

説明対象と関連語に対する説明のための類推における既知概念の抽出 Source concept extraction in analogy for explanations to target and related words

高橋 凌, 寺井 あすか
Ryo Takahashi, Asuka Terai
公立はこだて未来大学
Future University Hakodate
g2122039@fun.ac.jp

概要

未知の事柄に対する理解を促すには、既知っている概念と結び付ける類推による説明が有効である。本研究では、特に関連性のある2つの概念を説明対象として、それらを説明するための適切な類推表現を得ることを目指し、類推に用いる既知の概念の抽出を目的とする。類推のための概念抽出には格フレームを用いる。説明対象と共通する格フレームを持つ語を候補群として、使用される頻度を考慮し順位付けを行う。また、提案手法に対して評価実験を行い、その妥当性を検証した。

キーワード：類推, 比喻, 格フレーム

1. はじめに

私たちは文章を読むことによってさまざまな知識を得られる。しかし、その中で意味の分からない言葉や理解の難しい概念に遭遇することがしばしばあり、文章を理解するうえで大きな妨げとなる。単純にそれらの意味を調べることもできるが、意味の説明自体が難しい場合、理解に時間がかかってしまう。

その解決策の一つとして、未知の言葉や表現に既知の概念を対応させ、それに喩えてわかりやすく言い換えるという方法が挙げられる。例えば、中学生に対して電気回路における電流のふるまいを説明する際は、身近でその仕組みが既知である水の流れに喩えることで理解を促すなどの工夫がなされている。このように既知の事例と対応づけて、直面する新奇な状況に関する考え方を得ることを類推という。類推を用いた説明を行うことで、より柔軟かつ迅速に新たな知識の獲得が可能になる。

また、複数の単語が共通の概念で結びついており、強い関連性を持つことがある。そういったとき一つの単語のみを説明の対象とするのでは不十分である。例えば、電流のふるまいは水の流れに喩えられるが、「電流は水のような」という喩えを生成するだけでは

なく、電流における電圧や抵抗をその四項類推に基づき説明することが望ましいと考えられる。四項類推とは、 $(a : b)$ と $(c : d)$ の対からなる4つの単語や概念における a, b, c のみが与えられたとき、 $(a : b)$ の関係と $(c : d)$ の関係から d を推定するというものである。例えば、(電流 : 電圧) という関係に対して、電流に当たるものとして水が与えられたとき、電圧に当たるのは水圧である、という形式で類推を行う。このように、四項類推は単語間の関係性を表現することに適している。そこで本研究では、ある説明対象とそれに関連する語を説明する四項類推生成システムの構築を目指し、類推に用いられる既知概念の抽出として説明対象を喩える語の抽出を目的とする。

先行研究として、近村ら [1] は、説明対象の単語に対して共起する述語表現を利用し、喩える語を抽出・評価する手法を提案した。しかし、述語表現利用の有効性は示されているが、手法に関する詳細な検討が行われておらず、改良する余地があると考えられる。同様に述語項構造を用いた類推に関する研究として、Mason [2] が提案した、WordNet を利用し入力で指定された2つの概念間の比喩的關係を発見するシステムや、Shutova ら [3] の、文章を入力として受け取りその文脈を考慮して喩えを用いた動詞の言い換えをするシステムが挙げられる。しかし、前者は入力の時点で対応させる2つの概念(例: 電気, 水)を指定する必要があり、後者は比喩理解に基づき入力の文章の意味が変わらないように別の表現を見つけることが目的である。そこで、本研究では、近村ら [1] の手法を改良したシステムの提案を行う。

2. 提案手法

2.1 システム概要

類推の生成は2つの段階に分けて行う。まず、入力として「電流」と「電圧」のような関連性のある組合

せの単語が与えられる。次に、説明対象である「電流」を説明する既知の概念（喩えの語）として「水」などの候補単語を抽出する。その後、入力単語と抽出した候補単語の組合せから四項類推を生成する。本稿では類推に用いる既知概念である喩えの語の抽出方法について検討を行った。

