

共参照解析のための 事象間関係知識の文脈化

井之上 直也♣,♠

杉浦 純♠

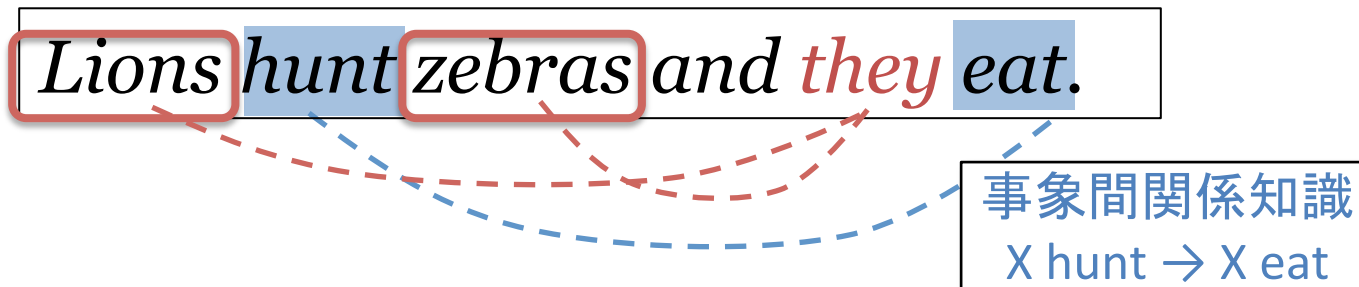
乾 健太郎♠

♣(株)デンソー 基礎研究所

♠東北大学

本研究の目的

- 行間の認識を必要とする共参照解析問題 [Levesque+ 11] に取り組む
 1. “easily disambiguated by human”
 2. “not solvable by simple techniques”



先行研究と前発表 [杉浦+ 14]、 本発表の関係

Lions hunt zebras and *they* eat.

事象間関係知識
X hunt → X eat

【先行研究の主張】
知識を使っても、性能が
大きく向上しない。原因は
知識の規模。

【前発表の主張】
NO! 知識獲得の周辺
文脈を考慮していない
ことが大問題!

先行研究
[Rahman&Ng 12]
[Inoue+ 12]
:

A5-2

共参照解析における事象間関係知識の適用

○杉浦純, 井之上直也, 乾健太郎 (東北大)

A5-3

共参照解析のための事象間関係知識の文脈化

○井之上直也 (デンソー・東北大), 杉浦純, 乾

①正解を支持する推論事例が得られていても誤る

周辺文脈を考慮できていない

入力問題文: John fired Dave because he had too many **absences**.

既存の項共有事象ペア > fire X ~ X have

適切な項共有事象ペア > fire X ~ X have **absence**

目的語

入力問題文: Gorden got mad at his brother because he did **not** get good sleep.

既存の項共有事象ペア > X mad at ~ X get

適切な項共有事象ペア > X mad at ~ X **not** get sleep

否定表現

入力問題文: Justin Bieber sold many more albums than Jack Johnson, **but** he is also very unpopular among older adults.

既存の項共有事象ペア > X sell ~ X is unpopular

適切な項共有事象ペア > X sell ~ **but** X is unpopular / X sell ~ X is popular

談話関係

先行研究と前発表 [杉浦+ 14]、 本発表の関係

Lions hunt zebras and they eat.

事象間関係知識
X hunt → X eat

【先行研究の主張】
知識を使っても、性能が
大きく向上しない。原因は
知識の規模。

【前発表の主張】
NO! 知識獲得の周辺
文脈を考慮していない
ことが大問題!

先行研究
[Rahman&Ng 12]
[Inoue+ 12]
:

【本発表の疑問】
実際に文脈問題に対
処すると、どれぐらい
効果があるのか?

参照解析における事象間関係知識の適用
井之上直也, 井之上直也, 乾健太郎 (東北大)
参照解析のための事象間関係知識の文脈化
井之上直也 (デンソー・東北大), 杉浦純, 乾

✓ おさらい

□ 文脈問題への対処方法

□ 出現文脈付き知識の表現方法

□ 知識の出現文脈を考慮した知識適用の
枠組み

□ 評価実験

【本発表の疑問】
実際に文脈問題に対
処すると、どれぐらい
効果があるのか？

先行研究における事象間関係知識の表現方法

7/20

知識獲得の対象コーパス:

... I'm thrilled that **Google** has **acquired** Zagat. ...

