

# 文脈解析を用いた推理小説の犯人推定と難易度評価

平岡 聖也<sup>†</sup> 奥村 紀之<sup>‡</sup>

香川高等専門学校 情報工学科<sup>‡</sup>

## 1. はじめに

推理小説には多くのジャンルが存在し、その推理の難易度も様々である。推理小説を読み慣れた人とそうでない人では、面白いと思う難易度に差が出るため、推理小説の難易度を評価することで読み手の取捨選択の一助とすることができる。

本研究では、青空文庫の推理小説を対象に文脈解析を行い、登場人物名の出現頻度情報から犯人を推定し、その情報から推理の難易度を評価する方法を検討する。

## 2. 推理小説

推理小説とは、推理作家江戸川乱歩によって「主として犯罪に関する難解な秘密が、論理的に徐々に解かれていく経路の面白さを主眼とする文学である」と定義されている。基本的に、推理小説で起こる事件は殺人事件である。

推理小説では、物語の流れを追いかけてながら登場人物の行動やセリフなどから犯人を推定する。西島らの研究<sup>[1]</sup>では、文単位での繋がりを調べるために物語を結末から追跡し、命題間の論理的関係を抽出する方法を用いている。

## 3. 犯人推定方法

推理小説の難易度は、事件の犯人を当てる難易度と直結している。そのため、推理小説の難易度を評価するためには、その犯人を推定する必要がある。本研究では、他の研究では見られない、登場人物の出現回数と物語の進度におけるその推移から犯人を推定する。

### 3.1. 人名抽出方法

登場人物名の抽出は、JUMAN-7.0 と MeCab-0.996 の出力結果から行う。JUMAN では表 1, MeCab では表 2 のパターンで出力されたものを人名とする。なお、JUMAN, MeCab 共に既存の辞書だけでは対応可能な人物名が少数であったため、新しく「松茸用人名テキスト」から約 80,000 語の人名をユーザ辞書として追加した。

表 1 人名として抽出する JUMAN 出力パターン

フィールド	内容
品詞	名詞
品詞細分類	人名

「A Culprit Estimation and A Difficulty Evaluation using Context Analysis」

<sup>†</sup> 「Seiya HIRAOKA」

<sup>‡</sup> 「Noriyuki OKUMURA」

Kagawa National College of Technology, Department of Information Engineering

表 2 人名として抽出する MeCab 出力パターン

フィールド	内容
品詞細分類 1	固有名詞
品詞細分類 2	人名

### 3.2. 代名詞等の変換

代名詞や役職名等に対応する登場人物名に変換しなければ、小説テキストから人名の抽出を行っても正確な出現回数を求められない。そこで、KNP の照応解析を用いて機械的に対応付けを行う。小説テキストを JUMAN で解析し、その出力結果を KNP に以下のオプションを付与して解析する。

表 3 KNP の付与オプション<sup>[2]</sup>

オプション	内容
anaphora	ルールに基づく協賛賞解析
同上	述語項構造解析
ne	CRF に基づく固有表現解析

これらの解析結果より小説の代名詞や役職名を登場人物に変換する。また、結果を比較するために手動で同様の変換を行う。基本的に名字へ変換を行うが、家族等で同一名字の人物が複数登場する場合は KNP によって名字と下の名前の対応が取れれば、手動で下の名前へ変換を行う。

## 4. 比較実験

KNP での代名詞等の変換が正しく行われ、犯人の推定を行うにふさわしいかどうかを手動で代名詞等の変換を行った結果と比較して考察する。

### 4.1. 実験に用いた作品

本研究では、以下の 5 作品について実験を行った。

- ・「麻雀殺人事件」 浜尾四郎
- ・「愚人の毒」 小酒井不木
- ・「謎の咬傷」 小酒井不木
- ・「彼は誰を殺したのか」 海野十三
- ・「カンカン虫殺人事件」 大阪圭吉

## 5. 実験結果

上記の 5 作品の内、「麻雀殺人事件」の結果を例として示す。

### 5.1. 登場人物抽出結果

JUMAN と MeCab によって抽出された「麻雀殺人事件」の登場人物の抽出結果とそれぞれの人物の作中の役割は表 4 の通りである。

姓名合わせて 15 個の内、MeCab では 11 個、JUMAN では 6 個の抽出に成功した。抽出できなかった名前は、既存の辞書に人名として載っておらず、追加した人名辞書にも掲載されていなかった。

表4 登場人物の抽出結果

名前	MeCab	JUMAN
星尾信一郎	姓:○名:○	姓:×名:○
川丘みどり	姓:×名:○	姓:×名:×
園部壽一	姓:○名:○	姓:○名:×
帆村莊六	姓:×名:×	姓:×名:×
松山虎夫	姓:○名:○	姓:○名:×
舟木豊乃	姓:○名:○	姓:×名:○
雁金	姓:○	姓:×
本田	姓:○	姓:○
河口	姓:×	姓:○

