

Learning to Embed Words in Context for Syntactic Tasks

Lifu Tu; Kevin Gimpel; Karen Livescu
Toyota Technological Institute at Chicago

How To Capture Linguistic Characteristics Of Tokens?

1. Same syntactic category, different senses:
He robbed 9 **banks**. vs. It washed up on the **banks**.
2. Different POS tag and sense:
I was unable to **police** the situation. vs. I was unable to contact the **police**.
- ...

How to solve this? Each word type can have a different vector representation in different contexts!

Models and Loss Function

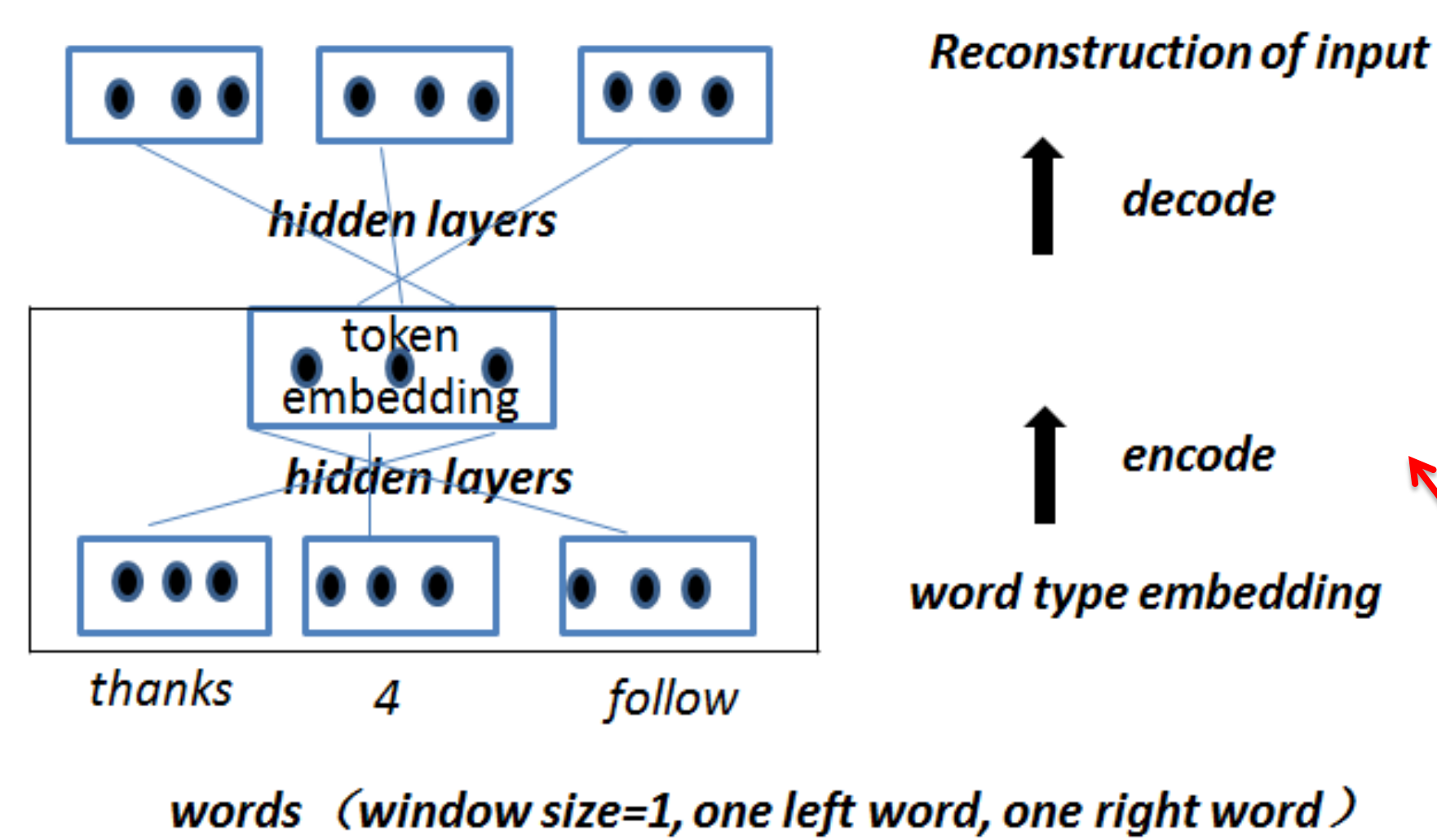


Figure 1. DNN token embedding model

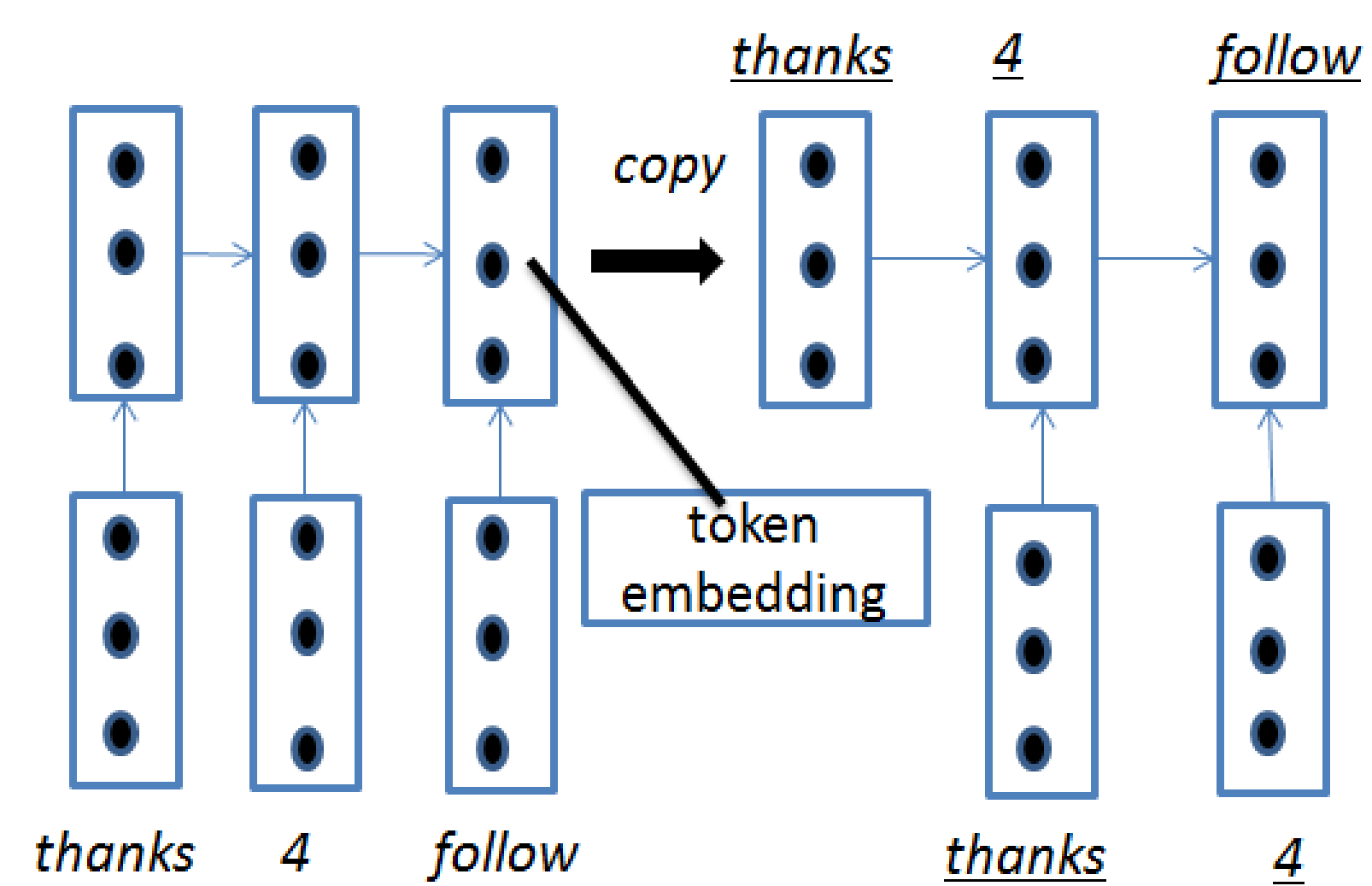
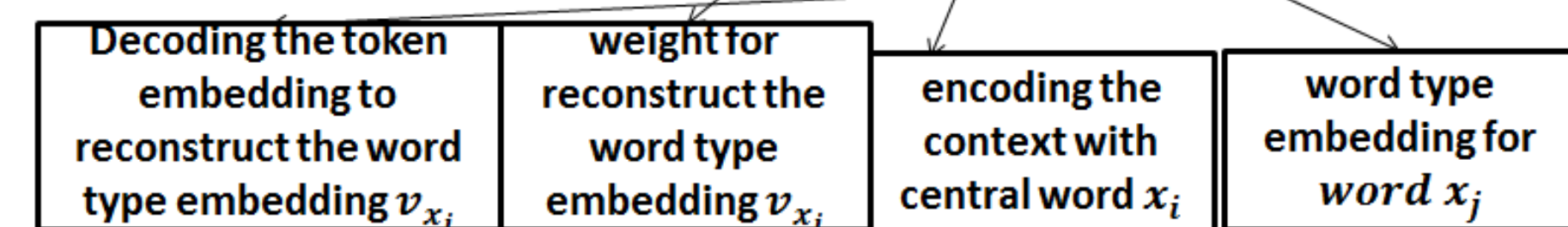


