

# マーケティングミックスとテキストマイニングを用いた 市場分析支援

## Marketing Analysis using Marketing Mixture and Text Mining

竹内広宜\*1\*2  
Hironori TAKEUCHI

杉山喜昭\*2  
Yoshiaki SUGIYAMA

太田千景\*2  
Chikage OTA

山口高平\*2  
Takahira YAMAGUCHI

\*1 日本アイ・ビー・エム株式会社 東京基礎研究所  
IBM Research - Tokyo

\*2 慶應義塾大学  
Keio University

In order to acquire valuable knowledge through text mining, it is generally critical to identify important expressions to be monitored and compared within the textual data. It is therefore very important to construct appropriate analysis viewpoints for each target domain. It is also not clear how we can effectively use the analysis viewpoints to get insights through the text mining. In this paper, we introduce the viewpoints for the marketing analysis based on the marketing mixture and propose the method to reach to the trends leading to the insights in the marketing analysis. Through the analysis of real textual data, we will show that we can effectively get insights through the proposed method.

### 1. はじめに

近年、サービスの基本構造を科学的に明らかにするサービス・サイエンスの取り組みが始まっており、その中の活動で、オントロジーの観点からサービスの構造や知識を「見える化」する試みがされている [8]。一方で、蓄積されたテキストデータを知識発見に活用するテキストマイニング技術が研究・開発されている。特に CRM(Customer Relationship Management) の分野では蓄積された顧客の声をテキストマイニング技術を通して分析し、ビジネスに活用することが行われている [10]。本研究では、テキストマイニング技術を用いてサービスの構造を「見える化」し、サービスやビジネスの改善につながる知見を得ることを検討する。

テキストマイニング分析の目的の一つは、テキスト内に出現する概念(キーワードや表現)について、出現頻度の時間的分布や特定の文書集合内における共起情報の中から特徴的なものを見つけ出し、知見につなげることである。その中でも、2つ以上の文書集合があった場合、集合間の差に相当する概念を抽出する差異分析は非常に有効な分析の一つである。例えば企業のコールセンターに蓄積されるデータ(コールメモ)にはエージェントが書くテキストデータの他に、問い合わせ対象製品やエージェント名等が定型データとして付与されている。このようなデータを分析する際、製品またはエージェントごとの文書集合に注目し、集合間の差を見つける分析は、特定の製品に固有の問題の同定や、優良エージェントが持つ know-how の発見・共有などにつながることから、テキストマイニングの魅力的なアプリケーションである。CRM の分野だけでなく、経営分析においてもテキストマイニングを用いた継続企業と倒産企業の差異分析が行われている [6]。市場分析に上記の差異分析を適用することで、購買者而非購買者の違いや、顧客満足度の高い顧客と満足度の低い顧客の違いに関する知見を得ることが期待できる。

しかし、分析観点の設定や分析時にどの分析観点間に注目すればよいかは、分析者の対象分野に対する知識やスキルに依存する。本論文では、市場分析の主な分析手法であるセグメント分析に対して、マーケティングミックスを考慮した分析観点とテキストマイニング分析手法を提案する。そして、実データ

を用いた分析を通して、提案した手法の有効性を検証する。

### 2. 市場分析のための分析観点

#### 2.1 マーケティングミックス

マーケティングミックスとは企業が、対象とする市場での目的を達成するために用いるマーケティング要素の組み合わせである [4]。製品提供のためのマーケティング要素として、Product(製品やサービス自身)、Price(価格)、Promotion(販売促進活動や広告)、Place(提供場所、立地、流通範囲)の4つが定義されていた(4P)。サービスの場合、さらに Physical Evidence(デザインや機能性)、People(販売員やサービスを提供する全ての要員)、Process(サービスを提供する手段)の3つを加えたマーケティング要素が組み合わせの対象となっている(7P)。

#### 2.2 顧客が感じる品質 (SERVQUAL)

マーケティングミックスはサービスや商品を提供する側の視点であるのに対し、サービスや商品の品質を測定するための顧客側感覚尺度として以下の5つの観点(5D)がSERVQUALとして提案されている [5]。