MeCab でのみ抽出された部分は、JUMAN では人名として登録されていなかった。全ての姓名の内、MeCab, JUMAN それぞれでどれだけの姓名を抽出できたかの精度は表5の通りである。

表5 MeCab, JUMAN の人名抽出精度

	MeCab	JUMAN
精度	0.7333	0.4000

### 5.2. 各人物の出現割合

以下の実験では、精度の高かった MeCab の出力結果を用い、抽出できなかった人名は手動でユーザ辞書へ追加し抽出した。

犯人と容疑者候補の2人、探偵役の出現割合の推移の仕方をまとめたのが以下の図1, 2である。横軸に章、縦軸に各人物の出現割合を取る。図1はKNPを用いた結果、図2は手動で変換を行った結果である。グラフは同様の傾向を示しており、KNPでの代名詞変換が犯人の推定を行うのに十分であることが分かる。

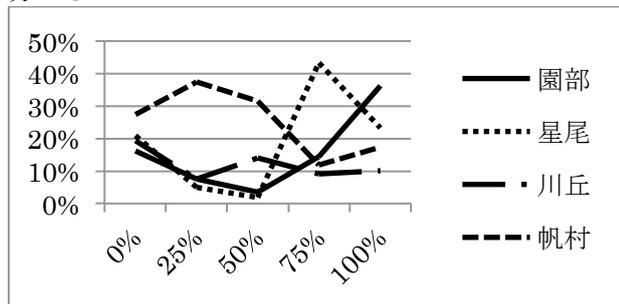


図1 KNPを用いた各章の各人物の出現割合

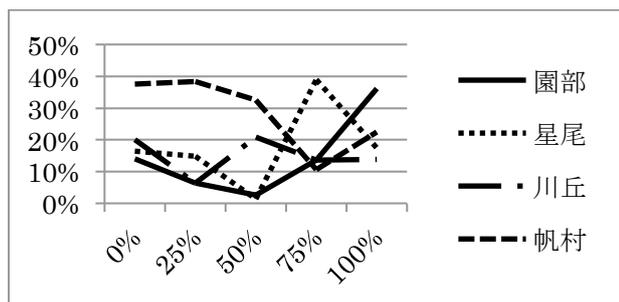


図2 手動で変換を行った場合の人物の出現割合

### 6. 考察

図1, 2を見ると、犯人は序盤の出現割合は高くなく、中盤から終盤にかけて徐々に高まり、最終章では最も高くなっている。以下の図3は、縦軸に出現割合、横軸に物語の進度をとった5作品の犯人の出現割合の推移である。「麻雀殺人事件」「カンカン虫殺人事件」「謎の咬傷」の3作品はこの傾向が顕著である。「彼は誰を殺したのか」は1章, 2章が犯人の目線で描かれているため、序盤の出現割合が高くなっているが、終盤の割合も低くない。

「愚人の毒」は終盤に探偵役が犯人に一对一で語りかけるため、出現回数が少ない。これらより、最終章の出現割合や登場時との変化の大きさから犯人を推定できると考えられる。

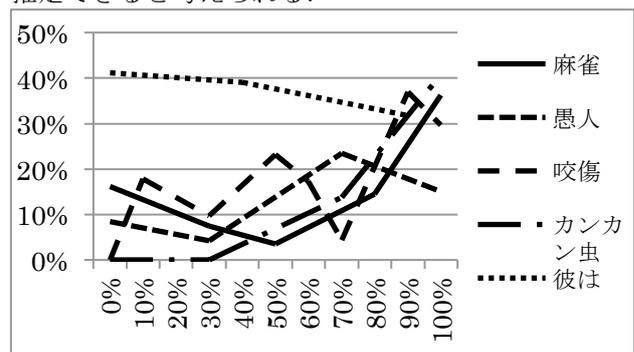


図3 各作品の犯人の出現推移

以上の結果から、推理小説の難易度は犯人と他の容疑者の推移から評価できると考えられる。犯人が序盤から出現し、上記のように終盤にかけて出現割合が増加する場合、推定は容易である。しかし、犯人が終盤まで登場しなかったり、他の容疑者と同様の傾向を示す場合は推定は難しくなると考えられる。同じ傾向を示した作品を同じ難易度としてよいかは今後他の要因と合わせて解析を行い判断する必要がある。

### 7. まとめ

本稿では、小説テキストを機械的に解析し、そこから得られる登場人物の出現情報をもとに犯人を推定する方法について述べ、難易度評価の観点から検討を行った。

今後は、発生した問題の解決と共に人名の抽出割合を更に向上させ、また他の要因と合わせてより精密な解析が行えるように進めていく。

### 参考文献

- [1] 西島恵介, 神山文子, 藤田米春. (1997). 短編推理小説の論理構造の分析と推理. 情報処理学会全国大会講演論文集. 第55回平成9年後期(2). pp57-58.
- [2] 笹野遼平, 河原大輔, 黒橋禎夫, 奥村学. (2013). 構文・述語項構造解析システムKNPの解析の流れと特徴. 言語処理学会第19回年次大会発表論文集. pp110-113.