

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

{katsuma, yotaro-w, junta-m, okazaki, inui} @ecei.tohoku.ac.jp

1 はじめに

含意関係認識の課題の1つに、一方もしくは両方の文が数量表現を伴う場合への対応が挙げられる。例えば、現状では以下のような例で含意関係を認識するのは難しい。

- (1) t : 21世紀半ばには最悪の場合、全人口の7割以上にあたる70億人が水不足に直面する。
 h : 近い将来、世界は深刻な水不足になると懸念されている。

このような文の認識は、先に述べた含意関係認識技術の応用先である質問応答や情報抽出などで重要である。しかし、Sammonsら[1]やLoBueら[2]も指摘するように、現状の含意関係認識の研究は数量表現を伴う文に対して十分な対応ができていない。例えば含意関係認識の評価型ワークショップであるRTE-6[3]において(1)のような問題を解決しているシステムは存在しなかった。含意関係認識における数量表現の扱いは既存研究でほとんど無視されてきた問題であった。

これを受けて成澤ら[4]では数量表現の取り扱いに関する問題の分類、解決に必要な処理の分析を行った。また数量表現を処理する上で最も基本的な処理と言える数量表現の抽出・規格化を行う手法を提案した。この手法により、数量同士の含意関係(「70億人」⇒「六十億人以上」など)が認識できるようになった。しかし、冒頭の「70億人が水不足」⇒「深刻な水不足」を導くために必要な数量と人間の主観の間の含意関係の問題は対処できなかった。すなわち「数量表現が示す数量の認識」(「70億人」という表現が、どんな数量を示すか)は可能になったが、「数量が示す意味の解釈」(70億人という数量が、どんな意味をもつか)が処理できなかった。

そこで本研究では、数量の解釈にむけた第一歩として、数量の大小を判定するという問題に取り組む。大小に着目するのは、多くの数量の解釈の根本は数量の大小の理解にあると考えられるためである。例えば、冒頭の例では「70億人が水不足」⇒「たくさんの人々が

水不足」⇒「深刻な水不足」という推論が妥当である。

本稿では大小知識の自動獲得手法、および大小知識を用いた数量の大小判定手法を提案する。ここで大小知識とは「数量表現」「数量が示す対象」「その大小情報」という3つ組の情報を指す。提案する抽出手法では、例えば「身長が150cmしかない」という文から「150cm, 身長, 少ない」という「身長が150cmだと少ない」ことを意味する知識を抽出する。大小判定では、獲得した知識を用いて、入力された文中の数量の大小判定を行う。例えば「彼は身長が2mある」の「2m」が多いのか少ないのかを判定する。

以下、大小知識の抽出手法、大小の判定手法の詳細を述べた後、最後に評価実験の結果とエラー分析を述べる。

2 大小知識の抽出

大小知識とは「数量表現」「数量が示す対象」「その大小情報」という3つ組の情報を指す。抽出のおおよその流れは以下のようになる。

1. 入力文中から数量表現を抽出・規格化
2. 数量が示す対象の抽出
3. 数量の大小情報の抽出

以下でそれぞれのステップについて詳しく述べる。

2.1 数量表現の抽出・規格化

数量表現の抽出・規格化ツールである `normalizeNum-exp`¹ を用いて入力文中の数量表現を抽出・規格化した。ここで規格化とは、数量表現から「値」「単位」を抽出することを指す。例えば「千二百円」という数量表現から「値：1200」「単位：円」を抽出する。

2.2 数量が示す対象の抽出

「全国で無免許運転により3人の学生が退学になった」という文における「3人」は「全国で無免許運転により退学になった学生の人数」を示している。我々はこれを「数量が示す対象」と呼ぶ。数量が示す対象をどのように計算機上で表現するか、またどのようにそ

