

消費者委員会 消費者安全専門調査会
(第29回)
議事録

内閣府消費者委員会事務局

消費者委員会 消費者安全専門調査会（第29回） 議事次第

1．日時 平成29年3月30日（木） 16:00～17:31

2．場所 消費者委員会会議室

3．出席者

（委員）

山本座長、藤田座長代理、相澤専門委員、市瀬専門委員、西田専門委員、
村田専門委員

（消費者委員会委員）

河上委員長、大森委員、長田委員、中原委員、樋口委員

（説明者）

（株）プラスアルファ・コンサルティング 浦坂グループマネージャー

（事務局）

黒木事務局長、福島審議官、丸山参事官、友行企画官

4．議事

（1）開会

（2）第28回消費者安全専門調査会における議論の整理

（3）テキストマイニングを活用した事故情報の分析等

（4）閉会

1 . 開会

山本座長 それでは、時間ですので、開始したいと思います。

皆様、本日はお忙しいところをお集まりいただきまして、ありがとうございます。ただいまから第29回「消費者安全専門調査会」を開催いたします。

それでは、配付資料の確認を事務局からお願いいたします。

友行企画官 配付資料でございますが、議事次第の下でございます、資料1、資料2ということでお配りしております。不足があります場合は、事務局までお願いいたします。

2 . 第28回消費者安全専門調査会における議論の整理

山本座長 それでは、最初の議題は「第28回消費者安全専門調査会における議論の整理」です。

事務局から、10分程度でお願いいたします。

友行企画官 それでは、資料1を御覧いただけますでしょうか。「第28回消費者安全専門調査会における議論の整理」ということでございまして、前回の専門調査会の議論の整理でございます。

まず最初は、「事故情報の公開における留意点等」とまとめておりますが、前回、東京大学の宍戸教授から「消費者事故情報の公開について」ということで、消費者事故情報の公開の位置付けですとか、事故情報のオープン化の意義ですとか、オープン化する際の留意点などについてお話をいただきました。

その際に主に議論された論点でございますが、まず最初は「事業者の不利益」等ということで、1のところでございますけれども、オープン化することによる企業への過度な責任追及のおそれにつきましては、この製品を使ったらこういうことが起きたというものであって、原因に対する評価を含むものではないので、直ちに事業者の責任追及に直結するものではないのではないかと。

あと、そもそも安全法や消費生活用製品安全法は、まず事故情報を一元的に集めて、それによって対策を打っていくという仕組みになっているので、事故情報の公開によって、直ちに過度の責任追及が進むものではないのではないかとという御意見でございました。

それから、2番目といたしまして、こちらも同じく「事業者の不利益」に関するところでございますが、事故情報を公開する際、あるいは公開を受けた人が利活用をする際に、情報の中に製品の型番等のメタデータを明らかにすることによって、その後の改善等によっ

て問題を解決したことなどの現在の情報が分かるようにする、あるいは一定の期間を過ぎた場合、非公開とするといったような工夫も必要ではないかという御意見がございました。

それから、「公開の相手方の限定範囲」の考え方でございますけれども、事故情報を公開することについては、原則は法の下での平等との関係があって、合理的な差別はありますけれども、合理的な差別的取扱いは許されるという考え方がある。そのため、事故情報を何のために公開して、それによってどのように消費者保護を達成していくかという制度設計の中で検討すればよいのではないかという御意見がございました。

それから、「公開の原則」の考え方でございますが、事故情報は原則として公開するという考え方もございますけれども、公開を望まない消費者もいるので、原則公開というのは難しいのではないかという御意見がございました。

それから、事故情報の公開は、事故の被害の拡大防止や原因究明の促進などのメリットがあることを前提にしつつ、公開する場合にはどのようなことに気を付けなければならないのかということについて議論を深めていくことが大事だろうという御意見がございました。

おめくりいただきまして、次の整理でございますけれども、「情報の詳細度」などにつきましてでございますが、事故情報については、もともと記述の詳細度が異なっている。そういうことから、できるだけ入力する時点で方式をそろえることが重要であるのではないか。それから、情報の入力、取得あるいは集約の局面、いろいろな面がございますが、そもそもデータをオープンにする。オープンデータ・バイデザインの発想でやっていくことが必要ではないかという御議論がございました。

それから、事故情報についてですが、車両については、例えば運転者は車について一定の知識があって、詳細な情報が入力されているものもあるけれども、必ずしもそういうものばかりではなくて、商品の知識があるものもないものも含まれている。詳細な報告ができないものについては、一定の加工や範囲を絞った上で公開するという工夫も考えられるということ。

次のところでございますが、事故情報は後日、分析したり、公表することも考えて、それがやりやすいような形で整えられることが必要ではないかという御意見がございました。

それから、今回はもう一つ、独立行政法人日本スポーツ振興センターというところから、事故情報の収集とか活用の取組について、お話をいただきました。

それを受けての議論でございますけれども、まず「情報の公開」につきましては、日本スポーツ振興センターにおいては、公開している情報については、死亡案件全てと障害見舞金を支払った全件について公開しています。その際には、名前、学校名等の情報は削除していますという御議論がございました。

それから、この日本スポーツ振興センター自体は数百万件というデータが蓄積されていますけれども、実際にデータベースで公開されているのは6,000件程度となっております。ただし、公開の請求があった研究者の方と守秘の契約を結びながら、実際にデータベースで公開している以上のデータを公開するというのもやっているというお話もございまし

た。

それから、ここが抱えている「分析の課題」といたしましては、日本スポーツ振興センター自体でもいろいろ分析しているけれども、まだ自分たちで分からない、把握していないような分析手法もあると思われるので、ワーキンググループなどで分析の手法について検討している。例えば、AIなども活用できないかという観点も関心を持っているというお話でございました。

それから、「取組の評価」につきましては、事故情報を分析して公開することによって、どれだけ事故が減ったかといった取組に対する評価も必要だと考えているということでございまして、評価の方法についても、専門家から意見をいただいて検討しているというお話もございました。

それから、「情報の発信」の部分につきましては、このセンターでは、学校に冊子を配布したり、全国の学校保健会、PTA連合会などの場で配布しているという発信の仕方をしていくというお話がございました。

それから、「その他」の議論といたしまして、このセンターのデータは、学校で起きた事故が全部把握されており、非常に有用であるといった御意見ですとか、あとは、JSCの関心事項と当専門調査会における関心事項は似通っているところがあるので、お互いに引き続き情報交換していくことが重要であるという御意見もございました。

以上でございます。

山本座長 それでは、ただいまの説明の内容につきまして、御質問、御意見のある方は御発言をお願いいたします。いかがでしょうか。

はい。

河上委員長 前回、出席できなかったのですが、お伺いしたいのですけれども、この学校の事故に関して、消費者事故であるかどうかということについて、特に、事故の報告義務が、これまで学校からは必ずしも十分行われていなかったということが前に問題になったことがございました。この日本スポーツ振興センターが集めている事故の中で、消費者事故としているものについては、消費者庁にもこの情報の報告が行われているものなのですか。

友行企画官 行っていると答えていました。

山本座長 そのほかにもございますでしょうか。

よろしければ、これはこれで確定と申しますか、また蓄積していきたいと思えます。

3. テキストマイニングを活用した事故情報の分析等

山本座長 続きましての議題は、「テキストマイニングを活用した事故情報の分析等」です。本専門調査会では、専門委員の皆様それぞれの専門分野の知見を活用して、事故情報データバンクに登録されている情報の分析をしていただいておりますけれども、民間の事業者におきましてテキストマイニング等の分析技術を活用した情報分析に取り組んで

おられる方々もいらっしゃいます。

本日は、企業等からの依頼を受け、テキストマイニング技術を活用した情報の分析などに取り組みられている株式会社プラスアルファ・コンサルティングにお越しいただいております。プラスアルファ・コンサルティングには、一般に公開されている事故情報データバンクに登録されている情報の分析、それから、SNS上の事故情報の分析などをお願いしております。今日は、そのために席の配置も通常と違っておりますけれども、30分程度でプレゼンをお願いして、その後、いろいろ質問等、あるかと思しますので、質疑の時間を設けたいと思います。

それでは、よろしく願いいたします。

(株)プラスアルファ・コンサルティング 甫坂グループマネージャー 株式会社プラスアルファ・コンサルティングの 甫坂と申します。本日は、お時間いただきまして、ありがとうございます。お手元の資料にもありますが、一部、こちらのスライドにしか入っていないものもありますので、基本的には前の画面で御覧いただければと思っております。

私のほうからお話しさせていただくのは、今回、事故情報データバンクのデータをお預かりして、私たちの持っているテキストマイニングと呼ばれる、簡単に言うと、日本語で書かれた、いわゆる文字情報の山というものを機械に解析させることによって、おおよそこんな話題があるとか、この年代の人にはこういう発言の傾向がありますみたいなものを抽出するような技術を使って解析した結果を御覧いただければと思っております。

