

属性つきキーグラフ

Attributed KeyGraph

伊藤貴一*1
Takaichi Ito

大澤幸生*2
Yukio Ohsawa

*1 慶應義塾大学院政策メディア研究科 *2 東京大学
Graduate School of Media and Governance, Keio University #1 Tokyo University #2

We have developed attributed Key-Graph to analyze free-answer text such as opinionaire. The attributed Key-Graph is able to visualize attributions by indicating pie-graph showing distribution of attributions to every node. With this visualization, it is capable to understand correlation between difference of opinion and difference of attribution. In this research, we take a free-answer opinionaire survey on bullying for example of the visualized analysis. As a result, visualization of correlation between experience on bullying and difference on bullying has been successful.

1. はじめに

情報技術の発達によって、従来の紙を使ったアンケート調査だけでなく、Web やケイタイを使ったアンケート調査が増えてきた。以前は、テキスト解析の困難さから、アンケート項目の定量調査がメインで、自由回答文は参考程度にしか扱われなかった。しかし、近年のテキストマイニング技術の発達によって、十分に扱えるようになってきた。

アンケート調査は、回答者の年齢、性別、職業などの人口統計的データ、そのアンケートで聞きたい質問項目、それに加えて自由回答文によって構成される。このようなデータというのは、ありふれているが、なかなか解析できていない。どういう人たちが、どういうことを考え文章を書いているか、ということを知ることは、社会分析、市場分析のうえで重要なことであり、分析する手法が求められる。この分析において必要なことは気付きを誘発し、直感的なイメージが描きやすいということである。

2. KeyGraph

気付きを誘発するテキストマイニング手法として KeyGraph[1]がある。本論文では、KeyGraphの機能拡張をする。

2.1 紙芝居キーグラフ

KeyGraphにおいて、複数の KeyGraph を重ね合わせて一面面で見ると紙芝居キーグラフ[2]という手法がすでにある。これも、属性ごとにデータをわけて、KeyGraph を作れば、属性ごとの特徴を読み取ることができる。しかし、紙芝居キーグラフは、KeyGraph の構造的差異を見るものであって、実データの量的な差異を見るのが難しい。グラフのベースとなる黒ノードの選択は、頻度の上位語を選択するため、ある属性によって、頻度が低いため、黒ノードが出現しなくなったという差異はわかる。しかし、それは、黒ノードの数の指定に依存しており、指定数によって決まってしまう。また、高頻度語同士であれば、差異があったとしても見えなくなってしまう。

3. 属性つきキーグラフ

アンケートの定量分析として、人の属性とアンケート項目とのクロス集計をすることは基本であるが、これと同様のことが、テキスト分析でできないのか、というのが、この手法の基本的なアイデアである。

