

依存構造に基づく 単語から語義の分散表現への細分化

芦原和樹* 梶原智之** 荒瀬由紀* 内田諭†

*大阪大学大学院情報科学研究科

**大阪大学データビリティフロンティア機構

†九州大学大学院言語文化研究院

目次

1. 研究背景

2. 研究目的

3. 提案手法

4. 評価実験

1. 実験設定

2. 意味的類似度推定タスク

3. 語彙的換言タスク

5. まとめ

研究背景

- 多くのNLPタスクで**分散表現**が使われている
- SGNS^[1] (Skipgram with Negative Sampling) や CBOW^[2] (Continuous Bag-of-Words) は**多義語**を1つの分散表現で表す



- 品詞ごとに異なる分散表現を割り当てる^[3]
 - ex) *well_noun*, *well_adv*
- トピックごとに異なる分散表現を割り当てる^[4]
 - ex) *bat_sport*, *bat_animal*

[1] Mikolov et al. : Distributed Representations of Words and Phrases and their Compositionality (NIPS 2013)

[2] Mikolov et al. : Efficient Estimation of Word Representations in Vector Space (ICLR 2013)

[3] Paetzold and Specia : Unsupervised Lexical Simplification for Non-Native Speakers (AAAI 2016)

[4] Fadaee et al. : Learning Topic-Sensitive Word Representations (ACL 2017)

品詞・トピックでは粒度が粗い

トピック : food
soft : 形容詞

I ate a **soft** cheese.



tender
(やわらかい)

I drank **soft** drinks.



non-alcoholic
(ノンアルコールの)

さらに細かい粒度で区別する必要がある

本研究の目的

品詞やトピックよりもさらに細かい粒度で多義語を区別

文中で依存関係にある単語 (context-word) ごとに
語義を区別する



それぞれを区別して扱う

データスパースネス問題の解決

- 各単語とcontext-wordの組み合わせ数は膨大
→各単語が多く分散表現を持つ
- 低頻度な単語とcontext-wordの組が存在
→学習がうまく進まない



- 各単語を原形化
- 事前学習と事後学習の2段階に分ける
 - 低頻度な組は初期値の近傍にとどまる
 - 大きく意味が外れない

提案手法

事前学習

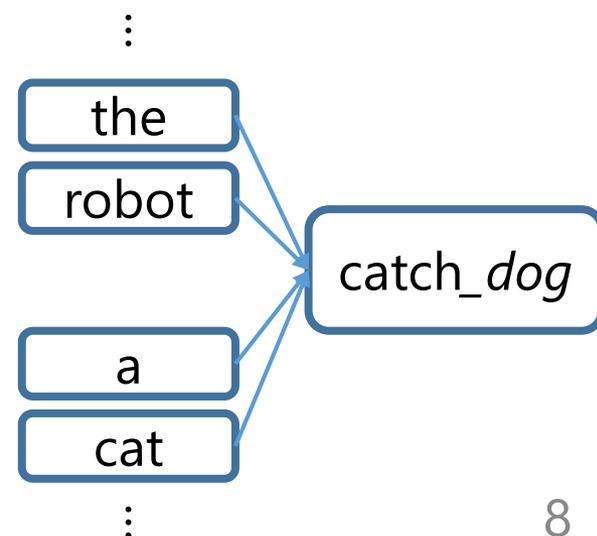
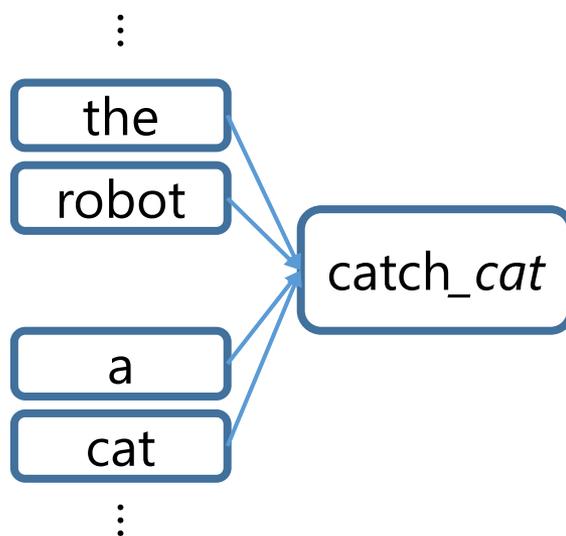
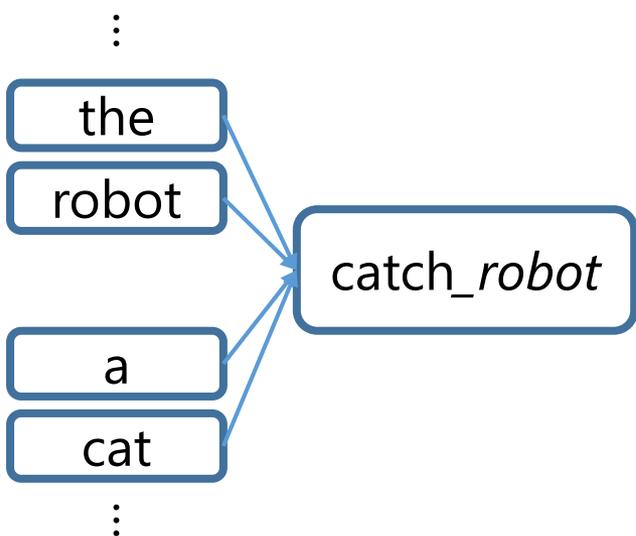
- 通常のCBOWを用いて分散表現の獲得