Google has **purchased** DeepMind, a British artificial intelligence ...

1. 述語間関係 (e.g., **acquire** ~ **purchase**)
– [Gordon & Swanson 09; etc.]
2. 述語間関係 + 項間関係 (e.g., **X acquire** ~ **X purchase**)
– [Lin & Pantel 01; Pekar 05; Chambers & Jurafsky 08; etc.]
3. 述語間関係 + 項間関係 + 項クラス (e.g., **X acquire** ~ **X purchase**, **X = <COMPANY>**)
– [Pantel+ 07; C&J 09; Melamud+ 13; etc.]

本研究の文脈捨象のレベル

知識獲得の対象コーパス:

The program could not **run** because it **has** syntax errors.

捨象する

述語 ● **run ~ have**

述語 + 項間関係 ● **run X ~ X have**

述語 + 項間関係 + 項クラス ● **run X ~ X have, X =**

【本研究の方針】知識獲得時、できるだけ一般化しない

- ・動機1: さまざまな一般化の粒度が考えられるが、事前に最適な粒度を一意に決めることは難しい
- ・動機2: なるべくシンプルな枠組みの上で、挙動の確認を行いやすくしたい

原文を完全に保持

The program could not be **run**
because it **had** syntax errors.
[program = it]

捨象しない

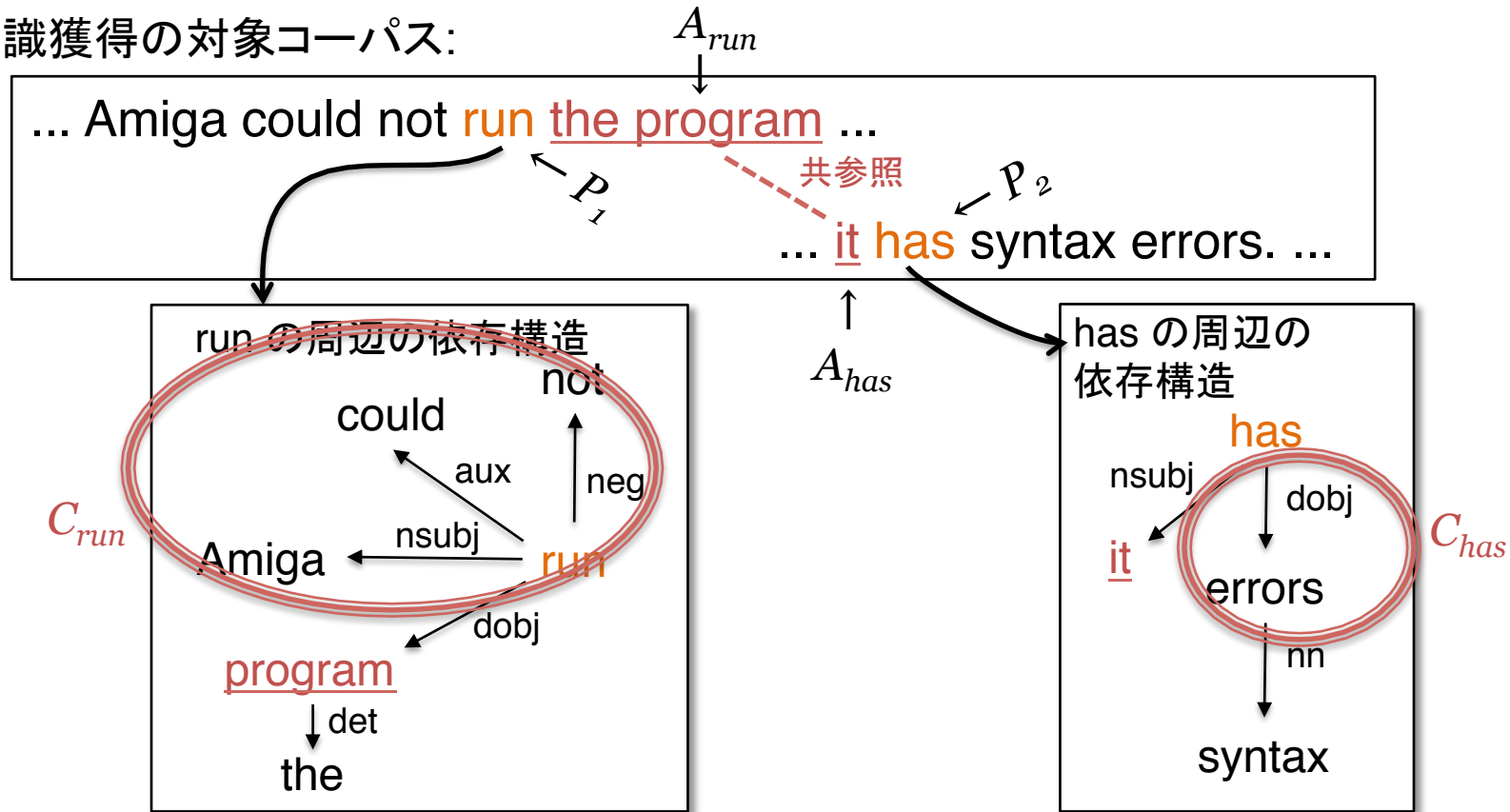
e errors

知識の表現方法

- 知識 $\equiv P_1 \sim P_2 [A_{P_1}, A_{P_2}, C_{P_1}, C_{P_2}]$
