

供元により「辞書サイズ」、「網羅率」、「平均バイト数」の傾向が大きく異なることは、マイニングを実施する際に注意を要する。異なる情報提供元どうしで事故情報を共有するだけでなく、事故情報の入力インタフェースや整合性チェックツールを共有することが、データの品質向上に結び付くと期待される。

事故情報データバンク中では、「子ども」、「高齢者」、「外国人」など、分析の中で重要な役割を果たす概念が明示的に定義されていない場合がある。このため図表2-12で試みたように、分析に先立ってデータの整理や不整合なデータへの目視による対応などが必要になる。テキストマイニングの目的に応じて辞書を整備することが必須であると考えられる。

本調査では、事故情報データベース中の分類項目を、「絞り込み項目」と「表示項目」の2種類に振り分けて分析を試みた。絞り込み項目で値が入力されていない欠損値について、自由記述項目のテキストを利用して情報を補完する仕組みを検討することも今後の検討課題である。

#### (4) 事故情報データ分析 (市瀬 龍太郎)

##### ア 分析目的

事故情報データバンクでは、大量の事故情報を保持している。起きる事故には、一定のパターンが存在する。そのようなパターンを発見することで、頻繁に起こる事故の特徴を抽出することを目的とする。

##### イ 分析手法

本分析においては、頻繁に起こる事故の特徴を抽出するために、データマイニングにおいてしばしば利用される頻出パターンマイニング手法を応用する。頻出パターンマイニングとは、事例同士で頻繁に共起する事象を抽出する手法であり、スーパーにおいて精算時の買い物かごの中に入っている商品の組み合わせから、購買傾向を把握することなどに使われる。例えば、ソーセージを買う人の多くがロールパンを購入しているといった傾向を抽出することができ、マーケティングなどにも応用される技術である。本分析においては、事故情報データで頻繁に共起する語を利用することで、起こる事故のパターンの抽出を試みる。

本分析では、2段階の手続きを取り、事故の特徴の抽出を試みた。最初は、事故のグループ化である。そのために、対象の事故データの「事故概要」、「商品など分類」から名詞を抽出し、抽出された名詞が頻繁に共起する事例をグループとして抽出した。次に、各グループ内の特徴を抽出するために、グループ内の事故の「事故内容詳細」から名詞を抽出し、抽出された名詞で頻繁に共起する名詞をグループ内の事故を特徴づける名詞として列挙した。

##### ウ 対象データ

事故は、被害の重大さによって異なる傾向を持っていることが予想される。また、子供に起こる事故と、高齢者に起こる事故も異なる特徴を持っていることが予想される。そのため、被害者の重傷度と年齢により、いくつかのカテゴリに分けて分析を行った(図表2-14)。図は縦軸に被害者の重傷度、横軸に被害者の年齢をとったものである。重傷度は、重いものから死亡、重傷(治療に1か月以上)、その他に分類されており、その中から、「死亡」「重傷・死亡」の場合に着目した。年齢は、9歳以下、10~59歳、60歳代、70歳代、80歳以上で分類を行い、子供と高齢者に着目した。重傷度と年齢の組み合わせにより、図表2-15に掲載した14個のカテゴリを設定し分析を行った。なお、データの総件数は、

198,135件である。

【図表 2-14】 重傷度と年齢による分類



図 1 : 重傷度による分類

図 2 : 年齢による分類

【図表 2-15】 年齢と重傷度の組み合わせによる14個の分析カテゴリ

1.	重傷・死亡	13,621件	8.	60歳代, 死亡	53件
2.	死亡	1,517件	9.	70歳代 (全て)	8,972件
3.	9歳以下 (全て)	3,889件	10.	70歳代, 重傷・死亡	1,541件
4.	9歳以下, 重傷・死亡	555件	11.	70歳代, 死亡	97件
5.	9歳以下, 死亡	74件	12.	80歳以上 (全て)	4,366件
6.	60歳代 (全て)	13,191件	13.	80歳以上, 重傷・死亡	1,037件
7.	60歳代, 重傷・死亡	1,916件	14.	80歳以上, 死亡	208件

