

テキストマイニングシステムの開発

顧客意見の分析をサポート

Development of Text Mining System

Support Tool for Analyzing Customer Opinion

(エネルギー応用研究所 お客さまネットワークG 情報通信T)

顧客意見などのテキストを分析し、顧客満足度向上やサービス施策に有用な知見の発見を支援するシステムの開発を行った。本システムでは、文書の自動分類と視覚化、評判分析、主題分析、トピックの時系列分析などの機能を実現している。

(Information and Communication Team, Customer Supply Network Group, Energy Applications Research and Development Center)

We have developed a system for analyzing customer opinion that supports to find an effective knowledge for improving customer satisfaction or service planning. The system provides the following functions: 1) automatic text categorization and their visualization, 2) reputation analysis, 3) theme analysis, 4) time series analysis of topic.

1 研究の背景と目的

近年ではコールセンターにおける問い合わせ記録や顧客からのメールなど、多様な情報源から「お客さまの声」を大規模に収集することが出来る。しかし、実際にこれらのテキストデータは数値的な解析には向いておらず、情報として蓄積されるだけでなかなか利用されないのが現状であった。

そこで、本研究では、顧客ニーズの分析やマーケティングを支援するためのテキストマイニング(Text Mining)技術を検討し、現場で広く利用できる利便性の高いツールの開発を行った。

2 システムの要件

本システムの開発に際しては、営業部、情報システム部等の部署にニーズ調査を行い、開発コンセプトを明確化した。その結果、機能要件としては、「文書内容把握」、「潜在ニーズ発見」、「評判分析」に重点を置いた開発を行うこととした。

3 システムの機能概要

(1) 文書のカテゴリ分類と視覚化

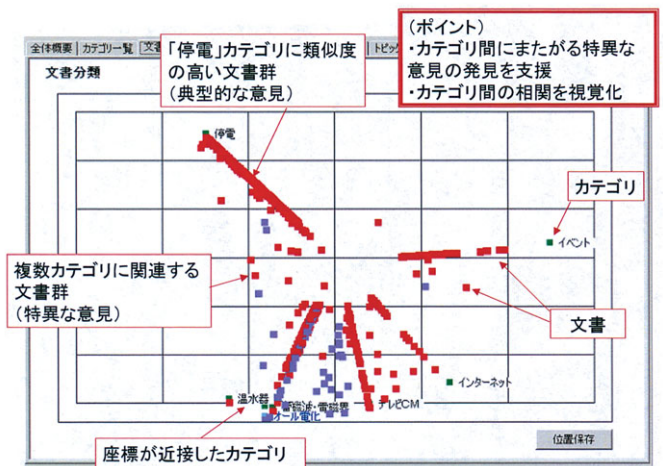
ユーザがあらかじめ設定したカテゴリに文書データを自動分類する機能を開発した。文書分類にはベクトル空間モデルを用いている。

さらに、上記カテゴリ分類機能の分類結果を見通しよく情報提示するための視覚化技術を開発した。提案手法では、カテゴリと分類結果の文書を2次元平面上にマッピングし、全体傾向を俯瞰すると共に、特異な意見の発見を支援するインタフェースを実現している(第1図)。

視覚化画面ではカテゴリと文書がそれぞれ色付きの

■のシンボルで表示され、平面上で原点からカテゴリのシンボルに近い文書ほど、そのカテゴリに類似していることを表している。また、カテゴリの座標は多変量解析の一手法である対応分析によって求めており、類似したカテゴリ同士が接近するため、カテゴリ間の相関も把握することができる。

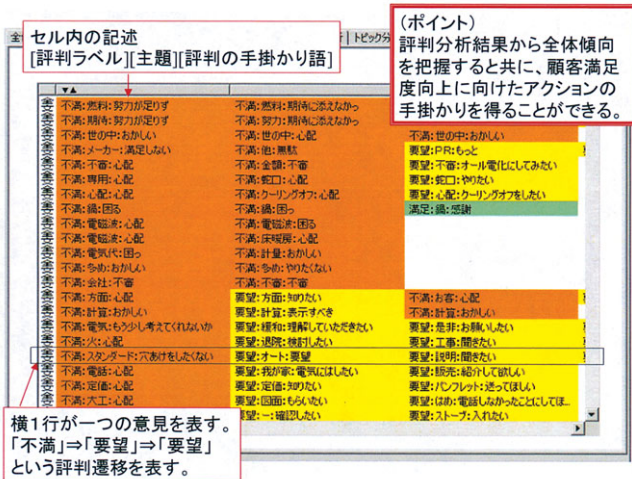
視覚化結果の解釈であるが、原点から、あるカテゴリに向かって外周方向に伸びているスペクトル状の文書群は、そのカテゴリのみに関連の深い話題であることを示しており、周囲に散らばっている文書は複数カテゴリにまたがる特異な話題であることを示している。