¹<http://www.cl.ecei.tohoku.ac.jp/~katsuma/software/normalizeNumexp/>

表 1: 数量表現の用法と抽出ルール

用法	説明	例文	抽出ルール
連体修飾型	数量表現が体言を修飾している場合	3人の学生が来た	数量表現の係り先の名詞、その名詞の係り先の動詞の原形、その動詞の格情報 ²
連用修飾型	数量表現が用言を修飾している場合	学生が3人来た	係り先の動詞の原形、その動詞の格情報
同格型	数量表現が名詞の直後にきて、後ろから名詞を修飾するようにふるまう場合	学生3人が来た	直前の名詞、その名詞の係り先の動詞の原形、その動詞の格情報
非修飾節の場合	以上の3つの用法以外の数量表現	来た学生数は3人です。 あの3人は学生だ	様々な事例が考えられ、数量が示す対象の抽出が難しいため、大小情報抽出の対象としない。

れを文中から抽出するかは難しい問題である。本稿では手始めに係り受けと動詞の格情報で数量が示す対象を表現し、これで捉えきれない対象がどれだけあるかを評価実験において分析する。

実際の抽出ルールを表1に示す。数量表現は文中の役割から4つの用法に分類され³、用法によって若干異なる抽出ルールが用いられる。冒頭の「全国で無免許運転により3人の学生が退学になった」の「3人」は連体修飾型の数量表現の用法なので、抽出される情報は「係り先名詞：学生」「名詞の係り先動詞：なる」「動詞のガ格：学生」「動詞のニ格：退学」「動詞のデ格：全国」となる。

2.3 数量の大小情報の抽出

このステップでは数量の大小情報、すなわち「その文章での数量の大小に関する話者の捉え方」を抽出する。話者が数量の大小に関してどんな捉え方をしているかを、文全体の意味を考慮して判断するのは非常に難しい。しかし、我々は取り立て助詞などの表現を手がかりとするシンプルな手法で、数量に対する話者の主観を抽出することを可能とした。我々が用いた手がかりは以下の4つである。

- 取り立て助詞「も」 例：「3人も来た」
- 「も」の名詞修飾形 例：「3人もの学生」
- 取り立て助詞「しか」 例：「3人しか来ない」
- 形容動詞「わずか」 例：「わずか3人の学生」

抽出処理の際には、数量表現が「も」「もの」を伴えば話者はその数量を「大きい」と捉えているとし、「しか」「わずか」を伴えば「小さい」と捉えているとする⁴。「も」「しか」は連用修飾型、「もの」「わずか」は連体修飾型の数量表現につく（同格型には大小を判断する手がかりが存在しないので、大小情報の抽出対象にしない）。

²「ガ・ヲ・ニ・ハ・デ・ヘ・カラ・ヨリ・ニテ・マデ格」を抽出した。今回は処理の簡略化のため、格情報の推定には述語項構造解析のような深い解析は用いず、係り受けを用いたヒューリスティックな手法で推定を行っている。

³連体修飾型、連用修飾型、同格型の3つの分類は現代日本語文法 [5] による。[5] では修飾節となる数量表現しか扱っていないため、その他の分類として「非修飾節の場合」という分類を用いた

⁴「も」には上で挙げた以外の用法が存在するが、これに対する処理は紙面の都合で省略する。

表 2: 「単位：cm」「動詞：ある」「ガ格：身長」と一致する知識を抽出したもの（一部）

値	大小	(抽出元の文章)
150	小	身長が150センチしかありません
154	小	中3男子です。身長が154cmしかありません
160	小	現在、高校一年生の男です身長が160cmしかありません。
161	大	不思議だったのは身長が161cmもあった事だ
162	小	僕は今高2なのですが身長が162cmしかありません
163	大	身長が163センチもあって体重39キロってありえますか
163	小	私は身長が163センチしかありませんから少し大きな小学生でもいけそうです^^
165	小	私高校三年の18歳男なのですが、身長が165cmしかありません
167	大	彼女は身長が167センチもあり、私が「背が高くなりたいな〜と(略)
168	大	ヒールがなくても168センチも身長があるのに、このサンダルはいたら...
170	大	身長が170cmもあってきれいな人でした。でもグラブプリの人には負けてたかな
173	大	また、身長が173cmもあり、16歳の女性キャラの中では1番背が高い。