データを見ていただくに当たって、今回、一部、性別・年代情報等、もともとのデータで入っているものを使用しているのですが、こちら、入力されているもののみを使用しておりますので、仮に全てのデータに今後、属性が登録されたなどの場合には、別の解析結果になる可能性もあるので、そこは事前に御了承いただければと思っております。

また、「傷病の程度」というデータがもともとのデータの属性の中に入っているのですが、今回、私たちのほうで「重大事故」というものを属性として作っておりまして、これは「傷病の程度」が「1か月以上」もしくは「死亡」となっている方を重大事故に位置付けて、今回は使用しています。

こちらがデータの概要です。今回、おおよそ1年分のデータをお預かりして、データ件数としては2万7,000件ほどの解析結果を御覧いただきます。テキストマイニングの対象としては、事故の内容が詳細に記載されているもの、自由記述の内容を使用しています。

実際の文章としては、例えばこんなものになります。乗合バスが運行中、乗客1名が転倒し、右足骨折の重症。共同住宅3階、窓の転落防止柵に布団を干していたところ、転落防止柵ごと転落して重傷を負いましたなどといった形で、通常であればこういったデータ、2万7,000件全部読んで内容を把握するというのが今までのやり方であったのですが、私たちはこれをテキストマイニングという技術を使っているということです。

こちらが対象のデータになります。

早速、ここから解析結果を御覧いただければなと思うのですが、テキストマイニングと

いう技術において一番最初にやることは何かというと、先ほど文章で書かれていたテキストデータを、全て単語ごとに分割していきます。そうすることによって何が分かるかといいますと、おおよそ何が話題になっているのかというのが件数で比較されるということです。

今回は、特に日本語の中でも、左側が名詞、右側が動詞のランキングを表示しています。実は、分けているのにも意味があって、左の名詞のランキングを御覧いただきますと、1位が火災、2位が走行中、4位がメーカー、業者等々ありますが、こうして見ていただくことで、2万7,000件全て読まなくても、一番登録が多い情報としては、火災についての記載ということが分かりますし、次に走行中。実際に中を見ていただくと、高速道路とか車、いろいろな走行中があるのですが、その話題が2番。3番以降は、業者やメーカーの対応についてのトラブルの話題が多いということも分かります。

今度、動詞を見ていきますと、消費者がどういった行動をとったときに、こういうトラブル、事故に見舞われるのかというのが分かりやすくなるという意味です。例えば1番は、何かを購入したとき。もしくは、焼損とか出火。火災が多いので、炎損・出火というのが比較的出ているところになります。

次です。こちらは、先ほどのページで御覧いただいた動詞のランキングと同じものが、今、左側に出ています。なぜこれを出したかといいますと、実は日本語のテキストマイニングというのは、海外の言語と比べて日本語は非常に難しい言語だと言われていています。これは何かというと、1位の買うという単語はまさにそうですが、買うという単語が4,800件ありましたと言われても、何を買ったのか、どこで買ったのか、はたまた何を使って買ったのかみたいなものが分からなければ、結局、全部読まなければいけないという作業とそんなに変わらないというネックがあります。

なので、重要な技術が「係り受け」と呼ばれる技術で、簡単に言うと、主語と述語の関係というものを機械に覚え込ませることがテキストマイニングではできるようになっています。

右のマップを見ていただくと、よく見ると矢印になっていて、矢印を追いかけていただくと、おおよそ何を買ったという話題が分かるようになっていきます。一例で、SNS広告を見てネット通販で買いましたとか。左のほうは、健康食品を買いました、店舗で買いました、通販で買いました、スーパーマーケットで買いました。このような形で、全件読まなくとも、おおよそ買うというものに対して、何を買ったのか、どこで買ったのかというものが簡単に視覚化されるということです。

もしSNS広告というものが気になれば、クリックしていただくと、具体的にSNS広告のネット通販で100円のお試しの痩身健康食品を購入したら3口分の契約になってしまった。実際、買うという単語はどこにも入っていないのですが、買うという単語と購入という単語は、今回、同じ単語として認識しているので、実はまとめて集計して出してくれているということになります。なので、全件読んで、そもそもどこで買っているかというチャンネル

の分析というのをやらずとも、単純にクリックしていただくだけで、こんな形で見えるようになるというものです。

また、この係り受けと呼ばれる主語と述語の関係、つなごうと思えば幾らでもつなげることが可能でして、今度は3単語つなげたものを単語のランキングにしています。そうすると、上位を見ていただくと、1位のパソコンの内部が焼損しましたとか、ACアダプターコードから出る。これは、恐らく煙なり火が出たのだと思いますが、何かが出ているというのが出ていたり。風呂釜のケーシングが変形しましたみたいな形で、ここまで来るとほぼ要約文に近いのですが、おおよそどういう話題かというのがさらに具体的に分かるようになります。

ここを分けているのは、左は主に熱によって変形だったり、煙が出たりというのが上位を占めているのですが、次は業者の対応が不満、メーカーの対応が不満、店の対応が不満といった形で、比較的、人の問題が絡むのが10位以下、固まった話題になってくるというのが分かるかなと思います。

今、単語のランキングを幾つか御覧いただきましたが、実はあのランキングにも別の見方があります。それは何かというと、今、見ていただいたのは全ての年代の方をまとめて単語のランキングにしているので、特に誰というのを絞らず、言ってしまえばごちゃごちゃに混ぜた状態でランキングにしているものです。ただし、より具体的に見ていただく上では、一例ですが、こんな感じで年代別に分けていただくことで、各年代でどういう事故、トラブルに遭いやすいのかというのが分かるようになります。

色が着いているのですが、同じ単語に色を着けていて、横に並んでいるのは、各年代にデータを絞ったときの単語のランキングです。そうすると、例えば一番左の10代のランキングを見ていただくと、ネット通販 - 買うという単語があるかと思えます。これは、20代になるとやや上がって、30代、40代、50代がネット通販 - 買うという情報が非常に多く登録されていることが視覚的に分かりますし、実は60代からはぱたっと消えるのですが、替わりに出てきているものがあります。これは何かというと、下の赤いものです。テレビショッピングで買うというのが60代から増えてきて、80代になると一気にこの数が増えてくるというものです。

なので、50代と60代ではちょっと境目があって、ネット通販とテレビショッピングというチャンネルの違いがあるかもしれませんが、実は健康食品を買うという話題は30代から出始めるのですが、急激に件数が増えてくるのは60代からで、健康食品のトラブルに遭いやすいということが分かるかもしれません。

今、幾つか単語のランキングで、おおよそどんな話題があるかというのを御覧いただいたのですが、ここまでやると、実はもう一つ、機械で自動的にやってくれたらいいのと言われることがあるのが話題の分類です。つまり、データを取り込んだときに、おおよそどんな話題があるのかというのが簡単に分かりませんかというのが、実は分かります。それをやっているのがこちらのマップになりまして、仕組みとしては、同じ文中で一緒に

発言されやすい単語同士を線で結んであげることによって、今、丸で囲っているような話題の島みたいな分類をテキストマイニングで自動的に行います。

そうすることで、例えばブレーキが不良でリコールの話があったり、化粧品についての肌トラブル、返品の話。あと、歯の治療費についての記述。あと、火災の発生源がどうこう。あと、さっきの走行中という単語がありましたが、これは高速道路という単語と一緒に走行中という単語が出やすく、エンジンから変な音がするとか、エンジンの不良ですというものが一緒に書かれやすいということが、このマップを見ていただくと分かるということです。

なので、まずこれを見ていただくことによって、おおよそ今回のデータの主要な話題としては、こういった話題がありましたというものが、単純にワンクリックしてもらっただけで、こんな感じでできるようになります。

次のページに行きます。こちらは、「重大事故」に限定した場合の話題分類と記載させていただいていますが、これは先ほどお伝えした傷病の程度が1か月以上、もしくは死亡という、結構重たい事故に遭ってしまった方だけに絞ったマップになります。見方は一緒ですが、ここで見ていただきたいのはこの2つです。左上の自転車走行中と化粧品の話になります。

例えば、今回、化粧品のデータを挙げると、先ほどの全てのデータに対してのマッピングでも、この化粧品という話題はありました。つまり、全体から見ても、この化粧品に関するトラブルというのは比較的件数の多い話題だということになりますし、実はこれを重大事故に絞った場合にも、まだここに表示されるということは、件数が多い、かつ消費者の方が重大事故に遭いやすいというもので、非常に危険なキーワードかもしれません。

ちょっと補足すると、時期的には白斑の問題があった時期なので、今回、出ている可能性が多分高いのですが、こういった形で、まず1つは全体のマップと比べてもらって、どっちにも出てくるものというのは、かなり危険度も高く、件数も多いものです。

