

永松健司 田中英彦†
 東京大学 工学部

1 はじめに

文章の理解過程において、読み手がその文章のどこに重要性を見出すかは、興味・目的に大きく影響される。つまり、要約処理等を考える場合には、利用者の興味・目的に応じて内容を変更できる方が質の高い処理を実現できると言える。我々の研究では、読み手が解釈に先だって踏まえているそのような立脚点を大きく“視点”と呼ぶこととし、テキストの理解に対して重要な前提として機能するものと見做している。

しかし、そのような視点に基づいた要約文生成を行う場合、最も基本的な処理として、文章のそれぞれの意味構造間に視点情報を前提とした距離関係を導入する必要がある。本稿では、そのような距離関係を導入する場合に必要な条件を考察し、それに基づいた距離関係を定義する。かつ、その定義により処理を行った場合の結果を評価すると共に、視点情報を利用することの有効性を考察する。

2 視点情報を前提とした意味構造間距離

[1]では、視点情報に求められる効果として、

- 概念間の距離の算出において、視点情報を前提とした経路が採用されること
 ⇒ 視点に依る概念距離関係の変化

というものを考え、視点情報を前提とした概念間の距離関係について考察した。

本稿では、これを格フレーム間の距離へと拡張する。

2.1 距離関係の種類と要請される条件

最初に、自然言語テキストの意味構造間に定義しうる距離を以下のように分類する。

- 概念的距離
 - テキストを構成する個々の単語概念どうしに定義される距離に基づいて、その単語が構成する意味構造間で定義される距離。
 - シソーラスによって算出される。
- 知識的距離

*A definition of the distance between two semantic structures and an evaluation of it

†Kenji Nagamatsu Hidehiko Tanaka

{naga,tanaka}@mtl.t.u-tokyo.ac.jp

Faculty of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo, 113, Japan

$$\begin{pmatrix} {}_1H \\ {}_1Q_1 \\ \vdots \\ {}_1Q_n \end{pmatrix} \leftarrow \text{距離} \Rightarrow \begin{pmatrix} {}_2H \\ {}_2Q_1 \\ \vdots \\ {}_2Q_n \end{pmatrix}$$

図 1: 格フレーム内のヘッドとその修飾辞

- シソーラスでは判断のつかない、より高次の関係 (相関、因果、etc.) に基づく距離関係。
- 知識ベースによって算出される。
- 論理的距離
 - 文章の論理構造 (逆接、例示、etc.) 間の関わりの度合によって定義される距離。

より上のものほど、より基本的な距離関係であり、またより一般的に適用可能になるが、逆に得られる結果があらゆる場合で適切とは限らないというものになる。

2.2 格フレーム間の概念的距離

格フレーム間での概念的距離関係を定義するにあたって、そこに要請される条件には、次のものが考えられる。

- 同じ格フィルター同士を比較する必要
- フレーム構造で表層にあるヘッドほど重要性が高い

が考えられる (例外も考えられるが)。そこで、図 1 のようなヘッド (H)、およびその修飾辞 (Q) に対して、

$$\begin{aligned} \text{Dis}_{\text{pov}}({}_1Q_{i\dots 2}^k, {}_2Q_{j\dots}^k) = & \\ \text{Dis}_{\text{pov}}({}_1H_{1i\dots 2}^{k+1}, {}_2H_{1j\dots}^{k+1}) + & \\ \alpha \cdot \max_{m,n} \text{Dis}_{\text{pov}}({}_1Q_{mi\dots 2}^{k+1}, {}_2Q_{ni\dots}^{k+1}) & \end{aligned} \quad (1)$$

という式で格フレーム間の概念的距離を定義する。ここで、 $\text{Dis}_{\text{pov}}(c_1, c_2)$ は視点 pov を固定した際の概念 c_1 と c_2 の距離であり、[1] で定義されている。また、 ${}_iH_{lmn\dots}^k$ の k はフレーム中のサブフレーム (句構造木での部分木に相当) の深さ (表層のものを 0 とする)、 l が同じ深さのフレームの中で文の先頭から数えて自分が何番目かを、同様に m は親が何番目かを、 n は…を意味する。また、 α は深い場所にあるものほど寄与が少ないことを示すために入れた、 $0 < \alpha < 1$ の範囲の値である。

