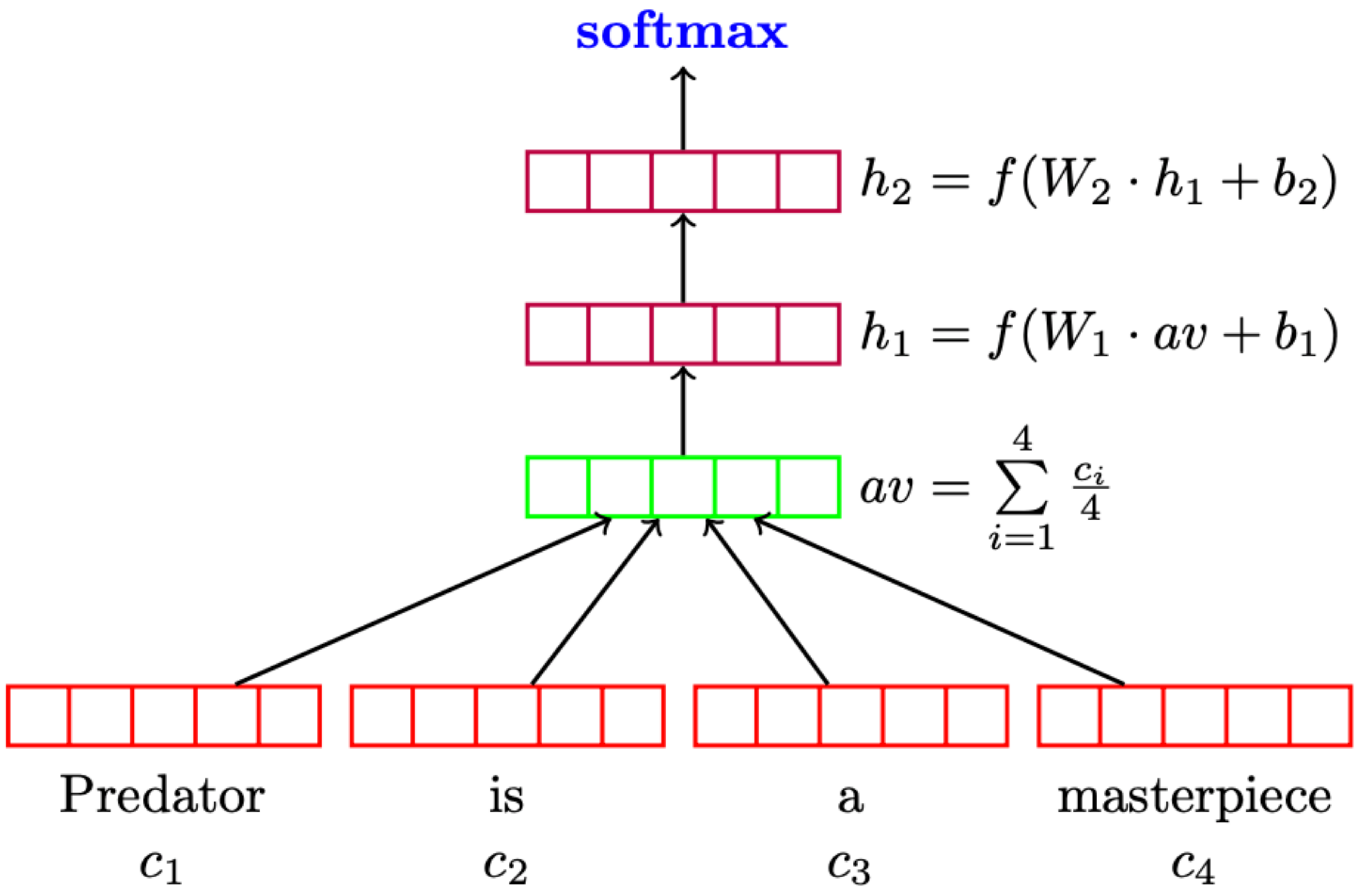
Neural Bag-of-Word Model

Neural Bag-of-Word model



Iyyer et al., (2015) Deep Unordered Composition Rivals Syntactic Methods for Text Classification