2.2 類推に用いる概念の抽出

類推に用いる概念の抽出には、格フレームを用いる。格フレームとは、述語とそれが格関係をもつ語（項）を記述したものである。「電流が流れる」、「水を飲む」のように、ある語に対して頻繁に使われる述語が存在し、それらが語のもつ構造や働きを特徴づけると考えることができる。よって、共通する格フレームを多く持つ語同士は構造が似通っているという仮説を立て、これを候補単語抽出の指標に利用する。候補単語の取得には、京大格フレームコーパス [4] を用いる。京大格フレームコーパスは Web 上から収集した日本語 100 億文をもとに構築されたデータベースを有し、格フレームに対する用例の頻度も付加されているため、使用頻度を考慮した候補単語の評価が行える。京大格フレームコーパスにおいて入力単語と共通の格フレームを持つ語を候補単語とした。なお、使用頻度が極端に少ないものは特徴の表現として不十分であるという考えから、格フレームごとに頻度が 100 未満の語は候補から除く。

2.2.1 関連度計算

京大格フレームコーパスから取得した候補単語の入力単語との関連性を評価するため、近村ら [1] が提案した関連度計算を改良した。この先行システムでは候補単語に対し入力単語と同様の格フレームが存在する割合をガ格やヲ格などの格ごとに分けて計算し、その総和を関連度としている。ただし、「ノ格」の格フレームは名詞と動詞との関係よりも名詞同士の関係を示す側面が強いため除外している。(1) 式に計算式を示す。

$$r0_{ij} = \sum_{n \in N} \frac{v_{ijn}}{u_{in}} \quad (1)$$

ここで、 $r0_{ij}$ は入力単語 i における候補単語 j の関連度、 N はノ格を除いたすべての格、 u_{in} はある格 n における入力単語 i で使われる格フレームの数、 v_{ijn} はある格 n における入力単語 i と候補単語 j で共通して使われる格フレームの数である。この手法では入力

単語と候補単語で共通する格フレームの数が評価値に影響するが、京大格フレームコーパスにはデータベース構築に使用した文章内の単語と格フレームの組合せの用例が出現する頻度が登録されている。出現頻度が高い用例ほどその単語の特徴が強く表れると仮定し、これを利用すればより適切な候補単語の評価ができると考えた。そこで本システムでは、入力単語と候補単語で一致する格フレームごとに出現頻度を加味した重み付けを行い、その総和を候補単語の評価値とした。なお、「ノ格」の格フレームは先行システムと同様に評価から除外している。重み付けの方法をさらに変更した派生形も含め、計 3 種の計算方法を提案する。(2)、(3)、(4) 式にそれらの計算式を示す。

$$r1_{ij} = \sum_{m \in M} \frac{f_m(i)}{\sum_{w \in W} f_m(w)} \times \frac{f_m(j)}{\sum_{w \in W} f_m(w)} \quad (2)$$

$$r2_{ij} = \sum_{m \in M} \frac{f_m(i)}{\sum_{w \in W} f_m(w)} \times \frac{f_m(j)}{\sum_{w \in W} f_m(w)} \times \frac{1}{g_{im}} \quad (3)$$

$$r3_{ij} = \sum_{m \in M} \frac{f_m(i)}{\sum_{w \in W} f_m(w)} \times \frac{f_m(j)}{\sum_{w \in W} f_m(w)} \times \frac{f_m(i)}{\sum_{m \in M} f_m(i)} \quad (4)$$

ここで、 $r1_{ij}$ 、 $r2_{ij}$ 、 $r3_{ij}$ は入力単語 i における候補単語 j の関連度、 M は入力単語 i と候補単語 j で共通するすべての格フレーム、 W はある格フレーム m を持つすべての単語、 $f_m(w)$ はある格フレーム m におけるある単語 w の頻度、 g_{im} は入力単語 i におけるある格フレーム m の順位である。 $r1_{ij}$ が単語と格フレームの組合せの出現頻度を利用して評価値を計算しているのに加えて、 $r2_{ij}$ 、 $r3_{ij}$ は入力単語を基準とした格フレームの重要度による重み付けをしている。先行システムも含めたこれら計 4 種の性能を評価実験により比較した。