- Tangible(有形性):物的施設・設備の内容、接客員の外見等
- Reliability(信頼性):サービスの遂行に関して信頼し、期待する結果が実際に提供されること
- Assurance(確実性):サービスの遂行に関して十分な技能を持つ、顧客への丁寧さを持つ、リスクを抱かない
- Responsiveness(反応性):素早いサービスを提供する意欲を感じるかどうか
- Empathy(共感性):十分な情報提供、積極的な顧客理解

#### 2.3 市場分析における分析観点の活用

2.1, 2.2 で述べた観点と、サービスや商品を提供される顧客の属性(性別、年齢、家族構成、年収など)と合わせることで図1に示すような分析観点(オントロジー)を構築することができる。顧客がサービスや製品について経験や知識を持っている場合、マーケティング要素の組み合わせによって顧客が期待する品質が形成される。また、顧客属性の組み合わせによっても顧客が抱く期待品質が変わると考えられる。従って、

- 7P $\iff$ 5D
- 5D $\iff$ 顧客属性

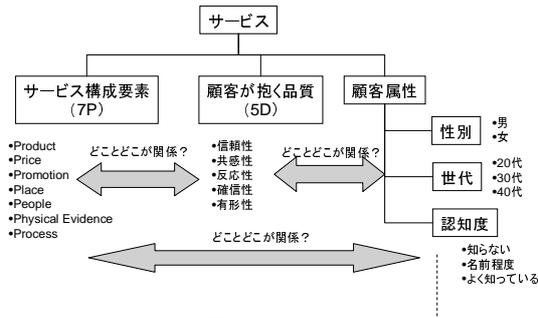


図 1: 分析観点

の 2 種類の関連を分析することで、顧客とサービスの間の関係を品質という観点で視える化することができ、マーケティングにつながる知見を導出することが期待できる。

一方、新規の製品やサービスの場合、顧客は対象製品やサービスに対して十分な経験や知識を持っていないと考えられる。この場合はマーケティング要素と顧客属性の間の関連より、マーケティングにつながる知見を得ることになる。

### 3. テキストマイニングを用いた市場分析

#### 3.1 テキストマイニングを用いた差異分析と課題

テキストマイニングでは通常、テキストデータから自然言語処理技術を適用し、分析対象となる表現を抽出し蓄積する。そして、蓄積されたデータに対してインデックスを張り、様々な観点からデータを対話的に俯瞰できるテキストマイニングシステムを構築し、知見を得る試みがされている [2]。

文書集合間の差を求める手法の一つとして、分析観点(カテゴリ)を事前に定義し、各観点到に属する概念を辞書として準備する方法がある [9]。例えば製造業のコールセンターでは部品名、苦情、要望、質問といった観点とそれぞれに該当する表現を事前に集め辞書に登録し分析に用いられている。こうすることで製品ごとに分けられた文書集合を分析する際、製品名を縦軸、設定した観点到に属する表現を横軸にとり、両属性を満たす文書数を表す 2 次元表を作成することで、文書集合間の関係を概観することができる。しかしながら、分析観点的定義や辞書の準備は手作業であり、サンプル文書中における頻出語を元に分析者の勘を頼りに作成されることが多いという課題がある。

一方、特定の文書集合に特徴的な表現を自動的に抽出する方法があり [3]、注目している文書集合の特徴を分析するために用いられている。また、分析だけでなく、文書分類においてもこういった指標を利用し、文書分類に有効な素性を選択することが行われており [7]、文書クラスタリングにおいても各クラスタのラベル付けに指標が用いられている [1]。しかしながら、こういった手法で抽出されるキーワードリストの上位に来る語は、注目している文書集合で非常に多く出るといった観点で特徴的であると判定されるだけであり、必ずしも有用な知見につながらないことが多いという課題がある。

