5Q-08

言葉の連想に基づいたなぞなぞ文の自動生成

長谷川　凱　　寺岡　丈博
拓殖大学工学部情報工学科

1　はじめに

コンピュータによる文生成の技術は, 身近なもので小説文の自動生成や音声対話システムなどに使われている. しかし, 文の自動生成は意味が通じないことや, 文脈が間違っているなど, 人間のようにうまく作成できない問題がある. このような背景の下, 本研究では, なぞなぞ文の自動生成を目的とする. なぞなぞは言葉遊びの一種であるが, 問題と解答の意味的な関係性を正しく反映させる必要がある. そのため文生成に関する基礎技術の 1 つであり, 今日までに, なぞなぞに対する面白さや問題の納得感などを重視した研究 [1] や, ほかの言葉遊びとして駄洒落を取り入れた研究 [2] などがある.

本研究では, 従来の研究で質問文を生成するために人手で作成していた因果関係情報を自動で抽出し, なぞなぞ文を生成する. そして, 回答者の問題に対する納得感を調べることで, 生成されたなぞなぞ文がどれくらい人手で作成されたものと近いかを評価する.

2　関連研究

濱田・鬼沢の研究 [1] では, 人が作るような面白いなぞなぞの生成を行っている. なぞなぞの生成方法は動詞と因果関係データベースを参照し, それから「C を A すると C は B しますが, では A しても A しても B しないものはなんでしょう？」のテンプレート文に単語を当てはめて作成されている. そして, その際に利用する因果関係データベースを自動的に構築する方法が今後の課題として挙げられている.

金久保の研究 [2] では, 動詞の同音異義語を用いることで駄洒落を基盤としたなぞなぞの生成を成立させている. なぞなぞの生成は任意の単語を入力すると, 駄洒落になる別の語を検索し, 双方の上位概念を辿って両方を含む共通文型から作成される仕組みとなっている.

Automatic Generation of Riddles using Word Associations
Kai HASEGAWA and Takehiro TERAOKA
Department of Computer Science, Faculty of Engineering, Takushoku University

3　提案手法

3.1　手法の概要

なぞなぞ文は問題文に対する答えとセットで生成しなければならない. 問題文は,「<A1>は<B1>すると<C1>ですが, では<B1>は,<B1>でも,<C2>なものは？」と「<B1>と<C1>で、かつ<B2>と<C2>ものは?」のようなテンプレート文の言葉を当てはめることで生成する. その際に用いるデータとして因果関係データを使用する. 図 1 のように, まず動詞と名詞の連想概念辞書 [3,4] から複数の単語からなる因果関係情報を自動で抽出する. 次に,Word2Vec[5] による単語の分散表現から単語間の類似度を利用して, 因果関係が成り立っていないデータを取り除く. そして同音異義語を元にテンプレート文を使用し, なぞなぞを生成する. 本研究では, 連想のされやすさや, 分散表現から得られる類似度を用いて, 問題に対する解答の意外性向上を図る.

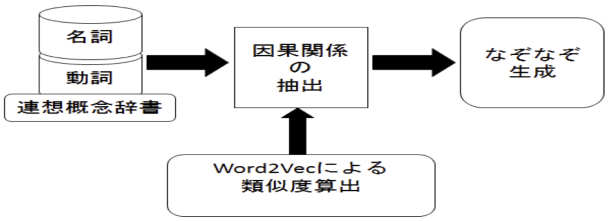

図 1: 手法の概要図

3.2　因果関係の抽出

因果関係データを図 2 のように動詞連想概念辞書と名詞連想概念辞書の連想課題の「対象」,「属性概念」から単語を抽出することで作成した.「対象」は動詞の対象を表し, 例えば「飼う」という動詞に対して「犬」などが記載されている.「属性概念」は名詞の属性にあたるもので, 例えば「犬」という名詞に対して「かわいい」などが記載されている.

なお, 因果関係が取れていないデータを取り除くために単語間の連想距離を使用した. その際に各連想概念辞書から引用した単語に関しては単語間の近さを表す連想距離が使用できるが, 名詞連想概念辞書の属性概念から引用した単語と動詞の単語間に関しては連想距離のデータがない. そこで Word2Vec による単語の分散表現から類似度を抽出し, 各連想距離と組み合わせて, 下

記のように因果関係スコア (Score) を計算する.

$$Score = (\frac{1}{d_n} + \frac{1}{d_n} + S_w) \times \frac{1}{3}$$

d_n は名詞の連想距離,d_v は動詞の連想距離,S_w は Word2Vec の類似度を表している. 数値が一定数以下 (Score<0.2) であれば因果関係の取れていないデータとみなしてデータを整理する. これにより質の高い因果関係データを作成する.