 - $P_1 \sim P_2$: 述語間関係 + 項間関係 (述語 + スロットのペア)
 - e.g., run X ~ X have
 - A_{P_i} : P_i の項のクラス (名詞で表現)
 - e.g., $A_{P_1} = \text{program}$, $A_{P_2} = \text{it}$
 - C_{P_i} : P_i の出現文脈 (依存構造上の隣接要素)
 - e.g., $C_{P_1} = \{\rightarrow:\text{aux:could}, \rightarrow:\text{neg:not}\}$,
 $C_{P_2} = \{\rightarrow:\text{dobj:error}\}$

知識獲得の例

知識獲得の対象コーパス:



獲得する知識:

述語 (P_1, P_2)	項 (A_{P_1}, A_{P_2})	出現文脈 (C_{run})	出現文脈 (C_{has})
run <u>X</u> ~ <u>X</u> has	program, it	{→:nsubj:Amiga, →:aux:could, →:neg:not}	{→:dobj:error}

【本発表】

文脈問題に対処すると、どれぐらい効果があるのか？

✓ おさらい

□ 文脈問題への対処方法

✓ 出現文脈付き知識の表現方法

□ 知識の出現文脈を考慮した知識適用の
枠組み

□ 評価実験

k近傍法による共参照解析

入力問題文 The man shot his friend and he died.

× X shoot ~ X die
 (the man - he)

● shoot X ~ X die
 (his friend - he)

凡例

×

X shoot ~ X die
 {(a girl-she), (he-he), ...}

●

shoot X ~ X die
 {(Bob-Bob), (he-he), ...}

■

Others

項共有 事象ペア

X shoot ~ X die
 (a girl - she)

