

「暇」ツイートテキストマイニングによる潜在的ニーズ分析 Potential Needs Analysis by Text Mining the "Boring Tweets"

町田悠貴[†] 大槻明[†]
Yuki Machida Akira Otsuki

1. はじめに

総務省「平成27年情報通信メディアの利用時間と情報行動に関する調査[1]」によると、全年齢におけるソーシャルメディアの利用時間が調査開始の平成24年から年々増加傾向にある。また、内閣府の「平成27年度 青少年のインターネット利用環境実態調査[2]」では、18歳未満の青少年のインターネット利用目的の第3位がコミュニケーション(メール、メッセージ、ソーシャルメディアなど)と示されており、若い世代のソーシャルメディア利用率が極めて高いことが伺える。このような状況下において、第2章に示すように、最近では、ユーザの潜在的ニーズを明らかにするために、ソーシャルメディア上の消費者の投稿からサイレントカスタマー分析を行う研究や、サイレントマジョリティー分析を行う研究、さらには、ソーシャルリスニング分析などの研究が盛んに行われており、消費者の潜在的な需要の発見に役立てられている。しかし、「暇」と対外的に意思表示をしているユーザが何を求めているのか、という観点でユーザの状況を分析することで、ユーザの潜在的なニーズを明らかにするようなアプローチは未だに提案されていない。ゆえに、本研究では、Twitterを題材として、「暇」と意思表示をしているユーザのツイートを収集し、「暇」と共起関係がある語彙(名詞)を分析することで、ユーザの潜在的なニーズを明らかにし、マーケティングや地方創生に資する知見を得ることを目的とする。

2. 先行・関連研究

熊坂ら[3]は、ソーシャルな時代において潜在的ニーズを把握する方法論として、東京ガールズコレクションや森ガールなどを題材として、「柔らかい構造化手法」を活用し、つぶやき等のサイレントなロングテールが示す微細な行動を分析することによって、ユーザのライフスタイルの組み立て方や、社会文化事象を解析する方法論を提示した。ここでいう「柔らかい構造化手法」とは、ユーザのつぶやきや趣向などの非構造化データを、レイヤー分割、クラスタリング及びツリー化して全体を構造化することでコミュニティ関連図を作成することである。なお、非構造化データについては、後述の3.1節の中で述べる。また、消費者の潜在的な需要を分析する手法としてサイレントマジョリティー分析[4]やソーシャルリスニング分析[5-6]が挙げられる。これらの分析アプローチは、消費者がソーシャルメディアやブログに投稿する会話や文章を、ポジティブ・ネガティブ判定、キーワード抽出及びグラフ化といった手法を用いて分析することで、サイレントカスタマーやサイレントマジョリティーの状況を明らかにしようとするアプローチで

表2 暇ツイート数の人口比割合(「割合(abe/人口)」の上位10件を表示)

都道府県	市区町村	2015年(暇ツイート数)				2016年(暇ツイート数)								ave (暇ツイート数 平均)	人口	割合 (abe/人口)
		9月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月			
大阪府	東大阪市	247	835	636	613	660	423	406	648	712	666	791	492	594.1	488,613	0.00122

ある。しかし、本研究が提案するアプローチは、「暇」と対外的に意思表示しているユーザの状況を明らかにしようとするアプローチであり、本章で概観した先行研究とは内容的に異なるものであると考えられる。

3. 提案コンセプト

3.1 SNS などの非構造化データ

前章の先行研究では、ツイートなどの非構造化データを分析していたが、本研究でもユーザのつぶやきという非構造化データを分析するため、本節では、構造化データとの比較(表1)を通じて、本研究における非構造化データの特徴について述べる。

表1 本研究における非構造化データと構造化データの比較

非構造化データの例	構造化データの例		
日大太郎さんは東京都出身で 35 歳。経済次郎君は若くて 28 歳で神奈川県生まれ。	名前	出身	年齢
	日大太郎	東京都	35
	経済次郎	神奈川県	28