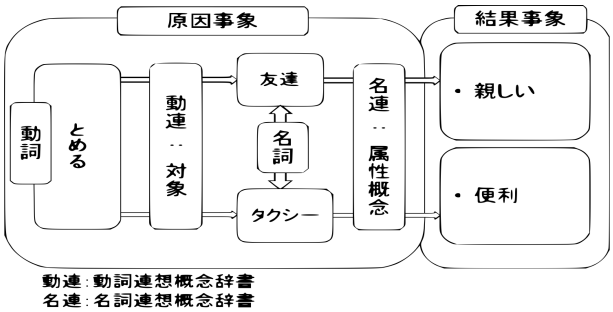

図 2: 因果関係データ作成図

3.3 なぞなぞ生成

なぞなぞの文は 2 タイプあり,Type 1 は動詞の同音異義語を使用して作成する. 因果関係データから問題に使われるデータを検索し, 見つかったデータの動詞に対して同音異義語となるデータを検索する. この 2 つの因果関係データを使用してなぞなぞの文生成を行う. Type 1 のテンプレートは先行研究を参考にしているが, 答えに使用している因果関係データの結果事象を問題文にも用いることで, 答えに該当する単語をより類推しやすくしている. もう一方の Type 2 は, 答えとなる名詞が同じである因果関係データを 2 つ用いることで答えを連想しやすくしたものである. 図 3 は, 生成されたなぞなぞの一部を表している.

```
・Type1
[メダル, 取る, 軽い, とる,0.239012668]
[花, 撮る, きれい, とる,0.399014543]
Q,<メダル>は<とる>と<軽い>ですが、では<とる>は、<
とる>でも<きれい>なものは？
A, 花

・Type2
[いす, 触る, かたい, さわる,0.340451071]
[いす, 下げる, 低い, さげる,0.322164077]
Q,<触る>と<かたい>で、かつ<下げる>と<低い>ものは?
A, いす
```

図 3: 自動生成したなぞなぞ文

4 評価と考察

生成されたなぞなぞの中から 10 問選出し,10 人の学生（18 歳〜22 歳）を回答者としてアンケートを実施した。アンケートは問題に対する「難易度」「意外性」「納得感」「面白さ」「人手かコンピュータのどちらで作成されたか」を調査した. その結果, 全体的には一定の評価を得ることができたが, 一方で「問題から読み取れる解答範囲が広いため答えが類推しにくい」や「答えへの納得感が薄い」などの意見が見られた. これは問題生成の際に使用される 2 つの因果関係データが同音異議語で一致したものだけで問題を作成したため, 答えに使用される因果関係の結果事象から連想できるものが多いデータが選ばれてしまったためである. この点を改善するためには, 同音異義語の一致だけでなく, 答えに使用する因果関係データの結果事象から最も連想距離の近い名詞のあるデータを選ぶことで答えをより連想しやすくなると考えられる.

5 今後の課題

Score の値が高いデータの中には, 因果関係が曖昧なデータも見られる. そのため今後の課題としては, これらのデータの整理とともに因果関係を判断する基準となる Score の検討が求められる.

参考文献

[1] 濱田真樹・鬼沢武久 : 同音異義語の意味の多様性を構造に持つなぞなぞの生成, 知識と情報, 日本知能情報ファジィ学会誌, Vol.20, No5, pp.696-708 (2008)

[2] 金久保正明 : 一般的な概念辞書を用いたなぞなぞ質問文生成システム, 情報処理学会第 73 回全国大会講演論文集 (2), pp.73-74（2011）

[3] T. Teraoka, J. Okamoto, and S. Ishizaki, “An Associative Concept Dictionary for Verbs and its Application to Elliptical Word Estimation”, In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pp.3851-3856 (2010)

[4] 岡本潤・石崎俊 : 概念間距離の定式化と既存電子化辞書との比較, 自然言語処理, Vol.8, No.4, pp.37-54 (2001)

[5] T.Mikolov, K.Chen,G.Corrado, and J.Dean, “Efficient Estimation of Word Representations in Vector Space”, CoRR, abs/1201.3781 (2013)