項共有 事象ペア

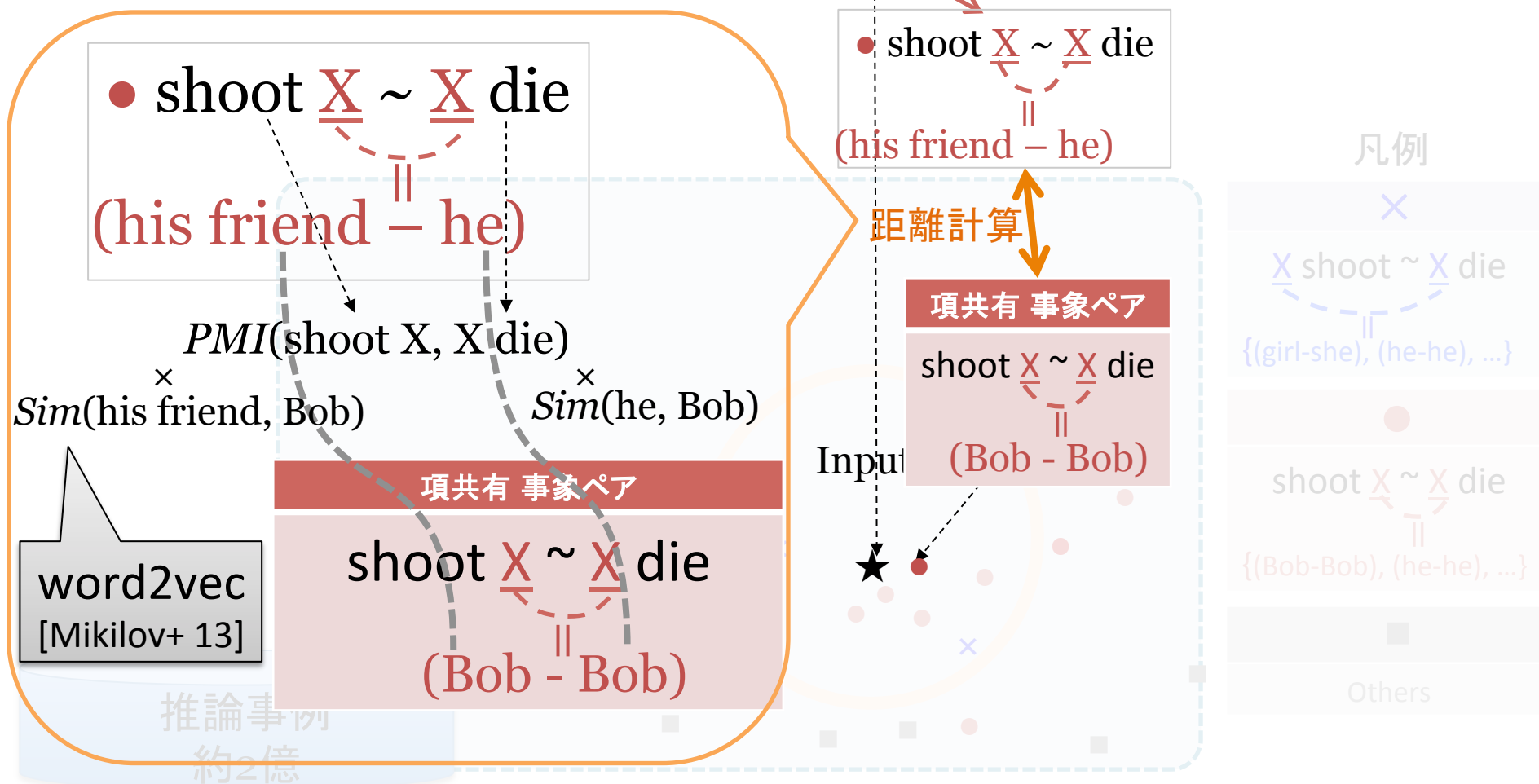
shoot X ~ X die
 (Bob - Bob)