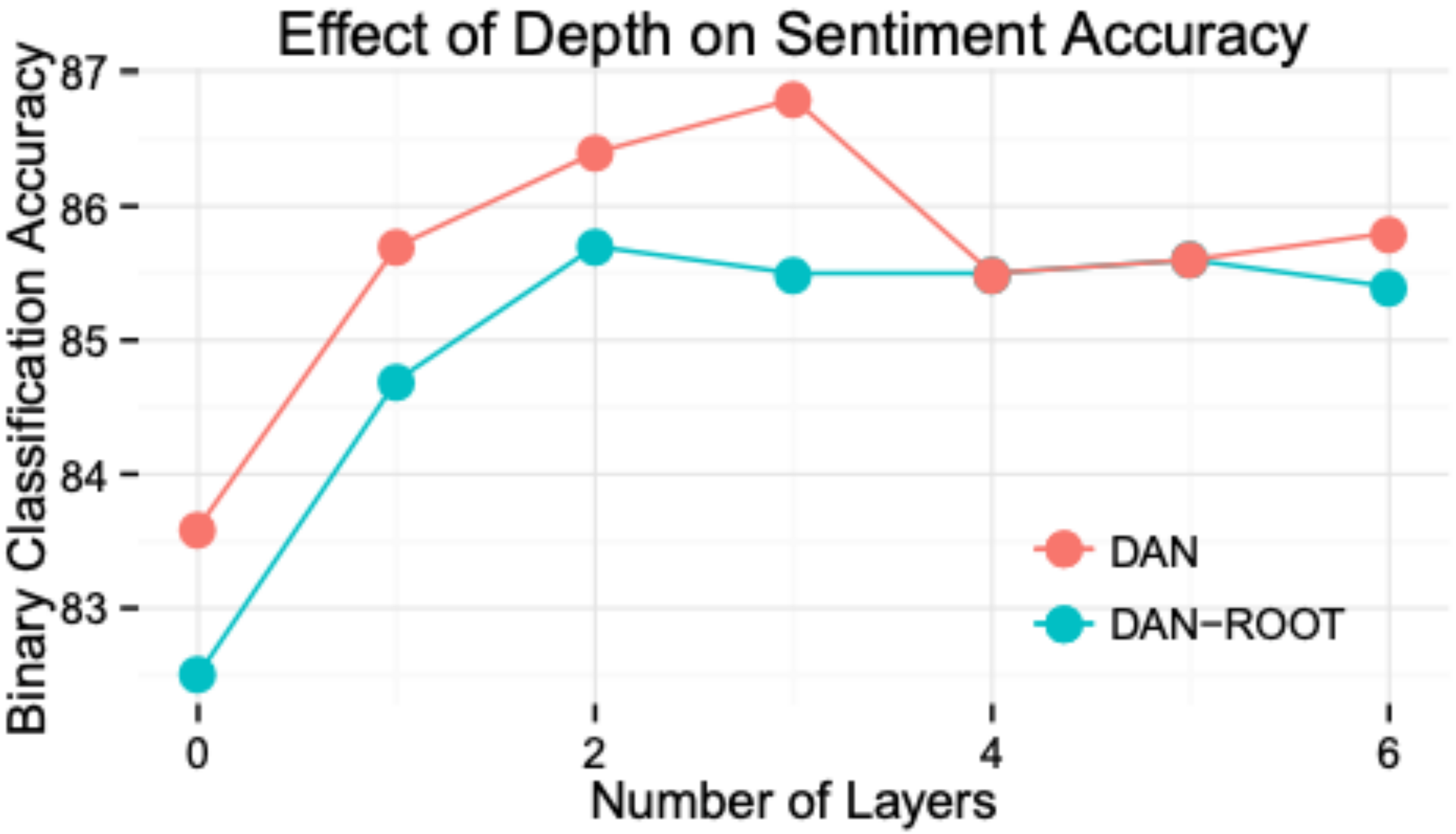
Neural Bag-of-Word model



Model	RT	SST fine	SST bin	IMDB
DAN-ROOT	—	46.9	85.7	—
DAN-RAND	77.3	45.4	83.2	88.8
DAN	80.3	47.7	86.3	89.4
NBOW-RAND	76.2	42.3	81.4	88.9
NBOW	79.0	43.6	83.6	89.0
BiNB	—	41.9	83.1	—
NBSVM-bi	79.4	—	—	91.2
RecNN*	77.7	43.2	82.4	—
RecNTN*	—	45.7	85.4	—
DRecNN	—	49.8	86.6	—
TreeLSTM	—	**50.6**	86.9	—
DCNN*	—	48.5	86.9	89.4
PVEC*	—	48.7	87.8	**92.6**
CNN-MC	**81.1**	47.4	**88.1**	—
WRRBM*	—	—	—	89.2