2.2.2 入力単語と同カテゴリの単語の除外

関連度計算により得られた候補単語には、入力単語と同じ系統の単語が多く見られる（「電流」であれば「電気」、「電源」など）。同じ系統のものが似通った構造を持つのは妥当だが、異なる概念に結び付けて説明するための喩えの語としては適切ではない。その対策

として、Wikipedia 日本語版 [5] の記事をもとに構築されたデータベースである DBpedia Japanese[6] を利用した。各記事にはカテゴリという項目が存在し、同カテゴリに属する記事は同じ系統となっている。入力単語の記事と同カテゴリの記事を取得し、見出しの語を候補単語から除外するようにした。

2.2.3 難しい単語の除外

本システムは入力単語に対して類推によって説明を与えるため、前提として類推に用いられる概念は入力単語よりも難しいものであってはならない。そこで、京大格フレームコーパスから得られた候補単語の中から難易度の高いものを除外する方法について述べる。

初出学年

単語の難易度の指標の一つとして教科書コーパス語彙表 [7] を使用する。このコーパスには日本の小学校、中学校、高等学校で使われている教科書の中に出現する単語が約 250 万語登録されている。各単語には「初出学年」という項目が存在し、これはその単語が教科書内で初めて出現する学年を表している。学年は小学校低学年、小学校高学年、中学校、高等学校の 4 段階に分かれており、初出学年が高い単語ほど難易度が高いとして、初出学年が入力単語以下のものを候補単語として残すようにした。

単語親密度

教科書コーパスによるフィルタリングは難易度の区分が 4 つしかないため、かなり粗い方法である。そこで、NTT の単語親密度データベース [8] を使った方法も検討する。このデータベースには、語のなじみ深さを評価実験により評定した単語親密度がおよそ 16 万語分登録されている。親密度は 1 から 7 の数値で、値が大きいほどなじみ深い語となる。なじみ深い単語ほど意味をよく理解できているため難易度が低いと判断し、入力単語よりも親密度が高いものを候補単語として残した。

さらに、単語親密度を候補単語の評価値に利用する方法も検討した。入力単語より親密度が高い候補単語の中でも、よりなじみ深いものの方が喩えの語として望ましいと考え、関連度と親密度それぞれの値の範囲を 0~1 に正規化して足し合わせた。計算式を (5) 式に示す。

$$s = r + \alpha t \quad (5)$$

ここで、 s は候補単語の評価値、 r は関連度、 t は親密度で、 α は任意のパラメータである。

3. 評価実験

3.1 事前実験

評価実験では候補単語の抽出方法の性能比較を行うが、難しい単語の除外方法や単語親密度を足し合わせる際のパラメータを考慮すると比較する数が膨大になり、評価が煩雑になる。そこで、データセットを用いて機械的に事前実験を行い、本実験で比較する対象を絞り込んだ。データセットには SAT Analogy Questions を使用した。これは実際に人に対して課された試験問題を参考に Turney ら [9] が作成した類推用のデータセットで、クエリの単語ペアの関係と同様の関係を持つものを 5 つの候補から選ぶという形式の問題が 374 個用意されている。

実験方法

SAT Analogy Questions をシステムの性能比較に利用するために、まずクエリと正解のみを抜き出し、単語がすべて名詞の問題を抽出した。これは、説明対象やそれを喩える概念は名詞となるためである。次に、本研究で提案するシステムは日本語を扱うため、すべての単語を日本語へと翻訳した。翻訳した際に使用することが難しいと判断した問題を除外し、最終的に 189 個の問題が残った。各問題のクエリと正解それぞれの単語ペアの右側同士、左側同士を説明対象と喩えの語のペアとして、説明対象に対する候補単語の抽出を行って抽出方法の評価値を計算した。評価値 score の計算式を (6) 式に示す。

$$\text{score} = \sum_{e \in E} h_e \quad (6)$$

ここで、 E はすべての説明対象、 h_e は、説明対象 e に対する喩えの語の候補単語における順位の逆数であるが、候補単語に喩えの語が含まれていない場合は 0 とする。また、本システムは喩えの語が説明対象よりも簡単である必要があると前述したが、単語ペア同士が同じ関係を持つかを判定させるタスクのために作られた SAT Analogy Questions ではその前提は成立しないため、説明対象と喩えの語を入れ替えた場合の score も計算して合計した。