本論文で主張する属性つきキーグラフとは、キーグラフのノードに各属性の割合を円グラフで表示する。円グラフで表示するこ

とによって、一つ一つのノードに関する属性の分布が一目でわかる。そのため、この言葉は、こちらの属性の人たちが主に言っていることだ、ということが分かる。

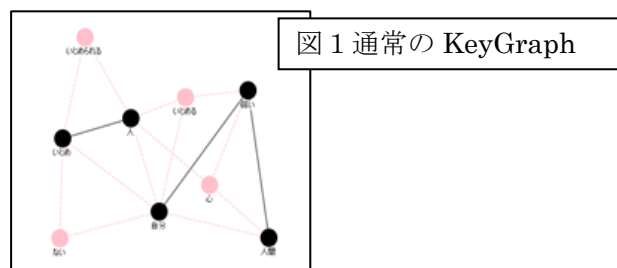


図1 通常の KeyGraph

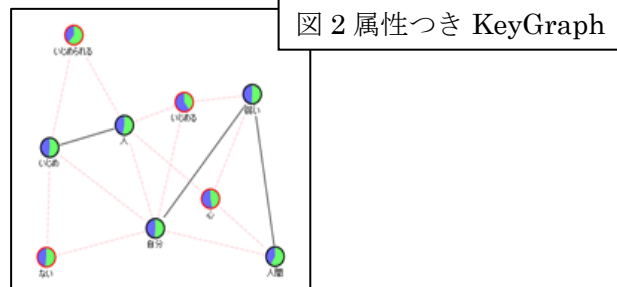


図2 属性つき KeyGraph

3.1 入力データの形式

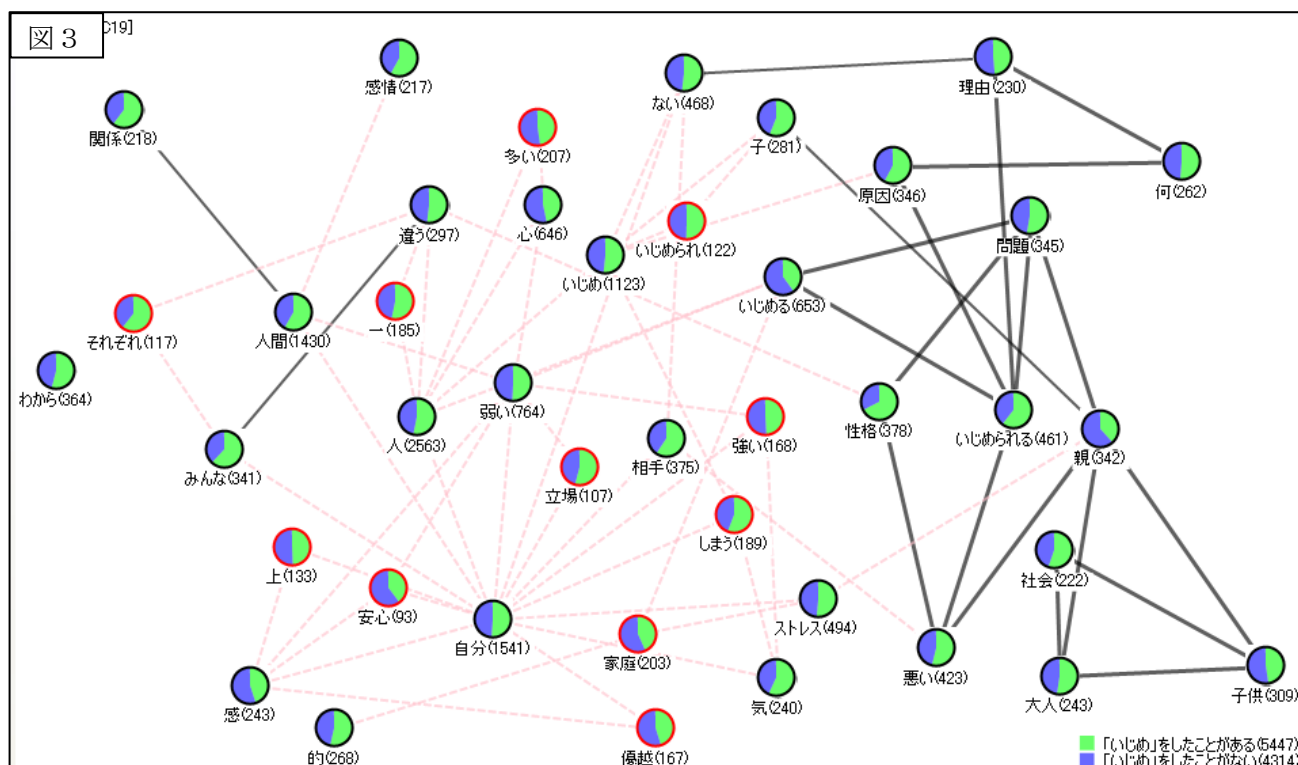
入力データの形式は運用上の利便性を考えて、改行区切りで文章に<>のなかにカンマ区切りで属性データを埋め込めるようにした。このように文字列で属性を入力できるようにすることで、文章に付随する属性を、離散的にすれば、取り込めるようにした。そのため、時系列的変遷にたいしても、たとえば、前期、中期、後期といったような属性をあたえることで分析することができるようにした。

入力データの形式

文章<属性 1, 属性 2, 属性 3, 属性 4 ···>

例

- だと思えます<男, 大学生>
- なのが好き<20代女姓, イチゴ味派>
- ですよ。<前期、イノベーター>



4. 結果

図3、図4はケイタイを使って「あなたは、何故「いじめ」が起こると思いますか？」という設問に対する回答文を分析した結果である(回答数 10907 人)。このアンケートでは、同時に、「あなたは「いじめ」をしたことがあるのか？」という設問もしているので、図3では、それを属性として採用し、円グラフを描画させた。緑色は「いじめをしたことがある」、青色は「いじめをしたことがない」である。図4は、図3と同じ配置の通常の KeyGraph である。

図3をみると、多くのノードは、「いじめをした」「したことがない」で大体、均衡し、同じ割合である。しかし、割合が著しく崩れているものを見ていくと興味深い。

「いじめをしたことがある」で特徴的なのは、「いじめられる」「性格」「みんな」というキーワードであろう。これを解釈すると、いじめられる側にとって、いじめられるのはいじめられるやつの性格に問題があるというコンテキストが浮かび上がってくる。「みんな」というのは何かしらの集団心理に関係するのであろう。

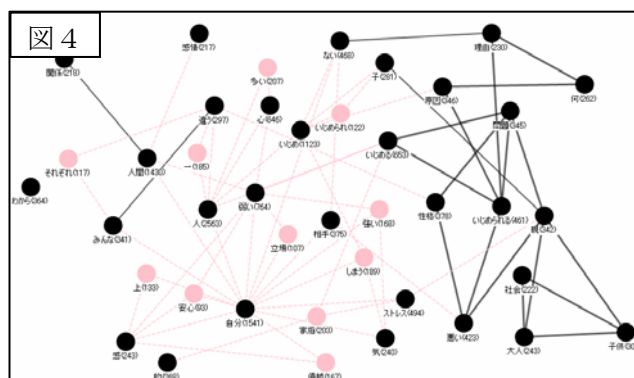
また、「いじめをしたことがない」で特徴的なのは、「いじめる」「親」「安心」「優越」「家庭」というキーワードであろう。解釈をすると、いじめる子の親や家庭の問題について、いじめをしたことがない子は考えていることが見えてくる。

このような解釈は、図4のような、通常の KeyGraph では行なえない。言葉のつながりによって、一通りの解釈はできるが、書いている人の立場の違いに基づいた解釈はすることができない。また、先ほど特徴的だと書いた、「性格」や「親」といったキーワードについての重要性は図4からは見えてこない。

5. まとめ・考察

KeyGraph のノードに属性の円グラフを書き加えるというただそれだけで、立場の違いによっての意見の相違が読み取れることが分かった。

先述で、紙芝居キーグラフは構想的差異、属性つきキーグラフは量的差異の比較であると述べた。



この構造的差異、量的差異の二つをうまく組み合わせると可視化することができるであろう。これは研究課題だ。

また、この手法は、属性の表示を行なうことができるが、どの属性を選択したらいいのか？ということは教えてくれない。最適な属性を探すためには、何度も繰り返さなければならない。これについても、AI 的な補助は必要であり、これからの課題でもある。

参考文献

- [1] 大澤幸生: チャンス発見の情報技術—ポストデータマイニング時代の意志決定支援, 2003, 東京電機大学出版局
- [2] 大澤幸生: チャンス発見のデータ分析—モデル化+可視化+コミュニケーション→シナリオ創発, 2006, 東京電機大学出版局