KerBS: Kernelized Bayesian Softmax for Text Generation

Ning Miao, Hao Zhou, Chengqi Zhao, Wenxian Shi, Lei Li

Motivation

- Softmax based generation models assume context embeddings of a word should concentrate around its word embedding as in Figure 1.a.
- However, embeddings of some words actually exhibit properties such as **multi-sense** (Figure 1.b) or **varying variance** (Figure 1.c).
- We need to explicitly model the properties for more accurate word prediction!

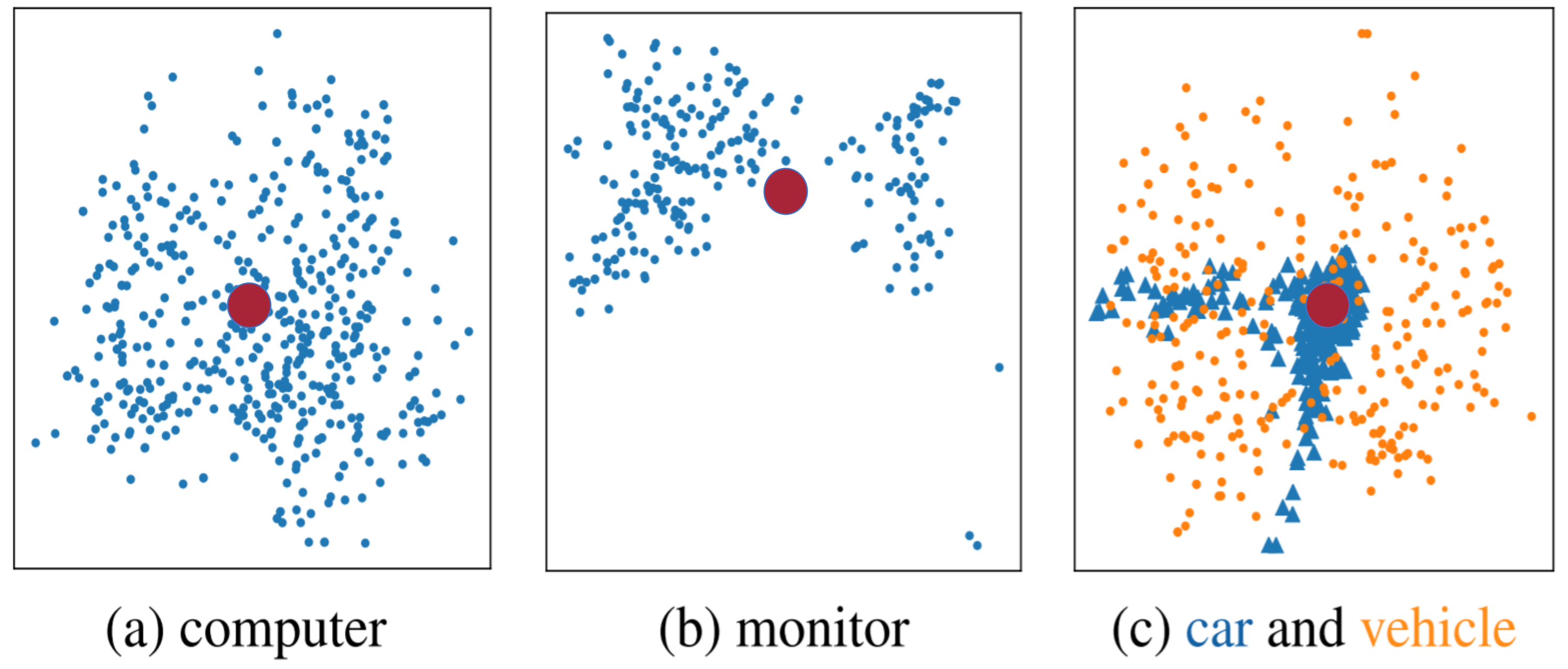


Figure 1: Context embeddings and word embeddings (●) of different word.

ByteDance AI Lab
字节跳动人工智能实验室

Method

- Multi-sense
  - We replace word embeddings with **sense embeddings**, and sums all sense probabilities of a word to get the probability of the word.