## エ 分析結果

(ア) カテゴリ 1 (重傷・死亡): 抽出グループ数122個

5語以上の名詞によりグループ化されたものは30個。

グループ例: 「洗顔、石鹸、化粧、衛生——顔、皮膚、受診」

「住宅、設備、建物、構成、材——救急、転倒、骨折」

「住居、生活、用品——製品、建物、火災」

ここで、グループ例の最初に挙げた語は、グループを作成する際に使われた「事故概要」「商品など分類」に見られた頻出語の例であり、後に挙げた語は、グループの特徴をみるための「事故内容詳細」の頻出語の例である。

このカテゴリでは、医療、保健関連の事例などが見られる。

【図表 2-16】 カテゴリ 1 (重傷・死亡)

ID	出現数	語1	語2	語3	語4	語5	語6	語7
24	492	洗顔	衛生	保健	品	石鹸	類	化粧品
41	612	機器	医療	保健	福祉	サービス	衛生	品
64	886	理	美容	保健	福祉	サービス	衛生	品
9	448	器具	理	保健	用品	美容	衛生	品
46	618	機器	医療	保健	衛生	品	サービス	
65	887	理	美容	保健	衛生	品	サービス	
66	896	理	美容	保健	福祉	サービス	品	
42	613	機器	福祉	保健	サービス	衛生	品	
71	916	美容	福祉	保健	サービス	衛生	品	
78	1012	医療	保健	福祉	サービス	衛生	品	
34	534	洗顔	衛生	保健	品	化粧品	類	
90	1231	石鹸	衛生	保健	品	化粧品	類	
4	435	美容	医療	保健	福祉	サービス		
43	613	機器	医療	保健	福祉	サービス		
47	619	機器	衛生	品	保健	サービス		
72	917	美容	衛生	品	保健	サービス		
79	1018	医療	保健	衛生	品	サービス		
87	1168	理	美容	保健	福祉	サービス		
115	2883	化粧品	類	品	保健	衛生		
28	504	構成	材	住宅	設備	建物		
10	449	器具	理	保健	用品	美容		
7	447	用品	福祉	保健	サービス	品		
11	450	理	美容	保健	用品	品		
59	846	機器	医療	保健	衛生	品		
67	897	理	美容	保健	サービス	品		
68	904	住	生活	住居	用品	品		
74	926	美容	福祉	保健	サービス	品		
76	962	理	美容	保健	衛生	品		
80	1049	医療	保健	福祉	サービス	品		
100	1629	福祉	サービス	衛生	品	保健		

(注) ID : カテゴリ内グループの識別子  
出現数 : 語が含まれる事故の数 (語は順不同)

(イ) カテゴリ 2 (死亡) : 抽出グループ数102個<sup>17</sup>

6語以上の名詞によりグループ化されたものは24個。

グループ例 : 「ベッド、介護、生活、用品、住居——すき間、ボード、首」

「体育、指導、課外、部、活動——(体操、練習)(野球、バッティング)(サッカー、部、活動)」

「医薬品、医療、サービス——投与、過剰、中毒」

「住居、開放、式、石油、ストーブ——火災、全焼」

表記方法は、カテゴリ 1 と同様であるが、2番目の例では、グループ内で体操、野球、サッカーといくつかの代表的なケースが見られたので分けて記述した。

<sup>17</sup> (イ) 以降の図表については、第31回消費者安全専門調査会 市瀬消費者安全専門調査会専門委員資料参照。

[http://www.cao.go.jp/consumer/kabusoshiki/anzen/doc/031\\_170612\\_shiryou5.pdf](http://www.cao.go.jp/consumer/kabusoshiki/anzen/doc/031_170612_shiryou5.pdf)