一方で、左上の自転車はちょっと違います。これは何かというと、先ほどのマップには出てきませんでした。つまり、事故情報データバンクに登録されている全体の件数の中では、自転車というキーワード自体はそこまで件数が多いものではないのですが、重大事故に限ったときだと、こうして出てくるということは、事故自体は少ないのですが、仮に事故に遭ってしまったときの傷病が重たくなりやすいという可能性があるかもしれません。なので、こうした形で全体と特定の集団に絞ったものを並べて見ていただいたりすることで、こういった傾向が見えてくるということになります。

こちらは、重大事故だけに絞って、また単語のランキングを御覧いただいています。先ほど、火災だったり焼損だったりという単語がランキングの上位に出ていたかなと思うのですが、重大事故に絞ると、転倒、骨折、死亡するという形になりまして、右側が一番最初に御覧いただいた単語のランキングになります。なので、実は火災という単語は件数としては多かったのですが、重大事故にはそんなに結びついていない可能性があるかもしれ

ないということになります。

また、別の集計です。こちらは、登録されているデータの件数を集計したものになりまして、今、月別で横に並べています。これは何別に並べているかといいますと、施設別、施設用途別の時系列にしています。こうして見ていただくことで、まず青とピンクのグラフが基本的には上位を占めているので、住宅と店舗・商業施設についての登録が基本的には多いということがわかります。

ただ、このグラフを御覧いただくと、大抵の皆さんが気になるのが緑のグラフの2月です。つまり、その他が急に増えている。これが何か気になれば、ここも実際には中を見ていただくことによって、勝手にテキストマイニングツールの中で、今は月別なので、今回で言うと前月と比較して、急に増えた属性値が何かというものを自動的に見つけてきてくれるというものです。

なので、何年分のデータを取り込んだとしても、この急に伸びている日は、傷病内容として一番増えたのは食中毒についてで、情報提供元はどこですかというのは、厚生労働省のシステムですと。実際にこれが特に多く発生した県はどこかということ、7番の茨城県というのが、グラフをただクリックしてもらっただけで、こんな形で自動的に見せてくれたりするようになります。

なので、実はこういったツールの上で、まずグラフを見つけた後の作業を全件目視でやっていたりしますと、こういった属性を見つけるというのは、全件を読んだ後にまた属性の集計を一からやらなきゃいけないということもありまして、頭を使うパートと、集計するという作業をするパートが、基本的には分析作業が交互に発生してしまうのが、分析担当の方からすると非常にやりにくいものだったりするのですが、機械に任せられるところはある程度任せてしまったほうが、こんな形で見やすくなるということになります。

また、ちょっとこれも別の見せ方になりますが、先ほどの年代別で話題のランキングを御覧いただきました。今回は、男女別の話題のランキングを御覧いただいています。そうすると、緑の色が着いている乗合バスが女性特有の単語だったり、病院という単語は比較的男性のほうが多いですみたいなことが分かるというものです。

ただし、実は1番の本生徒とか、女性の6番の被介護者みたいなもので言うと、明らかに年代が偏っているような可能性もありますので、こちらについては、集計画面も併せて見ていただくことで、今回、登録されているデータ自体が、そもそも男性については10代の内容が非常に多く、男性と比較して、女性は80歳以上のデータが多いというのがわかりますので、今回に関して言えば、データに偏りがあるので、話題の変化が出ている可能性はありますが、データがそろえば、こういった形で男女ごとにトラブル、事故に遭いやすいものが何かという話題の比較もしやすくなるということです。

また、これもちょっと別の見せ方になりますが、これは何かといいますと、左にあるものがもともとのデータに入っている属性になります。何のジャンルなのか、製品カテゴリーみたいなものだと思っていただければと思いますが、それに対して、右の軸、契約トラ

ブル、火災、異音・異臭、皮膚障害などといった形で、これは本文中に出てきた単語で、こういったフラグ立てをテキストマイニングツールの中で行ってしまうということです。

仕組みとしては、契約という単語が入っていれば、これは契約トラブルに分類してあげてください。1 というフラグを立ててあげて下さいみたいなことをやってあげることで、各製品カテゴリーと、それぞれどういったトラブルに遭いやすいのかというものを、こんな形のヒートマップで表現しています。

一番分かりやすいところで言うと、健康食品・サプリの商品カテゴリーは、契約時のトラブルというものが非常に多いということが分かりますし、実は、美容で言うと、美容のカテゴリーは皮膚障害とかが多そうに思うのですが、実は数字だけ見れば、いまだに契約トラブルのほうが多かったり、火災についての話題は、家電と自動車・エンジンのあたりが結構集中しているということが分かるというものになっています。

なので、実際、消費者の方が仮にこういったデータを見るときには、自分が普段買う商品。自動車も家電も美容も健康サプリも全部買う消費者の方は、いないとは言いませんが、比較的少数かなと思います。そういったときに、自分が普段買う商品で、他の方がどういったトラブル、事故に遭っているのかというのが知りたいという際には、こういった製品カテゴリーと、それぞれ遭いやすいトラブルというものを簡単に分類してあげることで、よりデータとしては見やすくなるかなと考えています。

幾つか御紹介させていただきましたが、ここまでがざっくり事故の概要についての分析になります。

そして、2つ目、ここからは見せ方の話になります。消費者の方が目にしたときに、より見やすい公表資料の作成をするにはどうしたらよいかというところで、幾つか御覧いただければと思いますが、太文字で書いてあるところは、私たちが普段、一般消費財系のメーカーさんに、社内でお客様の声の情報というものを届ける、見てもらう上でどうしたらよいかという話をさせていただくときに、よく言うポイントですが、今回のデータでも全く同じことが当てはまります。

つまり、データを見る側からしたときに、興味を持てるレポート、内容はどのようなものかということ、1つ目の条件としては、情報が最新だということです。例えば、自分が事故に遭った、ネット上でこういった事故が話題になっていた。ただし、それについてのレポートが出てくるのが半年後だとなってしまうと、それだけで情報の鮮度が落ちていて、余り見られることがないというのがネックになったりします。ただ、薄い文字で書かせていただいているのですが、このデータに関しては常に最新の情報が、それこそ毎週発信されているので、データの鮮度としての課題は特にないかと考えています。

2つ目、変化が直感的に分かることです。つまり、今回のデータであると、例えば単語のランキングで火災が多いとか、お化粧品だと白斑の問題があったから、こんなものがありましたみたいなお話しはさせていただきましたが、このデータを見るときに皆さんが大体知りたいことは何かということ、その前の年、2014年がどうだったのかということです。

特に、最新の情報が知りたいということにつながるのですが、去年と比較して何が増えたのか、減ったのかというものを、変化を見せてあげないと、単発でレポートを見ても、これがどれだけ特殊なことなのか、それとも、いつもある話なのかというのが分かりにくいということです。なので、直感的に分かるようにしてあげるとというのが2つ目になります。

そして、3つ目です。これは、見る情報が取捨選別されていること。メーカーさんなどと言うと、例えばお客様相談室で欲しいのは、お客さんからのクレームであったり、商品についてのリスクになるような話題というものが知りたい。一方で、メーカーさんのマーケティング部だと何が知りたいかという、何がお客様から褒められているポイントで、どういう人がこの商品について買って、意見をしてきているのかといった情報が知りたいということになります。

ただし、これが一つのレポートに入っていると、これはお客様相談室の人もマーケティング部の方も、見たい情報としては、お互い50%しか情報がないレポートになりますので、言ってしまうと非常に雑音の多いレポートになってしまうということです。なので、実際にレポートを作る際には、各部門ごとにレポートの内容を出し分けしている企業様が非常に多いのですが、今回のデータでデータの出し分けというのをどうするかといいますと、例えば、ここに書かせていただいたような、男性に見てもらいたいものは男性用のレポート。例えば、お子さんがいる家庭で、お子さんが学校で遭いやすいトラブルはこれですよみたいな形で、具体的に自分に該当するものだと思ってもらえるような切り口でレポートを作ってあげたほうが、より閲覧されるようになるかなと考えています。

あとは、それだと余りに細か過ぎるという話であれば、先ほど見ていただいたヒートマップのようなカテゴリーをざっくり分けてしまうというものでもいいかなと考えています。例えば、家電で多いトラブルはこれ、化粧品の通販だとこういうトラブルが多いみたいなものですね。

今みたいなポイントがあって、第1パートで、性別、年代別というのを御覧いただいたのですが、もう一つあるのが、そもそもアウトプットとして見やすくないと、というところもあたりるので、幾つかここでは、さっきお見せしなかったアウトプットを並べて御紹介させていただいています。

例えば、左側で言うと、これはパレート図と呼ばれるものですが、もともと入っている事故のカテゴリーを横にずらっと並べて、縦のグラフが件数です。そして、折れ線グラフで出ているものが、左から足していったときに、全体の件数の何%を占めるのかというものです。つまり、見方としては、例えば火災事故から発火みたいな、左から4つの青いグラフまでを全部足していただくと、ここまでの件数で全体の73%を占めていますよということが分かるということです。つまり、他にもいろいろなトラブルがあったりするのですが、主要なところはこの4つですという見方ができたりするものとか。