事後学習

- 学習データ中からcontext-wordの集合を獲得
- context-wordごとに分散表現を与える

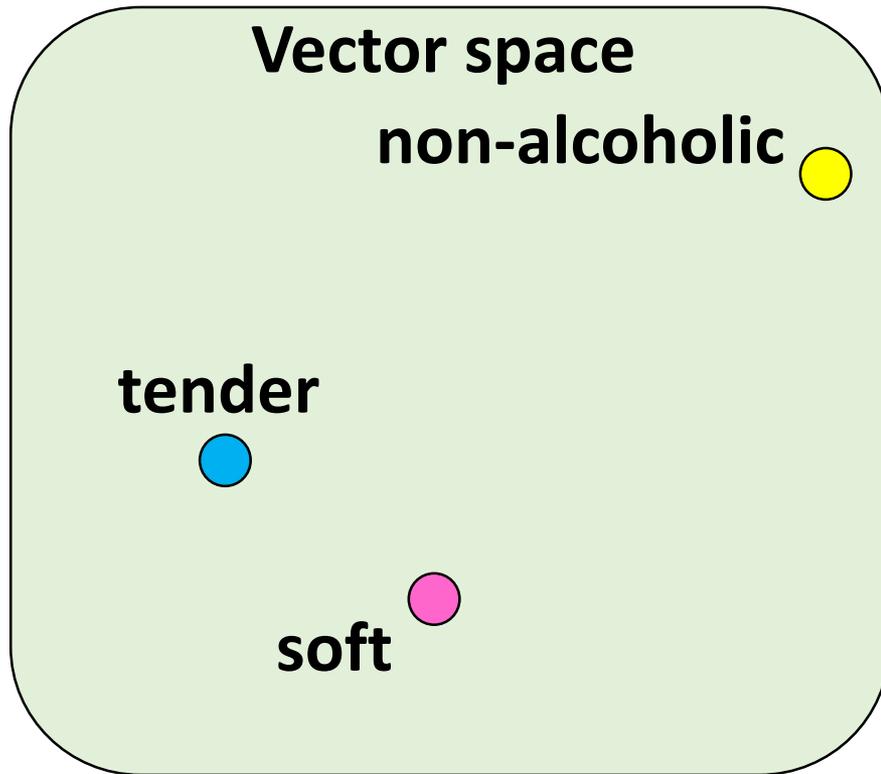
提案手法（事後学習）

学習データ中の各単語からcontext-wordの集合を獲得



提案手法

word : soft
context-word : cheese, drink, iron

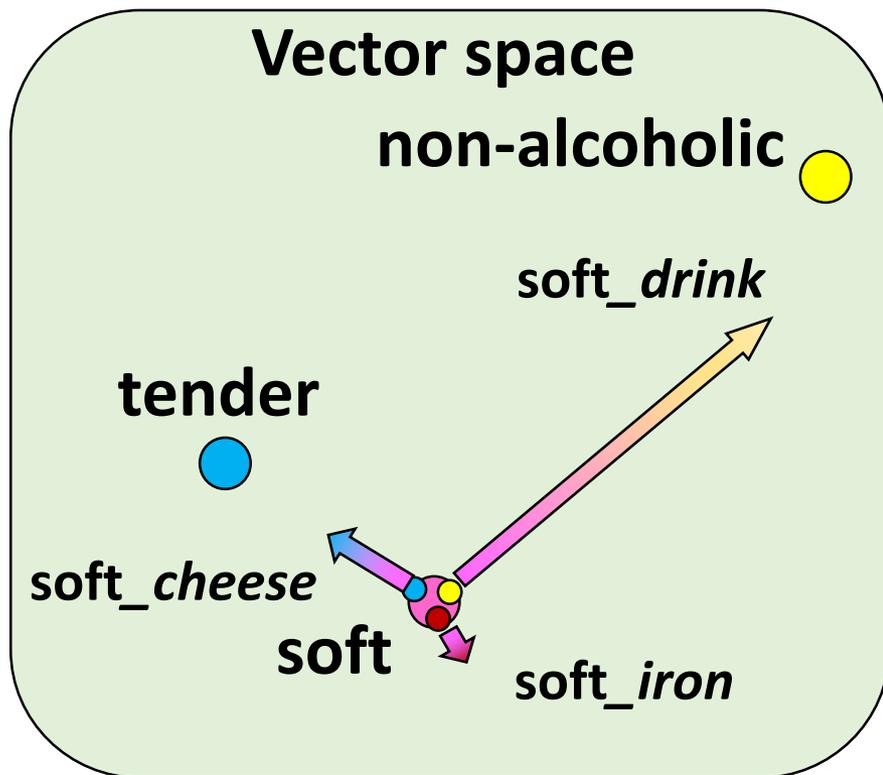


soft_cheese : tender
soft_drink : non-alcoholic
soft_iron : 低頻度語

事前学習

提案手法

word : soft
context-word : cheese, drink, iron



soft_cheese : tender
soft_drink : non-alcoholic
soft_iron : 低頻度語

事後学習

評価実験

- 実験設定
- 意味的類似度推定タスク
- 語彙的換言タスク

前処理

- English Wikipedia^[5] の本文約10億文を利用
- 品詞タグ付け及び依存構造解析のために Stanford Parser^[6] を使用
- 出現回数200回以下の単語を < *unk* > タグに置き換え
→ 約11万語を使用
- 事前学習・事後学習はともにCBOWのアルゴリズム

[5] <https://dumps.wikimedia.org/enwiki/20170601/>

[6] Manning et al. : The Stanford CoreNLP Natural Language Processing Toolkit (ACL 2014)

比較手法

CBOW

- CBOWのアルゴリズムを利用
- 各単語に1つの分散表現を割り当てる

SGNS

- SGNSのアルゴリズムを利用
- 各単語に1つの分散表現を割り当てる

MSSG^[7]

- 文脈の類似度に基づくクラスタリング
- 各単語に複数の分散表現の割り当てる

POS

- 各単語に品詞ごとの分散表現を割り当てる

TOPIC

- 各単語にトピックごとの分散表現を割り当てる

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

評価実験（意味的類似度推定タスク）

Stanford Contextual Word Similarity (SCWS) [8]

In 1955 the Soviet Union forwarded \$ 100 million in **credit** to Afghanistan, which financed public transportation, airports, etc.