表1の「非構造化データの例」列をコンピュータが理解できるようにRDB形式に書き換えたものが、「構造化データの例」列である。従前のコンピュータでは、非構造化データそのままでは「日大太郎」が名前であることも理解できなかったが、現在では、AIやテキストマイニング等の技術が発達したことで、この問題も徐々に解決されてきた。しかし、2015年現在においても、つぶやきなどの分析に適さない非構造化データは約8割存在していると指摘されており、これら非構造化データを分析することで新たな市場が生まれる可能性があるという指摘されている[7]。

ゆえに、本研究では、「暇」を含むつぶやきを分析することで、ユーザの潜在的なニーズを明らかにすることを主なコンセプトとしている。

3.2 ツイート取得

Twitter APIを用いて、2015年9月~2016年8月の1年間で、全国で「暇」というキーワードが含まれるツイートを約80万件取得した。そして、自治体ごとに「その自治体における暇ツイート数/自治体人口」を計算して、暇ツイート数の人口比割合を調べた(表2)。表2に示すとおり、暇ツイート数の人口比割合の上位5つの自治体は、大阪府岸和田市、大阪府和泉市、京都府伏見区、大阪府東大阪市、愛知県豊橋市であったため、本研究ではこれら自治体のツイートを対象として次節の分析を行うこととした。

[†] 日本大学

大阪府	岸和田市	166	634	526	597	643	457	398	639	556	406	494	294	484.2	201,701	0.00240
大阪府	和泉市	119	462	380	415	354	312	252	366	292	228	247	199	302.2	182,678	0.00165
京都府	京都市伏見区	171	640	470	521	549	407	371	515	467	326	435	321	432.8	274,201	0.00158
愛知県	豊橋市	106	439	375	688	606	381	339	374	359	224	462	277	385.8	365,143	0.00106
東京都	新宿区	122	438	169	334	276	271	241	375	363	308	401	266	297.0	281,783	0.00105
大阪府	八尾市	134	429	344	379	310	236	202	309	264	226	263	230	277.2	265,518	0.00104
大阪府	枚方市	162	469	338	376	332	249	242	296	323	226	292	192	291.4	406,253	0.00072
宮崎県	宮崎市	76	280	156	289	311	274	296	343	306	235	324	277	263.9	372,032	0.00071
愛知県	豊田市	67	361	186	284	301	270	257	372	330	276	396	218	276.5	406,821	0.00068

注) 人口は、総務省サイト：「市町村別の人口及び世帯数」[8]のデータを用いた。

3.3 「暇」と共起関係がある語彙の抽出

暇ツイート数人口比割合が高かった上位5つの自治体のツイートを対象として、T値及びMI値を求めることで、「暇」と共起関係のある名詞を抽出した。T値は次式のように表される。

$$T = \frac{Co - \frac{NoN \times CoN}{AllNo}}{\sqrt{Co}} \quad (1)$$

NoN(NodeNo)は注目している語の頻度、CoN(CollocateNo)は注目している語と共起関係にある語の頻度、AllNo(AllNode)は総語数、Co(Collocate)は共起頻度をそれぞれ表す。つまり、T値は、共起頻度を総語数や期待度数で調整した値である。

次に、MI値は次式のように表される。

$$MI = \log_2 \frac{Co \times AllNo}{NoN \times CoN} \quad (2)$$

NoN(NodeNo)は注目している語の頻度、CoN(CollocateNo)は注目している語と共起関係にある語の頻度、AllNo(AllNode)は総語数、Co(Collocate)は共起頻度をそれぞれ表す。つまり、MI値は、共起頻度をその期待値で割り、さらに2を底にした対数を取った値である。MI値は、低頻度の単語であっても共起関係を抽出できるという特徴を持つ。

石田[9]によると、言語コーパス学において、共起関係が認められる判断できるT値とMI値の閾値は、「T値>=1.65」かつ、「MI値>=1.58」であるため、本研究では、この閾値を参考に、3.2節のツイートを対象として、「暇」と共起関係が認められる名詞を抽出して分析を行った。