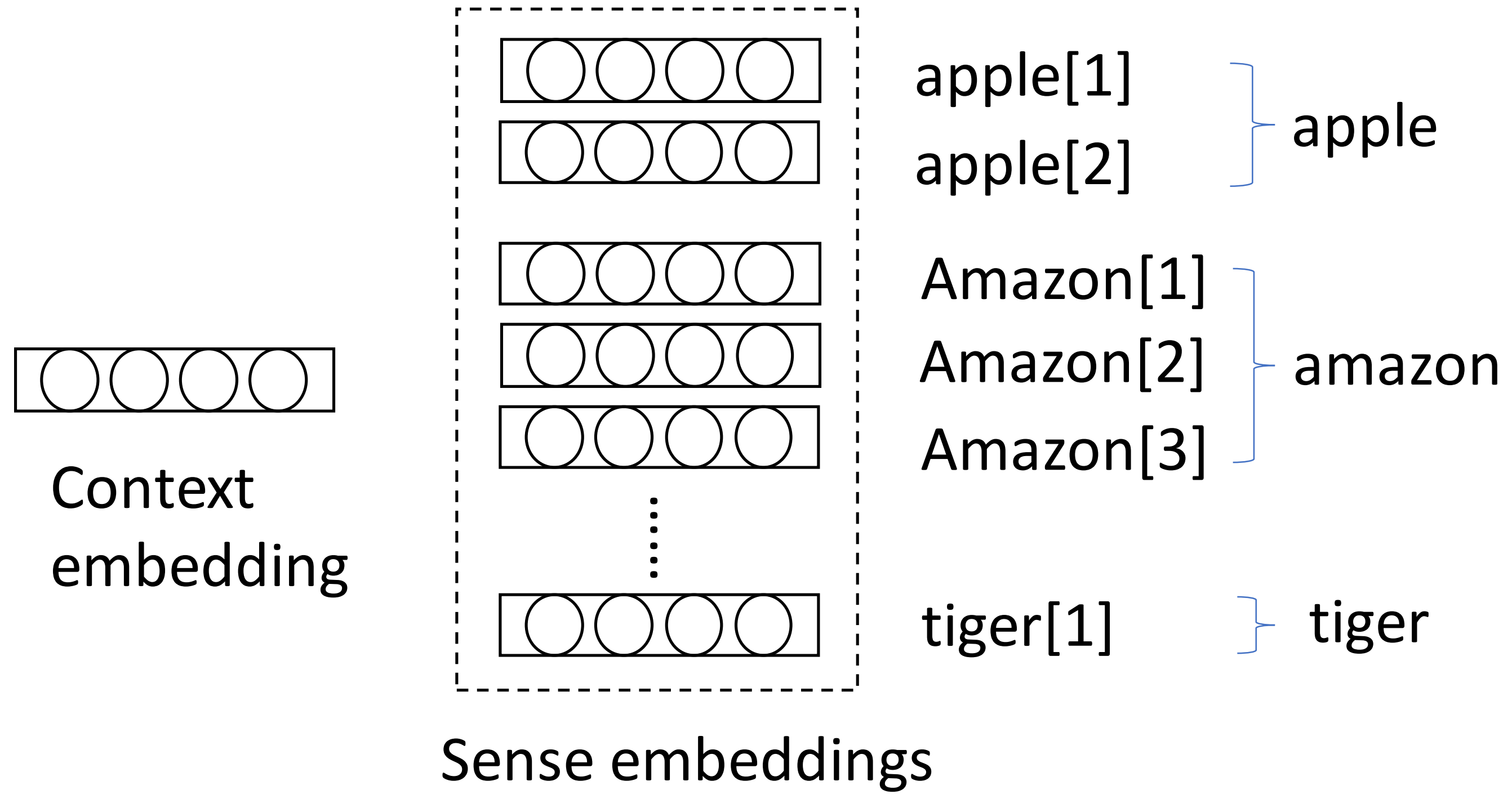


Figure 2: Illustration of sense embedding.

- Kernelized Variances
  - Since Gaussian distribution is instable in high-dimensional space, we designed a **kernel function** to model **different variances** of each word's context embeddings.