Only congress has the authority to coin this **money** that should be used by the States.

creditとmoneyの類似度を推測する



多義性を考慮した類似度推定が必要

意味的類似度の推定

V : context-wordで与えられた
分散表現の集合

$$\bullet S_{avg} = \frac{1}{|V_i||V_j|} \sum_{v_i \in V_i, v_j \in V_j} \cos(v_i, v_j)$$

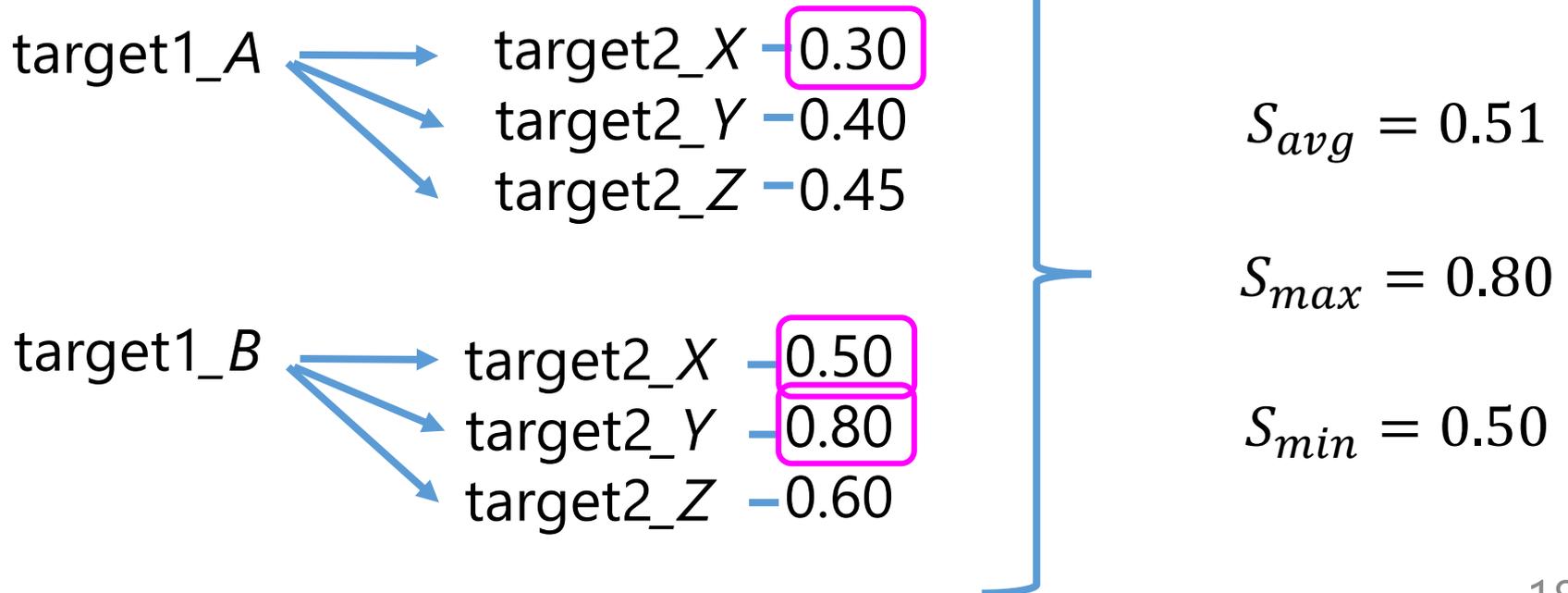
$$\bullet S_{max} = \max_{v_i \in V_i, v_j \in V_j} \cos(v_i, v_j)$$

$$\bullet S_{min} = \min_{v_i \in V_i, v_j \in V_j} \cos(v_i, v_j)$$

語彙にないため類似度を出せない場合は
2単語間の単純な余弦類似度

意味的類似度の推定

	context-word
target1	A, B
target2	X, Y, Z



評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

実験結果（意味的類似度の推定）

表：スピーアマンの順位相関係数

Model	Spearman's ρ
<i>SGNS</i> [9]	0.59
<i>MSSG</i> [9]	0.61
<i>TOPIC</i> [9]	0.61
<i>CBOW</i>	0.63
<i>POS</i>	0.64
<i>Ours</i> (S_{avg})	0.64
<i>Ours</i> (S_{max})	0.64
<i>Ours</i> (S_{min})	0.63
<i>Ours + POS</i> (S_{avg})	0.65
<i>Ours + POS</i> (S_{max})	0.64
<i>Ours + POS</i> (S_{min})	0.65