Iyyer et al., (2015) Deep Unordered Composition Rivals Syntactic Methods for Text Classification

ต้องลึกแค่ไหน

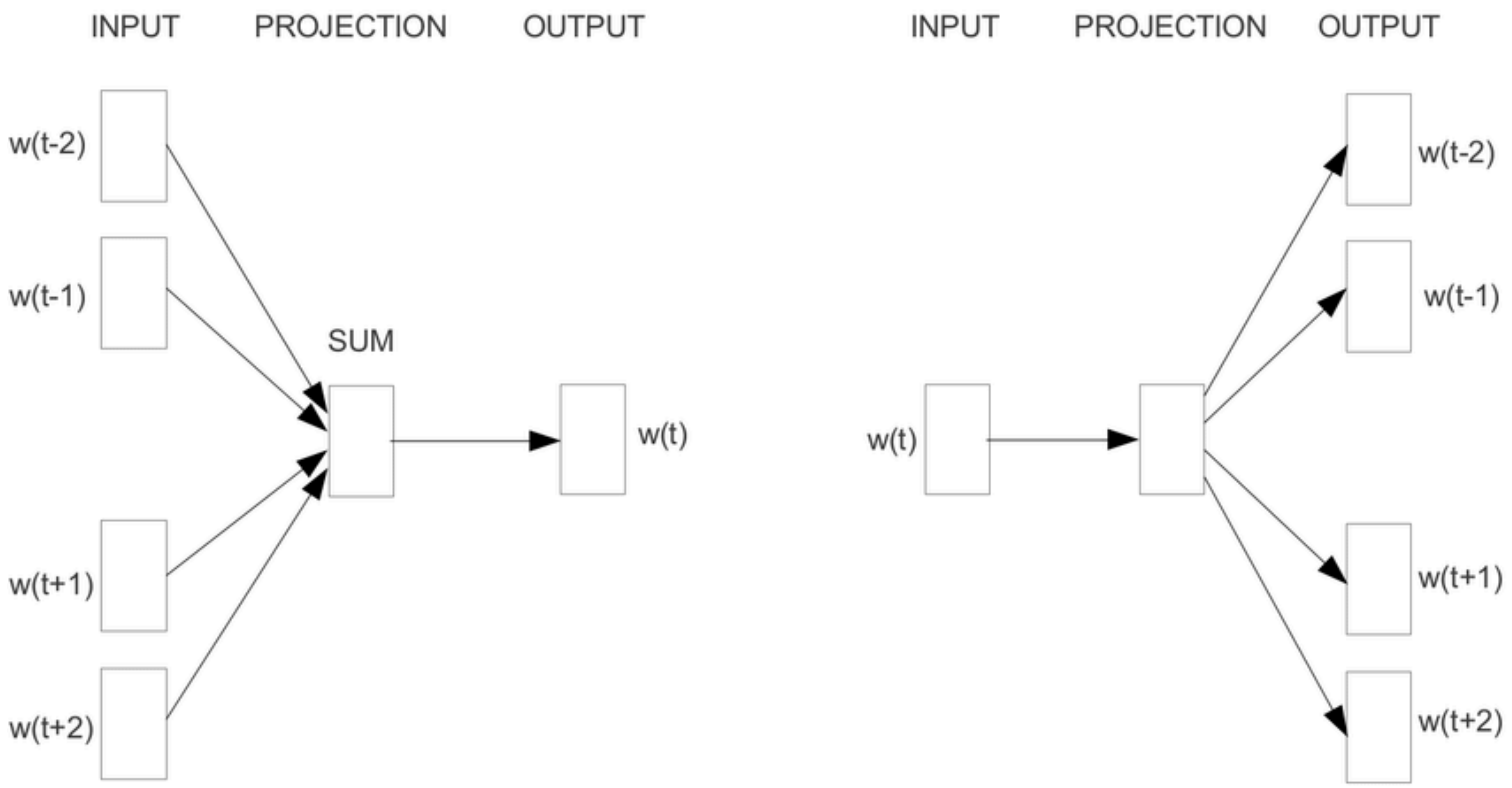


Effect of Depth on Sentiment Accuracy
Binary Classification Accuracy
87
86
85
84
83
0
2
4
6
Number of Layers
DAN
DAN−ROOT

