

数量表現と比較に着目した意味解析に向けて

佐々木 翔大 田 然 乾 健太郎 (東北大学)

背景・導入

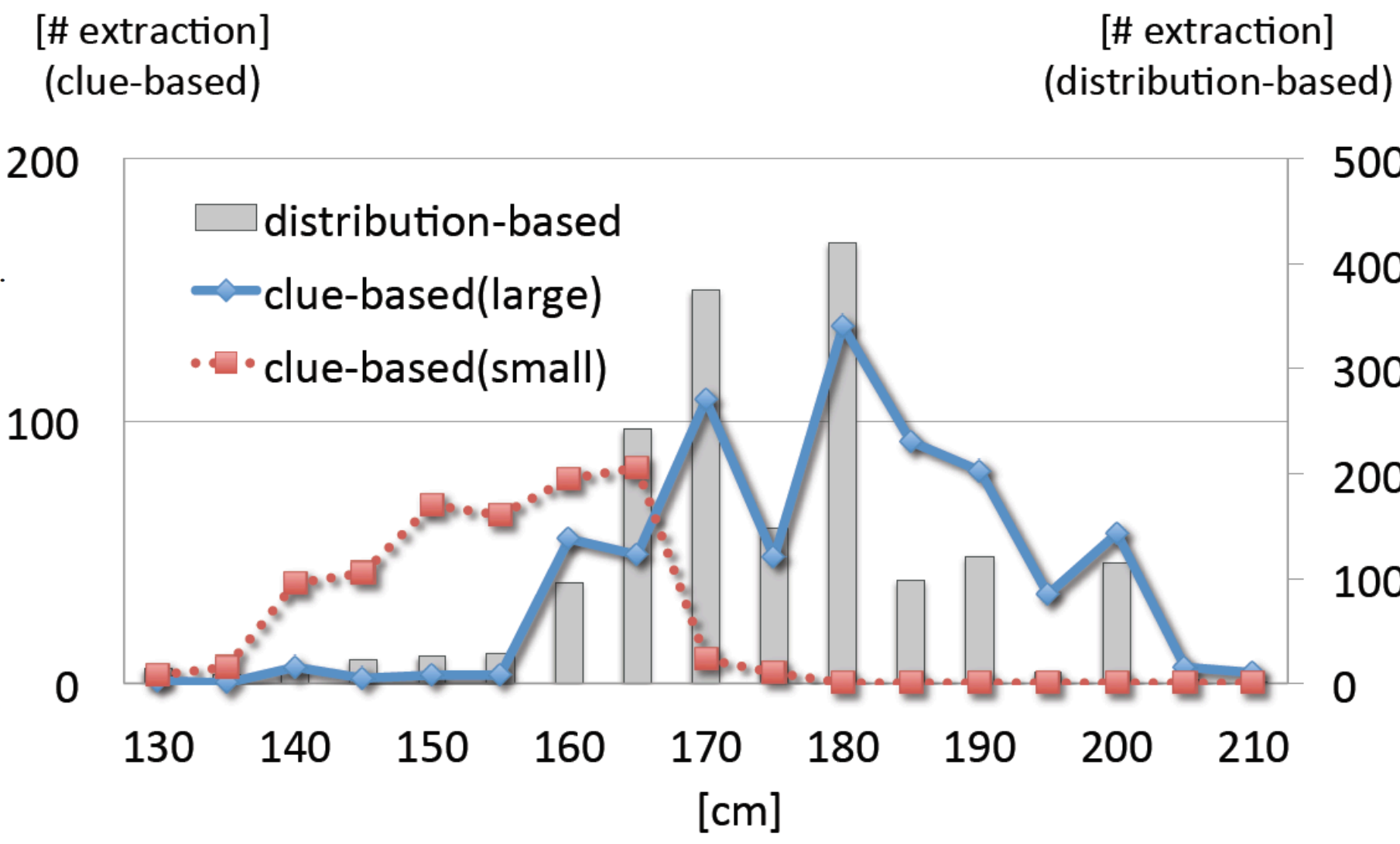
文章理解における数量に関する常識の必要性

例文: 「身長2m体重180kgの小錦は俊敏な動きで相手のパンチをかわした」

Narisawaら(2013)は数量の大小の常識を獲得

- 数量の分布を単位と文脈ごとに計算

English: He gave \$300 to a friend at the bank.
 Number: {value: 300; unit: '\$'}
 Context: {verb: 'give'; nsubj: 'he'; prep_to: 'friend'; prep_at: 'bank' }



しかし

- 記号ベースの文脈エンコードは困難であるとの報告

Point 1 事前実験

ナイーブな手法で数量の文脈をエンコードできるか?

- 数量表現の周辺のベクトルを平均する

入力: Andorra's tourism services an estimated [10.2 million] visitors annually.

類似度上位:

類似度	数量表現を含む文
0.85	An estimated [10.2 million] tourists visit annually, attracted by Andorra's duty-free status and by its summer and winter resorts.
0.79	In the 1970s and 1980s, however, tourism was an important industry, drawing an average of [150,000] visitors annually.
0.77	With over [10 million] foreign tourists annually, tourism generates revenue in excess of \$7 billion.
0.73	An estimated [20 million] tourists visit the state each year.

