

Wikipedia を利用した概念辞書における属性情報の獲得と 物語自動生成ゲームでの利用

Attribute Information Acquisition in a Conceptual Dictionary by Using Wikipedia and the Use in an Automatic Narrative Generation Game

小野淳平*¹
Jumpei Ono

小方孝*²
Takashi Ogata

*¹ 岩手県立大学大学院
Graduate School of Iwate Prefectural University

*² 岩手県立大学
Iwate Prefectural University

By “attribute information”, the authors mean the information that represents the attributes, namely characteristics and properties, to be stored into each noun concept in the noun conceptual dictionary in an Integrated Narrative Generation System (INGS) that the authors have been developing. In this paper, the authors present a method for acquiring automatically attribute data using the Wikipedia by recursively analyzing the texts. A role in the INGS, especially in the story generation phase, is to concretely or in detail explain the noun elements in an event. This paper shows a real usage method in the story generation of acquired attribute data to show the effectiveness of this idea.

1. はじめに

[影山 12]は恒常的に持続する性質を記述する文を属性叙述文と呼んだ。筆者らは属性叙述文の解析により、ある概念の特徴・性質に関して記述した情報である属性情報を自動的に獲得することを試みた[小野 14]。属性情報は、統合物語生成システム(Integrated Narrative Generation System: INGS)[Ogata 16]における使用を目的として獲得した。INGS では、動的な事象を含む構造によるストーリー生成を行う。一方で、ある物事の特徴や外見の明ないし描写、ある時間における状態等の静的な内容の生成は検討中である。本稿では特に説明の挿入を行うため、ある一般名詞を説明するテキストから構造的な属性情報の獲得を試みる。獲得した属性情報は、理解しづらい概念について、受け手の理解を補助するために使用する。

2. INGS における属性情報

本稿における属性情報とは、ある物事の特徴・性質に関する記述である。以降では属性情報を属性と記述する。

INGS における属性は、名詞概念辞書に格納された要素それぞれが持つ。名詞概念辞書は、一般名詞及び固有名詞に関する概念を体系的に格納した辞書である。特に一般名詞概念の体系は、名詞概念を意味的に分類する中間概念と、実際にストーリーに現れる要素と結び付く終端概念により木構造を成す。固有名詞概念は、上記の終端概念と結び付く。ストーリーに登場する人・物・場所は、終端概念に基づき生成されたインスタンスである。一つの名詞概念は、一つ以上の属性を持つ。ある一つの名詞概念に関して、全ての属性を構造的にまとめた集合を属性フレームとする。ストーリーに現れるあらゆるインスタンスは、is-a 関係により名詞概念から継承した属性フレームに加え、その時々における状態を記録するフレームを持つ。以上の二つのフレームに記述された情報は、ある時間における状態を表現する事象や説明を生成するうえで利用する。

本稿では特に説明について考える。物語生成において説明

は、説明する内容から 2 種類のパターンに分かれる。一つはある物事について、辞書に記述されているような、一般的な事実を述べる説明であり、もう一つはある物事について、ストーリーの中で具体的にどのような状態・内容を持つのかを説明する説明である。例えばある男性に関して説明するとき、前者は「男性とは、ヒトの性別のこと。」というクラスそのものを説明し、後者は「その男性の名前は、太郎である。」というインスタンスを説明する。特に理解が難しいものは前者の説明を必要とし、ストーリーの筋を理解するためには後者の説明を必要とだと考えた。

3. 属性情報の獲得

この節では、属性情報の記述形式の説明した後に、属性情報の獲得手続きとその結果を示す。

3.1 システム内における属性情報の記述形式

ここで提案する方法は 2 節で述べた一般的な真実について述べたための属性を獲得する。提案する属性獲得システムは、ある名詞概念 X に関するテキストを入力とし、名詞概念 X に関する属性フレームを出力する。なお本稿では、ある名詞概念に関するテキストは、一つ以上の属性が記述されると考え、また複数の属性を記述する場合は、任意の助詞を基準として、任意のテーマごとに属性をまとめると仮定した。