語彙の網羅性

事後学習で得られた分散表現をどの程度利用できているのか
(語彙外の場合は事前学習した分散表現を利用)

表：意味的類似度推定タスクにおける語彙の網羅性

In vocabulary / All tokens		
Ours	7297 / 8772	83.2 %
Ours + POS	7058 / 8772	80.5 %

8割以上の分散表現が語彙に含まれている

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

評価実験（語彙的換言タスク）

表：語彙的換言タスク例

ターゲット	clear
言い換え候補群	improve, change, remove, take, free, above, empty, ...
TEXT1	..., I am well taken care of until the weather finally clears .
言い換え可能単語	improve (4), change (1), ...
TEXT2	When he stands , he begins to clear the dishes from the table.
言い換え可能単語	remove (4), take (1), ...

文脈を考慮しつつランク付けをする必要がある



文脈中での単語の意味を考慮する必要がある

データセット

LS-SE_[10]

- 201種類のターゲット
- 各ターゲットに10種類の文脈
- 5名のアノテータが最大3種類ずつ言い換えを付与

LS-CIC_[11]

- 15629個のインスタンス
- 6名のアノテータが最大3種類ずつ言い換えを付与

言い換え候補を付与した
アノテータの人数を重みとする

[10] McCarthy and Navigli : SemEval-2007 Task 10: English Lexical Substitution Task (SemEval 2007)

[11] Kremer et al. : What Substitutes Tell Us - Analysis of an "All-Words" Lexical Substitution Corpus (EACL 2014)

言い換え候補のランク付け

ターゲットの分散表現 : v_t

言い換え候補群の分散表現 : v_p

COS

- $\cos(v_t, v_p)$
- 単純な2単語間の余弦類似度

balAddCos_[12]

- $|s| \cos(v_t, v_p) + \sum_{\omega_i \in s} \cos(\omega_i, v_p)$
- s はターゲットを含む1文
- 文脈との類似度も考慮する

言い換え候補のランク付け

意味的類似度推定タスク

TEXT1	In 1955 the Soviet Union forwarded \$ 100 million in credit to Afghanistan, which ...
TEXT2	Only congress has the authority to coin this money that should be used by the States.

語彙的換言タスク

ターゲット	clear
言い換え候補群	improve, change, remove, take, free, above, empty, ...
TEXT1	..., I am well taken care of until the weather finally clears .
TEXT2	..., I am well taken care of until the weather finally improve .
TEXT3	..., I am well taken care of until the weather finally change .

共通のcontext-wordのみを考慮可能

言い換え候補のランク付け

S_{avg} , S_{max} , S_{min} に加え、 . . .

$$\bullet S_{avgc} = \frac{1}{|C_t|} \sum_{c_t \in C_t} sim (vec(w_t, c_t), vec(w_p, c_t))$$

$$\bullet S_{maxc} = \max_{v_i \in V_i, v_j \in V_j} \cos(v_i, v_j)$$

C_t : context-wordの集合

$$\bullet S_{minc} = \min_{v_i \in V_i, v_j \in V_j} \cos(v_i, v_j)$$

- $sim(\cdot, \cdot)$: CosまたはbalAddCos
- $vec(\cdot, \cdot)$: context-wordから得られる分散表現

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

実験結果① (語彙的換言タスク)

Model	LS-SE		LS-CIC		
	Cos	balAddCos	Cos	balAddCos	
全組み合わせの context-word	<i>Ours</i> (S_{avg})	47.0	47.2	45.7	45.9
	<i>Ours</i> (S_{max})	48.4	48.5	46.4	46.5
	<i>Ours</i> (S_{min})	45.0	31.1	44.4	44.6
	<i>Ours</i> + POS (S_{avg})	46.7	47.0	45.7	45.9
	<i>Ours</i> + POS (S_{max})	47.7	48.1	46.3	46.5
	<i>Ours</i> + POS (S_{min})	45.4	30.6	45.3	45.5
	<i>Ours</i> (S_{avgc})	47.4	47.7	46.1	46.2
共通の context-word	<i>Ours</i> (S_{maxc})	48.4	48.5	46.4	46.5
	<i>Ours</i> (S_{minc})	45.5	30.4	45.2	45.5
	<i>Ours</i> + POS (S_{avgc})	46.9	47.4	47.7	47.8
	<i>Ours</i> + POS (S_{maxc})	47.7	48.0	48.1	48.2
	<i>Ours</i> + POS (S_{minc})	45.5	30.1	46.9	47.1

