

数量の大小の自動判定：

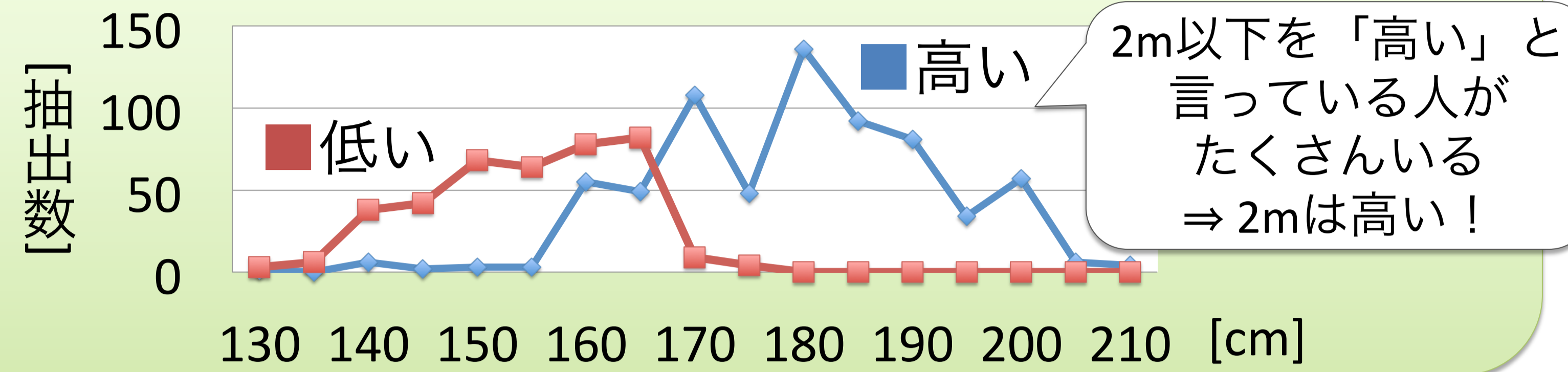
「彼は身長が2mある」は高いか低い

成澤 克麻 渡邊 陽太郎 水野 淳太 岡崎 直観 乾 健太郎 (東北大学)

概要

- 大小知識の獲得手法と大小判定の手法を提案
- 大小知識：[数量, 対象, 大小についての意見]
これをウェブから自動抽出 例 [2m, 身長, 大きい]
- 結論: 幅広い対象について高い精度で大小判定できた

➤ 例：「身長」についての大小知識の抽出結果



背景

- 含意関係認識において数量表現に関する問題がある [成澤 2012] 例: 「リンゴ3つとオレンジ5つ」⇒「果物8個」
- 課題は様々あるが、課題の1つは数量の大小判定

t: 近い将来、70億人が水不足に直面する
h: 近い将来、深刻な水不足が懸念される

関連研究

- [Aramaki 2007][Davidov 2010] 長さ、重さなどの情報を抽出。part-of関係などの推定、質問応答に利用
- ⇒ 客観的な数量の情報抽出 (大小の判定=主観的)

