

消費者事故情報の活用等の あり方について

言語解析技術

相澤彰子（国立情報学研究所）

発表の構成

1. 目的
2. 分析する事故情報の分野
3. 分析に用いられた技術の概要
4. 分析の試行を通して明らかになった点
5. 分析の試行を通して課題と感じられた点

1. 目的

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、

- 分類項目の網羅性や整合性を調査
- 自由記述項目に対する言語解析の適用可能性を調査



効率的なデータ収集・管理・分析に向けた課題の整理を行う。

2. 分析する事故情報の分野

- 分類項目ごとのデータ登録状況に関する俯瞰的な調査
 - 事故情報データベースに登録されたすべてのレコード198,135件
- 言語解析の適用可能性に関する調査
 - 《情報提供元》が《国民生活センター》で、《事故種別》が《事故情報》である85,778件
 - 分析例では、上記からさらに「子ども」の「重大事故」にかかわる4,459件を選択

3. 分析に用いられた技術の概要

(1) 分類項目ごとのデータ登録状況に関する俯瞰的な調査

- 全事故情報 198,135件に対する83個の分類項目について、以下の数値を調査
 - 「辞書サイズ」：項目ごとの登録文字列の異なり数
 - 「網羅率」：値が登録されていない、いわゆる欠損値を除くデータの割合
 - 「平均バイト数」：項目ごとに登録されているテキストの長さの平均で、欠損値を除いて計算

3. 分析に用いられた技術の概要

(1) 分類項目ごとのデータ登録状況に関する俯瞰的な調査

調査対象（合計198,135件）

情報提供元	登録件数
国民生活センター	121,363
製品評価技術基盤機構（NITE）	20,060
消費者庁	17,653
厚生労働省	8,308
消費者庁・経済産業省・農林水産省	7,901



- 分類項目ごとのデータ登録状況は、情報提供元によって大きく異なる

3. 分析に用いられた技術の概要

(1) 分類項目ごとのデータ登録状況に関する俯瞰的な調査

■ 人手により分類項目をタイプ別に分類

種別	説明	分類項目の例
I	識別番号	《事故情報ID》など
P	個人情報で分析対象としないもの	《通報者情報-氏名》など
L	場所	《発生場所》など
D	時間・時期	《発生日時》など
T 0	リスト選択で辞書が固定されているもの (データベース運用期間を通して語彙が変化しないもの)	《事故種別》など
T	リスト選択で対象が固有名であるもの (データベース運用期間中に語彙を追加する可能性があるもの)	《商品など名称》など
T S	リスト選択に近いが一般名詞であるもの	《傷病内容詳細》など
S	自由記述文	《事故内容詳細》など

3. 分析に用いられた技術の概要

(2) 言語解析の適用可能性に関する調査

■ 85,778レコードを抽出

- 《情報提供元》が《国民生活センター》で、《事故種別》が《事故情報》
- 自由記述項目の平均バイト数が約890バイト（言語解析の適用に十分な分量）
- 《傷病内容》や《傷病の程度》の網羅率が1.0（比較的均質なレコード群）
- 網羅率等を踏まえて、合計で14個の分析用分類項目を設定

3. 分析に用いられた技術の概要

(2) 言語解析の適用可能性に関する調査

- 14個の項目をさらに以下の2つに振り分け
 - 「絞り込み項目」：値を指定して条件に合致するレコードを絞り込む。
L（場所）、D（時間・時期）、T O（リスト選択）
 - 「表示項目」：辞書サイズが大きいことから絞り込みには適さない。
T（固有名）、T S（一般名詞）、S（自由記述文）