4. 実験

前節のT値及びMI値を用いて「暇」と共起関係が認められる名詞を抽出した結果を図1に示す。

まず、図1の①に示されているとおり、季節や曜日関係の名詞(例：土日-暇、クリスマス-暇)が最も多く抽出されていたが、しかし、週末やクリスマスなどのイベント時を対象とした販売促進等に関するマーケティングは現状でも既に一般的に行われていると考えられるため、本研究では考察の対象からは除外した。図1から、現状一般的に行われていないと考えられる名詞は、愛知県豊橋市でみられた「新幹線-暇」や「お風呂-暇」、さらには、多くの自治体でみられた「待ち時間-暇」や「夜勤-暇」などであったため、次章で「夜勤-暇」など、マーケティングや販売促進などには直接つながりにくいようなものを除き、「新幹線-暇(5.1節)」及び「待ち時間-暇(5.2節)」を取り上げてマーケティングや販売促進、さらには地方創生に関するアイデアの可能性について一考察を行う。

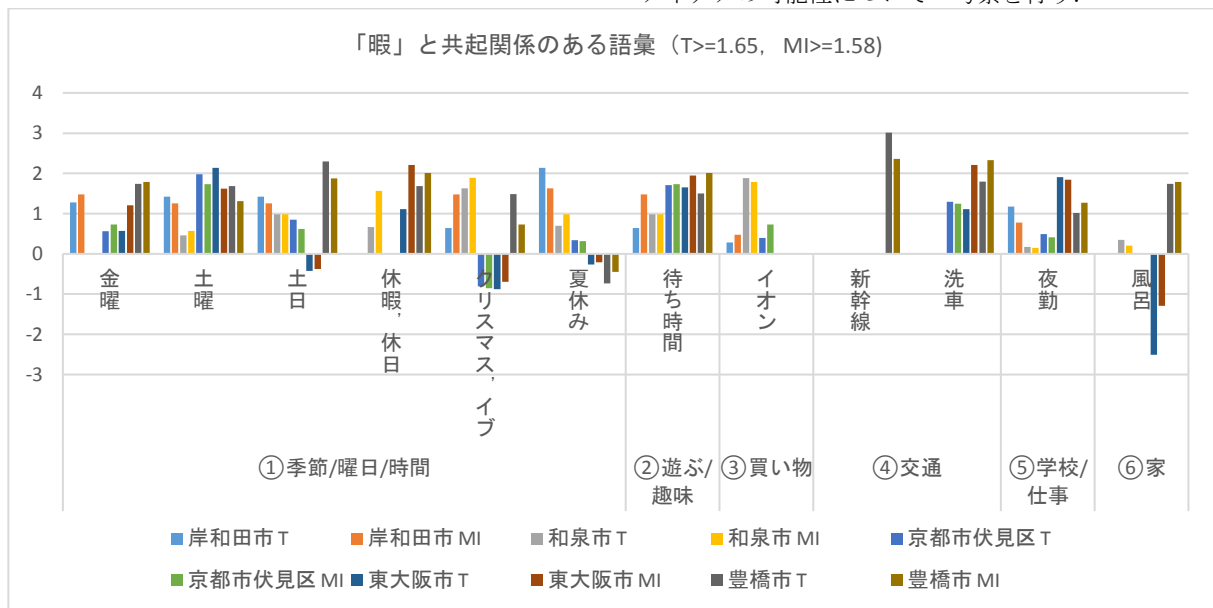


図1 「暇」と共起関係が認められた名詞 (暇ツイート数の人口比割合の上位5つの自治体)

考察

5.1 愛知県豊橋市でみられた「新幹線-暇」

「新幹線」を含むツイートを確認したところ、「新幹線の中で暇している」という内容が一番多かった。確かに、愛知県は、地理的に、東京、関西、九州など多方面に新幹線を使って出張などをする機会が多いと考えられるため、愛知県において「新幹線-暇」の共起関係が認められたのではないかと考えられる。そこで、地方創生に係る一考察を次のように行った。

飛行機のように、新幹線の前席にモニターを設置して、例えば、名古屋(豊橋)から博多に出張する人がいた時に、博多までの通過駅(各都市)の紹介・PR動画を流すなどすれば、地方創生(地域活性化)にもつながるような効果が期待できるのではないかと考えられる。あくまでイメージでの一例であるが、例えば、新幹線が新倉敷-福山間と通過しているときに、福山市のPR動画(例:スポット紹介)を流す、というように、次の駅に到着するまでの間に、その次の駅のPR動画を流すことにより、次のような効果が期待できると考えられる。