このカテゴリでは、火災、介護、部活動、医薬品などの事例が見られる。

(ウ) カテゴリ3 (9歳以下) : 抽出グループ数112個

5語以上の名詞によりグループ化されたものは25個。

グループ例 : 「建物、設備、サービス——(カビ、アパート、部屋)  
(スイミング、スクール)」

「娯楽、遊具——(滑り台、骨折)(ショッピング、センター)」

「外食、サービス——アナフィラキシー、メニュー」

このカテゴリでは、アレルギーや遊んでいる最中などの事例が見られる。

(エ) カテゴリ4 (9歳以下、重傷・死亡) : 抽出グループ数60個

4語以上の名詞によりグループ化されたものは27個。

グループ例 : 「建物、設備、構成、材——自動、ドア、怪我」

「乗り物、車両——車、指」

「娯楽、遊具——テーマパーク、滑り台」

このカテゴリでは、乗り物、遊んでいる最中などの事例が見られる。

(オ) カテゴリ5 (9歳以下、死亡) : 抽出グループ数24個

4語以上の名詞によりグループ化されたものは9個。

グループ例 : 「体育、教科——(運動、場)(プール、水泳)」

「認可、外、保育、施設——就寝、幼児」

このカテゴリでは、学校活動、保育中などの事例が見られる。

(カ) カテゴリ6 (60歳代) : 抽出グループ数99個

4語以上の名詞によりグループ化されたものは33個。

グループ例 : 「住宅、設備——骨折、転倒、治療」

「調理、食品——冷凍、ピザ、農薬」

「美容、サービス——毛、染め、皮膚」

このカテゴリでは、医療、保健、サービスなどに関する事例が見られる。

(キ) カテゴリ7 (60歳代、重傷・死亡) : 抽出グループ数60個

4語以上の名詞によりグループ化されたものは28個。

グループ例：「住宅、設備——（シックハウス、頭痛）（転倒、骨折）」  
「医療、サービス——（歯医者、抜歯）（視力、矯正）」

このカテゴリでは、医療、保健のサービス、住宅に関する事例などが見られる。

(ク) カテゴリ8（60歳代、死亡）：抽出グループ数37個

4語以上の名詞によりグループ化されたものは23個。

グループ例：「住居、生活、用品、手すり——ベッド、すき間」  
「食料、食品——のど、窒息」

このカテゴリでは、福祉、医療などの事例が見られる。

(ケ) カテゴリ9（70歳代）：抽出グループ数65個

4語以上の名詞によりグループ化されたものは24個。

グループ例：「医療、サービス——イン、プラント」  
「石鹸、化粧——皮膚、科、小麦」

このカテゴリでは、医療、保健、サービスに関する事例などが見られる。

(コ) カテゴリ10（70歳代、重傷・死亡）：抽出グループ数62個

5語以上の名詞によりグループ化されたものは17個。

グループ例：「保健、福祉、医療、機器——（イン、プラント、歯科）  
（採血、注射、しびれ）（金属、アレルギー）」  
「乗合、バス、サービス——バス停、停車、転倒、骨折」

このカテゴリでは、医療、保健のサービス、バスに関する事例などが見られる。

(サ) カテゴリ11（70歳代、死亡）：抽出グループ数52個

5語以上の名詞によりグループ化されたものは20個。

グループ例：「食品、嗜好、品——（団子、喉）（パン、喉）」  
「電動、車いす、移動——（転落、水田）（踏切、列車）」  
「医療、過誤——透析、入院」

このカテゴリでは、介護、医療、車いすなどに関する事例が見られる。

(シ) カテゴリ12（80歳以上）：抽出グループ数95個

5語以上の名詞によりグループ化されたものは19個。

グループ例：「住居、生活、用品、サービス——介護、大腿、骨」

「医療、機器、サービス——入れ歯、歯科」

「乗合、バス、サービス——バス停、発車、転倒、骨折」

このカテゴリでは、医療、保健、サービス、バスなどに関する事例が見られる。

(ス) カテゴリ13 (80歳以上、重傷・死亡)：抽出グループ数62個

4語以上の名詞によりグループ化されたものは27個。

グループ例：「乗合、バス、運輸——バス停、発車、転倒、骨折」

「建物、設備——(スポーツ、クラブ、階段)(スーパー、床、水)」

このカテゴリでは、バスや福祉などに関する事例が見られる。

(セ) カテゴリ14 (80歳以上、死亡)：抽出グループ数48個

4語以上の名詞によりグループ化されたものは23個。

グループ例：「医療、機器——電位、治療」

「養護、老人、ホーム——ご飯、のど」

「乗り物、移動——電動、車いす」

このカテゴリでは、介護、保健などに関する事例が見られる。

#### オ 分析における課題・留意点

分析の際に一番大きな課題となったのは、記入方法、入力用語が統一されていない点である。日付など、入力の際に標準化ができるものは、なるべく標準化することで、機械的な分析により、類似事例の把握などが容易になると考えられる。また、発生場所など未入力 of データも多く、統一的な傾向の分析が行えない属性も見られた。分析の際に、どのような分析をするかを視野に入れながら、必要に応じて必須入力事項などの設定をすることも考えられる。