私たちのほうでは、例えばTwitterのデータを併せて取ってくることによって、今回、登録されている食中毒に関する報告の月別の件数推移と、ネット上の食中毒についての話題

の件数を並べていただくことによって、重ね合わせると、これに相関があるのかわからないのか。登録されている月に、Twitter上でもそういう話題が本当に増えているのかどうなのかというのが、見て分かるものになります。

そして、単純に年代の集計をしていただいたり、右側のほうは先ほど見ていただいたヒートマップです。こちらはさくっと進めさせていただきます。

あとは、こういったグラフと中の内訳を掛け合わせるような見せ方が、この左のグラフになります。ここについて補足させていただくと、まず見ていただきたいのは、青い折れ線グラフです。こちらは、先ほど定義した重大事故の件数を折れ線グラフで表示しています。それについて、月ごとに、その重大事故というものが、1か月以上、入院が必要なものなのか、それとも死亡してしまったものなのかというのを赤いグラフと緑のグラフで出しています。

こうして見ていただくと、例えば2015年12月、重大事故が153件あって、翌月、1月には111件まで減少したので、3分の2に重大事故が減ったように一見見えるのですが、こうして見ていただくと、入院1か月以上の事故は減ったのですが、死亡事故はむしろ1、2月は増えているということが、グラフを見ていただくと分かると思います。増えている自体は数件に見えるのですが、重大事故の中の死亡の割合で言えば非常に増えている月だったりする。なので、単純にざっくり集計するだけではなく、集計と併せて、この中の内訳も出していただくような見せ方も1つあるのではないかと考えています。

あと、右側のもともともデータに入っている品名の集計です。こちらでも、上から順に件数順に並べていくとこのような形になって、基本的には自動車とかパソコンというものが多いのですが、上から5番目の健康食品というものは、ちょっと違うジャンルのものが比較的上位に入ってきているものになります。こちらのデータに関しては、現状は入っているとおり集計していますが、例えば幾つかある自動車関係のカテゴリーをまとめ上げて、自動車として集計するというのも、場合によってはやってみてもいいかもしれません。幾つかアウトプットを御紹介させていただきました。

次は、事故情報データバンクから外れて、SNSのデータを幾つか御覧いただければなと思います。今回、テーマは、先ほど全体マップの中で出てきた、件数は少ないのですが、大きな事故になりやすい自転車事故に関するようなキーワードで拾ってきています。

データの概要については、こんな形になります。誤字・脱字ではなくて、「自転車」「怖い」で取ってしまうと、自転車、怖かったという単語が収集できなかつたりしますので、「怖」という単語で止めることでデータをまとめてくるというデータの取り方をしています。

また、こういったキーワードでやると、体重が何kg増えて、すごく怖い思いをしたので、自転車をこぎますみたいなダイエット系の話題が非常に入ってきてたりするので、ウォーキング、ジョギングといったキーワードは外しています。

また、Twitterだと、右下に書かせていただいたのですが、ニュースサイトの投稿が非常に多いのです。例えば、自転車でこういう事故に遭って、誰々が亡くなってしまいました

みたいなニュースが、新聞社の公式版とかをたくさん発信していたりというものがありますので、そちらについては、今回除外させていただいているものになります。

最初に見ていただきたいのは、そもそもこういった発言がどれだけされているのかというものになります。2月18日から1か月間のデータを日別にグラフにしています。そうすると、1つあるのが、2月20日に山ができていることが見えるかと思います。こちらは、埼玉県で自転車で亡くなった方がいて、結構大きいニュースになっていたものに対して、一般の方がロードバイクが危ないみたいな話とか、いろいろしていたというのがこの山だったりします。

3月に入ってから緩やかに伸びているものは、テレビ番組のほうで、あるタレントさんが過去に自転車の事故でこんなにひどい目に遭いましたみたいな話をしていたのですが、よくよく話を聞いてみると、そのタレントさんのほうが自転車で危ないところを結構逆走していて、相手方に迷惑をかけたにも関わらず、自分がさも被害者のように語っていることにネットで炎上といいますが、騒がれていたのがここです。なので、自転車というのは基本的に左側通行とか、ルールがあるにも関わらず、一般道を平気で逆走している人がいるということに対して、様々大分議論がされたのがこの時期だったりするということになります。

また、ソーシャルメディアのデータというのは、基本的に毎日、物すごい量が発言されますので、日別で見ていただくのも1つポイントですが、もう一つは曜日別で見ていただくという方法です。大抵のテーマの場合、通常は土曜日・日曜日、いわゆる週末とか祝日の件数が多くなりがちですが、今回の自転車、危ないとか、自転車、死ねみたいなキーワードでとってみると、なぜか月曜日が多いみたいです。今回、ここは理由が分からなかったのですが、もうちょっと見ていただいてもいいかもしれません。

また、別の切り口で見ると、時間別です。1時間ごとに、どの時間帯に自転車についての発言がされやすいのかを見ていただくと、こんな形になりまして、山として見えるのは、8時、12時、夜の22時というものかと思います。20時以降に件数が結構増えているように見えるのですが、これは大抵のテーマの場合、こうなります。これは何かというと、皆さん、家に帰ってから、今日あったことを投稿するという方も一定数いらっしゃいますので、他のテーマでとっても、ここは件数が急騰するような時間帯になります。

また、今回、自転車について、危ないとか怖いという発言をしている方がどういう人かというものを、性別と年代の情報をこちらでは出しています。見ていただくと、30代が一番多そうですが、今回のテーマをとってみて分かったことなのですが、自転車はこういう乗り方が危ないとか、自分が自転車に乗って、こういう危ない思いをしたというよりは、逆に、危ない乗り方をしている人、人の行動を見て、あれが危ないから止めさせてほしいとか、すごく危険で怖い思いをしたみたいなことが比較的多い。

他のテーマでもよくあるのですが、自分の悪いことよりも人の悪いことのほうが目に付きますし、発言されやすいという傾向です。その発言が、比較的30代に偏って発言されて

いたかなということで書かせていただいています。

ただ、ちょっと補足させていただくと、ここに書いているのですが、そもそもTwitterのデータというのは、こういった性別・年代情報を登録されていない方のほうが圧倒的に多いです。なので、通常のマーケティングでTwitterのデータを活用する場合には、誰が発言しているというものは余りにせず、何が発言されているかというものを先ほどのテキストマイニングみたいな形で見えていただくことが多いのですが、実はこちらは一部、私たちの独自の技術みたいになってしまうのですが、テキストマイニングを応用すると、大体どういう人か分かるというのがユーザプロファイルというものです。

何かといいますと、例えば私のTwitterのアカウントがあったとして、そのアカウントをざっと洗っていくと、10年前に大学を卒業しましたという投稿がありますよ。4年前に奥さんと旅行に行ったという投稿を見付けてくると、少なくともこの人は例えば30代前半の男性みたいなことが、過去の投稿を一気にテキストマイニングで解析することによって分かるという仕組みになります。

同時に、ローカル線の遅延の話が非常に多いアカウントは、恐らく何県に住んでいる人だろうとか、地域特有の話題というものがありますので、過去の投稿を解析することによって、このアカウントの方は大体こういう方ですよというものを推定するというのがプロファイルの技術になります。なので、先ほど見ていただいたものは、自己申告ではなくて、あくまで推定のデータになるということです。

データの概要について、最後に1つお伝えすると、Twitterのデータは、基本的には一般の方が書き込んでいるものになるのですが、大きく3つあります。実は、この3つを分けて考えないと、分析の上では非常に危険なテーマになりますので、ここに記載させていただきました。

3つは何かというと、1つは通常と書かれているもの。これは、通常ツイートという、普通に一般の方が自分の思ったことを書いて投稿するようなものです。実は、他にも2つあって、1つはリツイートと呼ばれるもので、自分の考えは何も書かずに、そのまま転送するようなイメージになります。もう一つは、リプライ。これは、誰かが投稿した内容に対して、私もそう思いますとか、それは具体的にどこであった話ですかみたいなことを返信として書き込むみたいなものという、3つがあります。

実は、これをなぜ分けているかということ、例えばさっきの単語のランキングです。これは、リツイートという、そのまま転送しているものを全部まとめて解析するとどうなるかといいますと、たくさん拡散した単語ばかりがランキングに出てきてしまって、全くテキストマイニングの意味をなしません。なので、ネット上でこれが話題になっていますよというものを見つける。

何が拡散しているかというのを見つける分析では、リツイートを含めて分析していただくのもいいのですが、例えば一般消費者の方が、自分が自転車に乗っていてとか、こういうものを目撃して怖い思いをした。もしくは、危うく何々しそうだったというヒヤリハッ

トみたいなものを見つけないという話になったときには、リツイートというのは全て分析対象から外してしまったほうが、一般消費者が自分で発言したデータだけを分析できるという意味で有効だったりするということです。ここは、今みたいな内容で補足させていただきます。