3 数量の大小判定

文が入力された際に、入力された文中の数量の大小の判定を行う手法について説明する。これは以下の処理で達成される（例として「身長が190cmある」という文が入力された際の「190cm」の大小判定の処理を述べる）。

1. 数量が示す対象を抽出（「動詞：ある」「ガ格：身長」）
2. 数量表現の単位と数量が示す対象が完全一致するものを、知識中から取り出し、「値」「大小」のペアで構成されるリストを作成（表2）
3. リストを用いて大小の閾値を算出
4. 閾値を用いて大小判定（「大」「小」「普通」の3値）を行う。（「大」が出力される）

3の処理について詳しく述べる。表2においては、人が判断すれば背が低いとみなせる範囲はおおよそ160cm以下、背が高いとみなせる範囲はおおよそ170cm以上であると言える。このように対象の大小の閾値を「値」「大小」のペアで構成させるリストから抽出したい。これを達成するために様々な方法が考えられるが、本研究では「1人も意見が割れていない範囲」を大小の閾値として定めた。例えば表2では、167cm以上は全員が「高い」範囲だと捉えているが、165cmで「低い」と捉える人が表れるため、背が高いとみなせる範囲は167cm以上とする。もし意見が割れていない範囲をう

まく求められなかった場合（一致するものがない場合など）は、「普通」と判定する。

4 評価実験

4.1 実験設定

提案手法を評価するため、提案手法による大小判定と人手による判定が一致するかどうかを比較する評価実験を行った。評価用データとしてはウェブページから数量表現を含む文を抽出したものをを用いた。ただし、大小情報抽出の対象としなかった同格型、非修飾節の場合の数量表現は対象外とした。また、知識が十分に獲得できなかった連体修飾型についても今回は評価を行わなかった（次節で詳しく述べる）。

人手の判定においては、システムの出力と同じ「大きい」「小さい」「普通」の3値に「どちらかと言えば大きい」「どちらかと言えば小さい」「判定不能」を加えた計6値を、それぞれの問題文にアノテートした。そもそも数量には大小判定が微妙な範囲が存在すると考えられるため、「どちらかと言えば大きい/小さい」という値を設定するのは妥当である。本来であればシステムもこの値を出力させるべきでだが、適切なアルゴリズムを設計できなかったため、システムは3段階、人手では5段階の大小評価を行う。精度を評価する際は、「どちらかと言えば大きい(小さい)」は「普通」「大きい(小さい)」のどちらかと一致すれば正解とした。

また、「判定不能」は大小判定ができない場合で、これは評価対象としない。「判定不能」に該当する場合は以下の通りである。

- 大小が文脈に強く依存する場合：「2000円払った」という文は「何に払ったのか」という文中に書かれていない情報に「2000円」の大小が強く依存するため、評価ができない。
- 意味不明な文章
- 数量表現抽出ツールや係り受け解析器のミス

なお、評価を行う文中に既に「も」「しか」がついてる場合があったが、大小判定の際にこれを手がかりとして用いることはしない。

4.2 実験準備

実験準備として、ウェブページ約1500万文書から大小知識の抽出を行った。この際「も」「しか」などの大小の手がかりとなる語句がどれだけ表れるかを調査するため、大小を示す語句がつかない数量表現についても獲得対象とした。表3が獲得された知識の内訳である。連体修飾型は連用修飾型に比べ出現頻度が高かったが、大小の手がかりとなる表現を伴って表れる場合は非常に少なかった。連体修飾型については知識が十

表3: 連体修飾型、連用修飾型それぞれの抽出数

	連体修飾型	連用修飾型
「も」「もの」	87,827 (0.7 %)	346,826 (8.2 %)
「しか」「わずか」	11,825 (0.1 %)	251,820 (5.9 %)
何も伴わない	11,785,269 (99.2 %)	3,642,460 (85.9 %)
計	11,884,921	4,241,106

