青山学院大学　理工学部　情報テクノロジー学科

平成１７年度卒業論文要旨

意味解析を用いたテキストマイニングツール STM2005 の開発

青木　大輔（15802002）瀧口　健太郎（15802046）初見　省吾（15802057）
原　田　研　究　室

１．背景・研究目的

既存のテキストマイニングではキーワード中心による解析しか行っていない。結果として、語意や語間の関係を把握できない、「何がどうだ」「何をどうした」などの複数の語からなる関係を把握できない、表現のゆれによる差異を吸収できない、という問題点があった。本研究では意味解析を用い、表現が異なっても同じ意味を表す表現を同意見と見なせるテキストマイニングツール STM の開発を行う。

２．システム概要

STM は入力された CSV 形式のアンケートデータに対して意味解析を行い、その結果を用いて句を作成する。意味解析の結果や生成された句を全て Access データベースに保存し、このデータベースを基にクラスタリング分析や時系列分析を行う。クラスタリングの対象は句と文である。図１にシステム図を示す。

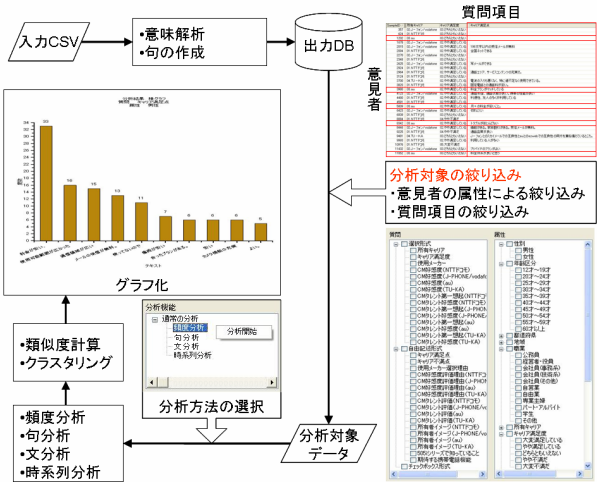

図 1. STM2005 の処理の流れ

３．句単位での分析

本システムでは意見を句という単位で扱う。

意見　∋　文　∋　句　∋　文節　∋　形態素

句とは、文中に並列節や副詞節があるときこれらを主節から切り離した各節のことである。句を分析に用いることで、単純な語の頻度分析よりは複数の語間の関係からなる回答者の意図をベースにした分析ができ、また、複文を単位とするよりは文を短くするので類似性の高いクラスタリングを行うことができる。

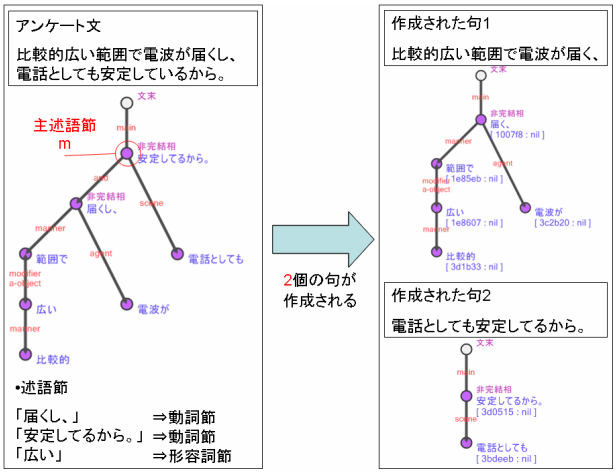

図 2. 句の作成例

４．分析対象の絞り込み

アンケートデータの入っているデータベースには無数のデータが入っている。そのため分析を行う際、前処理として分析対象を絞り込む必要がある。本システムではアンケートの回答者の属性を絞り込む、アンケートデータの質問を絞り込む、という２つの方法で分析対象を絞り込んでいる。

５．結論

本研究では、意味解析を用いることで表現の揺れによる差異を吸収することができ、句を分析に用いることでアンケート回答者の意図を把握することができ、また文類似度によるクラスタリングで同意図を集約することができるテキストマイニングツールの開発を行うことができた。