実際にさっきのキーワードが、Twitterのデータの単語のランキングになります。見ていただくと、基本的にデータの見方としては、テキストマイニングで扱うものは、アンケートであろうと、ソーシャルメディアであろうと、事故情報であろうと、テキストデータに属性情報が付いているものという意味では、データとしては何も変わりません。なので、分析の切り口も基本的には同じで、単語のランキングやマップで見ていただくという方法になります。

今回で言えば、危ないという単語が5位にあって、マップで見ていただくと、白くて分かりにくいので危なかったですみたいなものが出てきます。ラジオドラマで自転車の危ないものを解説している方がいて、それが大量に拡散されているものが入っていたという話題の分類になっていたりします。

この白くて分かりにくくて危ないというものは、実際に見ていただこうかなと思います。左の投稿ですね。マジでびっくりしたんだが、何で自転車レーンの上に置く。白くて分かりにくいから危ないという投稿です。これは動画がありますので、ちょっと動画を御覧いただきます。これが実際の投稿ですね。

自転車に乗って走っているのを動画で撮影していた方がいて、それを投稿していたみたいです。白くて危ないというのは、走っていたときに、いきなりこの白い物が道路上に置かれていて、正直、近づかないと全く分からず、このままぶつかったら大変な事故になってしまうというところで投稿していたものがたくさん拡散されていて、話題になっていたというものです。正直、僕も見えていて、これは近づくまで全く分からなかったの、確かに危ないなと思ったのですが。

ソーシャルメディアのデータのいいところは、こういった情報を集めたときに、その場の写真とか動画を一緒に投稿してくれることです。文字情報だと、体験した方は確かに危ない思いをしたかもしれないですが、伝わりにくい情報が結構あります。なので、動画で見ていただくと、今のものを見て危なくないという方はそんなにいらっやらないと思いますので、こういった情報は非常に有効かなと思います。

さらに、今の動画の投稿についてコメントを見ていただくと、文字が小さいので、実際の投稿のほうでお見せします。例えば、パシフィコに行くときに路駐して乗り降りさせるなよというので置いているんじゃないですかとか、投稿を読んでいくと、具体的に場所がどこで、どういう目的で置かれている。さらに下のほうへ行くと、朝7時ぐらいだと実は置いてないですよみたいな、時間的によって置かれている、置かれていないみたいなものがあるとか。こういう車が止まっているから置いているみたいですよみたいな情報交換をかなりされている。夜だともっと見づらくて危ないみたいな話を、別な写真を出していた

り。

こういった形で、一つの投稿に対して一般の方があれこれやり取りをしますので、1つ危ないものを見つけたときに、現場の特定というのは非常に早かったりするのがメリットかなと思います。SNSのものを幾つか御紹介させていただきました。

最後は、補足くらいの感じになってしまうのですが、先ほど例えばお子さんがいる方に限定してデータを見たほうがいいみたいな話をさせていただいたのですが、特に高齢者の事故、お子さんについての事故、2つのデータを御覧いただくかなと思っております。

1つは、ここで言う高齢者は60歳以上という扱いにしています。話題マップ、さっきと同じマップですね。見ていただくと、テレビショッピングの返品をしたいとか、病院に行ってお医者さんがみたいな話は結構あるのですが、全体の話題と比べて結構特徴的なのはこちらです。体調が悪いという単語自体は、出てきても特に違和感がない単語ですが、60代以上になってくると、臭いという単語と合わせて体調が悪いというものが一緒に発言されやすいということが分かります。これは、何かの臭いを嗅いだら具合が悪くなったという話もあれば、体調が悪いときに変な臭いを感じた、いろいろあったのですが、臭いがきっかけで体調を崩すというのは、高齢の方だとあるかもしれないというものになります。

今度は、10代、特に男の子に限定したマップがこちらになります。幾つか分類していますが、例えば左下は、教室という単語があって、手とか足があるのですが、教室内で起きた生徒同士の事故みたいなものが左下にある。右下だと、野球部という単語があるので、部活動周りの話題です。特に、野球部、ボールという単語と、右眼、左眼という単語と一緒に矢印で結びついています。つまり、野球部で目に当たってしまったときの事故というものが比較的多いということが分かると思います。

また、面白いのが左上です。口、歯、前歯という単語があるのですが、何が面白いかというと、具体的にどこでというものがセットで発言されていません。データとして入っていない可能性もあるのですが、逆に言ってしまうと、どこでも起き得る事故だったりします。よく見てみると、恐らく廊下で生徒同士がぶつかったのではないとか、階段から落ちたのではないとか、下校中だったのではないかみたいな推測はできたのですが、特に歯の事故というものは場所を構わず発生しやすいので、ちょっと危険かなと考えています。意外と体育の授業中というのは、そんなになかったです。

ここでまとめに入らせていただければなと思いますけれども、一旦、SNSについて簡単にお話しさせていただくと、一般のメーカーさんとか一般の企業さんがお客様相談室も持っていて、今までたくさんお金をかけてお客様の声を拾ったにも関わらず、なぜソーシャルメディアのデータを取りたいのですかというのと、こういったことを挙げたりします。一番下はネガティブですけれどもね。

調査とかお客様相談室みたいな、お客様が能動的に声をくれるようなものと比べて、データが非常に多いです。調査と比べて、お金がかからずに手軽に集められるので、非常に使い勝手がいいチャンネルだというのは、よく言われています。

2つ目が意外と重要でして、消費者が感じた「その時」の声を聴くことができる。これは何かといいますと、例えば何かトラブルがあって、すごく嫌な思いをしましたというのを家に帰ってから電話をかけてくる人だったり、何か商品を買った後の2か月後の調査で、この商品についてどう思いましたかと言っても、基本的には思い出して書いたり、よほど気に入らなかったことについて発言するものになりますので、ふっと思った、こんなものがあつたらいいのにとか、こういうものは本当はやっちゃいけない、危ないのではないかという、軽く疑問に感じたものというのが声としては非常に拾いづらかったりします。ただ、SNSは、今、感じた「その時」に発言してくれるので、比較的ライトな話題が拾いやすいというところで、メリットを感じている方もいます。

また、3つ目、これも調査と比べて言われるのですが、設問に左右されないという言い方をします。例えば、価格についてどう思いますかという設問を作った時点でどうなるかといいますと、私が価格について別に何とも思っていないくても、安いか高いか答えなきゃいけないという発想になってしまいます。そうすると、何とも思わないという選択肢があったとしても、大抵の方はもっと安いほうがいいという回答をします。これは結構あって、アンケートの設問設計をする上で、こっちが能動的に聞いてしまうと、ある程度そっちに先導してしまうというのが調査の難しいところになるのですが、ソーシャルメディアというのは、基本的に自分が思わないことは発言しませんので、こういった設問の影響を受けずに、本当に生の声を聞ける。

飲料メーカーさんとかお菓子メーカーさんだとあるのですけれども、ふだんお子さんに何を食べさせていますかと聞くと、七、八百円ぐらいのケーキですとか、どここの結構いいお菓子みたいな名前がよく出てくるのですが、後々、グループディスカッションみたいな感じで参加者同士でしゃべっていただくと、誰もそんなものは食べていなくて、結構見栄を張って回答される方もいらっしゃったりします。なので、本音で書いてくれるという意味では、非常にいいデータかなというのが3番目になります。

ただ、一番最後だけが難しく、性別・年代が分からないので、これがどういう人なのかというのは特定が難しいというのが、SNSだとちょっと課題になっていたのですが、3年くらい前から、さっきのプロファイリングの推定みたいな技術によって、ここもクリアできてきて、ますます使いやすいデータソースになっているというのが現状になります。

事故情報分析における、今度、SNSと併せて使っていこうという話になったときに、注意点といいますか、ポイントで箇条書きを幾つかさせていただきました。

1つ目が、今回、一番やりたかったのは、事故情報のデータとSNSのデータの相関です。事故情報として登録が多い月は、その事故もSNS上でたくさん発言されているというのがあればよかったのですが、結果、出すのがなかなか難しかったです。というのも、今回、いただいているデータは発生日ベースではなかったのです。

つまり、登録された日になると、実際にその事故に遭った方とか食中毒になった方というのは、その日付がいつかが分からずに、よくよく中のデータを見たりすると、一日二日

ずれるだけだったらいいのですが、実は月をまたいでずれていたり、先月、こういう事故に遭ってみたいなのが結構あって、件数の分析が登録日ベースだと大分難しそうな内容でした。なので、もしこれが発生日ベースで比較することができれば、SNSとの相関分析は今後できるのではないかと考えています。