① 出張時に新幹線で閲覧した店などのPR動画に興味を持った人が、

② 後日プライベートでその店を訪れる、

また、自治体側の視点に立てば、「PRする場が増える」ことを意味し、この場の活用法は幅広く考えることができる。さらに、ユーザがこのモニターを操作して情報検索(ネット検索)等ができる仕組みを準備すれば、さらなる効果が期待できると考えられる。

5.2 多くの自治体でみられた「待ち時間-暇」

「待ち時間」に関する全てのツイートを確認したところ、「待ち時間-暇」の内訳、つまり、何の待ち時間が暇なのかという内訳が表3に示すとおり明らかとなった。

表3 「待ち時間-暇」の内訳(一例)

自治体	「何」の待ち時間が暇なのか
大阪府岸和田市	眼科
大阪府和泉市	遊園地のアトラクション、 免許更新
京都府伏見市	免許更新
大阪府東大阪市	病院、みどりの窓口
愛知県豊橋市	病院、カラオケ、電車、授業

表3から、本研究では特に「遊園地のアトラクション」を取り上げて一考察を行う。例えば、ディズニーランドのアトラクションであれば、待ち時間4時間以上ということも珍しくないが、図2に示すとおり、ディズニーランドの各アトラクションにはスポンサーがついている(図7だと新日本石油)。また、ディズニーの某アトラクションでは、ユーザがそのアトラクションに向かうスタンバイ通路に、そのスポンサーの商品をディズニーのイメージに合わせる形で展示しているケースがあるが、このように、スタンバイ通路にデジタルサイネージやモニターを設置し、そのスポンサー関係のPR動画等をディズニーのイメージを崩さないように流すことで、マーケティング等に資する効果が期待できるのではないかと考えられる。



図2 ディズニーのスポンサーの例
(出典:著者撮影)

以上、本節では、「暇」と共起関係が認められた名詞を取り上げて一考察を行ったが、このようにユーザシチュエーションを分析することで、マーケティングや地方創生に資することが期待される知見を得ることができたため、本研究のアプローチの有用性について確認することができたと考えられる。

5. おわりに

本研究では、「暇」と対外的に意思表示をしている人が何を求めているのか、という観点で、ユーザ(消費者)のシチュエーションを分析することで、ユーザの潜在的なニーズを明らかにするアプローチについて提案した。第4章の実験及び第5章の考察では、暇を含むツイートを80万件以上取得して、「暇」と共起関係のある名詞を抽出して考察した結果、マーケティングや地方創生に資することが期待される知見を得ることができた。今後は、名詞以外の品詞についても分析することや、また、表3に示した、「暇」と「何らかのサービス」を結び付けた分析について、さらに掘り下げて研究していきたいと考える。

参考文献

- [1] 総務省情報通信政策研究所:「平成27年情報通信メディアの利用時間と情報行動に関する調査報告書」,平成28年8月。
- [2] 内閣府サイト「平成27年度青少年のインターネット利用環境実態調査」:<http://www8.cao.go.jp/youth/youth-harm/chousa/h27/net-jittai/pdf-index.html>。
- [3] 熊坂賢次,山崎由佳:「ソーシャルな時代・柔らかな構造化手法そしてライフスタイル論」,特集「ネットワークが創る新しい社会」,AD STUDIES,Vol.40,pp.22-27,2012。
- [4] 神田晴彦:「新しい顧客価値を生むソーシャルCRM-サポートサービスと顧客の声収集の変容-」,ITソリューションフロンティア,Vol.45,No.2(通巻338号),pp.14-15,2012。
- [5] KOBIT ブログ:「顧客の"声なき声"に耳を傾けよう!ソーシャルリスニングの事例まとめ」:<https://kobit.in/archives/2181>。
- [6] Marketing Bank ソーシャルリスニングってなに?おすすめツール3選:<https://www.marketingbank.jp/special/cat01/15.php>。
- [7] 大河原克行:「図解ビッグデータ」,中経出版,pp26-27。
- [8] 総務省サイト「参考資料4 市区町村別の人口及び世帯数」:<https://kobit.in/archives/2181>。
- [9] 石田基広:「Rによるテキストマイニング入門」,森北出版,2008。