

# 因果関係・推移関係コーパスの生成とそれを用いた未来事象の予測

富田 拓実<sup>†</sup> 那須川 里紗<sup>†</sup> 谷津 元樹<sup>†</sup> 原田 実<sup>†</sup>

青山学院大学理工学部情報テクノロジー学科<sup>†</sup>

## 1. はじめに

因果関係の知識は、質問応答や未来事象の予測への利用が期待できる。質問に対する回答を大量の文書から探し出す質問応答システム Metis[1]や NAZEQA[2]では、文に含まれる因果関係を元に回答を推論している。

文書から因果関係を抽出する研究として、文節間の深層格に基づいた西脇ら[3]や、手がかり表現に基づいた乾ら[4]や坂地ら[5]の研究が挙げられる。また、明示的に文書に書かれていない因果関係の抽出を目的とした研究として、小野ら[6]の研究が挙げられる。

本研究では、新聞記事から因果関係・推移関係を自動的に抽出してコーパス化し、生成した辞書を利用して入力文に対して未来事象の予測を行うシステムを作成した。意味解析システム SAGE[7]によって文を解析し、その際に決定された文節間の深層格を因果関係・推移関係の抽出に利用する。

## 2. 因果関係・推移関係の抽出

本研究では、1998-1999年の毎日新聞の記事 219,955 記事を対象として、因果関係・推移関係の抽出を行った。

### 2.1 意味解析システム SAGE

因果関係・推移関係を抽出するため、まず、新聞記事を意味解析システム SAGE で解析する。図 1 は SAGE で「点訳がパソコンでできるようになって、仕事は楽になった。」という文を解析して得られた意味グラフである。「cause」や「main」など、枝上に赤で記述されているものが深層格で、[]内に書かれているものが語意である。深層格とは、他の文節との役割関係のことである。語意と深層格の決定には、EDR 電子化辞書を利用している。

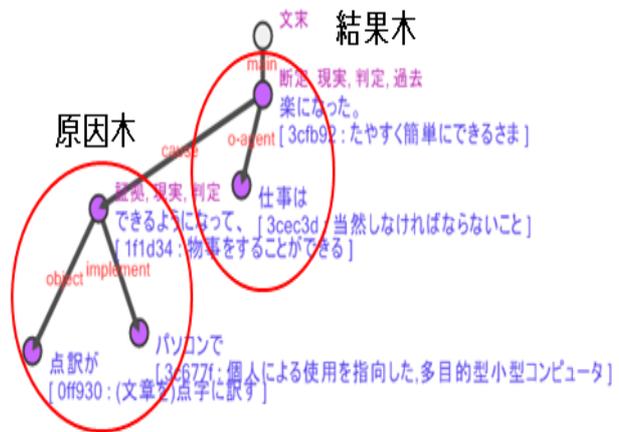


図 1. SAGE から得られた意味グラフ

### 2.2 因果関係・推移関係の抽出方法

因果関係を抽出する場合、各ノードの深層格を順に調べ、述語間の因果関係を表す深層格「cause」で係り側となっている原因ノード(例:できるよようになって)を見つけ出す。原因ノードで意味グラフを分割し、原因ノードを根とする木を原因木、原因ノードの cause 辺を介した受け側となっている結果ノードを根とする木を結果木とする(図 1)。原因ノードから直接係り側となっているノードを取得し、これらのノードからなる小さい木を、新たに原因木とする。結果木も同様に更新する。作成した結果木が主語や目的語を含まない場合は辞書には含めない。このようにして因果関係の抽出を行なう。抽出した因果関係の例を表 1 に示す。最後に原因ノードをキーとして、原因木の意味グラフ、結果木の意味グラフ、新聞の記事番号を含む因果辞書を作成する。

表 1. 因果辞書の例

原因ノード	原因木	結果木
発覚したため、	詐欺が発覚したため、	同社を懲戒解雇されている。
反対して	強制送還に反対して	暴動を起こした。
加えていくので	新しい技術を加えていくので	環境の負荷も大きい。

推移辞書を作成する場合は述語間の推移関係を表す深層格「sequence」で繋がるノードを見つけ

Generation of causal relation / transit relation corpus and prediction of future event