2つ目、これは先ほど一般消費者の方は本当に思ったことしか書かないということに近いのですが、ヒヤリハットみたいなものは比較的取りやすいです。事故に遭う一歩手前みたいな、間違った使い方をしてしまう方とかです。数日後、覚えていないような内容も、当日投稿されているのも結構あるのですが、ちょっと煙が出たりとか、バッテリーとかスマホが熱を持ったりしても、正直、わざわざ電話をかけたりするような方はいないのですが、そういうものもネット上に多く書き込まれていたりしますので、そういうちょっとしたものは拾いやすいかなというのが2つ目です。

3つ目が、これは難しいところですが、ニュースの投稿が非常に多いです。誰々がどこで事故に遭ったというのを大量に拡散している方がいるので、データのクレンジングをちゃんとしていただいて見ていただかないと、ちょっとミスリードしてしまう可能性があります。なので、正直、Twitterの検索というのは、Yahoo!でも何でも、検索窓に検索を打ったら件数がずらっと出るのですが、そのキーワードに引っかかったもの全部を見ていても、そのうちの何割がニュースサイトかというのは分かりませんので、ちゃんとしていったツールのテキストマイニングみたいなものをかけて、中でちゃんとデータをきれいにさせていただくというのは、やっていただいたほうがいいのかなというのが3つ目になります。

4つ目、ここも結構難しいところですが、検索キーワードをお客様の言葉に合わせないとだめだというのが難しいポイントです。「炎損」という表現はほとんどない。とはいえ、「火災」があるかということ、実は「火災」という言い方も一般の方はそんなにしないです。では、火事なのか、が燃えるという表現なのか。これは、実際にお客様の声を中で分析しながら、確かにこういう表現があるのだなということを見つけていただくというフェーズが一旦必要かなというのが4つ目になります。

最後、重大事故になればなるほど当人からの投稿は少ないです。先ほどの高齢者のマップなどを見ていただくと、夫とか妻という単語はあったのですが、自分が重大事故に遭っているときは投稿している場合じゃないですね。なので、基本的に当人からの投稿は少ないです。ただし、親戚のおじさんがこうとか、自分の身近な方とか関係者の方からの投稿は比較的多くあったので、テーマによっては、ここは重大事故でも取れるようになるかなと思っています。

こちらは、最後になります。今回のデータを通じて良かった点と悪かった点です。

1つは、一般企業のコールセンターみたいに企業で仕切られていませぬので、幅広く、一つの製品カテゴリーの中で見られるというのは非常に魅力的なデータでした。さらによかったのは、「傷病の程度」というデータというものは、入っている企業がそんなになくて、それがどれぐらい重たい話なのかというものは件数の大小だけで見られないところになり

ますので、「傷病の程度」や「傷病の内容」というのは非常に有効なデータかなと考えております。

2つ目、これも今、ちょっと言ってしまいましたね。企業、業界を横軸でデータが見られるというのがメリットの2つ目になります。

3つ目、できればもうちょっとやりたかったのが、年代とかエリアの情報というのが、登録されている数自体がそもそも少なかったのが、これがもっとデータに入っているような状態になれば、かなり有効な分析ができるようになるのではないかとというのが所感になります。

私からの御紹介は以上になります。

山本座長 ありがとうございます。

大変興味深いプレゼンテーションでしたけれども、いろいろ御質問、御意見あるかと思っておりますので、お願いしたいと思っております。いかがでしょうか。

私から単純な質問をしてよろしいでしょうか。先ほど、SNSの分析ですか、自転車の走行するところに危ないものが置いてあるという話ですけれども、例えばこれを分析するのにどれぐらいの時間がかかるというか、あるいはどれぐらいのノウハウが必要なものなのでしょうか。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー 今回、このデータを見つげるところまでで言うと、データの取得自体に30分~40分ぐらい。画面を御覧いただいたほうが分かりやすいかなと思っておりますが、データを見付ける自体は10分ぐらいでできました。やっている作業としては、これは今回、実際御覧いただいたデータになるのですけれども、取り込んでいただいて、開くとその場で全て単語のランキングにできるので、さっき見ていたランキングはこのままです。

私がやっていたのは、ここで「危ない」という単語があると見て、同じマップが出てきて、「白くて分かりにくい」というのは何だろうというのを押すと、この投稿がたくさんリツイートされているというのがありましたので、今回御覧いただきました。なので、やっていることはこれだけなので、データの取得を除けば、ものの数分でこういうものが見つけられます。

山本座長 ありがとうございます。

他にございませんでしょうか。

村田委員、お願いします。

村田専門委員 大変興味深い内容で、技術的なこととかも少しお伺いしたいのですが、まず手元の資料の9ページですけれども、共起関係の単語を結んでクラスタみたいな感じにしているのは、そういう分類的な手法を使っているのかなと分かるのですが、そのクラスタのラベル付けというのも自動的にできるのでしょうか。今、PCの発煙とか、青でラベルが付いているものですが、

(株)プラスアルファ・コンサルティング甫坂グループマネージャー このラベル付

けは、基本的には人の手でやるものになります。なので、私のほうで、正直、中を読んでなくて単語を拾っていっただけのものが多いのですが、基本的には目視でやるようなものになります。ただ、お客様からも結構御要望をいただいていたので、今後、技術的には可能です。例えば、ブレーキ、不良という言い方はちょっとあれですけども、火傷と子どもの怪我とか、風呂釜のトラブル、風呂釜の異音とかも話題になるかもしれませんが、ある程度こういうラベル付けが今後是可以できるようになるかなと思います。

山本座長 他にございますでしょうか。

どうぞ。

村田専門委員 続けていいですか。私のほうでもデータの分析をしているのですが、一点、非常に気になったものがありまして、分析するときに登録日ベースではなくて発生日ベースで比較すべきというところですけども、私も登録日と発生日、かなり日付がずれているなというのが気になっていまして、それを考えると、17ページの食中毒の報告件数とTwitter上の食中毒の件数ですけども、私の目から見て全然合っていないように分布が見えるのですが、これは登録日で棒のほうは作られているということなののでしょうか。その辺の影響はどの程度あるとお考えなのか、教えてください。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー 実は、ここも中身について余り触れなかったのは、まさにおっしゃっていただいたとおりで、正直相関があるとは到底思えないですね。ただ、Twitterのほうで出ているものが、実はその後にもまた件数がちょっと増えていたりするので、前後で出ているので、時期的には多いのかなくらいには分かるので、今回はこのデータについては、相関はないという見方しかできないという認識で間違いはないと思っています。ただ、できればこれは発生日ベースでやりたかったということはあるということになります。

村田専門委員 ありがとうございます。

山本座長 お願いします。

相澤専門委員 どうもありがとうございました。

御説明の途中で、キーワードの精緻化をされた。つまり、購入したとか同じような言葉でやるということをお話しいただいたと思いますけれども、具体的にはどういう辞書で、どのぐらいの規模のものを使っているのでしょうか。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー 辞書については幾つかあるのですが、1つは、一般的な日本語としての辞書はデフォルトで入っているので、「転倒」というのが「転ぶ」という単語と「倒れる」という単語に分かれる。そういうことは当然ないです。

ただ、「香り」と「臭い」という単語があるのですが、ぱっと見は同じ単語に見えるのですが、実は化粧品メーカーさんは「香り」と「臭い」は明確に分けるのです。「臭い」というのは、若干ネガティブな要因を含みますので、同じ単語として扱ってはいけなないと。ただし、お菓子なんかで言うと、逆に「すごくいい臭いがする」という言い方をしますので、

これは一緒にしてもいい。業界によって、一緒にしていい、悪いが別になるものについては、私たちのほうでは手は入っていません。なので、今回、追加でやっているのは、その辞書の表現というものを、私のほうでチューニングを20分くらいやっています。

具体的には、こういった画面を見ながら、例えば「危ない」というのと「危険」というのは同じ単語でいいですという話であれば、重ねていただいて一語にするという作業をペタペタとしていったような感じになります。

相澤専門委員 ありがとうございます。

山本座長 他にいかがでしょうか。

西田委員、お願いします。

西田専門委員 プレゼンテーション、ありがとうございました。非常に面白い御発表で、分かりやすかったです。

今のものに対する質問にならないかもしれないですけども、1つが、行政がやった施策の効果評価というか、評価できるといいなと思っているんですけども、ある時期に消費者庁さんでやったものがどう拡散していったかみたいなものが、SNSの分析で出せる可能性はありますか。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー そこは可能です。プレスリリースなどの分析とかは企業さんだとよくやられているのですが、そういったものは発表時からデータを集めていただいて、分析としては幾つかあるのですが、それが本当に投稿されているのか、拡散されているのかという数字で見るというやり方が1つ。

あと、そのプレスの内容を投稿しつつ、自分の意見を書き込む方がいらっしゃるんで、その意見が具体的にどういう意見だったかというのを、さっきみたいなテキストマイニングで見ていただくというのが2つ目ですね。