Input

推論事例
約2億

k近傍法による共参照解析

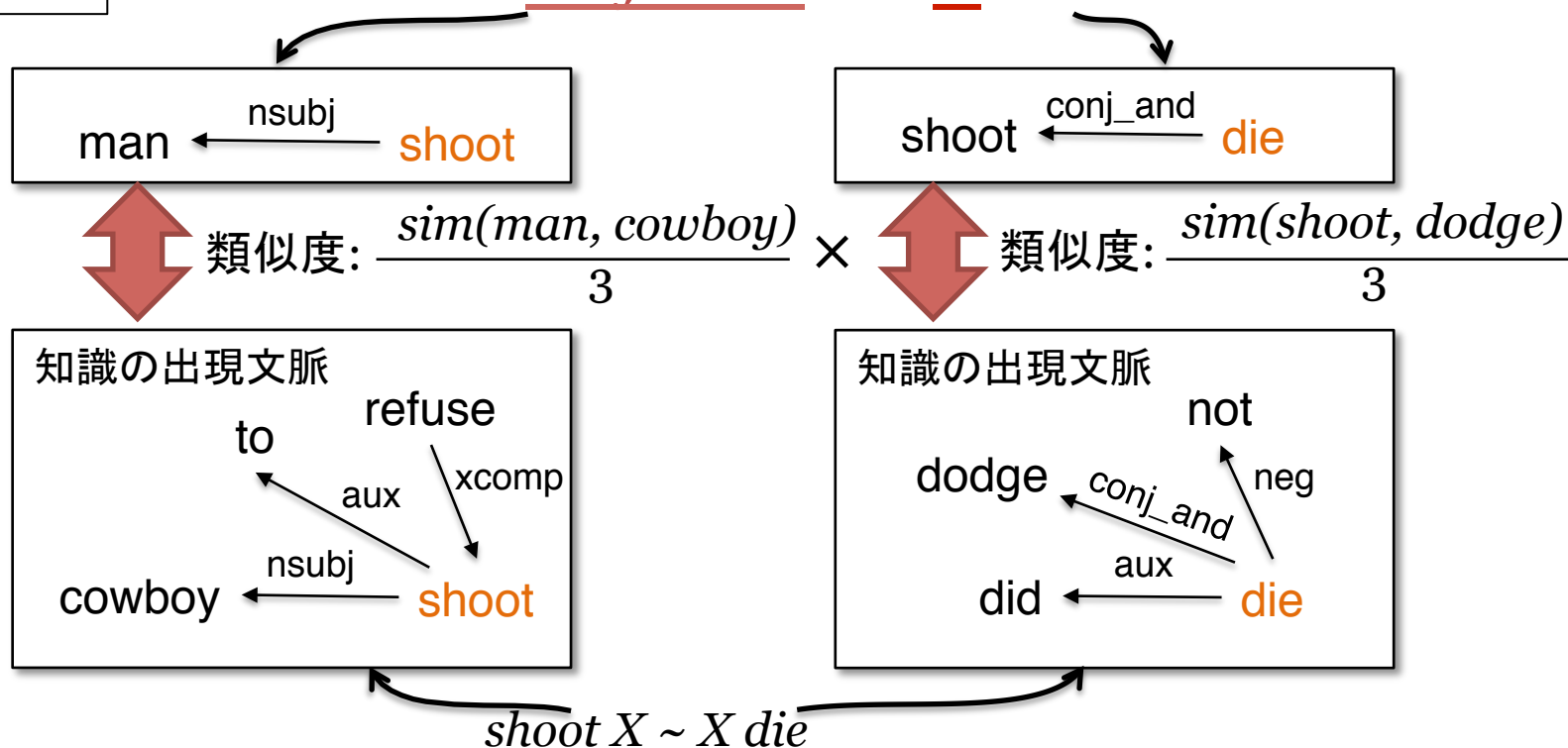
入力問題文 The man shot his friend and he died.



文脈類似度の計算方法

- 文脈の依存構造間の類似度を計算
 - 各依存関係における単語一致の割合

入力 *The man **shot** his friend and he **died**.*



【本発表】

文脈問題に対処すると、どれぐらい効果があるのか？

✓ おさらい

✓ 文脈問題への対処方法

✓ 出現文脈付き知識の表現方法

→ 依存構造上の隣接要素

✓ 知識の出現文脈を考慮した知識適用の

→ kNN のスコア関数へ

枠組み

□ 評価実験

実験設定 (cont'd)