実験結果

表1にSAT Analogy Questionsによる事前実験の結果を示す。難しい単語の除外方法は「フィルターなし」,「初出学年」,「親密度」,「初出学年と親密度」の4パターンで、親密度を加算する場合は初出学年と親密度の併用に固定し、パラメータ α は0.25, 0.5, 0.75, 1.0の4種類を設定した。

表1 事前実験の結果

	フィルターなし	初出学年
r_0	18.20	16.78
r_1	21.17	20.47
r_2	18.81	18.66
r_3	19.99	19.61
	親密度	初出学年と親密度
r_0	15.68	16.59
r_1	14.76	16.31
r_2	12.44	14.18
r_3	14.75	15.87
	親密度加算 ($\alpha = .25$)	親密度加算 ($\alpha = .5$)
r_0	16.75	16.81
r_1	16.31	15.51
r_2	14.18	13.93
r_3	15.42	15.33
	親密度加算 ($\alpha = .75$)	親密度加算 ($\alpha = 1.0$)
r_0	16.97	17.03
r_1	16.09	16.09
r_2	13.85	13.78
r_3	15.17	15.23

すべての条件において、 r_1 が r_2 , r_3 よりも高い評価値を示した。よって本実験では、先行研究の手法である r_0 と提案手法 r_1 の比較を行う。フィルタリングの条件についてはフィルターなしを除いた3つを比較する。また、親密度を関連度に加算してパラメータを変更しても、 r_0 と r_1 の間の評価値の傾向がほとんど変わらなかったため、ただ足し合わせただけの($\alpha = 1.0$)を用いる。

3.2 本実験

事前実験では1つのクエリに対して1つの正解単語しか与えられないが、類推に用いられる概念は実際には複数存在する。そのため、システムの評価で上位になった候補単語を複数提示し、人手で評価してもら

うこととした。事前実験の結果に基づき、比較対象を絞って本実験を行った。

実験方法

本実験はアンケート方式とし、クエリに対してシステムが出力した喩えの語を「非常に適切である」、「適切である」、「どちらでもない」、「適切でない」、「全く適切でない」、「喩えの語の意味が分からない」の6つの選択肢から評価してもらった。あるクエリに対して比較対象となる8通りの各条件ごとに上位30個喩えの語を取得した。クエリを9個設定し、各クエリから取得された喩えの語を8分割することで8種類の設問セットを作成した。1セットあたりの設問数は90程度であった。アンケートを作成した後、7人の男子大学生・大学院生に対して1セット分の予備実験を行い所要時間などに問題がないことを確認した後、クラウドソーシング仕事依頼サイトLancers[10]にて参加者を募集した。

実験結果

8つの各セットについて100~109人、合計821人分の回答が得られた。6つの選択肢に対して以下のように数値を割り当て、喩えの語それぞれの評定の平均値を算出した。

- 5: 非常に適切である
- 4: 適切である
- 3: どちらでもない
- 2: 適切でない
- 1: 全く適切でない
- 1: 喩えの語の意味が分からない

選択肢の「喩えの語の意味が分からない」が選ばれた場合、喩えの語として適切でないと言えるので、「全く適切でない」と共通で1を割り当てている。

閾値を3.0に設定し、平均値が閾値を上回ったものを正解単語とみなして、各条件ごとに評価値を計算した。評価値は事前実験と同様、正解単語の順位の逆数を合計している。ただし、1つのクエリに対して複数の正解単語が存在する点が異なる。表2に結果を示す。

最も評価値が高いのは提案手法と初出学年の組合せだった。親密度加算を除いた他2つの条件でも提案システムが先行システムを上回る結果となった。

また、各喩えの語に対して「喩えの語の意味が分からない」が選ばれた割合を計算し、それが5割を超え

表 2 本実験の結果

	初出学年	親密度	初出学年と親密度	親密度加算
先行システム	15.22	14.96	13.94	14.57
提案システム	18.42	17.32	16.36	13.98