$$\mathcal{K}_{\theta}(h,e) = |h||e|(a\exp(-\theta\cos(h,e)) - a)$$

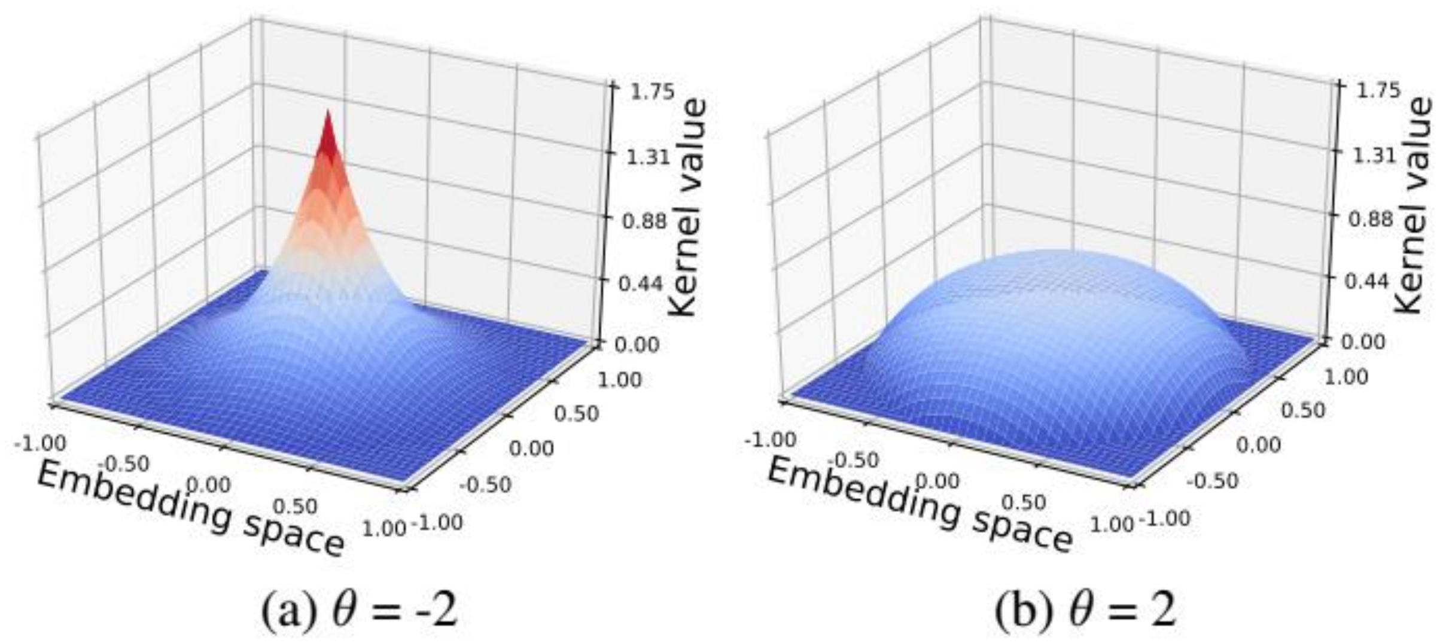


Figure 2: Kernels with different θ.

- Dynamic Sense Allocation
  - During training, we record the **usage** of each sense and **log prediction accuracy** of each word.
  - After every M steps, we **reallocate** the least used senses to the poorly predicted words.

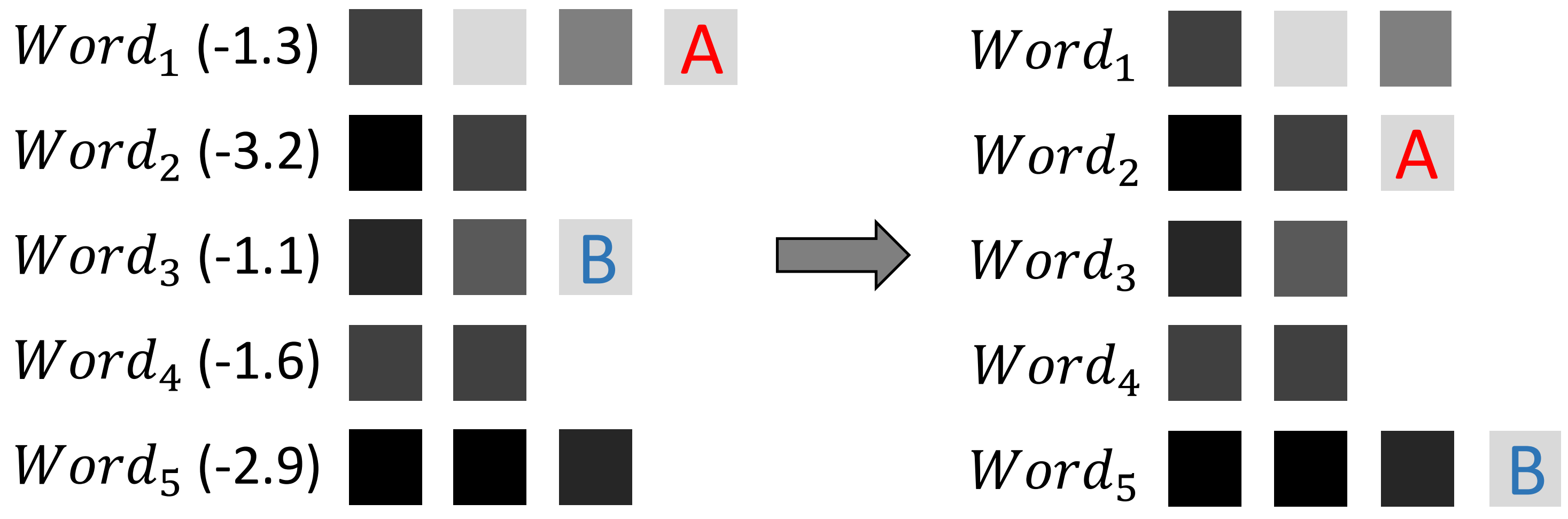


Figure 3: Illustration of sense reallocation. Numbers in the brackets are log prediction accuracies of each word. We use squares to represent senses and darker color means higher usage.