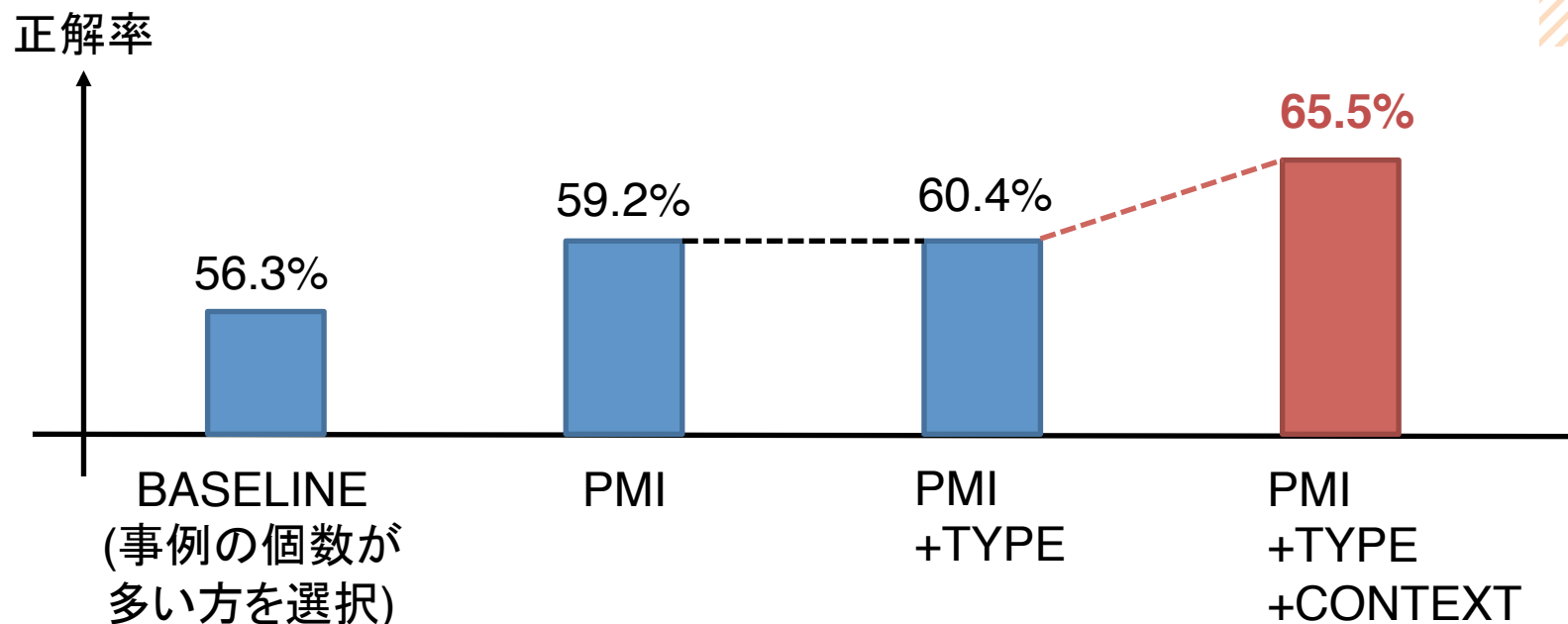
- 知識獲得
 - 獲得手法: [Chambers & Jurafsky 08]
 - 談話内の述語の項共有に基づく Event Chain 抽出
 - 対象コーパス
 - ClueWeb12 の一部 (約2億文書、7億文)
 - Stanford CoreNLP [Lee+ 11] にて共参照解析
 - 獲得できた知識の総数: 約2億 (事例単位)
- 単語ベクトル
 - Skip-gram [Mikolov+ 13] により構築
 - 次元数: 300
 - 訓練: 1000億語分の Google News 記事

実験設定

- テストセット

- Winograd Schema Challenge 問題集
[Rahman & Ng 12] 561 問のうち、正例と負例の先行詞候補の両方を支持する知識が見つかった 316 問
 - e.g., *The man shot his friend and **he** died.*
 - 「shoot X ~ X die」「X shoot ~ X die」の両方が知識 DB 内に見つかれば、対象とする
- i.e., 文脈考慮の効果が見えうる 316 問

実験結果



- 知識の出現文脈の考慮が解析に有効であることが確認できた
 - (次スライドでもう少し詳しく考察します)
- 項クラスの情報ほとんど効いていない
 - データセットの性質上、先行詞候補に同一クラスの名詞が来ることが多い

文脈考慮による改善例: 😊

- 目的語が因果関係成立に重要な問題

*The man gave the beggar some money because
he was very generous.*

- 問題を解くには、“X give Y money ~ X is generous” の関係を捉える必要あり
- 実際に、入力に近い事例が見つかった:

*They've given Dave cash just because
they are so generous.*

(前発表も含めた全体の)まとめ

- 知識に基づく共参照解析の先行研究において、性能向上幅が小さい原因を分析した
 - 結論: 知識規模の問題よりも、知識の出現文脈の捨象レベルの方が大きな問題
- 文脈を考慮した知識表現・知識適用の枠組みを構築し、文脈化の有効性を確認した
 - 結論: 文脈考慮の効果を実際に確認できた
- 今後の課題
 - 距離関数への述語の類似度スコアの導入
 - 項クラス・出現文脈にもとづく推論事例のクラスタリング
 - 論理推論の枠組みを用いた他の知識との組合せ的利用
[井之上+ 言語処理学会, 12; Inoue+, COLING, 12]