

# お客様の声分析システム VOCAnalyzer の開発

The Development of the System to Analyze the Voice of Customer

近年 CS (顧客満足度) 向上をスローガンに掲げる企業が多く見受けられる。CS 評価の方法としてアンケートを取ることが一般的である。数値情報として得られる CS の分析は容易に集計することができる。しかし大量に寄せられた自由記述文は機械的集計の困難さから文章量の規模が大きいかほど活用度が低くなる。この自由記述文を有効活用したいというニーズが高まってきている。そこで大量に寄せられたアンケートの自由記述文章へテキストマイニングを適用し、「お客様の声」を半自動で分析するシステム VOCAnalyzer を開発した。本報告では、VOCAnalyzer の説明と適用事例について述べる。

宮内 秀彰 Miyouchi Hideaki  
渡邊 まり子 Watanabe Mariko  
塚原 朋哉 Tsukahara Tomoya  
佐藤 俊也 Sato Syunya

## 1. はじめに

近年コンピュータの大容量化が進み、多くの情報を容易に蓄積できるようになってきた。中でも文章形式で記述された情報の増加が目立っているが、情報を整理しようすると基本的には人が読まなければならない現状がある。

企業でも社内にノウハウを貯めるため、情報共有のためにコンピュータに多くの情報を蓄積している。お客様に関わる文章情報としては、営業日報やコールセンターへの問い合わせ、お客様アンケートなどを日々蓄積している。CS 向上を掲げている企業としては、顧客の要望を把握するためにも、これら文章情報を有効活用させたいと考えている。

しかし大量のテキスト文章を一つ一つ読んで分析することは非常に時間がかかり非現実的である。そこで昨今アンケートなどの自由記述文章を分析し新たな知見を得るためにテキストマイニング技術が用いられており、CS 向上に役立てられている。<sup>1)</sup>

## 2. VOCAnalyzer 開発の経緯

日立東日本ソリューションズ(以下、日立 TO)の顧客である H 社も CS 向上を掲げている企業である。H 社ではユーザに、製品に関するヒヤリングを行ってお客様の声を収集している。このお客様の声を製品設計に活用する

方法として QFD(Quality Function Deployment)を検討しており、その有効性を確認している。

QFD は市場のお客様の声を分析し自社技術と対比させ、重要性の高い要素は何かを明らかにする手法である。

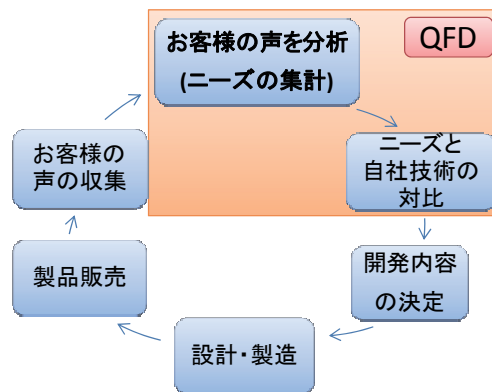


図 1 QFD を用いた製品開発の業務フロー

お客様の声の分析方法を図 2 に示す。まずひとつひとつの文章を読み解き、内容を要約し、要約内容から文章をグループ化する。グループ化した文章から分類体系を作成し、文章数を集計することでニーズの傾向を分析する。

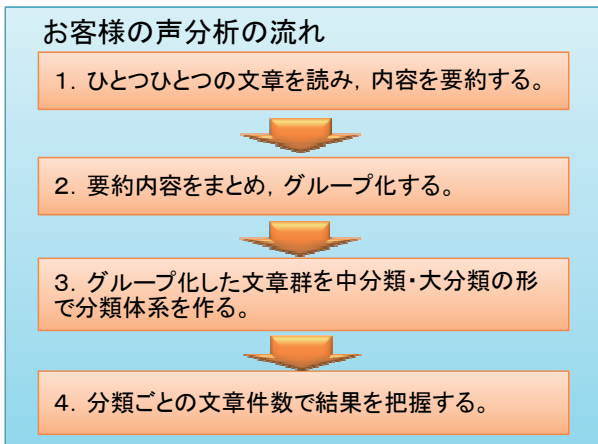


図2 手作業によるお客様の声分析の流れ

このお客様の声を手作業で分類する作業の負荷が非常に高い。QFDにおいて、お客様の声を分類する作業負荷が最も高く専門性が問われるため、この部分の作業負荷を低減させることが大きな課題だった。

