

# 新聞記事の内容把握図作成における重要要素配置方法の一提案

情報科学科 高瀬 恭子

指導教員：山村 毅

## 1 はじめに

新聞記事はそのほとんどが文章にて書かれている。ただ、文章には長所と短所があるので、より効率的な理解を求めるためには、内容を表した図があると便利である。

文章の可視化手法には、トレンド情報に焦点をあて、形態素解析とパターンマッチングによって抽出、重要度やキーワード間の各指標を計算することで多面的に解釈し、グラフ構造で可視化する研究 [1] などがある。

本研究では新聞記事の迅速な内容把握を支援するため、新聞記事から重要な要素とそれらの関係をまとめたリストを生成し、これを用いて要素を配置する内容把握図作成手法を提案する。

## 2 内容把握図の作成手法

### 2.1 要素間関係リスト

新聞記事から内容把握図を作成するには、先ず、図に用いる要素とそれらの関係を抽出する必要がある。本研究では、これを2つの要素とそれらの関係(要素間関係)という形で表現し、リスト化する。このリストを要素間関係リストと呼ぶことにする。

#### 2.1.1 要素間関係

2つの要素間(要素1, 要素2)の関係として、「包含」「発言」「方向」「場所」の4つを考える。

- 「包含」:要素1の説明または内容を要素2で表す。
- 「発言」:要素1が話した言葉を要素2として表す。
- 「方向」:要素1が要素2に対してどうしたかを表す。
- 「場所」:要素1は要素2がどこで起きたかを表す。

#### 2.1.2 要素の優先順位

要素に値を与え、優先順位を決める。これを用いることで、新聞記事の重点となる要素を判断する。

### 2.2 図の作成

新聞記事から要素間関係リストを生成したあと、このリストを用いて内容把握図を作成する。具体的には、リスト中の要素をその優先順位に基づいて配置し、要素間には2.1.1で述べた要素間関係を配置する。

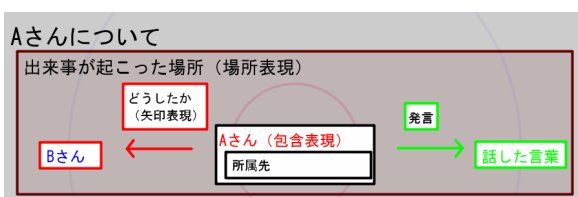
#### 2.2.1 要素の配置位置

要素の配置は、優先度の高い要素の周りに、それと関係のあるより優先度の低い要素が円形になるように行う(最重要要素は、図全体の中心に来るようにする)。

#### 2.2.2 要素間関係の表現

要素を配置したあとに、要素間の関係を図に配置する。2.1.1で述べたように、要素間の関係は4つあるが、これらを包含表現、矢印表現(「発言」と「方向」)、場所表現の3つで表現することにする。(図1)

図1 要素間関係の表現



### 2.3 配置位置の調節

矢印表現を用いるとき、配置の仕方によっては要素が重なって表現されることがある。そこで、重なりを判断する処理を付けることで、重ならない位置になるまで矢印表現に關係する要素を円上にて動かし、重ならない配置を実現させる。

## 3 評価実験

新聞150記事に対し、人手で要素間関係リストを生成し、これを用いて2.2～2.3で述べた方法で内容把握図を作成した。これらの内容把握図を5人の被験者に5段階の意味差判別法で評価してもらった。質問事項は以下の3つである。

1. 要素の配置位置は適切であるか
2. 新聞記事の内容を反映しているか
3. 要素に用いた文字の数は適切か

結果を表1に示す(値は平均値を表している)。

表1 実験結果

被験者	質問1	質問2	質問3
A	3.85	3.9	3.81
C	4.43	4.33	4.48
D	4.16	4.18	3.95
E	4.11	3.53	3.86
合計	4.14	3.99	4.02

ここで被験者Bに関しては、仮説検定(母平均の差についての検定)を行った結果、他の被験者の値と比べて有意な差があることがわかったため、表には載せていない。

この結果から、配置、反映、文字数ともに4点近い値となっており、バランスのとれた図が作成できたとと言える。ここから、今回提案した内容把握図の作成手法は妥当であると考えられる。しかし、詳細に結果を見ると低評価となる図もあった。被験者からは、「各要素の文字数の差によるバランスの崩れ」、「要素数が少なすぎて内容を反映しきれていない」、「要素の使用文字数が多すぎる」などの意見があり、このことから、要素の文字数が配置や反映に影響を与えていることが分かる。実際、図としての流れは分かるが、一目で理解するには要素の文字数が多すぎると感じられる部分があった。

## 4 まとめ

新聞記事から抽出した要素とその関係をまとめた要素間関係リストを用いて、自動で図を作成する内容把握図作成手法を提案した。結果から妥当な手法であると考えられるが、特に用いる要素の文字数に関しては改善していく必要が見られる。また、一目で理解できる図となるよう、図としての良さを考慮に入れていく必要がある。

今後の課題としては、3で述べた問題点を改善していくこと、また要素間関係リストの自動生成が挙げられる。

## 参考文献

- [1] 白井康之, 小関悠, "テキストマイニングによるトレンド情報抽出環境の構築", 三菱総合研究所 技術レポート, 2009