Takumi TOMITA, Risa NASUKAWA, Motoki YATSU and Minoru HARADA

Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University

出し、その後は因果辞書と同様に作成する。抽出した推移関係の例を表2に示す。

表 2. 推移辞書の例

原因ノード	前	後
多発、	抗議デモが多発、	市民多数が死傷した。
加え、	景気低迷の長期化に加え、	倒産を増加させている。
抱えて	約480億円の負債を抱えて	破産宣告を受け、

### 3. 未来事象の予測

入力文を SAGE によって意味解析し、述語節のノードを取得しそれを検索ノードとする。検索ノードの直接係り側となっているノードを取得し、それらのノードの集合を検索木とする。検索ノードをキーとして辞書から因果関係のペアを取り出す。検索木と各原因木の文の類似度を算出し、最も類似度が高い原因木に対応する結果木を取り出し、それを予測木の作成元とする。この元になる木における主語、目的語のそれぞれが、検索木の主語、目的語と閾値以上の類似度を持つ場合、前者を後者に置換し、それを最終的な予測木とする。未来予測の処理事例を図2に示す。

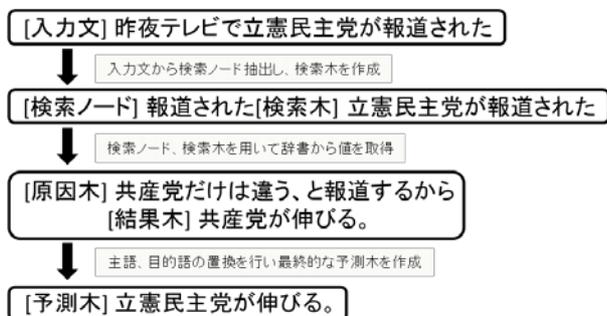


図 2. 未来予測の処理事例

因果関係・推移関係の抽出に使っていない2000年以降の新聞記事を入力文としたとき、出力された予測文の例を表3に示す。

表 3. 未来予測の結果例

検索木	予測木
大きな山火事が起きた。	注意が必要だ。
外資系メーカーの現地生産車などに押され、	利用客は減る一方。
病院や診療所が破壊されている。	具体的な支援が急がれる。

### 4. 評価実験

作成されたコーパスの中から無作為に選んだ50個の因果関係のペアを被験者に提示した。被験者は、理工学部にも所属する学生5人である。被験者は提示されたものが因果の関係としてふさわしいかを0から4の5段階評価を行なった。その平均値に25を掛けてスコアを算出した。各被験者におけるスコア（適合率）の平均値は63.9となった。

数値があまり高くならなかった原因として、因果関係かどうかを判断するのに個人差が大きいことがあげられる。さらに、本研究では、未来予測の再現率を上げるために、取得するノードの数を制限し、原因木を小さくする必要性があった。そのため、日本語として不完全な文が生成されてしまったことにより、辞書の適合率が低くなってしまったと考えられる。

### 5. まとめ

本研究では、意味解析システム SAGE を用いた因果・推移関係の抽出と辞書の作成を行った。また、作成した辞書を利用し入力文に対して未来事象を予測するシステムを作成した。

### 参考文献

- [1] 加藤裕平、蒲生健輝、古川勇人、韓東力、原田実、” WEB 検索による知識文の獲得と意味グラフ照合推論による質問応答システム Metis ”、情報処理学会第 67 回全国大会、2005
- [2] 磯崎秀樹、東中竜一郎、” パターンマイニングを用いて「なぜ」に答えるシステム ”、言語処理学会第 14 回年次大会、2008
- [3] 西脇剛、保立哲志、原田実、” 意味解析に基づくテキストマイニングツール STM ”、情報処理学会第 69 回全国大会、2007
- [4] 乾孝司、乾健太郎、松本裕治、” 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得 ”、情報処理学会論文誌、vol.45 no.3、2004
- [5] 坂地泰紀、増山繁、” 新聞記事からの因果関係を含む文の抽出方法 ”、電子情報通信学会論文誌、vol.J94 no.8、2011
- [6] 小野博紀、内海彰、” イベントの時系列分析による因果関係知識の獲得 ”、人工知能学会論文誌、vol.30 no.1、2015
- [7] 原田実、水野高宏、” EDR を用いた日本語意味解析システム SAGE ”、人工知能学会論文誌、vol.16 no.1、2001