本分析においては、カテゴリ内の事故に対して、いくつかのグループに分類を行った。しかし、語の共起頻度に基づく分類のため、グループ内のすべての事故が同じ形態とは限らず、グループ内においても複数のパターンが存在する可能性がある。また、同一の事故が複数のグループに分類される場合もあるため、似たグループの多さが同様の事例の多さを意味するものではない。このように分析手法により、固有の特性があるため、データ分析の際には、それを理解した上で、目的に応じた分析手法を設計する必要がある。データ分析は、全体の傾向を掴み、さらに

深い分析を行うための糸口であり、詳細な分析には、人間の目による個別事故の確認が必要である点にも留意が必要である。



## (5) テキストマイニングを用いた事故データ分析の試行と考察（西田 佳史）

### ア はじめに

事故対策や傷害予防で効果的な方法の一つは、関連する製品を特定し、その改善策を開発することである。近年、子どもや高齢者の製品に関連した事故が多発しており、データを活用した製品改善の方法論の開発が求められていることから、こうした生活機能変化者（心身機能や認知機能に変化しやすい子どもや高齢者）や製品関連事故を今回の分析の対象とする。

本調査の目的は、事故情報データベースに登録されている事故情報に関して、テキストマイニング技術等を活用した分析を行うことにより、事故対策（事故に起因する傷害の予防）を行うべき対象の明確化や、事故対策の効果評価に対してデータを活用する可能性を検討する。

本稿では、事故状況が記載された自由記述文を分析することで、子どもと高齢者が関与した重症事故と軽症事故の比較分析することで、重要領域の選定の可能性の検討をおこなう。また、今回、当委員会を通じて入手したデータ以外に、関連するデータベースの分析事例として、日本スポーツ振興センターの災害共済給付データを活用したトレンド分析（介入の効果評価）の可能性についても関連分析のケーススタディとして示す。

以下、試行結果とその考察を述べる。

### イ テキストマイニングを用いた分析結果

#### ○ 子どもの事故と高齢者の事故の分析

事故情報データベースに登録されている、子どもと高齢者の事故データについて分析を行った。

今回は、「事故内容詳細」の項目について、テキストマイニングを行い、名詞（製品名）を抽出した。その上で、各製品名が含まれているインシデントの数（全体数）と、そのうち、重傷1名、死亡1名、又は重症1名以上が含まれているインシデント数（重症以上数）をカウントし、その比率（重症以上数／全体数）を算出した。

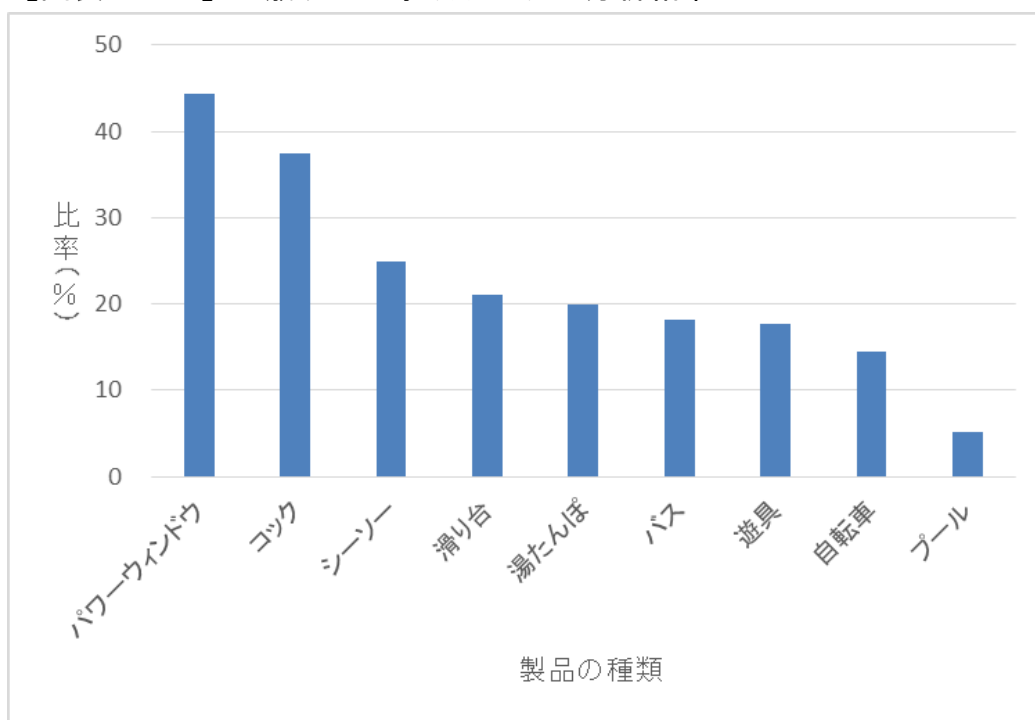
子どもについては、事故データの項目の「被害者の情報-年代等-年代」が「0歳以下」、「1～4歳」、「5～9歳」における全事故に出現する製品名に対して、重症以上数で出現する製品名の比率を算出した。

