

VOC マイニングと Analyze-Psychos

株式会社NTTドコモ 情報システム部 市瀬 眞 ichise@nttdocomo.co.jp

1. はじめに

多くの研究、製品が近年躍進になる自然言語解析技術の活用により、アンケートの自由文章の回答分析から CRM 活動として、お客様の声 (VOC:Voice of Customer) を収集し、傾向分析やトピックス・アラートを早期に把握するなどのリスクマネジメントにも対応が可能であることから企業において「データマイニング」同様に考えられることが多くなってきた。しかし、手段としては適切ではあるが、その曖昧性を認識した「テキストマイニング」として導入目的や成果へのプロセスが確立されているか。また、analyst の理解する「テキストマイニング」は単なるテキスト情報の体系化で処理していないか。多くの企業の事例を調査・ディスカッションから、「成功していない」という悩みや相談も多い。

数社のテキストマイニング製品等を利用した経験から、役に立つ「テキストマイニング」とは、VOC の資源を有効に活用する手段として利用され、初期においては自然言語解析の可能性を実感できるひとつの例として、analyst の習熟性と感性から発生した新たな分析モデルの構築があることが一つの機軸となる。

本稿では、テキスト化された大量の情報から定量的に体系化することを分類 (主題分析) として、定性的にアラート・トピックス抽出を VOC の可視化と整理する。また、テキスト情報を自然言語解析することにおいて、人の **need** や **want** といった感覚単位でセグメントすることを VOC マイニングという。これらを実現するためにユーザ視点として自然言語解析をおこなう analyst の心理状況と思考課題に対して評価するものである。

2. VOC マイニングにおける課題と対策

自然言語解析においては、概念、学習、共起、相関、構文解析等の何れの手段や技術を活用するとしても、成果を求めるためには多くのアイデアを必要とする。前述した **needs** と **want** は目的として分析する語彙の目標レベルでもある。

「VOC の可視化」への達成レベルの例として、2つの **needs** と **want** がある。

image1:1 件の VOC からの **needs** と **want**

「i モードの絵文字を自分で自由に作成できないですか？」
これを理解推測すると

テーマ	何をしたいか (want)	どうしたい(needs)
i モード	メールの絵文字を作成して	送信したい。(推測可)

文章中に意味する **needs** と **want** がある。絵文字を作成できても送れないなら意味がない。という関係から成り立つ。逆に「絵文字が作れないから送信しない」にはならないことから定性的な観点と捉えられることがわかる。

image2:複数の VOC にある **needs** と **want**

① ニーズ性のある VOC レベル

VOC-A:「地下鉄で電波が入らなければ携帯を持つ意味はない。」→MUST なレベル

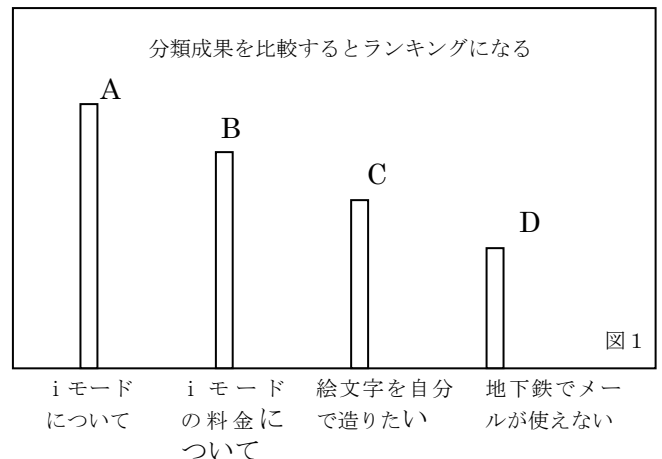
② ウォント性のある VOC レベル

VOC-B:「地下鉄で利用できればいいなあ」→希望レベル

これにより、テキストにある言葉と理由から顧客を理解した、VOC マイニングをおこなうためには、感情レベルでの主題分析によるマイニングが必要となる。これらを量の変化として定量的 (カテゴリとする) に時系列で分類する。また、少ない件数ではあるが内容に重要性があるものを VOC マイニングでは定性的な分類として整理する。