第1図 文書のカテゴリ分類と視覚化

(2) 評判遷移分析

顧客意見に含まれる評判(満足、不満、要望)という観点から分析を行う機能を開発した。また、ある意見の中では評判カテゴリは一つに限定されるものではなく、顧客の文書表現の中で、「不満」⇒「要望」のように心的状態が遷移する場合がある。そこで、評判分析結果を概観し、「不満」⇒「要望」などの因果関係を分析する閲覧機能を開発した(第2図)。



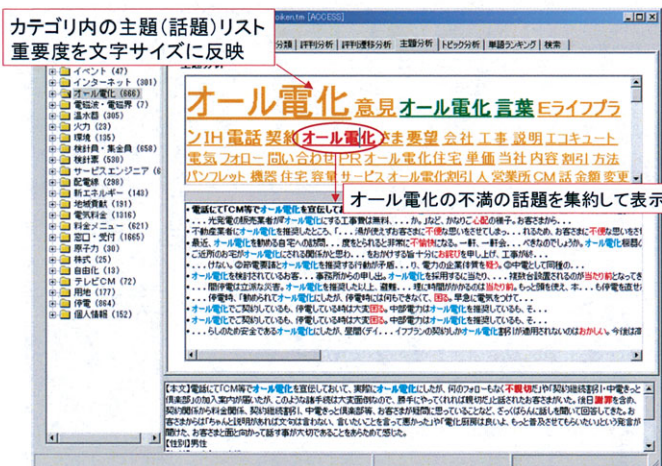
第2図 評判遷移分析

(3) 主題分析

上記評判分析においては、評判の対象は何か？という点が重要である。そこで評判の対象である「主題」を推定し、主題ごとに意見を集約して情報提示する機能を開発した。主題分析の手法であるが、以下の3つの評価基準を用いて文書中の名詞の中から主題と思われる単語を推定している。

- ・ 候補となる名詞と評判の手掛かり語の文書中での出現位置が近接している。
- ・ 候補となる名詞の直後の単語の品詞が格助詞(「が」、「を」、「に」など)である。
- ・ 候補となる名詞の文書中での重要度(TF*IDFスコア：出現頻度が高く、特定の文書に偏って出現する単語に高い重みを与える)が高い。

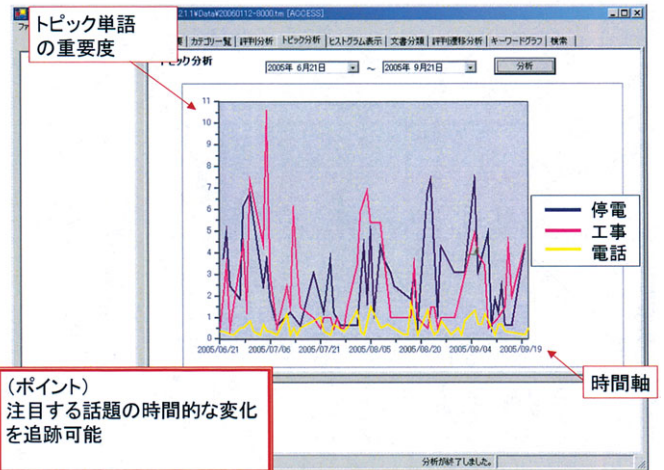
主題分析画面では、カテゴリごとに細分化された主題をタグクラウド風の画面で閲覧することが可能であり、カテゴリ内の話題集約を行うことができる(第3図)。ここでは、同義語のシソーラス辞書を用いて同一概念の話題を集約する処理を行っている。



第3図 主題分析

(4) トピックの時系列分析

注目するトピック単語の時系列変化を把握する機能を開発した。本機能では、ユーザが指定したトピック単語の重要度の時系列変化をグラフで表示することができる(第4図)。また、指定期間内に出現した特異な単語を提示する機能も備えている。



第4図 トピックの時系列分析

4 ユーザ評価

システムの有用性評価のために顧客意見データを用いた分析の試行を行った。以下に試用したユーザの評価を機能別にまとめる。

(1) 文書のカテゴリ分類と視覚化

カテゴリの相関や、カテゴリ間にまたがる特異な意見の発見に有用である。

(2) 評判遷移分析

評判の全体傾向を把握すると共に、顧客満足度向上のための施策の参考になる。

(3) 主題分析

カテゴリ内の細分化された話題集約に有用である。

(4) トピックの時系列分析

注目する話題の時系列変化の追跡や、特徴的なキーワードの発見に有用である。

5 今後の展開

本システムは営業部や情報システム部に試行導入され、実データを用いた評価を行っている。また、技術開発本部で実施したアンケートの分析に実適用されている。今後は設備保守の障害記録の分析などへの適用を検討し、応用システムの開発を行っていく予定である。(本システムは(株)中電シーティーアイとの共同研究で開発した。)



執筆/瀬川 修