た単語の数をフィルタリング方法ごとに集計して性能比較に用いようと考えたが、該当する単語は存在しなかった。

さらに、関連度計算の方法とフィルタリング方法によって喩えの語の評定値が異なるか、それぞれの喩えの語の平均評定値をサンプルとして統計的検定を実施した。各条件における評定値の平均と検定結果を図1に示す。なお、エラーバーは標準誤差を示している。

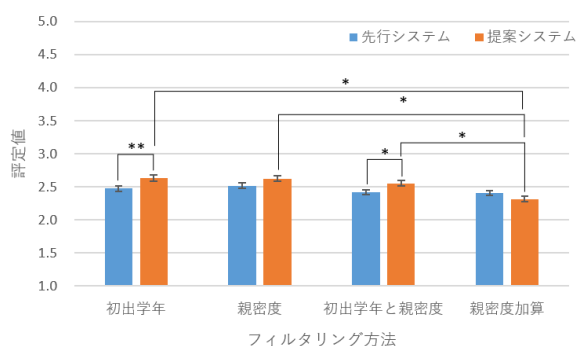


図 1 評定値の平均と検定結果

まず、被験者間 2 要因分散分析を行った結果、関連度計算の主効果は有意で ($F(1, 2152) = 6.44, p < .05$), フィルタリングの主効果も有意となり ($F(3, 2152) = 10.47, p < .05$), さらに交互作用も有意だった ($F(3, 2152) = 3.69, p < .05$).

交互作用が有意だったため、次に単純主効果の検定を行った。関連度計算の単純主効果は、初出学年で有意 ($F(1, 2152) = 6.71, p < .05$), 親密度では有意傾向 ($F(1, 2152) = 3.34, p = .068$), 初出学年と親密度の併用では有意で ($F(1, 2152) = 4.98, p < .05$), 親密度加算のみ有意ではなかった ($F(1, 2152) = 2.48, p = .115$). 図 1 からわかるように、親密度加算を除くフィルタリング方法では提案システムの方が先行システムより評定値が高くなった。フィルタリングの単純主効果は、先行システムでは有意ではなかったが ($F(3, 2152) = 1.51, p = .211$), 提案システムでは有意だった ($F(3, 2152) = 12.65, p < .05$).

提案システムでのフィルタリングの効果を詳しく検討するため、ライアン法による多重比較を行った。結果を表 3 に示す。

親密度加算と他 3 つの方法それぞれとの比較ではすべて有意な差があったが、親密度加算を含まない組合せは有意ではなかった。よって、提案システムにおいては親密度加算のみ評定値が低くなっていることが分かる。

4. 考察

まず、関連度計算について考察する。提案手法の中で最も性能が高いのは $r1$ で、次に $r3$, 最も低いのは $r2$ であった。 $r2$ と $r3$ は $r1$ をベースに、対象の格フレームを入力単語における重要度で重み付けしている。計算方法を考案した時点では、入力単語をより重視した方が適切に喩えの語を評価できるのではないかと考えたが、実際はそうではなかった。本システムでは入力単語と候補単語が共通した特徴を持つことを求められるので、入力単語に偏り過ぎると候補単語の特徴が反映されにくくなると考えられる。また、 $r2$ よりも $r3$ の方が性能が高いのは、順位の逆数よりも出現頻度で重み付けした方が重要度を詳細に反映しているからだと考えられる。先行システムと提案システムを比較すると提案システムの方が良い結果となったが、やはり単に共通する格フレームの割合で評価するよりも、出現頻度を考慮した方がより適切だと考える。出現頻度の高い用例はその単語の特徴を強く表すと考えられるので、入力単語と候補単語両方で出現頻度が高い格フレームを重視することが重要である。