高齢者に関しては、「被害者の情報-年代等-年代」が「60歳代」、「70歳代」、「80歳以上」における全事故に出現する製品名に対して、重症

以上数で出現する製品名の比率を算出した。

図表 2-17に、9歳以下の分析結果を示した。図中「コック」は、「ウォーターサーバー」のコックである。縦軸の比率は、上述したとおり、重傷1名、死亡1名、又は重症1名以上である事故を「重症以上」とし、これが全体に占める割合を意味している。上位の製品は、パワーウィンドウ、コック、シーソー、滑り台、湯たんぽ、バス、遊具、自転車、プールなどとなった。

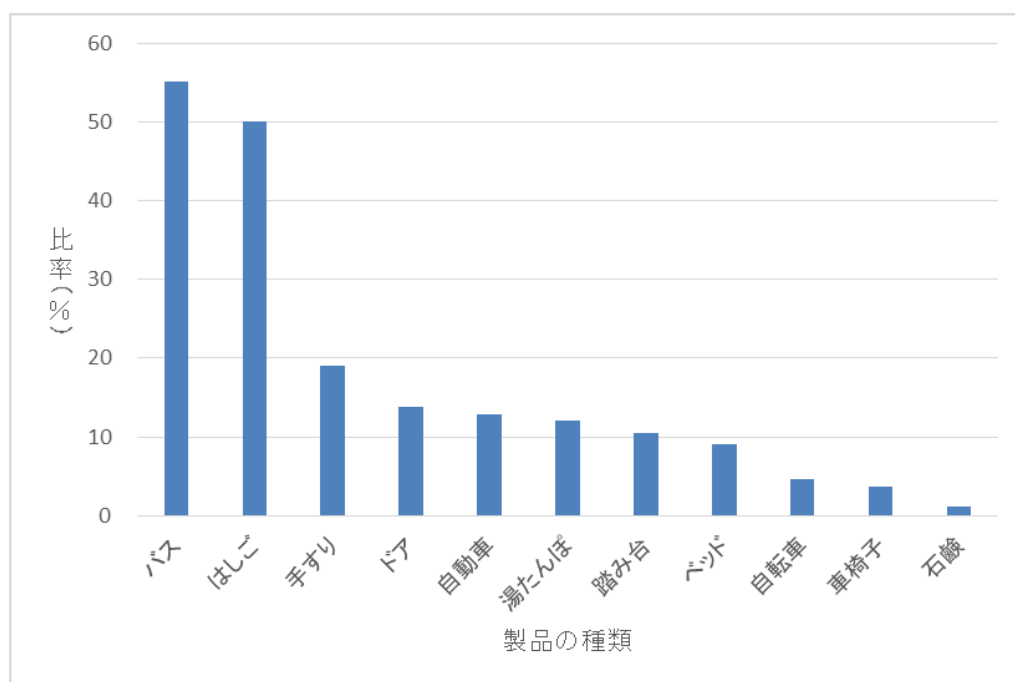
【図表 2-17】 9歳以下の事故データの分析結果



製品名	全体数(件)	重症以上数(件)	比率(%)
パワーウィンドウ	9	4	44.4
コック	8	3	37.5
シーソー	8	2	25.0
滑り台	38	8	21.1
湯たんぽ	10	2	20.0
バス	11	2	18.2
遊具	96	17	17.7
自転車	103	15	14.6
プール	39	2	5.1

図表 2-18に、60歳以上の分析結果を示した。上位の製品は、バス、はしご、手すり、ドア、自動車、湯たんぽ、踏み台、ベッド、自転車などとなった。

【図表 2-18】 60歳以上の事故データ分析結果



製品名	全体数(件)	重症以上数(件)	比率(%)
バス	388	214	55.2
はしご	8	4	50.0
手すり	58	11	19.0
ドア	203	28	13.8
自動車	62	8	12.9
湯たんぽ	33	4	12.1
踏み台	19	2	10.5
ベッド	277	25	9.0
自転車	149	7	4.7
車椅子	164	6	3.7
石鹼	1,498	17	1.1

