Unsupervised Learning of Style-sensitive Word Vectors

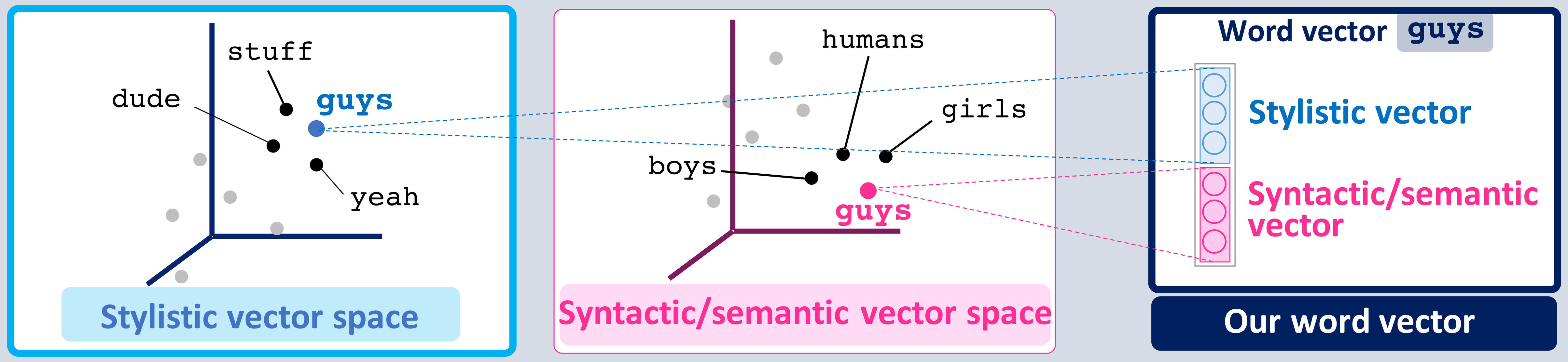
Reina Akama[1] (reina.a@ecei.tohoku.ac.jp), Kento Watanabe[2], Sho Yokoi[1,4], Sosuke Kobayashi[3], Kentaro Inui[1,4]

[1]Graduate School of Information Sciences, Tohoku University, [2]National Institute of Advanced Industrial Science and Technology (AIST), [3]Preferred Networks, Inc., [4]RIKEN Center for Advanced Intelligence Project

TOHOKU UNIVERSITY

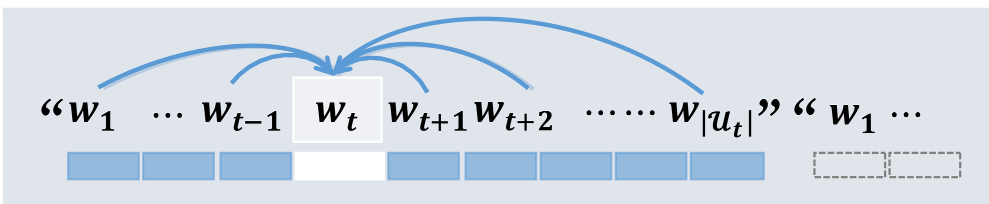
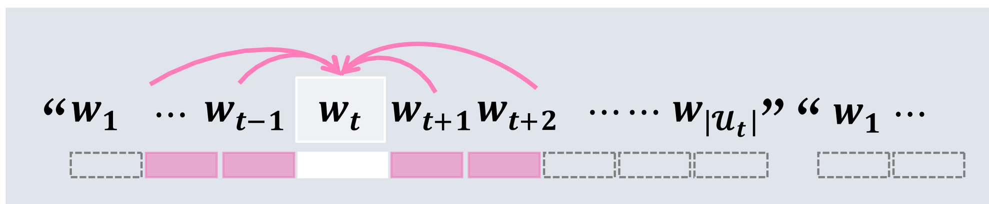
Contributions

- ✓ Proposed novel style-sensitive word vectors in unsupervised manner.
- ✓ Created word pair data stylistically similar for evaluation.
- ✓ Demonstrated that proposed methods capture the stylistic similarity between words.