定量的な主題分析をおこなうために、辞書・ルール (キーワード等) のチューニングをおこない、マイニングを実施した結果の量を比較することで VOC のランキングが成り立つ。これにより日々の変化を把握することで「VOC の日々の可視化」【1】とする。

以下 (図1) は主題分析で分析した4つのカテゴリになるが、初期の段階で発生しやすい問題点を整理する。



VOCのランキング可視化には2つの問題点が生じる。

(1) セグメンテーションにおける問題点

- ① A~B = 「~について」と大きなテーマや、まとめやすい大括りなカテゴリとなり、C~D = **needs**のある評価であることから、カテゴリA~Dまでの分類(主題分析)においてカテゴリを階層的(図2)にみてる。
- ② A、Bについては「テーマ」「大項目」レベルから、詳細化された「**needs・want-line**」では文章の意味あいが理解できるレベルまでになることから時系列や相関、共起性を把握するために必要な階層レベルとなる。このレベルでカテゴリ間を比較分析することがポイントになる。

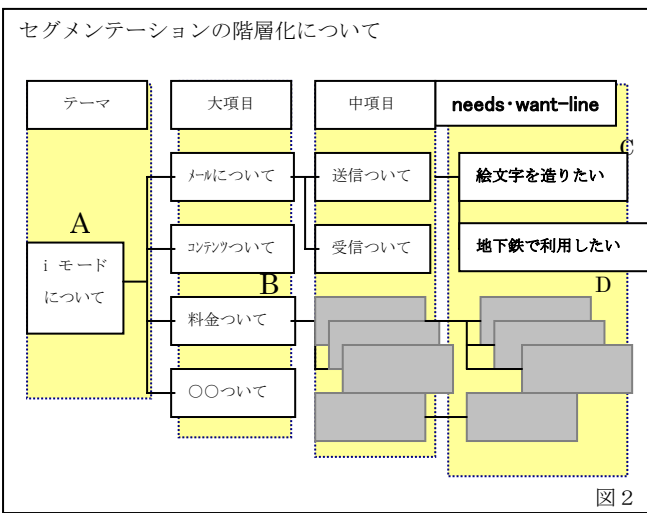


図2

(2) 品質と精度

VOCマイニングにおいて、マイニングの品質精度を考えたときに2つの精度を認識する必要がある。

①全体からの品質(取得できないVOC): 全体のVOCを母集団としたときに、「**needs・want-line**」のカテゴリに抽出されるべきVOCのうち何%の抽出率があったか。(図3)

②カテゴリの品質「**needs・want-line**」(誤取得したVOC): カテゴリに抽出されたVOCのうち異なる(誤抽出)カテゴリをひいたときに残りのVOCは何%になるのか。

この2つの品質を維持することで量的変化にも対応し、曖昧性を補完したテキストマイニングとして位置付けることが可能となる。(図3)