これらの結果からテキストマイニング技術を用いることで、事故に関連した製品のリストの作成や、各製品と重症事故の関係を分析可能であることが分かった。これらの技術は、基本的には、インシデント報告書の中に含まれる製品名を数え上げているため、事故の直接的原

因が製品にあったのか、そうでなかったのか、単に書かれていただけの無関係な製品なのかまでを自動で行うことは現時点でも課題があると考えられる。そのため、重要トピックの候補の作成後は、人の目による精査も必要である。しかしながら、膨大なデータに対して、これを行うことは困難であるため、解析者がこうした人工知能を活用して作業を進めることは有用であると考えられる。適用と限界に関しては、「エ 考察」で改めて議論したい。

#### ウ 他のデータを用いたスポーツ外傷（柔道）のトレンド分析

学校管理下で発生した事故による傷害を科学的に予防するためには、対策や予防につながる重要な情報を含んでいる自由記述文を分析することが求められているが、現状では、膨大な自由記述文の処理は解析者の手作業に頼っていることから、詳細な分析は困難である。本研究では、人工知能技術（テキストマイニング）を援用することで、介入の効果評価を行う技術の実現可能性を検討する。

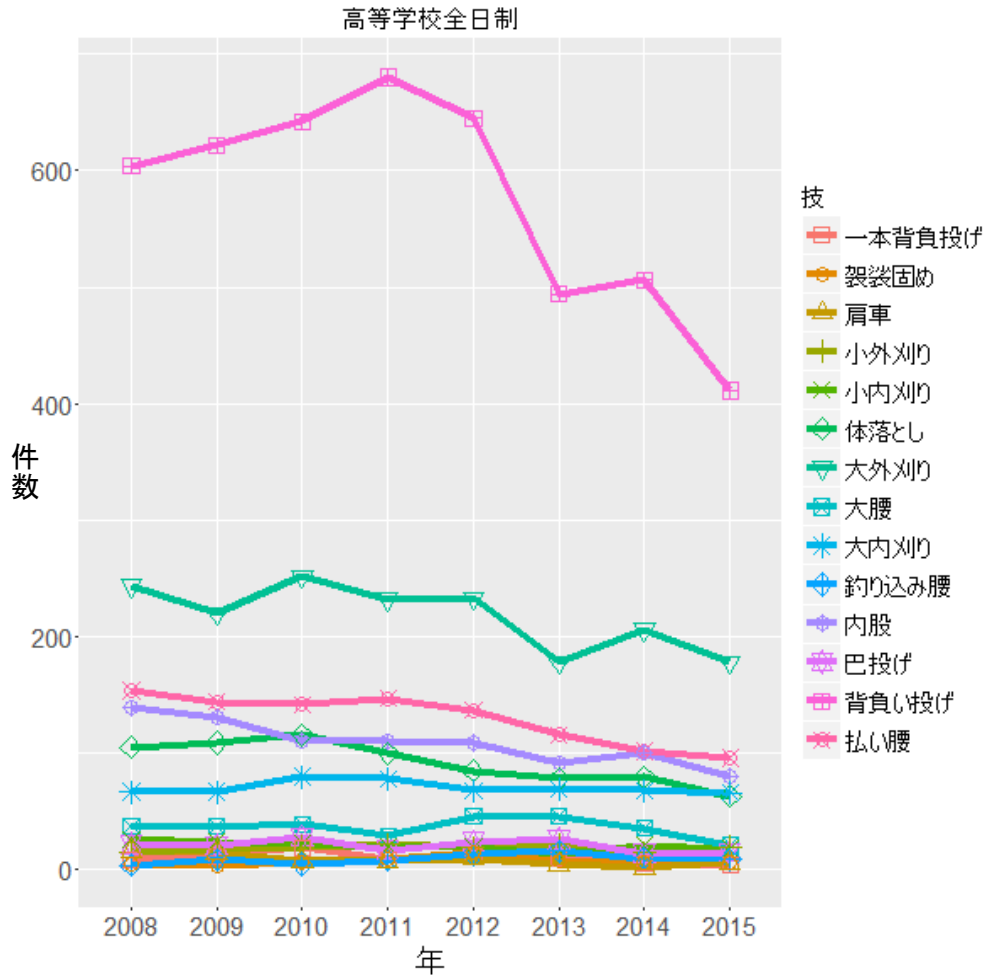
日本スポーツ振興センターの保有する災害共済給付データのうち柔道における災害事故ビッグデータ152,695件のうち被災年が2008年から2015年の8年間における高等学校の事故データ60,300件に対して、テキストマイニングを適用することで経年変化分析（トレンド分析）を行った。傷害の種類別の経年変化分析に加え、テキストマイニングを使わなければ困難な分析として、柔道の技に着目し、傷害に関連した柔道の技の経年変化を分析した。