（資料4-2の表1を参照）

3. 分析に用いられた技術の概要

(2) 言語解析の適用可能性に関する調査

処理の概要

- (ステップA) 絞り込み項目を使ったレコードのグループ化
- (ステップB) グループごとの項目要約表示

3. 分析に用いられた技術の概要

(2) 言語解析の適用可能性に関する調査

(ステップA) 絞り込み項目を使ったレコードのグループ化

- 絞り込み項目として指定した10個の項目の任意の組み合わせを指定して、値が同じレコードどうしをグループ化

例：《4月》に発生した《治療1週間未満の事故》

- 任意個の絞り込み項目を自由に組み合わせ、機械的に多数のグループを生成

試行したデータでは1,603,398グループを自動生成

- 得られたグループの中から注目すべきグループを選別するため、グループに特徴スコアを付与

スコアの計算方法は今後の検討課題

3. 分析に用いられた技術の概要

(2) 言語解析の適用可能性に関する調査

(ステップB) グループごとの項目要約表示

- (ステップA)で得られた各グループの項目ごとに、指定した数の特徴的な「キーワード」を表示
 - 《事故情報詳細》以外の項目については入力されている文字列そのもの
 - 《事故情報詳細》については自由記述文に対して言語解析（形態素解析と係り受け解析）を適用した結果から得られる名詞句、および名詞句と動詞の係り受けペア
- 「キーワード」は、統計尺度に基づく特徴度にしたがって、上位のものを出力

4. 分析の施行を通して明らかになった点

- 子どもの重大事故情報に、上記の分析技術を適用した例を以下に示す。
- まず、「子ども」に関する重大事故を抽出
 - ・ 「子ども」の定義は資料4-2の表2参照



《死亡》事故9件、《1カ月以上》の傷病にかかわる事故487件

4. 分析の施行を通して明らかになった点

■ 事故の頻度が高い事故グループの例

資料 4 - 2 表 3

■ 特徴的なグループ上位 3 件

1. 《中学生》の《化粧石鹼》による《皮膚障害》
2. 《住宅》における《ウォータークーラー》による《熱傷》
3. 《住宅》における《化学物質による危険》による《呼吸器障害》

4. 分析の施行を通して明らかになった点

まとめ

- 重大事故の分析は人手で詳細に行う必要があるが、その前段階として分析対象を絞り込み整理することが有効であると考えられる

4. 分析の施行を通して課題と感じられた点

- 情報提供元により「辞書サイズ」、「網羅率」、「平均バイト数」の傾向が大きく異なる

入インタフェースの共通化？

- 「子ども」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

目的に応じて辞書を整備することが必須

- 「絞り込み項目」で値が入力されていない欠損値の扱い

自由記述項目のテキストを利用して補完？

(参考資料) 名詞句のクラスタリング例

販売 会社
本部
バス 会社
販売 業者
本社
管理 会社
製造 業者
不動産 会社
製造 メー
カー
製造 会社
消費者 行
政 担当 省
庁
仲介 業者
工務 店
店舗 側
不動産 業者
タクシー 会
社
販売 社
ホテル 側
関係 機関
客 相

父 嫁
祖母
弟
子ども
孫 兄
義母
母親
息子
夫 姪
義父
甥
主人
子供
母 妻
娘
父親

対処 方法
関連 性
成分 等
非
危険 性
因果 関係
リスク
安全 性
有無
衛生 状態
店名
接客 対応
所在
品質
耐久 性
意見
害
知識
基準
可能 性

交渉 中
休み
平行 線
定休 日
同居
下痢 気味
傷 程度
連絡 待ち
年金 生活
原因 究明
中
便秘 気味
相談 中
保証 期間
内
調査 中
基準 値 以
下
健康 体
連絡 済
陰性
軽傷
敏感 肌

軸
支柱
取っ手
車輪
キャスター
パイプ
キャップ
金属 部分
ハンドル
ファスナー
ワイヤー
サドル
コルク
背もたれ
角
座 面
靴 底
縁石
ソール
フック

賃貸 マン
ション
廊下
居間
リビング
玄関
賃貸 アパー
ト
キッチン
風呂 場
台所
ベランダ
分譲 マン
ション
自宅 前
駐車 場
洗面 所
戸建て 住宅
脱衣 所
浴室
集合 住宅
道路
歩道

sns
テレビ
ショッピング
ネット 通販
テレビ 通販
tv ショッピ
ング
インター
ネット 通販
インター
ネット
割合
イベント 会
場
テレビ 番組
勧め
蕎麦 屋
催事 場
訪販
テレビ
ニュース
自力
1 回 払い
カタログ 通
販
代金 引き換
え
夫婦