#### 3.2 順序関係を持つ属性を考慮した分析観点的選択

ある概念  $k_j$  の対象テキストデータにおける言及の度合い(出現確率)を  $f_c(k_j)$  とし、セグメント数が  $n$  であるセグメント軸 ( $seg$ ) での分析を考える。セグメント軸の値が順序関係

$$seg_1 < seg_2 < seg_3 < \dots < seg_n$$

を持っている場合、

$$f_c(k_j|seg_1) > f_c(k_j|seg_2) > \dots > f_c(k_j|seg_n)$$

または

$$f_c(k_j|seg_1) < f_c(k_j|seg_2) < \dots < f_c(k_j|seg_n)$$

なる傾向を持つものを検出することを考える。このとき、 $f_c(k_j|seg_i)$  から

$$g_c(k_j|seg_i) = \sum_{h=1}^i \frac{f_c(k_j|seg_h)}{n f_c(k_j)} \quad (1)$$

を求め(正規化累積確率)、 $i/n$  ごとにプロットすると、図 2 で示すような結果が得られる。

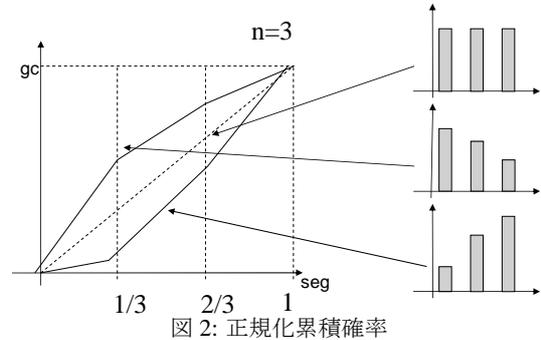


図 2: 正規化累積確率

$f_c(k_j|seg_i)$  がセグメント値によらない場合、 $g_c(k_j|seg_i)$  は傾き 1 の直線となる。この直線に対して、 $f_c(k_j|seg_i)$  がセグメント値の増加に対して減少する傾向を持つ場合は上に凸、増加する傾向を持つ場合は下に凸の折れ線となる。この折れ線と傾き 1 の直線で囲まれる面積を求め上に凸の場合は正、下に凸の場合は負の符号をつけ、選択したセグメント軸の特徴量 ( $S_{seg}$ ) とする。 $S_{seg}$  の符号および絶対値を見ることで、セグメント値の増加に対する注目している概念の言及の減少または増加傾向とその強度を判断することができる。これにより例えば市場分析では、“顧客属性〇〇が大きい(高い)ほど、マーケティング要素△△を気にする”といった傾向が得られれば、マーケティング戦略につながる知見を得ることができると考えられる。

### 4. 市場分析例

#### 4.1 未普及製品の販売促進のためのテキストマイニング分析

本分析例では、未普及製品を販売促進するための戦略につながる知見を顧客アンケートデータから得る。分析対象の製品は生ゴミ処理機とした。生ゴミ処理機は、電気と特殊な触媒で、生ゴミを分解し肥料にする機械であり、現在の普及率は 2% である。この製品を購入している人、購入していない人を対象に、アンケート調査を行いデータを 600 件取得した。テキストデータとして、“どういった課題が解決されれば生ゴミ処理機を買うか?(何が購入する前に気になったか?)”という質問に対する回答を自由記述形式で取得した。また、顧客の属性情報として、性別、年代、家族人数、年収、製品の認知度(全く知らない、名前だけ知っている程度、機能を把握している)を取得した。

アンケート中の順序関係を持つ顧客属性に 3.2 で提案した分析観点的選択手法を適用し、得られた特徴量  $S_{seg}$  を表 1 に示す。Price, Product, Promotion といったマーケティング要素への言及

表 1: 分析観点的特徴量 ( $S_{seg}$ )

	家族人数	世代	年収	製品認知度
Price	-0.00165	0.0148	-0.0366	0.0366
Product	0.0130	-0.00452	-0.0240	-0.0693
Promotion	0.0159	0.0748	0.214	0.0642