図表2-19に分析結果を示した。傷害に関連した柔道の技の経年変化の分析から、「背負い投げ」が、2013年から顕著に減少していた。2008年と比較すると、2015年は34%減少していた。「大外刈り」も減少していたが、2008年と比較すると、2015年は17%減少であった。柔道では、2013年に事故に関する書籍が発行されるなど、社会問題化し、注意喚起が行われた。特に、「背負い投げ」と「大外刈り」のリスクが指摘された。経年変化分析から、こうした注意喚起の効果が、「背負い投げ」と「大外刈り」に関して確認された。

以上の分析により、テキストマイニング技術を用いることで、あらかじめ項目立てされていない項目を事後的に設定し、その数を調査したり、数の経年変化を調べるということが可能であることが分かった。事後的な設定に関しては、事故報告書のフォーマットの修正などの際に、新たな項目の候補を作るなどの応用が考えられる。また、数の経年変化の分析に関しては、注意喚起や安全基準作成などの何らかの対策・介入を行った際

に、それがどのような効果に繋がったのかを分析することに繋がる機能であり、効果評価への応用が考えられる。

【図表 2-19】柔道の技と事故の関係の分析



エ 考察（テキストマイニング技術の適用可能性と限界に関する留意点）

今回、テキストマイニング技術をインシデントデータに適用することで、どのような分析が可能かの検討をおこなった。消費者事故データの適用結果からは、事故に関連した製品の抽出や、製品ごとに重症事故が占める割合などの分析が可能であり、重要なトピックの候補を抽出することに有効であると考えられる。基本的には、文章に含まれる単語を数え上げるため、厳密には、本当に事故を起こした製品なのか、単に書かれていただけかを自動で判断することは、今なお課題があると考えられるが、トレンド分析による効果評価など、トレンドを把握したい、もしくは、他の種別とのインシデントと相対比較をしてみたい、という応用に関し

ては、有効であると考えられる。

一方、基本的には、単語の数え上げなので、対象として取り上げたいインシデント以外のものが含まれてしまっている可能性がある。そのため、裁判のための資料など絶対数が重要な応用には適切ではない場合がある。少なくとも限界を示したうえで数値を使うことが求められる。

以上より、テキストマイニングのあるべき使用方法としては、以下のものが考えられる。重要トピックのフィルタリング、注目すべきトピックの候補リスト作成、頻度の経年変化を見ることによる効果評価、事故を記録する用紙やフォーマットの変更のための新たな項目の候補リストの作成、などである。

## 2 事業者による分析

ここでは、株式会社プラスアルファ・コンサルティング<sup>18</sup>による、事故情報データベースに集約されたデータの検討・分析結果を整理する<sup>19</sup>。

### (1) テキストマイニングによる事故内容詳細（自由記述）の分析

テキストマイニングの手法を用いることにより、自由記述に含まれている情報を、どのような形で把握することが可能か、検討・分析を行った。

#### ア 単語のランキング

まず、自由記述を単語レベルに分解し、名詞、動詞それぞれを、出現頻度別にランキングした（図表2-20）。

名詞ランキングを示すことにより、事故、トラブルとなる対象を把握することができる。

また、動詞ランキングを示すことにより、事故、トラブルの原因となる消費者の行動を把握することができる。

【図表2-20】 テキストマイニングにより抽出した名詞及び動詞ランキング

名詞ランキング					動詞ランキング				
No.	単語	品詞	件数	割合	No.	単語	品詞	件数	割合
1	火災	名詞	1,840	6.6%	1	買う	動詞	4,888	17.5%
2	走行中	名詞	1,293	4.6%	2	焼損する	動詞	2,948	10.6%
3	対応	名詞	1,289	4.6%	3	出る	動詞	2,495	9.0%
4	メーカー	名詞	1,247	4.5%	4	発生する	動詞	2,225	8.0%
5	業者	名詞	920	3.3%	5	使う	動詞	1,778	6.4%
6	発煙	名詞	918	3.3%	6	出火する	動詞	1,498	5.4%
7	エンジン	名詞	892	3.2%	7	する	動詞	1,424	5.1%
8	ネット通販	名詞	730	2.6%	8	含める	動詞	1,177	4.2%
9	情報	名詞	698	2.5%	9	食べる	動詞	1,109	4.0%
10	治療費	名詞	684	2.5%	10	求める	動詞	950	3.4%