洗顔 クリー
ム
基礎 化粧 品
セット
クレンジング
美容 クリー
ム
基礎 化粧 品
洗顔 料
ダイエットサ
プリメント
クレンジング
クリーム
ダイエット
食品
化粧 止め
ステロイド
軟膏
ビタミン 剤
健康 飲料
ロー ション
電気 治療 器
ファウンダー
ション
ダイエットサ
プリ
化粧 品 セッ
ト
青 汁
ダイエット
茶

入居 後
春
帰宅 後
購入 後
走行 中
飲用 後
食事 後
半月 前
夏 頃
使用 後
使用 直後
服用 後
終了 後
食事 中
入浴 中
運転 中
夕食 後
散歩 中
洗顔 後
乗車 中

耳鼻 科
内科
外科
救急 外来
総合 病院
口腔 外科
脳外科
他 病院
心療内科
脳神経 外科
精神 科
小児科
救急 病院
整形 外科
形成 外科
個人 病院
b 病院
大学 病院
耳鼻咽喉科
内科 医院

(参考資料) 名詞句のクラスタリング例

行きつけ
最寄り
市販
大手
専用
別
ショッピング
モー
ル内
馴染み
もうこ
系列
同メー
カー
市外
自主回収
中
同種
手持ち
なじみ
地元
実家近く
有名メー
カー
併設

動悸
頭痛
めまい
吐き気
咳
倦怠感
痺れ
かゆみ
鼻水
赤み
眩暈
耳鳴り
赤味
しびれ
胃痛
苦み
不快感
下痢症状
発疹
眠気

パスタ
お好み焼
き
肉まん
海鮮丼
鯖
いなりず
し
コーヒー
牛乳
きゅうり
アジ
盛り合わ
せ
うどん
日本酒
カップ麺
牛丼
茶碗蒸し
おでん
ブロッコ
リー
瓶詰
鶏
マグロ

手当て
解約返品
転院
返品交換
返金交渉
返品対応
原因調査
投薬治療
精密検査
話合い
無償交換
緊急入院
駆除
払い戻し
途中解約
開示
情報提供
測定
損害賠償
請求
賠償請求

レシート
納品書
名刺
控え
保証書
請求書
領収証
契約書面
契約書
誓約書
冊子
成分表
申込書
外箱
領収書
カルテ
合意書
書類
取扱い説
明書
パンフ
レット

入院費
タクシー
代
診断書代
薬代
治療代
病院代
治療費等
施術代
診察代
診療費
診察料
医療費
クリーニング代
初診料
通院費
修理代
購入代金
交通費
通院費用
検査費用

有害物質
物質
不純物
アルコール
小麦成分
ガソリン
加水分解
コムギ末
細菌
塩素
液
石油
薬液
黴菌
化学物質
科学物質
農薬
有機溶剤
毒
熱湯
小麦粉

一酸化炭素
中毒
原因だった
の
ノロウィル
ス
食中毒
中毒
不良品だっ
たの
胃腸炎
シックハウ
ス症候群
接触性皮膚
炎
中毒症状
シックハウス
外的要因
食あたり
急性胃腸炎
石鹸だった
の
アナフィラキ
シーショック
胃腸風邪
アナフィラキ
シー
熱中症
風邪

腹部
胸部
頬
手首
太もも
唇
お尻
首筋
手足
目の下
薬指
おでこ
肩
こめかみ
臀部
顔面
手の甲
首
両足
まぶた

使い捨てラ
イター
カセットコ
ンロ
電気温水器
石油ファン
ヒーター
空気入れ
電気ケトル
ヘアードライ
ヤー
オートバイ
ガスファン
ヒーター
自然冷媒
ヒートポンプ
給湯器
上着
チャイルド
シート
ストーブ
ハロゲン
ヒーター
ファンヒ
ーター
給湯器
石油ストー
ブ
自動車
加湿器
ガス給湯器