がそれぞれのセグメント軸でどういった傾向を示すのかを知る

には、マーケティング要素とセグメント軸の2次元の表を作成し、言及頻度を求め傾向を観測する必要があった。通常、どのセグメント軸を選択すればいいかわからないため、従来は必要と思われる全てのセグメント軸に対して分析を行う必要があった。それに対して、本研究で提案する手法を用いることによって、表1から **Promotion** については年収が高くなるにつれて言及が減少するという傾向を予測できる。また、製品認知度というセグメント軸では、製品を知っているほど **Price**, **Promotion** の言及頻度が下がり、**Product** の言及頻度が上がっていることが予想される。実際に、セグメント軸として年収と製品認知度を選択した場合の言及頻度の傾向分析結果を図3に示す。

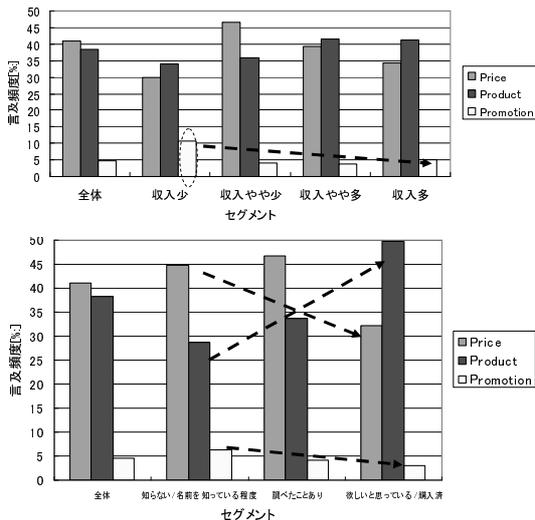


図3: セグメント軸を年収(上)と認知度(下)にした場合のマーケティング要素の言及頻度

一方、表1から家族人数、世代といったセグメント軸は特に **Price**, **Product** について顕著な増加・減少傾向は見られないと予想される。図4にセグメント軸として世代を選択した場合の言及頻度の傾向分析結果を示す。

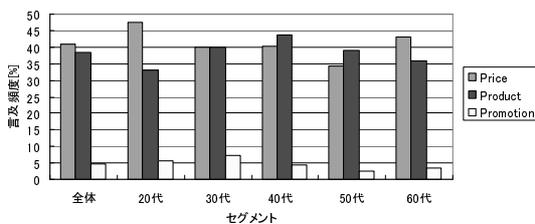


図4: セグメント軸を年代の時のマーケティング要素の言及頻度

分析結果から、マーケティングミックス(4P)と顧客属性の間には年収が低い顧客や製品を知らない顧客ほど、販売促進やキャンペーンや価格が気になるという傾向がわかる。製品の普及率を考えると消費者の大部分が生ゴミ処理機の詳細を知らないと考えられる。従って、補助金の存在や妥当な値段で購入できることを広く周知させる必要があると考えられる。また、製品を知っている消費者ほど、製品自体に対する不満点や期待を持っていることがわかった。

これらの知見から、販売促進のために、多人数向け広告や家電量販店などでは妥当な値段で生ゴミが処理できることを訴求し、興味を持った人が多く集まると思われる場所(ホームセンターなど)で実演販売を行う、といった戦略案が立てられる。

#### 4.2 サービス改善のためのテキストマイニング分析

ここでは利用したホテルに対する良かった点と悪かった点に関するアンケートコメントをテキストマイニング分析する。具

体的には、分析観点をを用いて、ホテルサービスの各サービス構成要素が持つ特性、顧客が抱く期待品質、顧客属性らとの間の関係から、ホテルサービスの改善につながる知見を得ることを試みる。

まずサービスのマーケティング要素が持つ特性として、各構成要素の品質分類を考える。品質の分類として魅力的品質・当たり前品質・一元的品質があり[11]、それぞれ以下の特徴を持つ。

- 魅力的品質：充足されると満足するが、不充足でも仕方がないと感じる
- 当たり前品質：不充足であれば不満を感じ、充足でも当たり前と感じる
- 一元的品質：充足されると満足し、不充足であれば不満を感じる