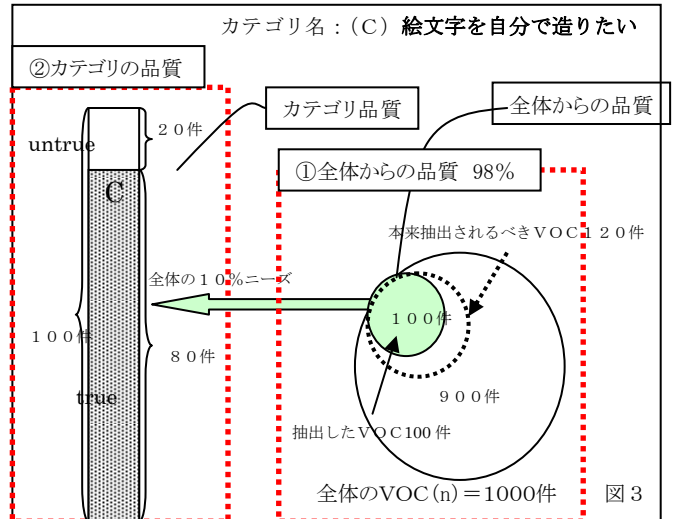


図3

analystの分析視点において、これらを標準化することでVOCマイニングと定義する。

3. Analyze-Psychos

VOCマイニングを成功するためマーケットやanalystにおいて目的や分析習熟過程に発生するメンタル面を中心とした課題等に問題と対策を整理し分析をおこなう。

(1) analyst視点によるVOCマイニングに向けた課題

①マイニングの前準備	VOCの質を見分ける ※「表記のゆれ」等含む	多くの表現を抽出し把握するルール、辞書メンテナンスの充実
②階層化が整理する(カテゴリ化)	カテゴリで悩む 「needs・want-line」の共有化	①イメージを共有する ②ロジカルシンキング手法 ③メンタル的な対応
③思考イメージからマイニングする	マイニング成果の感覚を得る	メンタル的な対応
④カテゴリを標準化する	VOCからカテゴリ化する	カテゴリイメージを伝える。 「末尾が要望・否定・称賛」等の感情表現が好ましい
⑤言語解析 ・ツール(製品)習熟する ・日本語の難しさ	トレーニングによる対応	関連度、頻度、相関性を意味、理解する ツール(製品)で遊ぶ
⑥専門用語への対応	辞書・ルールでの対応	新サービスに迅速に対応 未知語のマーケティング
⑦アナリストの標準化	業務スキル、マイニングスキル、操作性(製品)のスキルの習得	メンタル的な対応
⑧品質の維持	どれだけとれているか?	セグメントで仮説をたてる
⑨分析内容の充実	起因・改善案の考えが浮かばない	考えない。VOCの不満をカテゴリ化し、階層下位の具体性の高い要望カテゴリを起因・対策と考える。
⑩気づきを得る習慣をつくる	コミュニケーション(気づきの共有)	メンタル的な対応

これらは言語解析機能において、共起、相関、概念といっ
ても多くは理解しにくく、暗黙知として思考が理解している
ケースも多い。

これは、与えられたカテゴリの多くを目で読むことで言語処
理を体感し経験値から規則性や「VOCのあるべき姿が見
えているため、普遍的な要素に気づきが発生する。しかし、
経験や業務スキルだけでは考えにくい結果を見ることも多
く感性的な違いにより成果が異なることが多いのではない
かと推察する。また、最も困難な課題として、②②階層化が
整理する（カテゴリ化）「needs・want-line」の共有化⑦
analyst の標準化がある。⑩気づきのスキルや経験の違い
だけでなく「気づきの視点も違う」「深い理解力」等につ
いて改善するためには analyst のマイニング心理と把握による
対処のために検証をおこなう。

(検証1)

業務スキル保有者とスキルがない analyst の違い。

分類・分析のためには業務やサービスを理解しなければい
けないが、抽出能力だけで考えれば業務スキルは MUST
ではない結果がでた。

検証内容「名義変更」に関するVOCを抽出能力として母集
団数を分析する。

	平均抽出数(目標数:232件)
業務スキルチーム 12名	158件
業務スキル無チーム 10名	212件

「名義変更」のVOCを抽出する → このなかで固有名詞
の出現率や関連度を聞くこと(見ることで)結婚という名詞が
あることがわかることで、ルールの中に「結婚」で母集団を
増やすことは可能であるが、大量な VOC の中からは「名義
変更」カテゴリでは60%ぐらいの品質となる。analyst は品
質が見えないためこれで簡潔することになる。大切なのは
名義変更の理由が結婚で全てではないということに「気付き
」を得るか? 「離婚」という語彙が思考に展開されるかであ
る。これは関連度等や頻度に出現しないこともあり、出現し
ても関連性があると認識できないことが多い。

しかし、30 回程度実施するが、業務スキルの有無では決定
的な結論はでない。このことから、「気づき力」の重要性を
認識できるが、(検証2)個人別経験・習熟性等について更
に検証する。