3つ目が、比較的何回もプレスリリースを出す場合には、前回のプレスリリースと比べての件数の比較ということで効果を測定する方はいらっしゃいます。SNSで拡散したか、していないかというのは、指標として明確なものがあるわけではないので、例えば1万投稿されたから拡散したと言っていいですという指標があるわけではないというのが結構難しいところなので、あくまで自分たちのこれまでの発表と比較して、良かったか悪かったかという効果測定をするというのは、よくやられています。

西田専門委員 共起ネットとか、こういうものを拝見すると、これ自体には余り個人情報もないし、公開してまずいこともないような気がするのです。生まで行くとまずいかなと思うんですけども、やっている中でそういう理解でいいですか。まずそうなものが出てきたとか、そういうものはありますか。個人情報とか特定という面で。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー 今回のデータで言うと、ないと言ってしまって大丈夫だと思います。一般企業様の例で言うと、たまにお客様相談室とかで、今すぐ折り返していただきとって本文中に電話番号が入ったりというのがあったりすると、まれに単語のランキングとかに出てきたりする可能性があるんで、

そこは気を付けている会社さんは結構多いです。

会社さんによっては、口コミサイトをやっている会社さんとかのイメージに近いのですが、けれども、単純にお客様からの書き込みをだあっと一覧にするのではなくて、さっきみたいな感じで視覚化したい。けれども、基のデータは見せたくないみたいな会社さんは結構いらっしゃる。例えば、クリックしても、中のデータを見えなくして、広く公開するというのをやるというのは可能ではあるので、そういった感じで情報共有のツールとして使っていただく企業様とかはいらっしゃいます。

山本座長 続けてあれば、どうぞ。

西田専門委員 もう一つ、今回、動画がすごく役に立った例があったのですが、ああいう画像情報は相当役に立つシチュエーションがあるかなと思いますけれども、そういうものを自動的に引っ張ってくることも、典型的な共起を作る際にできますか。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー 最近、弊社でやっているのと言うと、画像の解析をやってしまって、何が写っているかを自動で判定してしまうというのをやっていたりします。特に、Instagramなんかは画像がメインになりますので、結構そういったものが重要だったりしますね。

なので、ごめんなさい、別の資料になってしまうのですが、例えば左の投稿はInstagramの投稿ですが、今から北海道へ行ってきますと写真を載せているのですが、見てもらうと、ハンドバッグと何か製品が写っていて、スターバックスですというのが、画像を自動で解析することで、こういったラベル付けみたいなものができたりします。

例えば、これはちょっとマーケティング寄りになってしまいましたが、各ビールと一緒に写っている料理は何料理なのかというのを分類すると、実はエビスが圧倒的に日本料理が多くて、アサヒビールは比較的ファストフードとか、ちょっと安い食べ物と一緒に写っている傾向がありますという。正直、ファストフードと一緒にアサヒビールを飲んでいまして書き込む人はいないので、画像を解析することで実はこういうものが見えてきたりというのはできたりします。

西田専門委員 ありがとうございます。

あと、コメントですけれども、さっきヒートマップがあって、何かのサービスを買う前に、どこに問題があるのかというのがぱっとみえるので、あれは行政サイドでも役に立つなと思ったのです。おっしゃられたように、消費者の目線に立ったときにも、今から利用しようとしているものがどういうトラブルが多いのかということを知って購入するのはすごくいいなと思いましたが、それに加え行政サイドでも役立ちそうだというコメントです。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー ありがとうございます。

この軸というのは、情報発信する上でも結構重要かなと思っていて、例えばさっきのTwitterで拡散しているものとかもそうですが、家電の情報サイトとか美容の情報サイトは結構あるのですが、事故情報のサイトというのはなかなかなくて、そういう軸で情報発

信してしまうと意外と拡散されづらいのですが、美容系のサイトに、最近、美容系のトラブルはこういうものがありますみたいなものは、実は意外と拡散しやすかったり。お客さんが興味を持っている軸で事故情報を発信してあげるとというのがうまくかみ合うと、非常にいろいろな方に見ていただけるのではないかなと思っています。

西田専門委員 まさに、これはそういう意義があるかなと思って。これは、海外の人に聞いたときに、これはちょっと発言に語弊があるかもしれないですけども、日本の消費者団体とか消費者を啓発するようなところの活動が、データに基づく情報発信という点でまだ十分ではなくて、海外の場合には、かなりこういうデータをNPOなりの組織が発信している、そういうものにデータを使っているということ聞いたことがあるのです。消費者が利用する情報だけではなくて、そういう消費者団体の方が利用する情報としても非常に役に立つのかなという印象を持ちました。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー ありがとうございます。

山本座長 その他にいかがでしょうか。

お願いします。

河上委員長 本当に面白いなと思いつつながら圧倒されていたのですけれども、よく分からないのは、SNSを使って分析しておられるところがあったときに、例えば事故情報で一定の傾向が出ているものと、SNSで出ているものを重ねるようなことをされていたと思うのですが、事実の重みづけの問題と。それから、言ってみれば、一般の人の関心动向というか、関心の強さみたいなもので左右されている数値と、それが重なり合って、その事実そのものとは違う重みづけがそこになされてしまうのではないかな。少なくとも人々の関心度の強弱と因果関係の存否は全く別問題です。そのゆがみみたいなものが起きないのかなというのが一つ、ちょっと気になりました。

もう一点ですけれども、幾つかキーワードになるものが関連付けられて出てくるときに、時々、注目すべきいろいろな推測をされているのですが、その推測とか解釈を関連付けの中からやっている。その解釈をするときの視点というのは、場合によると、その人の個人のバイアスがかかってくるのではないのかという気がします。例えば、「高齢者」で「臭い」、「体調」の話ということになると、私などはまさきに「加齢臭」を思いついたのです。

それはともかくとして、甫坂さんがおっしゃったのは、高齢者が異常な「臭い」を感じて気持ちが悪くなって体調を崩したという話だったのだけれども、それには、私が見たときに行った関連付けについて一定のバイアスがかかっていると、同じことがひょっとしたら甫坂さんのほうにもあるのかなという気がいたしました。

逆に、そういうことがあるとすると、キーワードでイメージが作られていったときに、例えば0-157によってカイワレ何とかによる食中毒事故という話がどこその地域で発生しているという分析結果が出たとしますと。あれは、本当は別の因果関係があることが後で分かったのですが、風評被害につながってしまうのですね。そう考えていくと、これを

ぱっと公表したときの負の結果というものがあるのかなというあたりも気になったので、お教えいただければと思います。

(株)プラスアルファ・コンサルティング南坂グループマネージャー ありがとうございます。

1つ目に関しては、ソーシャルメディアについて、扱う上で難しいのが、今回のデータで言えば、例えば事故に遭いそうで危ないという人もいれば、遭ったという人もどっちも入っているというのが難しいところですね。なので、実際に事故に遭った方とか本当にけがをした方というのが少なくても、投稿件数として増える可能性は非常にあります。なので、私たちの中で、遭ったという人と遭いそうだった人を分けるとするのは、細かくやろうと思えばできたりするので、そういった日本語の表現の仕方で、本当に遭った人だけを抽出するというのをやるというのが方法としては1つです。

ただし、これをやったとしても、お客様相談室のデータでもそうなのですが、本当にその事例に合った方100人が登録してくれるデータというのはほぼないので、どっちが正確かという議論からなかなか抜け出せないというのは、問題点としてはちょっとあるかもしれないです。なので、数字のどっちが正確かという議論よりは、そういうものがあつたということを抑えて対応するかしないかというのに各企業様は取り組まれることのほうが多いですね。最悪の事態を想定して動くというほうに使用されることが多いかなと思います。

2つ目のデータを見てみてということについては、まさに私たちが普段お客様に言っているのと同じことを実は言っていて、今回のマップを見て、これが全てかという、僕たちは100%、ノーと言います。なぜああいう書き方をしているかという、今回のデータで言うと、最終的には中をちゃんと全部読まないとだめなのです。テキストマイニングツールというのは、分類をしてくれるもので、こういうものがありますよというものではあるのですが、最終的には中を全部読んで、本当にそれで間違っていないかという確認を必ずしなければ、ちょっと間違える可能性の高いツールです。

ただ、2万7,000件、全部読むわけにはいかないもので、今回、この内容が本当に正しいかというのをチェックするのは、この三十何件見てもらえれば大丈夫ですよという当たりをつけるために使うのがテキストマイニングツールという認識でいたほうが、ここで出たから100%間違いないですというのは、ちょっと言い過ぎかなと思っています。

あとは、考察が浮かぶか浮かばないかという議論については、一番いいのは、普段、このデータを見ている方がこういう分析をやるのが間違いなくよくて、うちでよく言うのは、普通、分析業務を行っている方は2種類いて、分析するデータに詳しい方が分析操作をしているか、分析に詳しい方がデータを預かって、統計にのっかって分析するかという、大きくこの2つの担当者の方が多いのですが、一番いいのは、データのことをよく知っている方が分析できるほうが、テキストマイニングに関しては、正直、いいです。これは何かというと、この中の単語で、例えば今月、普段このデータだと滅多に見ない単語はどれですかというのは、分析に詳しい方はデータを知らないと分からないのです。