図5に良かった点と悪かった点のそれぞれのコメントにおけるマーケティング要素の言及頻度を示す。

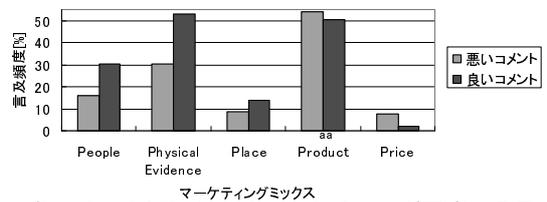


図5: 各コメント欄におけるマーケティング要素の言及頻度

各顧客のコメントには顧客が重視する点が定型項目として付与されており、それらは分析観点の顧客が抱く期待品質と結びつけられる。これにより、顧客が抱く期待品質ごとにマーケティング要素が良かった点、悪かった点に関するコメントでの言及頻度を求めることが可能になる(図6)。

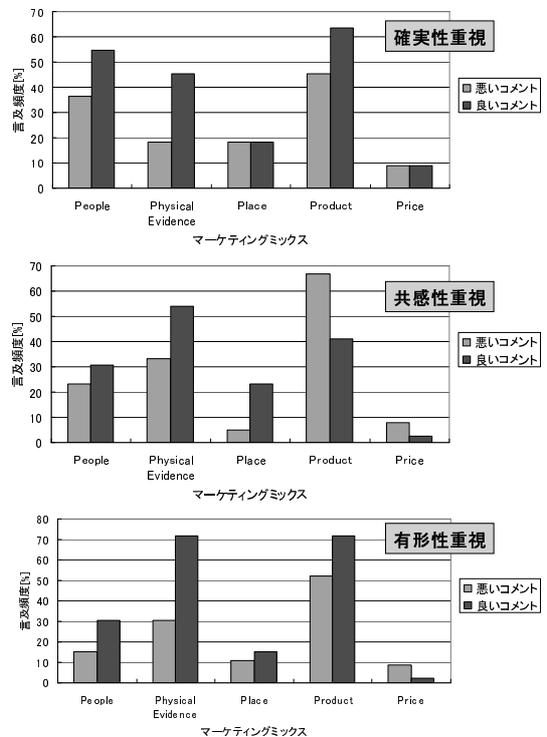


図6: 確実性(上)、共感性(中)、有形性(下)を重視する顧客コメントにおけるマーケティング要素の言及頻度

図5と図6に示した言及頻度の差から品質分類に関して以下の仮説が立てられる。

- 確実性重視の顧客にとって **Product** は魅力的品質である
- 確実性重視の顧客にとって **Place** は一元的な品質である

- 共感性重視の顧客にとって Product は当たり前品質である
  - 有形性重視の顧客にとって Product は魅力的品質である
- しかし、サービス提供者側がこれらの仮説をサービス改善に結びつけるのが難しい。なぜならば、サービス提供者側が顧客が5Dのうち何を重視しているかどうかを事前に知る事が難しいからである。サービス提供者側が得られる情報は顧客属性(世代, 性別, 人数)や提供している価格などである。

これらの顧客属性のうち順序性を持つものに対して、顧客の期待品質の重要視の増減傾向を持つかどうかを提案手法を用いて導出した。得られた特徴量  $S_{seg}$  を表2に示す。ここか

表2: 分析観点の特徴量 ( $S_{seg}$ )

	世代	価格帯
保証性	-0.0114	-0.00448
共感性	0.0391	-0.0465
有形性	-0.0144	-0.0801

ら, ”年代が低いほど共感性を重視する”, ”価格帯が高いほど有形性を重視する”, ”価格帯が高いほど共感性は重視しなくなる”という傾向が顕著であることが推測できる。これらの傾向は実際の分析結果(図7)で確認できる。この結果と品質