- 精緻なエンコードは困難

Point 2 比較の文に着目した意味解析

- 比較と数量表現は密接な関係にある
- 比較の文の意味解析を行い, 数量に関する常識を獲得
 - 自動アノテーション, 分散表現の学習

比較を構成する6つの要素

1. 比較対象
2. 比較範囲
3. 比較の視点
4. 比較の順位表現
5. 比較に関連する数量表現
6. 数量表現に紐づく述語

要素のアノテーション

アノテーションの例

• Alabama is the most religious states in the U.S. with 58% of the population attending church regularly.

対象: Alabama, 順位: the most, 視点: religious, 範囲: U.S., 数量: 58%, 述語: attending church regularly

• Algeria has the second largest defence budget in Africa (\$10 billion)

対象: Algeria, 順位: the second largest, 視点: defence budget, 範囲: Africa, 数量: \$10 billion

自動アノテーションの展望

1. 系列ラベリングモデル
2. 意味役割ラベリングモデル

アノテーションインターフェース

Overview

Workers are given an English text which contains a superlative comparison and a numerical expression to support that comparison. Workers are asked to identify spans in the text as different participants in this comparing structure.

Detail

For example, given the sentence:

"Alabama is the most religious states in the U.S. with about 58 % of the population attending church regularly."

Workers are asked to copy spans of this sentence into the following six categories:

1. The thing being compared

This is the focus of the sentence, a thing being compared to others and having the greatest degree in that comparison. In this example, it should be "Alabama".

....

Task

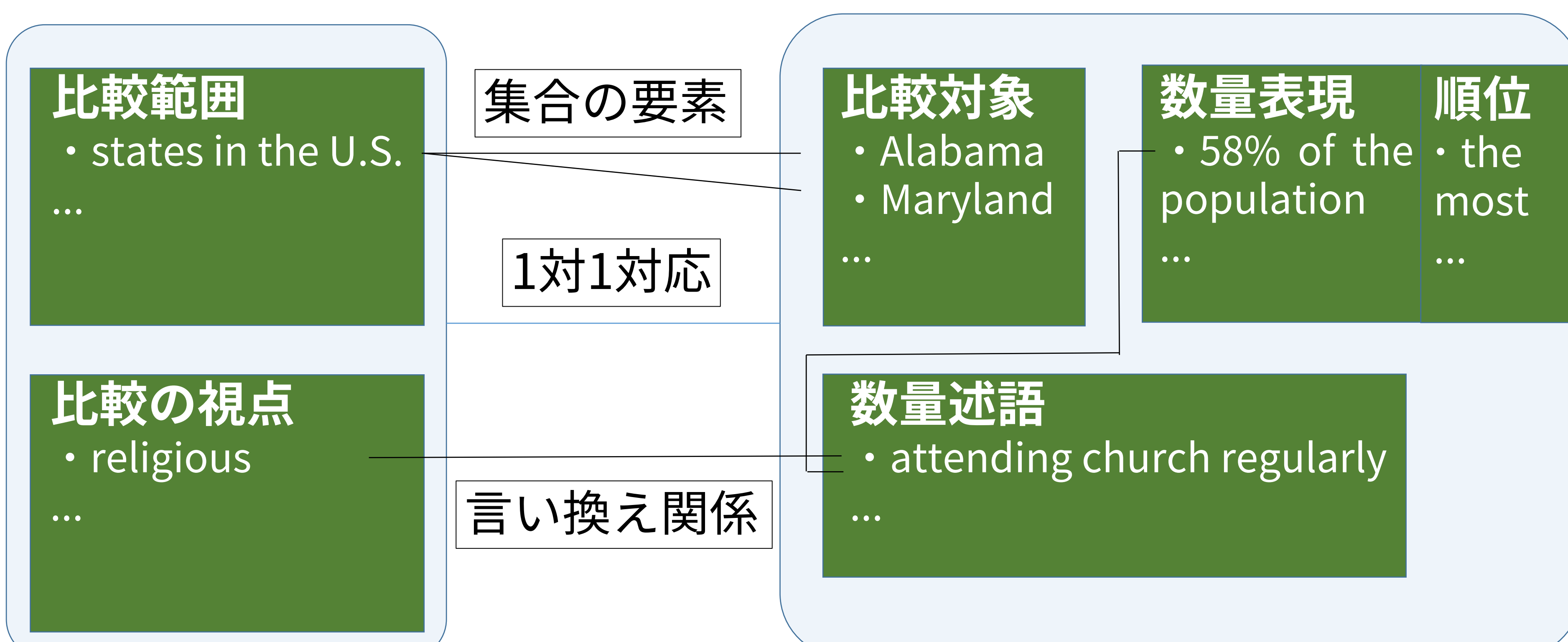
Target: "Alabama is the most religious states in the U.S. with about 58 % of the population attending church regularly."

1. The thing being compared

2. The class of things being considered for comparison

...

要素間の関係のモデル化



分散表現を用いた情報抽出

入力: 範囲 = destinations in Europe, 視点 = popular

出力: Amsterdam is one of the most popular tourist destinations in Europe, receiving more than 4.63 million international visitors.