Sentence	DAN	DRecNN	Ground Truth
a lousy movie that's not merely unwatchable, but also unlistenable	negative	negative	negative
if you're not a prepubescent girl, you'll be laughing at britney spears' movie-starring debut whenever it does n't have you impatiently squinting at your watch	negative	negative	negative
blessed with immense physical prowess he may well be, but ahola is simply not an actor	positive	neutral	negative
who knows what exactly godard is on about in this film, but his words and images do n't have to add up to mesmerize you.	positive	positive	positive
it's so good that its relentless, polished wit can withstand not only inept school productions, but even oliver parker's movie adaptation	negative	positive	positive
too bad, but thanks to some lovely comedic moments and several fine performances, it's not a total loss	negative	negative	positive
this movie was not good	negative	negative	negative
this movie was good	positive	positive	positive
this movie was bad	negative	negative	negative
the movie was not bad	negative	negative	positive

Word Embedding
จากโมเดลอื่น ๆ

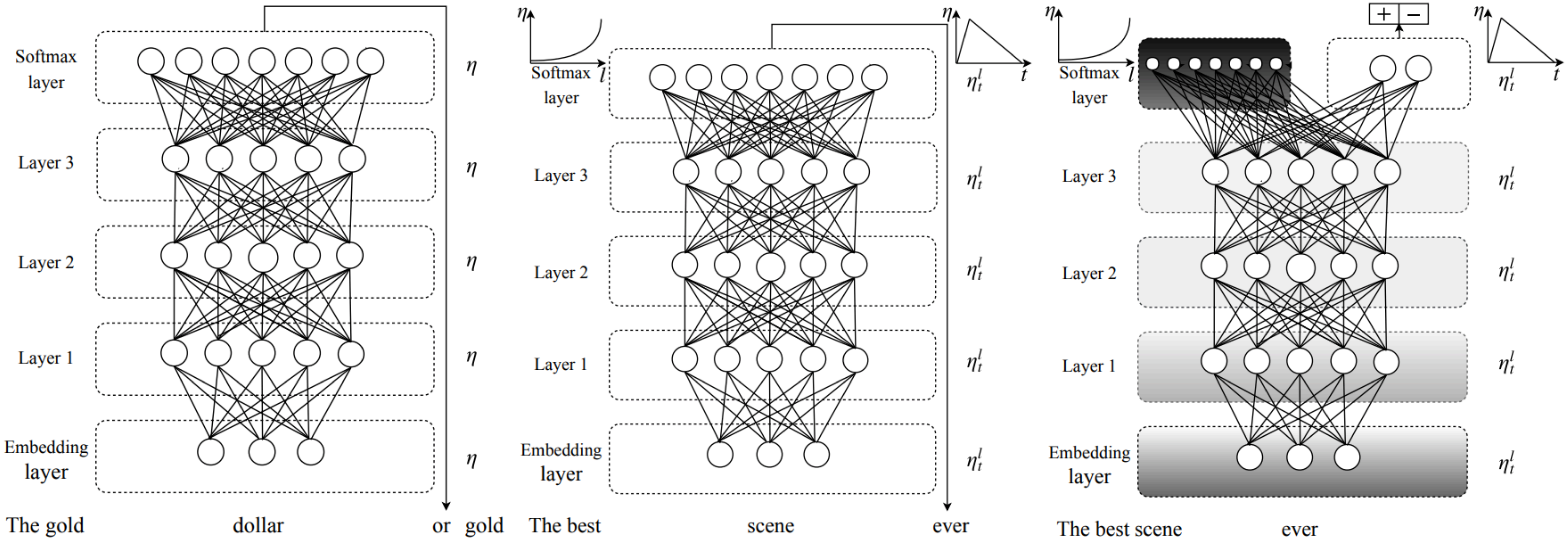
Continuous Bag-of-Word (CBOW)