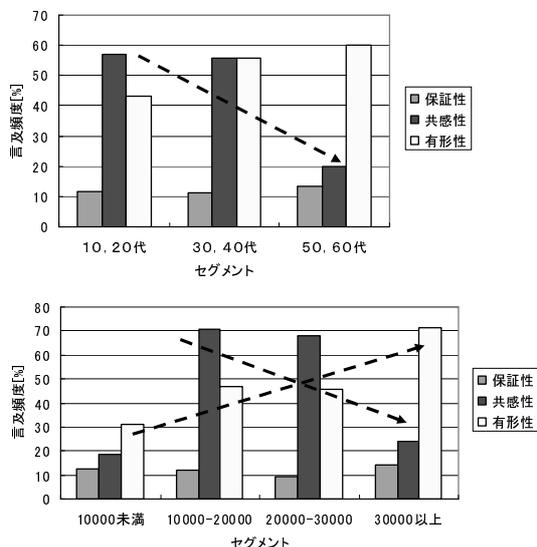


図7: セグメント軸を年代(上)と価格帯(下)の時の期待品質(5D)の言及頻度

分類から得られた結果を組み合わせることでホテルの形態にあった、サービス改善のための仮説を立てることができる。例えば、若い世代の利用者が多いホテルの場合、共感性重視の顧客が多いことが予想できる。共感性重視の顧客には Product が当たり前品質になっているため、設備が最低限の基準を満たしているかどうかを検証する必要があると考えられる。また、価格帯が高いホテルは有形性重視の顧客が多いことが予想される。有形性重視の顧客にとって Product は魅力的品質であり、Physical Evidence (眺望など) は一元的品質である。したがって、最新設備や他にない設備の考案が重要であり、眺望が良ければ積極的にアピールする価値があると考えられる。

## 5. まとめ

本研究では、テキストマイニングを用いた差異分析における課題となっている分析観点の利用手法について、対象にデータに付与されている定型データが順序特性を持つ場合にその性質を利用した分析手法を提案した。また、市場分析にテキストマイニングを適用する際に汎用的に使えるマーケティング視点を取り入れた分析観点を定義した。そして、2種類の市場分析を行い、収集した顧客の声に提案した分析観点と分析手法を適

用し、製品の販売促進やサービス改善につながる知見が得られることを示した。他のデータで同様の分析が広く行えることの検証や収集するテキストデータの種別に応じた分析手法の拡張が今後の課題である。

## 参考文献

- [1] F. Beil, M. Ester, and X. Xu. Frequent term-based text clustering. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 436–442, 2002.
- [2] R. Feldman and J. Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York, 2007.
- [3] T. Hisamitsu and Y. Niwa. A measure of term representativeness based on the number of co-occurring salient words. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 1–7, 2002.
- [4] P. Kotler, T. Hay, and P. N. Bloom. *Marketing Professional Services*. Pearson Education, 2002.
- [5] A. Parasuraman, A. Zeithaml, and V. A. Berry. Servqual: A multiple-item scale for measuring customer perceptions of service quality. *Journal of Retailing*, pp. 12–40, 1988.
- [6] H. Takeuchi, S. Ogino, H. Watanabe, and Y. Shirata. Context-based text mining for insights in long documents. In *7th International Conference Practical Aspects of Knowledge Management (PAKM 2008)*, pp. 123–134, 2008.
- [7] Y. Yang and J. O. Pedersen. A comparative study on feature selection in text categorization. In *Proceedings of the 14th International Conference on Machine Learning (ICML)*, pp. 412–420, 1997.
- [8] 高藤, 溝口. オントロジー工学に基づくサービス・サイエンスへのアプローチ. 第22回人工知能学会全国大会予稿集. 3B3-7, 2008.
- [9] 那須川. コールセンターにおけるテキストマイニング. 人工知能学会誌, 16(2):219–225, 2001.
- [10] 那須川. テキストマイニングを使う技術/作る技術—基礎技術と適用事例から導く本質と活用法. 東京電機大学出版局, 2006.
- [11] 具, 中條. 魅力的品質・当たり前品質を中心とする消費者品質要求のモデル化. 品質, 31(4):105–118, 2001.