Figure 2. seq2seq token embedding model

Loss Function

- Weighted Reconstruction Error:
- Input word sequence x , encoder f , decoder g

$$Loss(f, g, x, i) = \sum_{j=i-w}^{i+w} a_j \|g(f(x, i))_j - v_{x_j}\|^2$$



Contact

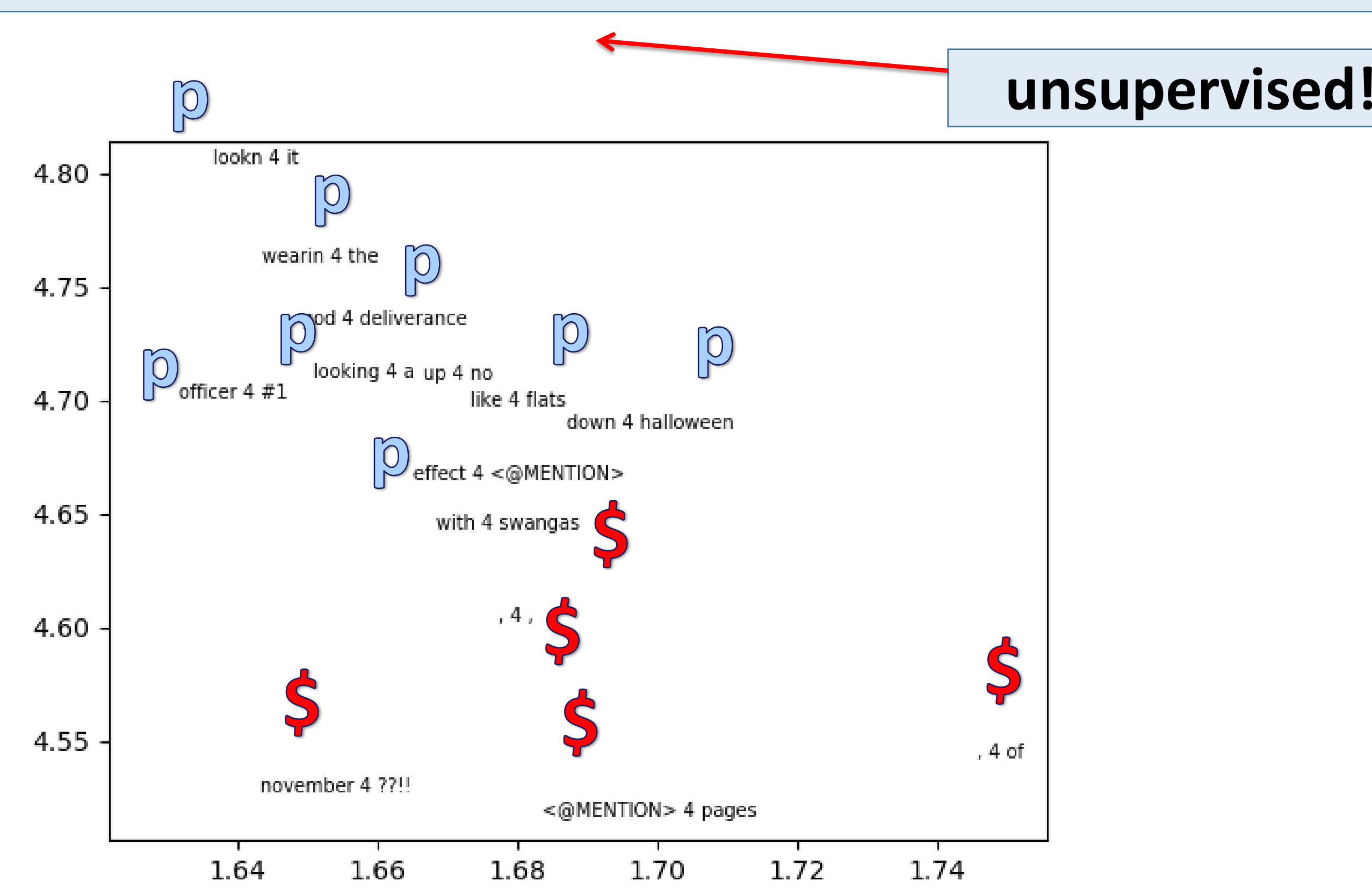
<Lifu Tu>
<Toyota Technological Institute at Chicago>
Email: lifu@ttic.edu

Qualitative Analysis

q	masters swimmers annual swim 4 your heart !	q	jus listenin 2 mr hudson and drake crazyness
1	so many miles loking for her and handing 1 way lol .	1	@mention deaddddd u go 2 mlk high up n bk
2	off to the rehearsal space for a weekend long	2	only a cups tho tryin 2 feed the whole family
3	on the inauguration for your enjoyment	3	are ya'll listening to the annointed one ?
4	#canucks now have a 4 point lead on the	4	@mention well could u come to mrs wilsons for
1	way lol . it's the 1 mile trail and then you off to the	5	i'm bored on mars i kum down 2 earth ... yupp !!
2	my first one was like 2 minutes long and	6	i am listening to amar prtihibi - black
3	my fav place- was there 2 years ago and	7	about neopets and listening to yelle (URL)
		8	high ritee now - ___ - bout too troop to the crib

Table 1. Nearest neighbours for token embeddings, where we consider neighbors that may have different word types from that in the query token

Visualization



- t-SNE visualization of token embeddings for word type "4".
- Each point shows the left and right context words (w=1).
- The tag "P" is preposition and "\$" is number.