図 1 は出力される属性フレームの記述形式である。<対象概念>はその属性フレームが対象とする名詞概念を示し、<テーマ n>は属性の種類を示す。例えば「所在地に関する属性のまとめり」、「外見に関する属性のまとめり」等である。<テーマ n に対する属性 k>は、構造化された文が入る。この属性フレームでは、利用する属性の量を調整することで、生成される文中に含まれる説明文の量を調整することができる。

```
(<対象概念>
 (<テーマ 1>
  <テーマ 1 に対する属性 1> <テーマ 1 に対する属性 2>
  ...<テーマ 1 に対する属性 n>)
 ... (<テーマ m> ...) ...)
```

図 1 属性フレームの記述形式

連絡先: 小野淳平, 岩手県立大学大学院ソフトウェア情報学研究科, 岩手県滝沢市菓子 152-52, 019-694-2000, g236m001@s.iwate-pu.ac.jp

3.2 属性情報の獲得手続き

属性フレームの構成手続きは大きく二段階に分かれる。まず属性獲得が容易な単位まで入力文を分割し、その後分割された文ごとに属性を獲得する。なお内部で行う形態素解析は、形態素解析器 MeCab (<http://taku910.github.io/mecab/>)を用いた。大まかな流れは以下の通りである

(1) 入力文の分割 ((A)句点による分割, (B)複文の分割): まず「。」を基準に入力文を分割する(処理(A))。その後、特定のパターン(「で」、「だが」、「であるが」、「であり」、「たり」、「が」、「と、)と一致した個所の前後で文を分割する(処理(B))。

(2) 分割された入力文の構造化 ((A)テーマ語の判定, (B)助詞・述語の判定): 分割された文ごとにテーマ語の判定を行う(処理(A))。特定の助詞(「は」「と」「までは」「には」「においては」「ものは」「が」「も」と一致した場合、助詞の前部分をテーマ語とし、その助詞に続く文を属性とする。

処理(A)にて属性と判断した文を対象とし構造化する(処理(B))。対象の文について、文末から文頭に向けて述語ないし助詞を探す。述語ないし助詞を見付けた場合、述語ないし助詞を切り取り、階層を一段下げ、文の残りの部分を再帰的に処理する。文頭まで処理を終えた場合、現在のテーマ語に関する構造を閉じ、次の文の処理(A)に戻る。全ての文を処理した場合、獲得処理を終了する。

3.3 属性情報の獲得結果

今回は Wikipedia が有する全ての記事のうち、名詞概念辞書に格納された概念に関する記事を対象とし、各々の記事の <Abstract> タグによる部分を入力テキストとした。一般名詞概念 115765 個に対して、属性フレームを獲得した名詞概念は 5615 である。表 1 に獲得結果の例として、「楼閣」に関するテキストと獲得された属性フレームを示す。

表 1 獲得された属性フレームの例

テキスト	楼閣(ろうかく)とは、重層の建築物をいう。たかどの、高樓のこと。塔と類義であるが、塔は本来仏塔を指し、tower の訳語としての塔は近代に入ってから用いられている。それ以前の高層建築は一般に楼閣、高樓という呼称が用いられていた。
属性	((「は」 "楼閣") ("と" " ") ("かゝ" "楼閣(ろう)") ((「は」 " ") ("をいう" "建築物("の" "、重層")) ("のこと" "たかどの、高樓") ("と" "塔") ("であるが、" "類義") ((「は」 "塔") ("と" "tower の訳語" ("を指し" "本来仏塔")) ((「は」 "しての塔") ("である" "用法" ("に入っ" "近代")) ((「は」 "高層建築" ("の" "それ以前")) ("と" "楼閣、高樓" ("に" "一般")) ("が用いられていた" "いふ呼称")