つまり、ある構造を持つ格要素間の距離は、ヘッドはヘッド同士で距離を求め (第 1 項)、修飾辞は同レベル (深さ) の修飾辞同士で全組合せで距離を取りその最大値 (第 2 項) を加えたものとなる (第 2 項は再帰的に求められる)。

2.3 知識的距離および論理的距離

この二つについては、現状ではまだ詳細が固まっていないために割愛する。

3 重要度の付与処理

実際の実験では、上述の距離関係を用いて、テキストのそれぞれの文章に対して重要度を付与し、その結果が妥当かどうかを判断する。

重要度の算出は [1] で行ったものと同様の構成を採るが、処理で用いる距離関係に個々の概念間の距離ではなく、上述の意味構造に対する距離関係を用いているところが異なっている。その過程は、次のようになる。

1. テキスト全体での中心となる意味構造 (格フレーム) の決定 (これを中心概念と呼ぶことにする)
2. 各々の文を構成する意味構造と中心概念との距離を算出
3. 同じ文内にある意味構造と中心概念との距離の最小値をその文と中心概念との距離と見做す

そして、平均距離の近い文ほど、テキスト内での重要度が高いものとする (今回の実験では、重要度を平均距離の逆数で定義した)。

中心概念の決定は [1] と同様 (ただし、対象が個々の概念ではなく、格フレームとなっている) の手法を採用している。

4 評価実験

4.1 実験方法について

以上のような文の意味構造間の距離の定義に基づいて、[1] で行った、視点情報を前提とした文に対する重要度の付与処理を試みる。

1. 入力テキストには天声人語などのコラムや社説を使用し、視点情報を与えない場合と与えた場合に、それぞれの格フレームに対して付与される重要度を求める。ただし、現状では解析は手作業で行い、概念の区分は EDR 電子化辞書 ver.1.0 の概念辞書に基づいている (図 2)。ただし、辞書中に存在しない固有名詞や数詞などは一般的な語で置き換えたり、省略したりしている。
2. 次に、同じテキストで、1. で分割したフレームごとくに下線をつけた上で、被験者に示し、そのテキストの中心的話題となる単語を一つ (文中から) と、下線をつけた中で最も重要だと判断される文を挙げてもらう。

現在、上記の方法によって、この方法で得られる重要度と、人間が判断する重要さの違いを実験しているところである。しかし、実際の評価結果は本稿に間に合わせることができなかった。

verb	登る	10324a
agent	人	103c4d
verb	住む	3ce729
loc	埼玉	0f2d2e
object	山	3ce994
verb	ある	1fa301
loc	愛媛	0e67af

図 2: 文「埼玉県所沢市に住む辻田功さんは、一九八二年に、故郷の愛媛県にある石鎚山に登った。」 ([2]) に対する格フレームの解析例

5 考察

評価結果がまだ出ていないため、ここでは結果として予想されることを述べる。

[1] では、概念間の距離のみに依った視点情報でも、関連性の判断に対して適切な指標となりうるという結果が得られている。これに対し、今回の意味構造間の距離に基づいた方法では、

1. 距離の算出が同じ種類の格フィラどうしに限定されるため、実際は似たような概念でも役割が異なれば計算の対象から外れてしまう。しかし、その一方で無制限な計算が抑制され、処理時間の減少が期待される。
2. ルールベースに基づく処理ではないため、どのような入力に対しても結果を与えることができる。しかし、結果の妥当性は必ずしも保証されない (今回の実験でその品質を求めているところだが) ため、知識的距離、論理的距離関係などにより、補完が必要だと思われる。

6 おわりに

[1] における、視点情報を前提とした重要度付与処理を、個々の概念から文を構成する意味構造 (格フレーム) を対象とするものへと拡張した。現在はこの手法による重要度付与処理の結果を、人による判断と比較するための実験を行っているところである。

このようなシソーラスの上下関係だけを拠り所とする距離関係は、どのような入力テキストに対しても一般的に利用可能であるという利点がある一方で、より意味の深い関連性が要求される場合に必ずしも妥当な結果が得られないことが予想される。これに対しては、知識や文章の論理構造に基づく距離関係を導入し、結果を補完するような方法を考えている。

参考文献

- [1] 永松健司, 田中英彦. 文章に対する重要度付与処理における視点情報の有効性の分析. 情報処理学会全国大会, 1995.
- [2] 朝日新聞社. 天声人語. 1994年6月10日.