Token Embedding Models

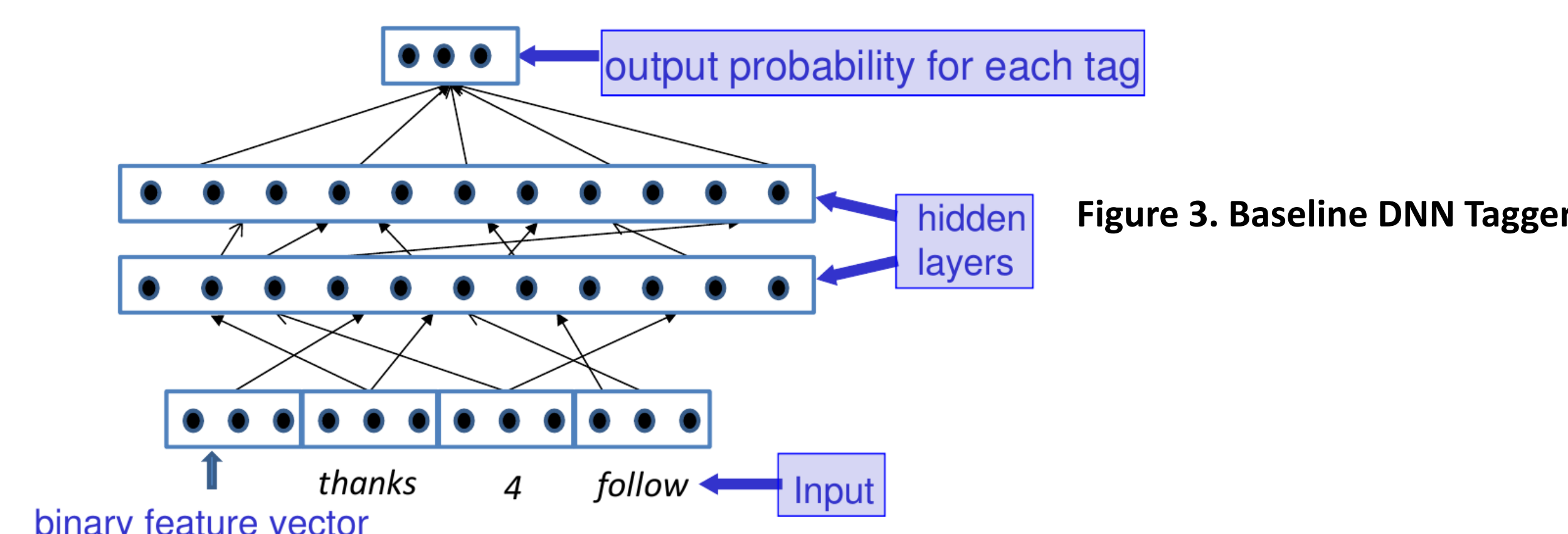