分に獲得できなかったため、今回は連用修飾型のみを対象として評価を行う。

また、正解データを1人のアノテータに作成してもらった。アノテートした文の数は1000問で、このうち判定不能とされた文の数は392文だった。評価実験ではこれ以外の608文について評価を行う。

4.3 実験結果、エラー分析

評価実験の結果は以下ようになった。上が「どちらかと言うと大きい/小さい」を含めた評価、下は含めなかった評価である。

表4: 実験結果

	適合率	再現率	F値
普通+	0.90(408/452)	0.79(408/516)	0.85
大きい+	0.63(42/67)	0.51(42/82)	0.57
小さい+	0.07(6/89)	0.60(6/10)	0.34
普通	0.86(266/310)	0.72(266/369)	0.79
大きい	0.41(15/37)	0.27(15/55)	0.34
小さい	0.01(1/82)	0.2(1/5)	0.11

実験結果を分析する。出力例を表5に示す。以下では誤りの主な原因として挙げられた「知識の不足」「数量が示す対象の抽出の失敗」の2つについて、また「小さい」の適合率が低くなったことについて分析を行う。

数量が示す対象の解析に失敗 計 57 / 152 事例

解析に失敗した多くの事例は述語項構造解析のミスであった。事例5のような基本的と思われる述語項構造解析から、非常に解析が難しい言語現象(事例6)まで幅広く見られた。

また項のヘッドだけではなく句全体を見る必要がある事例(事例7)や係り先の動詞の係り先を見る必要がある事例、更に漠然とした状況といったものを捉える必要がある事例(事例8)なども見られた。数量が示す対象の表し方そのものを再度設計する必要がある。

大小知識の不足 計 24 / 152 事例

判定を行う際、獲得された知識の中に入力文の数量が示す対象と同じものが1つも存在しない場合は322事例(全体の53%)あった(この際の判定は自動的に「普通」が出力される)。大小知識を更に増やす以外の解決策としては、数量が示す対象の粒度を粗くすること、数量が示す対象をより一般化して捉えること、などが考えられる。

表 5: 出力例と誤り分析

	ラベル	正解	文章	誤りの種類	分析
1	大きい	大きい	今日は 3 2°C以上 あった !?	(正例)	-
2	大きい	大きい	週 2、3 日働いて 1 0 0 0 万円 稼げる	(正例)	-
3	大きい	大きい	緑モス限定のスープで、単品で 4 6 0 円 するのだが、それも注文してみた。	(正例)	-
4	小さい	小さい	こんなの作れるのは世界中で <u>三人</u> しか いません。	(正例)	-
5	小さい	大きい	<u>十数人</u> 来たために 8 畳程の部屋は いっぱい になった。	対象の解析に失敗	述語項構造解析の失敗。対象を「来る人数」と解析してしまっただけ。「来る」の二格として「8 畳程の部屋に」を抽出し、「8 畳程の部屋に来る人数」と解析する必要がある。
6	小さい	普通	しかし、来週の新作が「X-MEN」とか、「バック・ダンサーズ」たらいうシロモノらしいので、ここはまとめて <u>二本</u> 観ておいた方がよからう。	対象の解析に失敗	述語項構造解析の失敗。対象を「観る本数」と解析してしまっただけ。「観る」のヲ格として「映画を」を抽出し、「映画を見る本数」と捉える必要がある(ヲ格の抽出には前半部で映画の話をしていることを推論する必要がある例)。
7	小さい	普通	ちょっと前に彼氏と別れた友達が 2 人 いる んだけど、今、そのコたちの恋愛進行 がとても楽しみです	対象の解析に失敗	項情報の詳細の欠落。対象を「友達の数」と解析してしまっただけ。「友達」を修飾している節を考慮し「ちょっと前に彼氏と別れた友達の数」として捉える必要。
8	小さい	普通	カルビを <u>一口</u> 食べさせてもらって料理 長 Y さんに感謝しました (笑)	対象の解析に失敗	状況推定の必要。対象を「カルビを口にした回数」と解析し、知識中に「この間の焼き肉ではカルビを一口しか食べられなかった」という文章から獲得した知識(一口=小さい)があつたため、判定を誤った。すなわち「(味見するような状況での) カルビを口にした回数」と「(普通は複数回食べる状況での) カルビを口にした回数」を区別する必要がある。
9	普通	小さい	歯科衛生士を <u>一人</u> おくだけで、無資格の 助手を置いていて、助手が歯石除去し ていたりするケース。	知識不足	対象を「歯科衛生士をおく人数」と解析したが、知識中に一致するものが 1 つもなかった。
10	普通	大きい	この会社は面接が <u>4 回</u> もあつたので、何 度も面接の練習をしていただきました。	知識不足	対象を「この会社の面接の回数」と解析したが、知識中に一致するものが 1 つもなかった。しかし「面接の回数」の知識はあつたため、対象をやや粗く捉えることで、正しく判定できる可能性がある。
11	普通	大きい	昔は一人で映画なんて寂しくてイヤだっ たけど、OD 1 を <u>2 5 回</u> 見に行った時は さすがに誰も付き合ってくれず、以来 一人で映画もすっかり平気になりました。	知識不足	対象を「OD 1 を見に行った回数」と解析したが、知識中に一致するものが 1 つもなかった。しかし「映画を見に行った回数」の知識はあつたため、「OD 1」を「映画」として一般化して捉えることで、正しく判定できる可能性がある。