GloVe

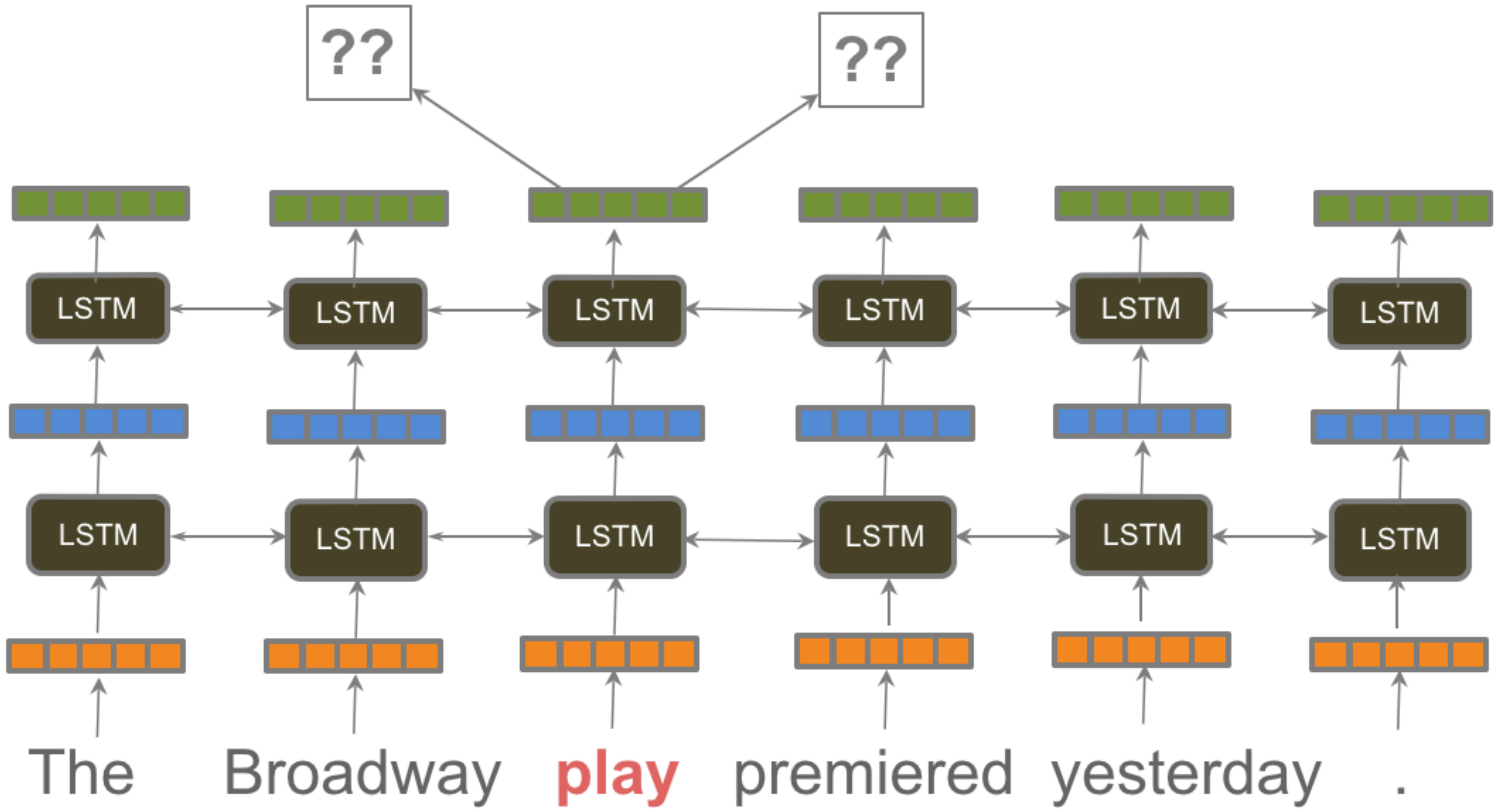
$$J(\theta) = \frac{1}{2} \sum_{i,j=1}^{W} f(P_{ij})(u_i^T v_j - \log P_{ij})^2$$

ULMfit



(a) LM pre-training (b) LM fine-tuning (c) Classifier fine-tuning

ElMo


??
??
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
The
Broadway
play
premiered
yesterday
.

ElMo

	Source	Nearest Neighbors
GloVe	play	playing, game, games, played, players, plays, player, Play, football, multiplayer
biLM	Chico Ruiz made a spectacular play on Alusik 's grounder {. . . }	Kieffer , the only junior in the group , was commended for his ability to hit in the clutch , as well as his all-round excellent play .
	Olivia De Havilland signed to do a Broadway play for Garson {. . . }	{. . . } they were actors who had been handed fat roles in a successful play , and had talent enough to fill the roles competently , with nice understatement .