そこでお客様の声の分類にテキストマイニングを用いることで分析負荷が軽減できるのではないか、との要望から日立 TO のテキストマイニング製品である CoreExplorer を用いた共同研究を開始した。その後テキストマイニングによるお客様の声の自動グループ化の有効性を示すことができた。しかし、CoreExplorer は単語の関連から探索的にテキストを分析することを主軸にしているため、文章分類の十分な機能を備えていない。そこでお客様の声の分類に特化させたツール、VOCAnalyzer を CoreExplorer のエンジンを用いて開発することとなった。

### 3. 課題に対する VOCAnalyzer の解決法と効果

これまで H 社が図2の手順で行ってきたお客様の声を分類する作業に対する課題と VOCAnalyzer を用いてどのように解決を図ったか、またその効果について述べる。

#### 3.1 手作業における文章のグループ化の課題

##### 【課題】

図2の1から2は、似た文章を抽出しグループ化する作業である。ひとつひとつの文章を読み解いて何が書かれているかを把握し、分類するための指標となるように特徴的なキーワードで要約を行う。そのキーワードごとにグループ化を行う作業である。

すべての文章を読んで、文章のグループ化を行う作業は、書かれている文章の分野に関してある程度の有識者でないと、グループ化を誤る可能性があり誰でも行える

というものではない。さらに、ひとつひとつの文章を読まなければならないため、非常に時間のかかる作業であることは明確である。

##### 【解決方法】

CoreExplorer のエンジンが持つ、文章クラスタリング機能を用いることで、類似した文章を自動的にグループ化することが可能である。文章で使用されている特徴的なキーワードを文章ごとに比較し、類似した文章でグループ化を行う機能を開発した。

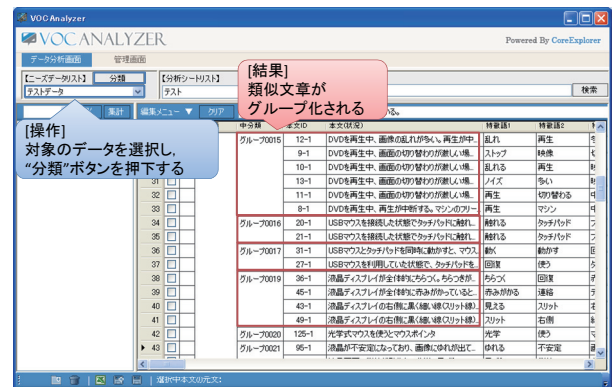


図3 文章自動グループ化機能

##### 【効果】

文章クラスタリング機能により自動的にグループ化することができ、より短時間に客観的な分析が可能となる。

##### 【作業時間における効果】

###### (1) 手作業の場合

##### 【条件】

- ・1000 件の自由記述文章
- ・要約に要する時間を 1 分/件と仮定
- ・グループ化に要する時間を 30 秒/件と仮定

##### 【作業時間】

要約：1000 分≒16 時間  
 グループ化：500 分≒8 時間  
 合計：24 時間 を要する。

###### (2) 自動グループ化の場合

##### 【条件】

- ・1000 件の自由記述文章
- ・CPU: Core2Duo 3.33GHz
- ・メモリ: 3GB
- ・HDD: SATA 7200rpm

##### 【処理時間】

自動グループ化処理時間：5 分

### 3.2 分類体系作成の課題

#### 【課題】

図 2 の 3 から 4 は、グループ化された文章群から分類体系を作成する作業である。分類体系を作るためには、まずグループの特徴を把握し名前をつけてから、グループごとの特徴を体系立てて整理して分類体系を作成する必要がある。手作業の場合、グループ化された文章は担当者の主観であることが多いため、明示されておらずグループ化された文章群を見ただけでは、グループの特徴を把握することは困難である。また、分類体系を体系立てて構築するにはグループごとの関連を把握する必要があるが、グループ同士の関連を判断する材料を見つけることは困難である。

このようにグループ化された文章群だけでは、分類体系の上位・下位のカテゴリを整理することが困難である。

#### 【解決方法】

- (1) 文章グループの特徴を把握するため、文章の特徴的なキーワードを抽出することを可能とした。この機能は、ある文章で使われており他の文章であまり出現しない単語は、重要度が高く設定され特徴語の上位に出現する。一方他の文章でもよく頻出する単語は重要度が低く設定され、特徴語の上位には出現しない。各文章に対して上位 15 個の特徴語を表示できるようにした。この機能により、グループ内で共通して使われるキーワードが明示される。
- (2) グループの関連を把握するため、検索ワードと関連性のあるキーワードを一覧表示することを可能とした。この機能は、ユーザが指定したキーワードを入力し検索をすると、そのキーワードを含む文章で使われている、特徴的なキーワードを表示する。この機能により、指定したキーワードがどういったキーワードと関連があるのかを把握することができる。

(その他分類体系作成の主な機能)