推論の例

70億人が水不足に直面
⇒ たくさんの方が水不足に直面
⇒ 深刻な水不足が懸念

数量の大小判定

今回は対象としない

大小知識の抽出



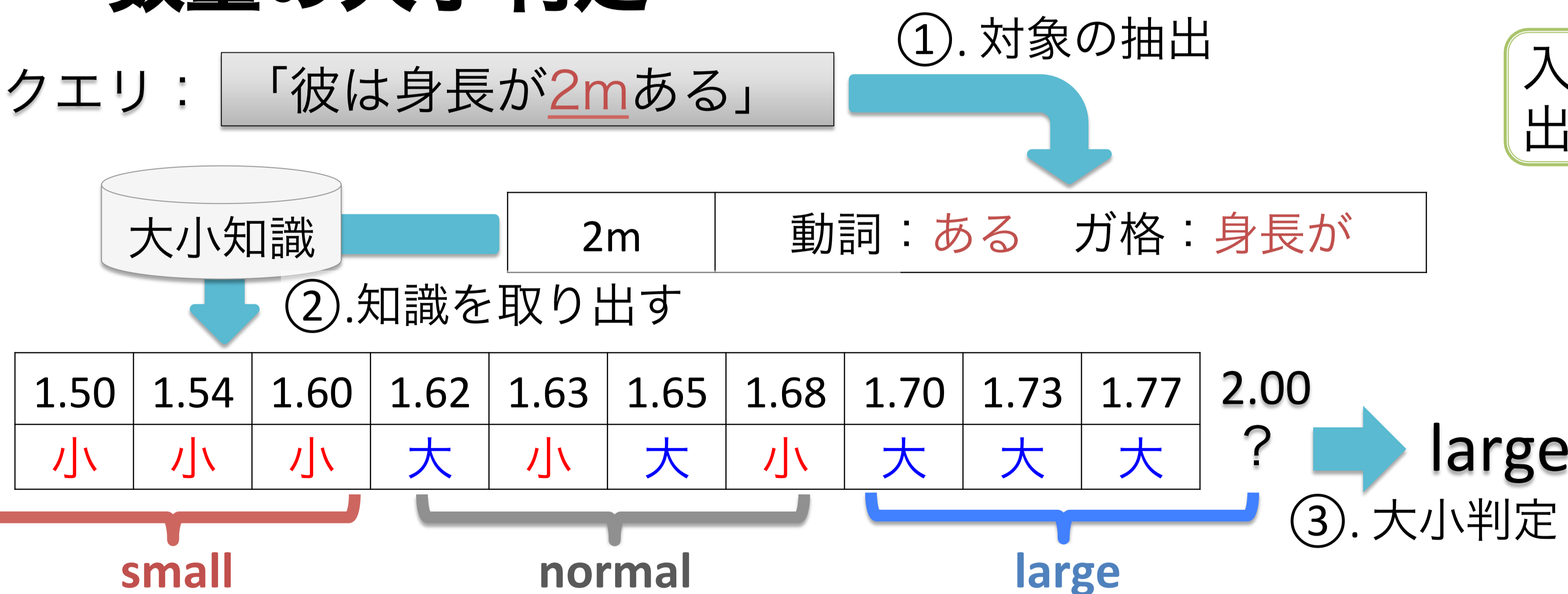
数量	数量が示す対象	大小
1.80m	動詞: ある ガ格: 身長が	大
1.52m	動詞: ある ガ格: 身長が	小
7匹	動詞: 飼う ヲ格: 猫を デ格: 家で	大

- 大小知識の抽出は、次の3ステップで行われる
- 1. 数量表現を抽出、規格化：漢数字や同一単位などを考慮して、数量表現から数量 (数と単位) を抽出。

- 2. 対象の抽出：対象を表す意味表現として数量表現の係りの先動詞、その項 (ヘッドのみ) を抽出。動詞に係っている数量表現のみを抽出対象とする
- 3. 大小情報の抽出：「も」「しか」に注目して抽出

数量の大小判定

入力：文と数量表現
出力：数量の大小を normal, large, smallの3値で出力



- ① 数量表現の規格化、対象の抽出
- ② 等しい単位・対象情報をもつ大小知識を取り出し、値と大小情報のペアからなるリストを作成
- ③ 「全員が大(小)と感じる範囲」を閾値とし、閾値を用いて大小を判定

評価実験、エラー分析

実験設定

- ウェブページから動詞に係っている数量表現を含む文 計2000文を抽出
- 数量に対して3人のアノテーターが5段階の大小評価を行った (小, やや小, 普通, やや大, 大)
- 文脈依存な場合など判定不能な文は無視
例) 「3000円払った」 (何に払ったのか?に依存)
- 2人以上の評価が一致 & 判定可能な文をゴールドデータとして、提案手法の評価を行う
- ※ウェブページ1億文書から知識を獲得

実験結果

Label	P	R	F1	Acc
large+	0.92	0.78	0.85	0.77
normal+	0.81	0.77	0.79	
small+	0.23	0.70	0.46	
large	0.90	0.66	0.78	0.62
normal	0.59	0.59	0.59	
small	0.16	0.55	0.36	

- 比較的良い精度で判定できた
- ※上は「やや小/大」を普通, 小/大のどちらかであれば正解としたもの。下は「やや小/大」を含めなかったもの

エラー分析

1. 小さい側の判定が難しい
例) 「理由は2つある」
→ 「普通」についての知識がないため
また、微妙な文脈を抽出する必要
2. 対象を詳しく捉える必要
例) 「彼氏と別れた友達が2人いる」
→ 「友達の数」として捉えたため
3. 知識不足による失敗
例) 「エヴァを10回見に行った」
→ エヴァについての知識はなかった