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

実験結果② (語彙的換言タスク)

表：GAPスコア②

Model	LS-SE		LS-CIC	
	Cos	balAddCos	Cos	balAddCos
<i>SGNS</i> [13]	40.5	40.9	32.1	36.1
<i>MSSG</i> [13]	41.1	NA	37.8	NA
<i>TOPIC</i> [13]	NA	42.8	NA	40.9
<i>CBOW</i>	41.0	41.1	44.1	44.4
<i>POS</i>	41.8	41.1	46.5	46.7
<i>Ours</i> (S_{max})	47.4	47.5	46.4	46.5
<i>Ours + POS</i> (S_{max})	47.7	48.1	46.3	46.5
<i>Ours</i> (S_{maxc})	48.4	48.5	48.4	46.5
<i>Ours + POS</i> (S_{maxc})	47.7	48.0	48.1	48.2

出力例

表：出力例

ターゲット	hard
TEXT1	... you are carrying on two conversations at once and you are required to <u>listen</u> hard .
出力	carefully (4), intensively (0), closely(1), intently(1), seriously (0), ...
TEXT2	One event in particular <u>hits</u> the platoon hard : the death of its platoon leader, ...
出力	badly (3), heavily (0), strongly (0), severely (1), firmly (0), ...

- 重みが大きい候補を上位にランク付けできている
- 異なる候補を出力できている

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

実験結果② (語彙的換言タスク)

表：GAPスコア②

Model	LS-SE		LS-CIC	
	Cos	balAddCos	Cos	balAddCos
<i>SGNS</i>	40.5	40.9	32.1	36.1
<i>MSSG</i>	41.1	NA	37.8	NA
<i>TOPIC</i>	NA	42.8	NA	40.9
<i>CBOW</i>	41.0	40.1	44.1	44.4
<i>POS</i>	41.8	42.1	46.5	46.7
<i>Ours (S_{max})</i>	48.4	48.5	46.4	46.5
<i>Ours + POS (S_{max})</i>	47.7	48.1	46.3	46.5
<i>Ours (S_{maxc})</i>	48.4	48.5	46.4	46.5
<i>Ours + POS (S_{maxc})</i>	47.7	48.0	48.1	48.2

実験結果

表：語彙的換言タスクにおける語彙の網羅性

		Coverage
LS-SE	Ours	82.2 %
	Ours + POS	79.7 %
LS-CIC	Ours	73.4 %
	Ours + POS	71.7 %

LS-SEの方が語彙の網羅率が高い



語彙の網羅性が高ければ性能がより大きく改善される

評価実験

- 意味的類似度推定タスク
 - タスク説明
 - 意味的類似度の推定手法
 - 結果
 - 先行研究との比較
 - 語彙の網羅性に関する考察
- 語彙的換言タスク
 - タスク説明
 - ランク付け手法
 - 結果
 - ランク付け手法の比較
 - 先行研究との比較
 - データセットの違いによる考察
 - 出現頻度の考察

出現頻度が精度に与える影響

Ours(S_{maxc}), *balAddCos*

表：context-word の出現頻度が与える影響

Model		LS-SE	LS-CIC
<i>CBOW</i>		40.1	44.4
<i>Ours</i>	5回以下	41.6	44.3
	100回以上	43.3	45.6
	全て	48.5	46.5
<i>POS</i>		42.1	46.7
<i>Ours + POS</i>	5回以下	42.8	46.7
	100回以上	43.6	47.6
	全て	48.0	48.2

依存構造に基づく 単語から語義の分散表現への細分化

- 1つの単語に複数の分散表現を割り当てる手法の提案
- 依存構造関係の利用
- 意味的類似度推定タスクと
語彙的換言タスクで高い性能を示す
- 事前学習による効果の確認