	言語解析スキル有	言語解析スキル無
業務スキル有 12名	1	3
業務スキル無 10名	2	4

この業務スキルの有無、言語解析の経験保有の有無を下
記の1~4のチームに分別し1回目「A:ドコモからNTTに電
話するときの通話料金」と2回目「B:NTT からドコモへ電
話するときの通話料金」を検証する。

	1回目 (正解数143件)	2回目 (正解数172件)	平均
1 (6名)	123(102件)	153(143件)	①正解率:87% ②抽出率:77%
2 (6名)	134(92件)	192(162件)	①正解率:78% ②抽出率:80%
3 (4名)	151(131件)	232(162件)	①正解率:77% ②抽出率:93%
4 (6名)	90(79件)	261(121件)	①正解率:57% ②抽出率:63%

この結果、高度なマイニングを要する分析には、経験・スキ
ルが有効であることは把握できたが抽出率と正解率を各グ
ループ個人別にみると、前提が崩れることが確認できた。
このためマイニングにおいて「感性」を前述したように直接
的にヒヤリングしてもパーソナリティに影響されるため深層
心理を中心に検証を実施する。

(検証3)感性を捉えた検証

マイニング 難度	1(高) ニーズ	2(高) サービ ス	3(低) サービ ス	4(高) ニ ー ズ	5(中) サー ビ ス	6(低) ニ ー ズ	7(低) サービ ス
Aさん (1チーム)	黄	深緑	黄	緑	黄	緑	青
Bさん (1チーム)	青	赤	緑	紫	青	青	緑
Cさん (2チーム)	灰	青	水色	青	黄	黄	水色
Dさん (3チーム)	黒	茶	灰	黒	黄	黒	灰

※太字は成果率が高い

成果率(太字)と色パターンとの関係に着目し全体的に得意
とする「サービス系」2、3、5、7は縦横比では変化がある。
また横に(一人一人)をみて成果率が高い色と引く場合の
色が比較的顕著に現れことがわかった。特に needs を中心
とした VOC マイニングでは、マイニング後に自信があるか
ないかとアナリストに聞いても個性による回答傾向にあるが、
メタファ【2】において検証すると、高度な分析での共通性と
パーソナル別での変化から補強ポイントを把握することが
可能ではないかと推測する。

これらの対策として、補強策としての機能開発やトレーニ
ングでの対応だけでなく、得意とするカテゴリごとに業務別
に担当させることや、アドバイザー的な存在、検証やルール
作成等を適材配置及び感性の高い analyst の有効利用

により困難である VOC マイニングを進めることは強化カリキュラムにより補強することで改善が図れる。

4. まとめ

企業が求めるVOCを活用した分析は多種多様ではあるが、目的を明確に提示し、言語解析で可能な許容範囲であれば曖昧マイニングから有効なVOCマイニングとしてとても有効な手段となることが評価できる。しかし、目的をさらに企業のマネジメントモデルやサービスの充実をはかるために、コールセンタ等では取得困難である。また、サイレントユーザの声から新たなPOSITIVE・Negative感情を定量的に捉えるために、Blogでの分析を研究・調査し希望性の高い素朴な要望「この色が良い」「おかしいと思う」や、こころの思い・葛藤や素直な意見を抽出し企業におけるトランスピアレンシー化を実現化することを課題とする。また、自然言語解析技術の発展に資することとして研究・調査・開発・業務が直接的に交流することで加速化したテクノロジーはVOCを反映したサービスや企業の在り方へと更に模索する素材となり社会に影響を与えるようなPDCAサイクルへと繋げる課題がある。

【参考文献】

- 【1】2004.10.18 財団法人 企業情報化協会（現 IT協会）経済産業省月間講演
資料名「顧客理解力によるコンタクトマーケティングの潮流」講演者 市瀬 眞
- 【2】2005.3.17 マーケティング協会講演資料名 「How Customers Think」
講演者 Pr.Gerald Zaltman