<sup>18</sup> <http://www.pa-consul.co.jp/corporate/index.html>

<sup>19</sup> テキストマイニングによる分析の対象は、「事故情報データベース」で公開されているデータのうち、2015年11月1日から2016年10月31日の間のデータ（登録日ベース、27,820件）とした。

イ 単語のランキングから係り受けの関係をマップ化

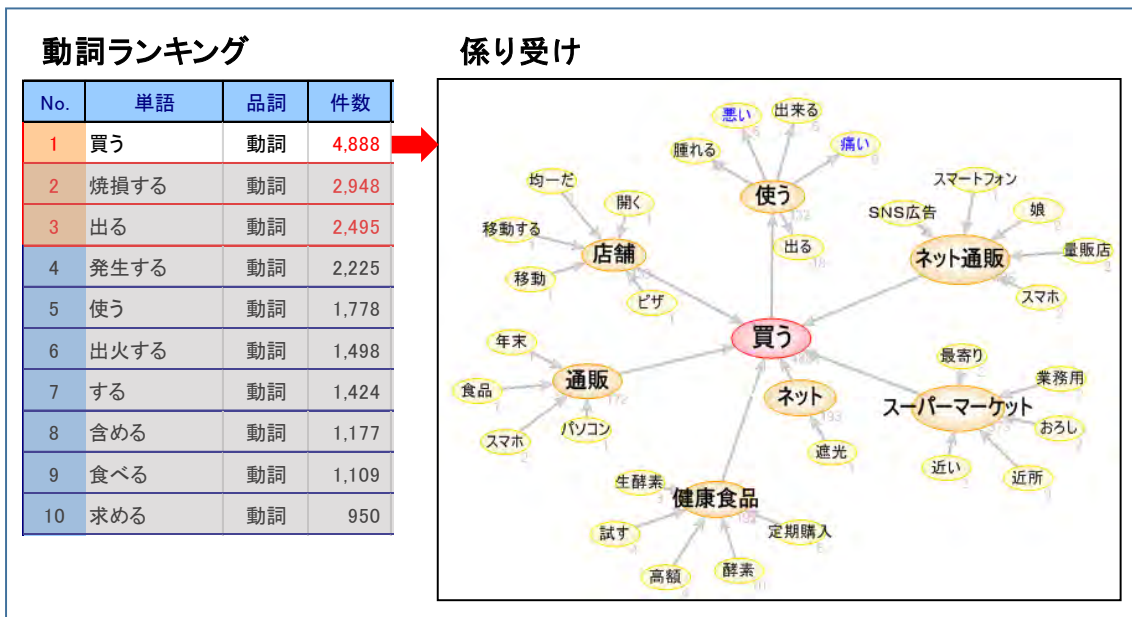
主語と述語の関係（係り受け）を認識させることにより、事故の内容をより詳細に把握することが可能である（図表2-21）。

ここでは、「買う」を中心において、購入製品、購入場所、その他の関連事項をつなげた。図表2-21には、「SNS広告」の「ネット通販」で「お試し」の「健康食品」を「購入」したら3口分の契約になってしまったといった、消費者トラブルが含まれている。

この係り受けをランキングすることにより、単語ランキングよりも、詳細に、どのような事故が多く発生しているのかを把握できる可能性がある。

例えば、この時期の事故情報を分析すると、「パソコン-内部-燃焼する」「タブレット端末用ACアダプター-コード部分-出る」等の係り受けが上位にあった。

【図表2-21】動詞から係り受けの関係をマップ化



ウ 共起の関係にある単語同士を線で結ぶことにより情報を分類

同じ文中に、一緒に発言される（共起の関係）単語同士をつなぐことにより、事故情報別のマップを作成することが可能である（図表2-22）。

この期間、子どもが足、指に火傷を負う、化粧品で肌トラブル、顔が腫れるといった問題等が生じていたことが分かる。