なので、私たち、基本のスタンスとしては、この商品の話になってしまうのですけれども、「見える化エンジン」という、この商品自体は、分析操作はとにかく簡単にできるように製品としては作るの、データをよく知っている方にこのテキストマイニングを使ってくださいという話をいつもさせていただいています。なので、考察がうまくできるかできないかは、そのデータのことをどれだけ知っているかということに比例することによりまでするので、正直、今回の僕のこの考察が全部合っているかということ、よくよく見ると、過去にこういうことがあってという背景知識があれば、違う結論になる可能性も全然あるかと思えます。

今ので、回答になっていますか。

河上委員長 ありがとうございます。

山本座長 どうぞ。

大森委員 すばらしい魔法使いのように見せていただいていたので、もっとこんな魔法が使えないかなという素人からのお聞きしたいことですが、例えば、赤ちゃんがお風呂で首だけ外に出るような浮輪があって、それですごく死亡事故が多いのですけれども、ネットとかで見ると、かわいいでしょと言って、そのぶかぶかしている赤ちゃんの写真をいっぱい投稿されているのですね。私たち、別に見る気がなくても、自動車の広告が入ったり、いろいろ入ることがあるわけじゃないですか。そういうものを投稿する人のところに、こういう事故がありますよというお知らせが何かできないかなと思います。

消費者庁ですばらしいデータバンクができたとしても、そこにアクセスしてくださる人というのは一般消費者のごく一部で、実際にそれを使って危険が及ぶ可能性のある人にどういったら伝えることができるかというのが大きな課題だと思って、その辺は何かマジックがないのかなと思います。

(株)プラスアルファ・コンサルティング南坂グループマネージャー やられている企業さんがいます。ただ、マーケティングとしてやっている会社様と、リスクの観点からやっている会社さんがいらっしやまして、イメージ的には、例えばここで危ないとか危険だと言っている人たちの一覧を抽出してきて、その方たちにさっきのTwitterでリプライという、直接その場でメッセージを送ってしまうというのをやられている会社さんがいる。

例えば、それがマーケティング寄りであれば、ニッセンさんとかがよくやられているのですけれども、ニッセンさんで商品を買って、商品が届いたと投稿すると、ニッセンの公式アカウントから連絡が来て、商品購入ありがとうございます。無事届いたみたいでよかったですみたいな挨拶が来るというのをやっちらっしやる。逆に、製品の間違った使い方をして非常に危ない方には、その使い方は非常に危ないので止めたほうがいいですみたいなものを、監視と言うとちょっとあれですけども、モニタリングしてアクションする会社さんがいらっしやったりします。

あと、海外の事例だと、直接的に危なくはないのですが、どこかの大学が一時的にやっていたのが、こういう発言を数か月にわたってする人は自殺する傾向にあるみたいなもの

を見て、警告するメッセージを個別に送ったりという取組をやっているというのは聞いたことはあります。なので、直接メッセージを送ってしまうというのが一番早い。

ただ、これは取組としては気を付けなければいけないところで、見られているつもりで投稿していない方も多分にいらっちゃって、トラブルになる可能性もあるので、そのあたり、対応マニュアルみたいなものをしっかり作られている会社さんは結構あるみたいですね。

山本座長 他にいかがでしょうか。

どうぞ。

市瀬専門委員 参考になるお話、どうもありがとうございます。

私のほうから、ちょっと聞きたかったのは、Twitterのところのお話ですが、SNSの活用の注意点というところで、データのクレンジングが必要だというお話があったかと思いません。この分析の際に、自転車ということでやられているのですけれども、ここら辺のデータをどういう具合にしてクレンジングしたのかということと、あと検索のワードをどうやって選んだのか。そして、除外のワードが使われていますけれども、その辺のところをどういうふうな形で選んだのかというあたりをちょっとお聞きしたいのと。

あと、先ほどお話の中で、データを分析するときには、領域のことを分かっている方が、実際にこういうツールや何かを使いながらやるというのが重要だというお話があったかと思えますけれども、御社の製品になってしまうのかもしれないですけれども、こういうものを実際にデータマイニングとかテキストマイニングが分からないような方が結構使われているのかといったあたりを少しお聞かせいただければと思います。

(株)プラスアルファ・コンサルティング甫坂グループマネージャー ありがとうございます。

後半の質問から先に回答させていただきますと、弊社のこのサービスを使っているお客様は、ほとんどがコールセンターで実際にお仕事されている方とか、マーケティングで他の仕事をしていて、それに使うのにテキストマイニングをちょっとやりたいみたいなお客様がほとんどです。

もともと、前職で分析担当者向けのテキストマイニングツールを作っていたのですが、分析担当の方が専門でいる会社さんはそんなになくて、現場で分析もやっていますという会社さんが非常に多かったので、そういう作りにした結果、私たちのお客様で言うと、分析だけをやっているお客様とか統計知識があるお客様のほうが圧倒的少数ですというのが、まず御回答になるかと思えます。

データのクレンジングに関して言えば、幾つか方法はあります。例えば、自転車で転ぶというのを、これは本当にあったのですけれども、1か月で4,000回投稿している人。間違っても1か月に4,000回、自転車で転ぶ人はいませんので、それはまず全部除外してしまうという、まず過剰に投稿しているアカウントを全部はじいてしまうというのが1つある。

あと、全く同じ文章を3分置きに投稿するという、ボットと呼ばれるものですがけれども、

ひたすらそういう自動投稿みたいなものがあつたりしますので、そういうものは全部ロジックではじいてしまうとか。あと、このボットの判定というものが意外と難しいのですが、さっきのプロファイリングの推定ですね。この人、男性、女性と判定する中で、この人は2週間に1回、この投稿しかしないねというのが判定できれば、これはボット、自動投稿機能だということを判定して取ってしまったりというので除外しているのが主立ったやり方になるかなということになります。

山本座長 よろしいですか。

他にございますでしょうか。

お願いします。

樋口委員 貴重な話をいろいろありがとうございました。

私は、テキストマイニングは実際に使っているんで、特に大学なので、うちの学生がいろいろ分析してしまっていて、中には消費者関係の委員会で消費者代表がどんな発言をしているか、テキストマイニングしたり。済みません、余談ですが、かなり実は普及しているのではないかという気はするのです。前段と後段に分けて、前段のお話のデータバンクのところは、私どもも学生が使って分析していますので、データバンクそのものの機能をもう少し拡充すれば、今お話のようなことをかなり多くの人ができるのかなと。消費者庁のデータバンクそのものの検索機能は、よくできていると思います。

後半のSNSのほうは、非常に可能性があつて、重要な点かなと感じました。お話の中でもちょっと触れておられましたが、プレスリリースの話がありましたけれども、今、河上委員長のお話にもありましたけれども、例えば事実と直ちに何う話が広がったり、風評被害と言われるものがあつたり、そういうときに行政サイドがSNSなどの分析をある程度機動的にして、消費者に対する安全情報の中身を変えていくとか、そんなことは実際には可能なのではないでしょうか。

(株)プラスアルファ・コンサルティング南坂グループマネージャー 正直、手段によるかなと思ってしまっていて、プレスリリースの情報を、プレスリリースのページを直接ツイートする方は、かなり少数です。大抵の方は、どこかのニュースサイトでそのプレスリリースの内容を見て、そのニュースサイトを拡散する方が非常に多いですね。最たる例はYahoo! ニュースだったりするのです。なので、第1段階として、そのプレスリリースがそういうニュースサイトに取り上げられるかどうかというのは、1つ大きなポイントかなとは思っています。

一旦拡散してしまえば、その後、ぱっと一気に流れるようなものではあるので、そこから関心事なのかどうかなので、一番大事なのは情報の速さですね。初動で間違えた情報が流れていると分かった時点で、すぐに流してあげないと、その話題が話題になっていないときに流してあげても、全く話題にならなかつたりというのはあつたりするので、そのスピードと、ちゃんとしたメディアに取り上げられるところがポイントになるかなと思います。

樋口委員 ありがとうございました。

山本座長 その他にいかがでしょうか。よろしいですか。

大体御質問いただいたようですので、本日は、プラスアルファ・コンサルティングから「テキストマイニングを活用した事故情報の分析」などについて、大変興味深いお話をいただきました。前半の事故情報データバンクのほうのお話は、今後、私たちが分析をさらに進めていく上で大変参考になるお話であったと思いますし、後半のSNSの話も、これからどういうふうはこのSNSを、いろいろな活用の仕方があると思うのですが、どのように生かしていけるかということについて、非常にいろいろな可能性を示していただいたプレゼンテーションだったのではないかと思います。

本当に本日はありがとうございました。

4 . 閉会

山本座長 それでは、本日の議題は以上になります。これにて閉会とさせていただきます。お忙しいところ、お集まりいただきまして、どうもありがとうございました。