Proposed Method

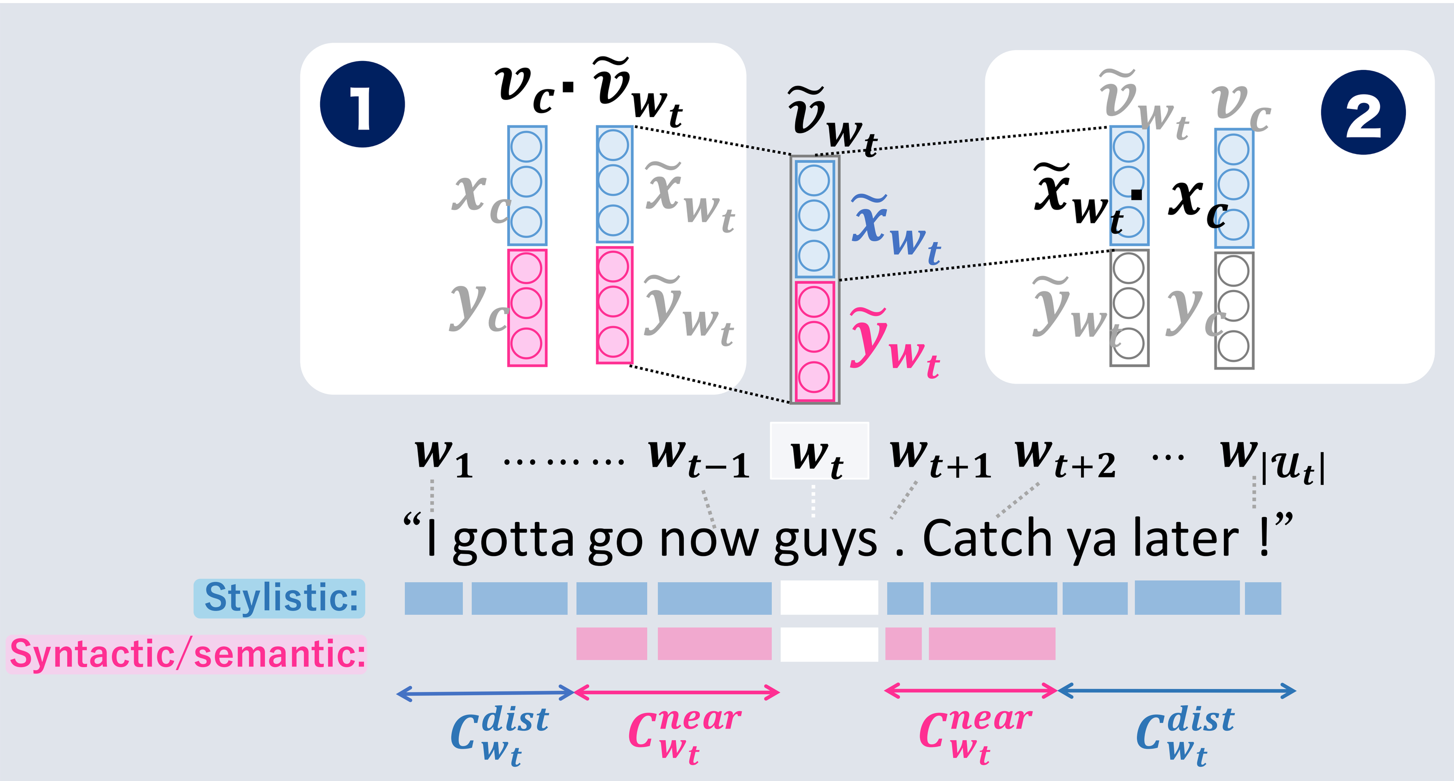
Key Idea: "The style of all words in one utterance is consistent"

Simple stylistic vector (CBOW-ALL-CTX)	**Word Vector** (CBOW)
Our hypothesis *"The style of all words in one utterance is consistent"*	**Distributional hypothesis** [Harris+ '54] *"You shall know a word by the company it keeps"*
"words with similar style will occur with similar words within an utterance"	*"words with similar meanings will occur with similar neighbors"* [Schütze+ '95]
	