4. ストーリー生成における属性情報の利用

この節では、獲得した属性フレームの利用方法とその有効性、及び物語自動生成ゲームにおける利用方法を提案する。

4.1 INGS における属性情報の利用

属性フレームを用いて、説明文を含んだストーリーを 100 個生成した結果、生成された約 60 種類のインスタンスのうち約 1 割に説明文が付与された。図 2 は生成したストーリーから生成された文の一部である。2 節で述べたように、ここで提案した方法は、ある名詞概念の辞書的な属性を獲得する。獲得した属性は、難解な名詞概念を含むストーリーの理解のし易さの調整に利用できる。例えば、現代では使用されていない概念(采女など)

や専門用語に関する概念(泉貨紙など)、馴染みのないカタカナ語の概念(ユマニテなど)等が使用される場合、説明はストーリーを理解する上で有効的に働く(図 3 はその一例である)。

<前略>…候が勇士に五山で「楼閣から沃地まで仕事に出る」ことを禁止する。「楼閣(ろうかく)とは、重層の建築物をいう。たかどの、高樓のこと。塔と類義であるが、塔は本来仏塔を指し、tower の訳語としての塔は近代に入ってから用いられている。それ以前の高層建築は一般に楼閣、高樓という呼称が用いられていた。」…<後略>

図 2 生成結果の一例(太字部分が説明文)

…采女が食事を運ぶ。「采女(うねめ)とは、平安時代以降は廃れ、特別な行事の時のみの官職となった」、日本の朝廷において天皇や皇后に近侍し身の回りの雑事を専門に行う女官のこと。」采女が食事を落とす。…

図 3 文の理解のし易さの調整例

4.2 物語自動生成ゲームにおける利用案

物語自動生成ゲームとは、我々が開発・構想中のゲームであり、アナログゲームである TRPG をモデルとしたストーリー生成を行う機構を持つゲームである[Ono 17]。このゲームでは、ストーリーの枠組みとその背景となる情報である世界設定を用意し、世界設定を利用しながら枠組みを肉付けないし拡張する。属性は 4.1 節と述べたようにこのゲームにおいても、難しい概念や非日常的・非現実的な内容を含むストーリーに利用可能である。

また本稿で獲得した属性は、世界設定の構築に利用する。世界設定は、あるストーリーにおける人・物・場所等を定義する情報である。そこで獲得した属性を編集することで、あるストーリーが生起する世界の仕組みや特徴を変化させる。これにより、ある概念が持つ情報に影響され、登場人物の行動も変化する。

5. おわりに

本稿では、INGS における説明文生成のため、属性情報の獲得を試みた。その結果、5615 個の名詞概念に関する属性フレームを獲得した。また獲得した属性フレームによる説明文生成を行い、その有効性について述べた。今後は、辞書的な説明に用いる属性は、固有名詞概念を含めた各々の名詞概念について属性フレームを獲得するため、テキストの範囲拡張方法を検討する。ストーリーごとの固有内容を説明する属性については、具体的な内容を獲得するのではなく、それを格納するための枠組みが必要となる。例えばある概念について、多数の実例からその概念が持ち得る属性を割り出すことで、その内容を定義できないか等を検討する。

参考文献

- [影山 12] 影山太郎: 属性叙述の世界, くろしお出版 (2012).
- [Ogata 16] Ogata, T.: Computational and Cognitive Approaches to Narratology from the Perspective of Narrative Generation in Ogata, T., and Akimoto, T. (eds.), Computational and Cognitive Approaches to Narratology, IGI Global, pp.1-74 (2016).
- [小野 14] 小野淳平, 小方孝: 名詞句の分類を利用した名詞概念の属性フレームの自動獲得—統合物語生成システムの一機構として—, 第 13 回情報科学技術フォーラム講演論文集 第二分冊, pp.335-338 (2014).
- [Ono 17] Ono, J. and Ogata, T.: Changing and Transforming a Story for a Framework of an Automatic Narrative Generation Game, Proc. of the 2017 International Conference on Artificial Life and Robotics, pp.643-649 (2017).