Experiments and Analyses

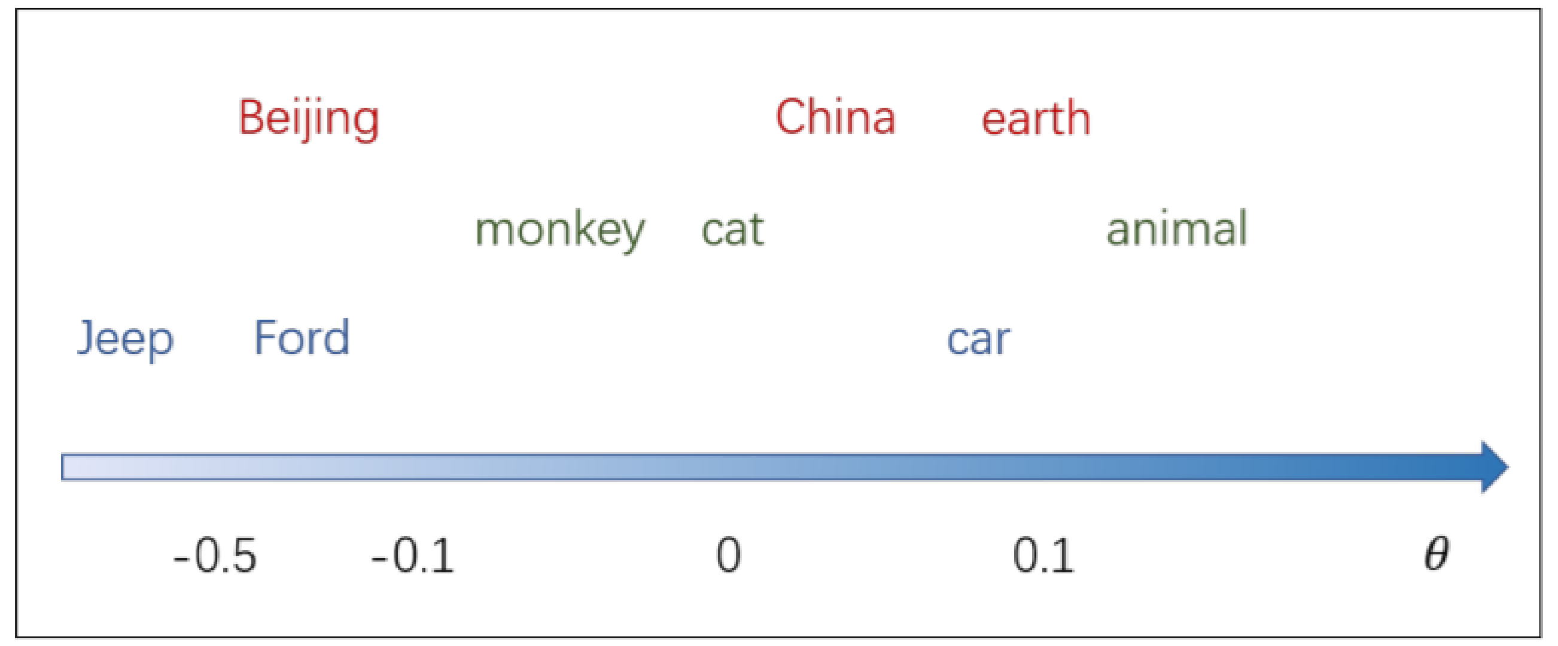
- Better performance than baselines.

Tasks	Base models	Metrics	Base	Base+Mos	Bas+KerBS
MT	Transformer	BLEU-4	29.61	28.54	**30,90**
	Seq2Seq	BLEU-4	25.91	26.45	**27.28**
LM	GRU	PPL	103.12	102.72	**102.17**
Dialog	Transformer	BLEU-1	10.61	9.81	**10.90**
	Seq2Seq	BLEU-1	16.56	13.73	**17.85**
		Human	1.24	1.04	**1.40**

- More sense are allocated to words with complex meanings .

Sense	1	2	3	4
word	Redwood heal structural theoretical rotate	particular figure during known size	open order amazing sound base	they work body power change

- θ reflects the sematic scopes of words.


Beijing
China
earth
monkey
cat
animal
Jeep
Ford
car
θ