- ・ ツリーによる分類体系管理
- ・ ドラッグ&ドロップによる文章の分類
- ・ ドラッグ&ドロップによる特徴語・関連語の追加
- ・ ドラッグ&ドロップによる分類項目の移動
- ・ 集計ボタンによる自動集計

#### 【効果】

- (1) グループ内で共通に抽出されている特徴語を見ることで、グループの内容を把握し、項目名を決める

判断材料とすることができる。

- (2) 検索ワードの関連語を見ることで、そのキーワードが他のどのようなキーワードと一緒に使われているかを知ることができる。分類項目をまとめる際に、有効である。

このように、文章の特徴的なキーワードを判断材料として分類体系構築の作業を支援することが可能である。グループから分類体系を構築するには、上記 2 つの機能によるキーワードを判断材料としてツリー表記された分類体系にドラッグ&ドロップで直観的に分類体系を作成できる。



図 4 分類規則作成の支援機能

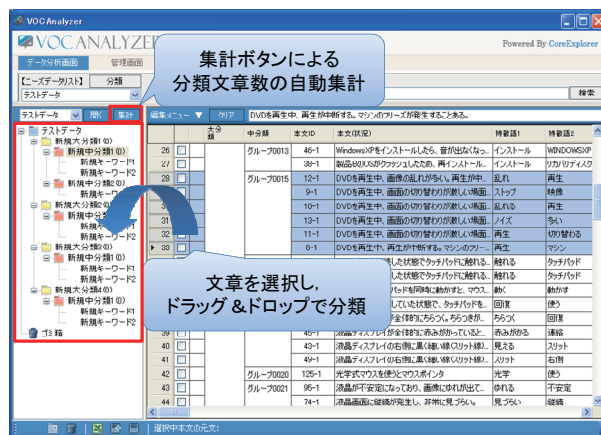


図 5 文章の手動分類

### 3.3 分類体系作成後の課題

分類体系は不変ではなく、その時々状況によって体系を変えて分類しなおすニーズがある。徐々に変化していくニーズに対して、分類体系もメンテナンスを行い、最新の体系にする必要がある。しかし、手作業で分類体系をメンテナンスしようとした場合、新たなデータを用いて最初から分類体系を作成しなければならない。

【解決方法】

分類体系構築そのものが大変な作業であるため、一度作成した分類体系を再利用することを可能とした。作成した分類規則に設定したキーワードを検索条件として、対象データに対し自動で分類を行う。分類規則に存在するキーワードが文章中に存在した場合、その文章は自動的に分類される。

【効果】

既に作成した分類規則でデータを分類できることで、毎回分類規則を一から作る必要がなくなった。既に作成した分類規則で文章は自動的に分類され、新しいキーワードを含む文章は「その他」に分類される。その他に振り分けられた文章から分類体系を作成しなおすことで、容易にメンテナンスすることができる。また、様々な分類規則を用いて分類し、別キーワードで作成した分類体系での集計結果を瞬時に得ることで、多角的に分析することが可能である。

4. VOCAnalyzer 自動分類機能の分類精度評価

H 社において、製品に関するヒヤリングをエンドユーザに対して行った結果を分析対象として、手作業による分類と VOCAnalyzer による自動分類の結果を比較した。

評価方法は自動分類の結果が手作業の分類にいかに対応したかで評価した。

【条件】

対象データ：H 社製品に関するヒヤリング結果

データ件数：1894 件

手作業によって文章を分類した結果を正解の分類する。

手作業によって作成された分類規則は、大項目が 15 項目、中項目が 66 項目であった。各中分類に自動で分類するためのキーワードは、H 社に依頼し各中分類の文章群から特徴的なキーワードを選出していただいた。

【結果】

手作業の分類結果に対して適合率の高い分類の例として 85.7%の割合で自動分類することを確認した。

しかし適合率の悪い分類では、11.1%しか分類されないことを確認した。全体の平均は 44.8%であった。

適合率の高い分類では、「デザイン」のキーワードが登録されているなど、分類そのものを示すキーワードが登録規則に含まれている場合が多い。逆に、適合率の低い分類の例としては、「排気ガス=排ガス」などの表記の違いで分類されない場合や、「環境」の分類に「環境」というキーワードがない、など分類そのものを示すキーワード