Figure 3. Baseline DNN Tagger

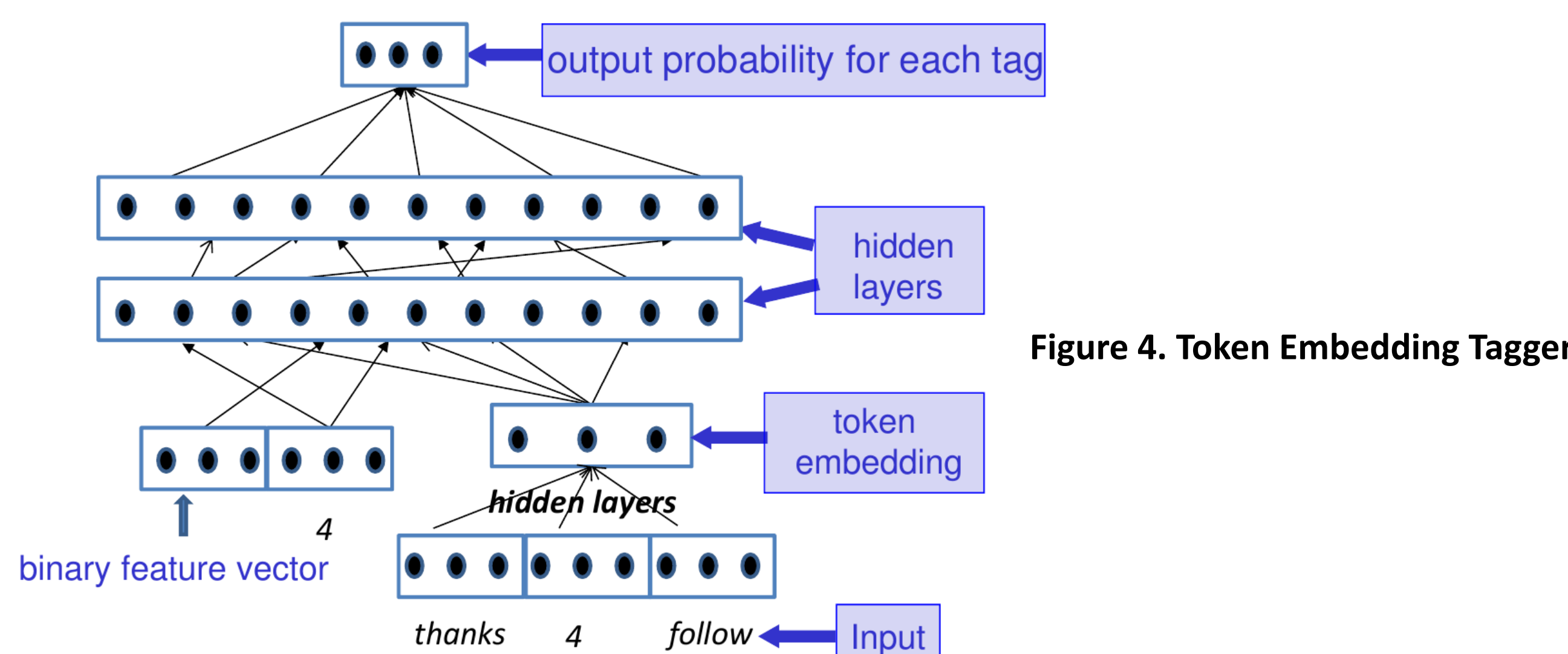