$P(w_t \mid C^{\text{all}}_{w_t}) \propto \exp\left(\tilde{v}_{w_t} \cdot \frac{1}{\lvert C^{\text{all}}_{w_t}\rvert} \sum_{c \in C^{\text{all}}_{w_t}} v_c\right)$	$P(w_t \mid C^{\text{near}}_{w_t}) \propto \exp\left(\tilde{v}_{w_t} \cdot \frac{1}{\lvert C^{\text{near}}_{w_t}\rvert} \sum_{c \in C^{\text{near}}_{w_t}} v_c\right)$
vectors capture **stylistic** word similarity	vectors capture **syntactic** and **semantic** word similarity

Separation of Style and Meaning by Sampling Strategy

PROBLEM: Simple stylistic vector also captures the syntactic/semantic similarity, due to the prediction of nearby contexts.

SOLUTION: Learn two vectors simultaneously while separating style and semantic information by using the distance between the target and the context as a clue.



1. Context words are near target word: Update both ○○○ and ○○○.

$$P_1(w_t \mid C^{\text{near}}_{w_t}) \propto \exp\left(\tilde{v}_{w_t} \cdot \frac{1}{\lvert C^{\text{near}}_{w_t}\rvert} \sum_{c \in C^{\text{near}}_{w_t}} v_c\right)$$

2. Context words are far from target word: Update only ○○○.

$$P_2(w_t \mid C^{\text{dist}}_{w_t}) \propto \exp\left(\tilde{x}_{w_t} \cdot \frac{1}{\lvert C^{\text{dist}}_{w_t}\rvert} \sum_{c \in C^{\text{dist}}_{w_t}} x_c\right)$$

Experiments on Fan-fiction Corpus

Training Setups

Training corpus: 30M utterances, vocabulary size 100K.
Model settings: nearby window width 5, vector size 600 (each part 300).

Examples of Similar Words

words \ spaces	Stylistic vector space	Syntactic/semantic vector space
guys	stuff guy bunch	boys humans girls
ninja	shinobi konoha genin	shinobi pirate soldier
俺 (I;male,colloquial)	おまえ(you;colloquial,rough) あいつ(he/she;colloquial,rough) ねぇよ(not;colloquial,rough)	僕(I;male,childish) あたし(I;female,colloquial) 私(I;formal)

☺ Two vectors captured stylistic and syntactic/semantic similarity, respectively.

Quantitative Evaluations

Stylistic sensitivity
Correlation with human evaluation about style using our dataset.

Created Japanese Stylistic Word Similarity Dataset
- Using crowd-sourcing
- including 399 style-sensitive word pairs & 5 scaled scores

Available on https://jqk09a.github.io/ style-sensitive-word-vectors/

	models \ metrics	ρ_{style}	ρ_{sem}	SYNTAX ACC @5	SYNTAX ACC @10
Baselines	#1	12.1	27.8	86.3	85.2
	#2	36.6	24.0	85.3	84.1
	#3	**56.1**	15.9	59.4	58.8
Ours	Syntactic/semantic vector	9.6	18.1	**88.0**	**87.0**
	Stylistic vector	**51.3**	**28.9**	68.3	66.2

Syntactic sensitivity
Concordance rate of syntactic features.

$$\frac{1}{\lvert\mathcal{V}\rvert N} \sum_{w \in \mathcal{V}} \sum_{w' \in \mathcal{N}(w)} \mathbb{I}[\mathrm{POS}(w) = \mathrm{POS}(w')]$$

Semantic sensitivity
Correlation with human evaluation. a.k.a. word similarity task.

☺ Stylistic vectors and Baseline #3 captured stylistic similarity effectively.
☺ Syntactic/semantic vectors and CBOW vectors captured syntactic similarity well.
☹ Stylistic vectors and CBOW vectors captured semantic similarity well, since topics are also consistent within an utterance.