が分類規則に含まれていない場合が多かった。

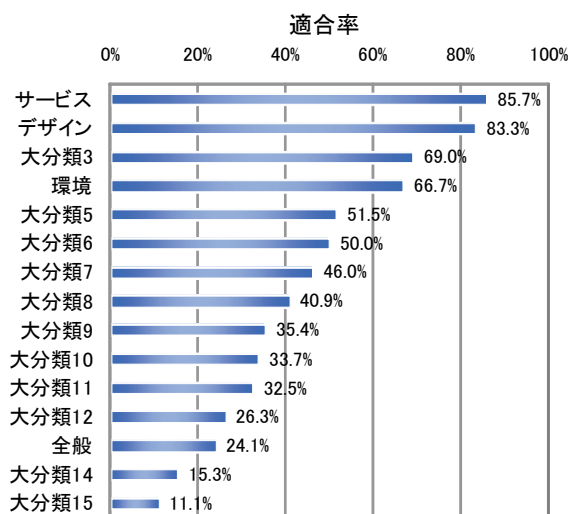


図 6 自動分類による大分類別適合率

この結果から、同義語辞書のメンテナンスと分類規則のキーワードを的確に選出することで、さらに適合率が向上することが示唆される。しかし分類規則のキーワードを的確に選出することは、分類規則作成者のスキルに依存してしまうため、容易ではない。分析者のスキルに依存しない、ユーザビリティの高いツールにするためには、分類規則のキーワード選定を自動もしくは半自動で行う機能が今後必要であることを示すことができた。

5. 今後の展望

5.1 VOCAnalyzer の自動分類における精度向上

手作業による分類体系構築に比べ、VOCAnalyzer の文章自動グループ化と分類構築の操作を工夫したことによって大幅に作業時間短縮が見込める結果が得られた。しかし、分類規則による文章の自動分類機能においては、自動分類の精度向上が今後の課題であることが示された。

5.2 QFD の作業支援ツールとしての拡張

今回作成した VOCAnalyzer において、QFD を行うプロセスの 1 作業である「お客様の声を分析する」フェーズにおいては、作業負荷を軽減させることが可能であることが示された。QFD には図 1 に示されるように、ニーズと自社の技術をマトリクスで比較し、開発すべき内容を分析するフェーズが存在する。このフェーズの分析を VOCAnalyzer の機能に加えることで、QFD の専用ツールとして拡張させることが可能である。



### 5.3 VOCAnalyzer の機能拡張

現在の VOCAnalyzer はアンケートなどの自由記述文章部分の、分類分け作業・集計作業に特化したツールである。アンケートには、自由記述文章だけでなく 5 段階評価や定められた内容にチェックを入れる定型項目の調査結果が含まれることが多い。よって、5 段階評価や定型項目の集計結果と自由記述文章の分類結果をクロス集計することで、文章の分類だけでなく多角的な分析が可能となる。またグラフ化・レポート機能を追加することで、お客様の声分析だけでなく、テキスト情報を含んだ様々なデータを分析することができる BI ツールとして拡張させることも可能である。

## 6. おわりに

今回 VOCAnalyzer を開発したことで、H 社が課題としていた、QFD におけるお客様の声分析の作業負荷を軽減させることが可能であることを示すことができた。

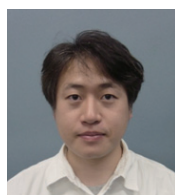
多くのデータが電子化されている現在、社内に蓄積しているテキストデータを有効に活用したいと考える企業は多く存在していると考えられる。VOCAnalyzer はテキストデータならば汎用的に分析することが可能なことからお客様の声に限らず、特許データや不良報告データ、プロジェクト報告データなどの技術的な文章の分析も可能である。今後は CoreExplorer との差別化を図りながら、製品化に向けて取り組んでいきたい。



宮内 秀彰 2008 年入社  
ナレッジソリューショングループ  
CoreExplorer, テキストマイニングツールの開発  
hideaki.miyauchi.01@hitachi-to.co.jp



渡邊 まり子 2004 年入社  
ナレッジソリューショングループ  
CoreExplorer を利用したテキスト分析・コンサルティング  
w\_mariko@hitachi-to.co.jp



塚原 朋哉 1997 年入社  
ナレッジソリューショングループ  
CoreExplorer, テキストマイニングツールの研究開発  
tomo@hitachi-to.co.jp



佐藤 俊也 1993 年入社  
ナレッジソリューショングループ  
CoreExplorer, テキストマイニングツール, 知識管理システムの拡販, コンサルティング  
shu\_sato@hitachi-to.co.jp

### 参考文献

- 1) 林田英雄, 脇森浩志: テキストマイニング技術とその応用, UNISYS TECHNOLOGY REVIEW, 第 84 号, pp.29-44(2005)
- 2) 長野徹, 武田浩一, 那須川哲哉: テキストマイニングのための情報抽出, 情報処理学会研究報告, No.91, pp.31-38(2000)
- 3) 那須川哲哉: テキストマイニングを使う技術/作る技術, 東京電機大学出版局 (2006)