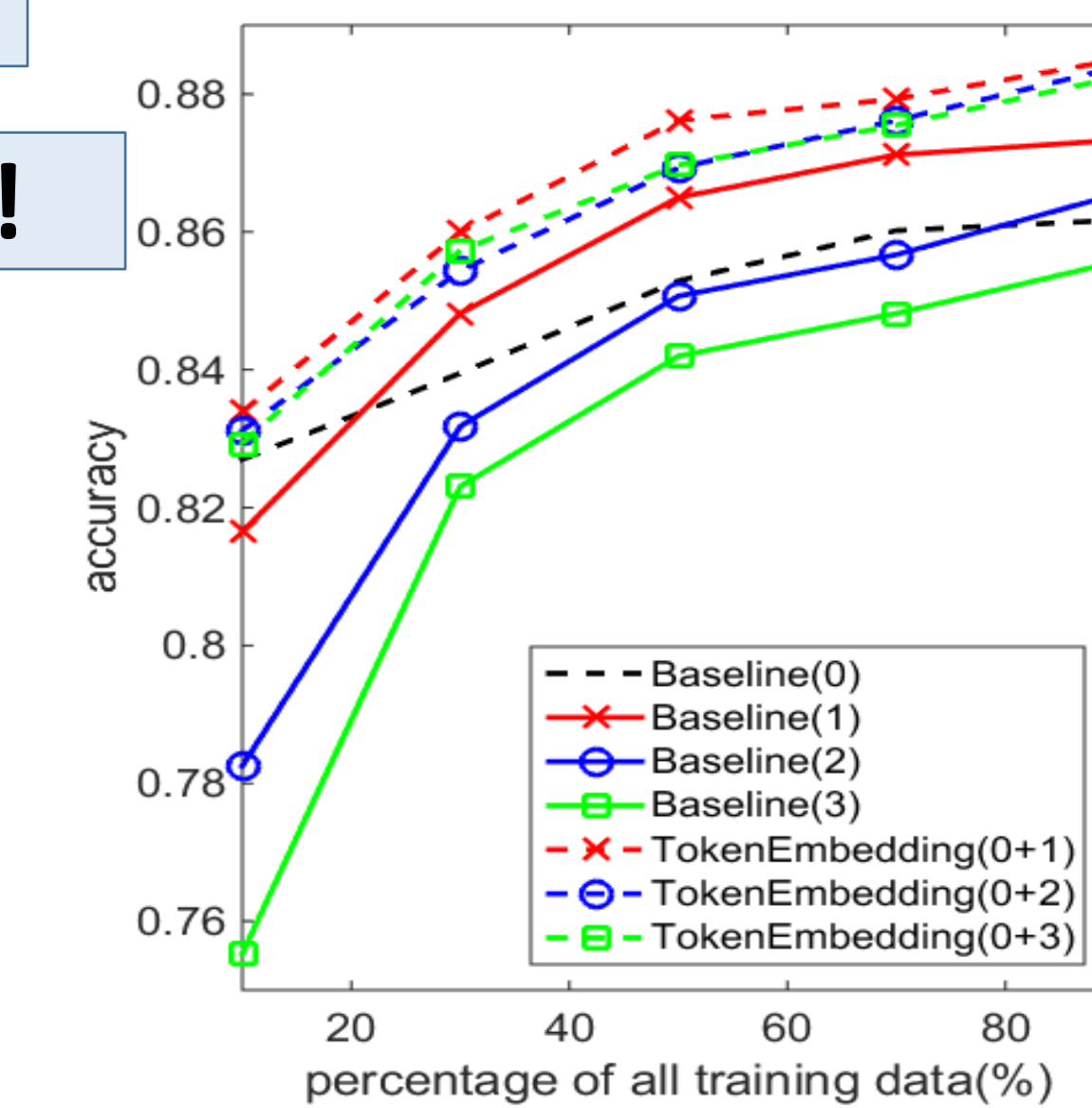