「小さい」の適合率について

誤って「小さい」と判定した事例の 9 割は「一度」「一回」「一つ」など、値が 1 である数量である。値が 1 である数量は入力文中で出現頻度が非常に高く (34%)、かつほとんどが「普通は 1 であるような対象の数量」(「一度やってみる」など) であつた。これは事例 8 で挙げたように「状況」などの難しい文脈を捉える必要がある問題であり、結果的にほとんど正解できず、精度が落ちてしまった。「大きい」に比べ「小さい」の判定はシビアであり難しい。

5 おわりに

本稿では数量の解釈に向けた試みとして、数量の大小知識を自動抽出する手法と、それを用いた大小判定の手法を提案した。評価実験の結果、「小さい」タグの判定に難があるものの、大小判定の第一歩としてはまずまずの精度が得られた。

今後の課題としては主に 3 つ挙げられる。1 つは数量が示す対象をより詳しく捉える事である。これは基本的には述語項構造解析の精度向上により達成されるが、「状況」のような非常に漠然とした情報を捉える必

要性もあり非常に難しい課題と言える。2 つ目は大小判定手法の洗練である。使用する知識の選び方に関して、数量を示す対象の粒度が細かくなればなるほど知識は疎になっていくため、これを解決するために知識の一般化などの処理が重要になっていくものと考えられる。3 つ目は評価実験を更に大規模に行うことである。本研究では「主観」という人によって差があるものを扱うため、複数人の評価者に手法を評価してもらうことは必須である。

参考文献

- [1] M. Sammons, V.G. Vydishwaran, and D. Roth. Ask not what textual entailment can do for you... In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1199–1208. Association for Computational Linguistics, 2010.
- [2] P. LoBue and A. Yates. Types of common-sense knowledge needed for recognizing textual entailment. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pp. 329–334. Association for Computational Linguistics, 2011.
- [3] H. Ji, R. Grishman, H.T. Dang, K. Griffitt, and J. Ellis. The sixth pascal recognizing textual entailment challenge. In *Proceedings of the Third Text Analysis Conference (TAC 2010) November*, 2010.
- [4] 成澤克麻, 渡邊陽太郎, 水野淳太, 岡崎直観, 乾健太郎. 数量表現を伴う文における含意関係認識の課題分析. 言語処理学会第 18 回年次大会発表論文集, 2012.
- [5] 日本語記述文法研究会. 現代日本語文法 2 第 3 部格と構文 第 4 部ヴォイス. くろしお出版, 11 2009.