フィルタリング方法に関しては、そこまで大きな差はないが初出学年が最も良い結果となった。学校の教科書で実際に使われている単語が対象となるので、多くの人にとって分かりやすい説明を提供できると考えられる。初出学年と親密度の併用では、2 つのコーパスで共通する単語でなければ候補単語に残らないので、有効な候補単語が除かれてしまうリスクが高まる。また、先行システムのみ親密度加算で評価値が上がったのは、京大格フレームの出現頻度と単語親密度の間に相関関係があるからではないかと考える。出現頻度が高ければなじみ深さを示す親密度も高くなるとすると、関連度の算出に出現頻度を使用していない先行システムに親密度を足すことでなじみ深い単語の評価が上がって評価値が上がるが、対して提案システムは出現頻度を関連度計算に利用しているので最初から

表 3 多重比較の結果

比較ペア	名義的有意水準	<i>t</i>	<i>p</i>
初出学年 - 親密度加算	0.0083	5.34	0.0000
初出学年 - 初出学年と親密度	0.0125	1.32	0.1880
親密度 - 親密度加算	0.0125	5.27	0.0000
初出学年 - 親密度	0.0250	0.08	0.9398
親密度 - 初出学年と親密度	0.0250	1.24	0.2146
初出学年と親密度 - 親密度加算	0.0250	4.03	0.0001

親密度の高い単語が上位に来ており、親密度を加算すると逆に関連度の低い単語の評価が上がってしまって評価値が下がると考えられる。よって、初出学年単体での使用が望ましいと考える。

5. まとめと展望

本研究では、ある説明対象とそれに関連する語に対して、関係性を説明するための類推の生成を行うシステムの構築を目指し、類推に用いる既知の概念（喩えの語）を自動的に抽出する方法を提案した。先行システムと提案システムを比較する評価実験を実施した結果、先行システムを上回る性能を示し、類推に用いる概念の抽出方法として有効であることが分かった。難しい単語を除外する方法に関しても検証を行い、より高い性能を発揮する組合せを特定した。

今後の展望としては、システムの次の段階として四項類推に関する検証を行うことを考えている。現状では単語埋め込みベクトルによる単語のベクトル化を用いた実装を行っている。入力単語ペアと提案システムにより抽出された既知の概念（喩えの語）を用いた出力例を表4に示す。単語埋め込みベクトルのモデルには学習済みの fasttext[11] を使用している。

表 4 四項類推の例

入力単語	類推結果
電流：電圧	血液：血圧
自動車：アクセル	バイク：ペダル
教師：生徒	医者：患者
王様：女王	男：女

この結果では入力単語の組の関係を説明するものとして妥当と考えられるものを人手で選出しているが、システムが自動で選出できるように評価指標を検討する必要がある。

謝辞

本研究は JSPS 科研費 (19K12121) の助成を受けたものである。

文献

- [1] 近村亮一, 荒木健治, (2011) "Web データを用いた単語に対する比喩的説明文収集手法", 日本知能情報ファジィ学会 講演論文集 第 27 回ファジィシステムシンポジウム, pp.112 - 115.
- [2] Mason, Z.J, (2004) "CorMet: A computational, corpus-based conventional metaphor extraction system", Computational Linguistics, 30(1), pp.23-44.
- [3] Shutova, E. et al, (2013) "Statistical metaphor processing", Computational Linguistics, 39(2), pp.301-353.
- [4] 河原大輔, 黒橋禎夫, (2006), "高性能計算環境を用いた Web からの大規模格フレーム構築", 情報処理学会 自然言語処理研究会 171-12, pp.67-73.
- [5] Wikipedia 日本語版, <https://ja.wikipedia.org/wiki/>
- [6] DBpedia Japanese, <https://ja.dbpedia.org/>
- [7] 特定領域研究「日本語コーパス」言語政策班, (2011) "教科書コーパス語彙表".
- [8] NTT 印刷, (2021) "令和版単語親密度データベース".
- [9] Peter D. Turney, Michael L. Littman, Jeffrey Bigham, and Victor Shnayder, (2003) "Combining independent modules in lexical multiple-choice problems", In Recent Advances in Natural Language Processing III, pp.101 - 110.
- [10] Lancers, <https://www.lancers.jp/>
- [11] Bojanowski, Piotr, et al, (2017) "Enriching word vectors with subword information", Transactions of the Association for Computational Linguistics 5, pp.135-146.