Figure 4. Token Embedding Tagger

Datasets

1. Part-of-Speech Tagging: from Gimpel et al. (2011) and Owoputi et al. (2013)
 - OCT27TRAIN, OCT27DEV, OCT27TEST
 - DAILY547
2. Dependency Parsing: from Kong et al. (2014)
 - 717 training tweets
 - 201 tweets TEST-NEW

Part-of-Speech Tagging



- Tagging results.
- "Baseline(w)" refers to the baseline tagger with context of w words
- "TokenEmbedding(w+w)" refers to the token embedding tagger with tagger context of w words and token embedding context of w' words.

	val.	test
(1)Baseline	88.4	88.9
(1)+DNN TE	+1.6	+0.9
(2)Baseline+updating	89.4	89.4
(2)+DNN TE	+0.6	+0.5
(3)Baseline+features	89.2	89.3
(3)+DNN TE*	-0.6	-1.0
(3)+DNN TE	+1.2	+1.0
(3)Baseline+features	89.2	89.3
(3)+seq2seq TE*	-0.6	-1.0
(3)+seq2seq TE	+1.3	+1.0
(4)Baseline + all features	92.1	92.2
(4)+updating	92.2	92.4
(4)+DNN TE + without updating	92.4	92.8
Owoputi et al. (2013)	91.6	92.8

- Tagging accuracies (%) on validation and test sets
- "updating" = updates type embeddings during training
- "features" = uses binary feature vector for center word
- * = omits center word type embedding

Dependency Parsing

Head Prediction Accuracy

W or w'	Baseline	DNN TE	Seq2seq TE
0	75.8	-	-
1	75.4	77.8	77.8
2	73.2	77.3	77.9
3	72.3	77.2	76.9

Adding an Additional Feature to Tweepoparser

(2) Kong et al.(2014)	80.6±0.25
(2) + Baseline parser(w=0)	80.5±0.30
(2)+ DNN TE(w=-1, w'=1)	81.5±0.25
(2)+ seq2seq TE(w=-1, w'=1)	81.0±0.17
(2)+ seq2seq TE(w=-1, w'=2)	80.9±0.33

- Attachment F1 (%) on validation set using different models and window sizes.
- For TE columns, the input does not include any type embeddings at all, only token embeddings.

- Dependency parsing unlabeled attachment F1 (%) on test sets for baseline parser and results when augmented with token embedding features.

References

1. Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for Twitter: annotation, features, and experiments. In Proc. of ACL
2. Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In Proc. of NAACL
3. Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archana Bhatia, Chris Dyer, and Noah A. Smith. 2014. A dependency parser for tweets. In Proc. of EMNLP