

大学における事故情報の 形態素分析を用いた解析

京都大学大学院工学研究科

長野 有希子

目次

- 1.背景と目的
- 2.使用した情報
- 3.分析概要
 - 3.1事故分類「その他」の再分類
 - 3.2事故情報の集計と因果の考察
 - 3.3事故分類のための構築学習と検証
- 4.考察
- 5.今後の課題

1.背景と目的

●過去の他大学における事故事例

- ・2009年にカリフォルニア大学にて、t-ブチルリチウムによる火傷により、研究員が死亡

Rex Dalron, University Fined after Researcher's Death, Nature (Online), doi: 10.1038/news.2009.447

- ・2011年にエール大学にて、工作機械(旋盤)に絡まった学生が窒息死

Richard Van Noorden, A Death in the Lab, Nature, 472, pp.270-271

●本研究の目的

事故情報を整理し、解析することで、発生した事故の傾向を理解し、事故低減に向けた計画の足がかりとする。

2.使用したデータ

対象:事故報告により集められた情報

期間:2010年度～2018年度半期

情報の種類

- ・個人に関する情報

(本解析では対象外)

- ・事故情報

分類、発生時刻、発生場所、物的損害、人的被害

事故の概要/対応に関する自由記述

3.分析概要

①事故分類「その他」の再分類

自由記述部にテキストマイニングを用いることで、「その他」の内容をさらに再分類する。

②事故情報の集計と因果の考察

再分類の結果から、事故報告件数を、月毎、年度毎、場所毎に集計し、傾向と原因を考察する。また、自由記述部を自然言語処理し、発生因果を考察する。

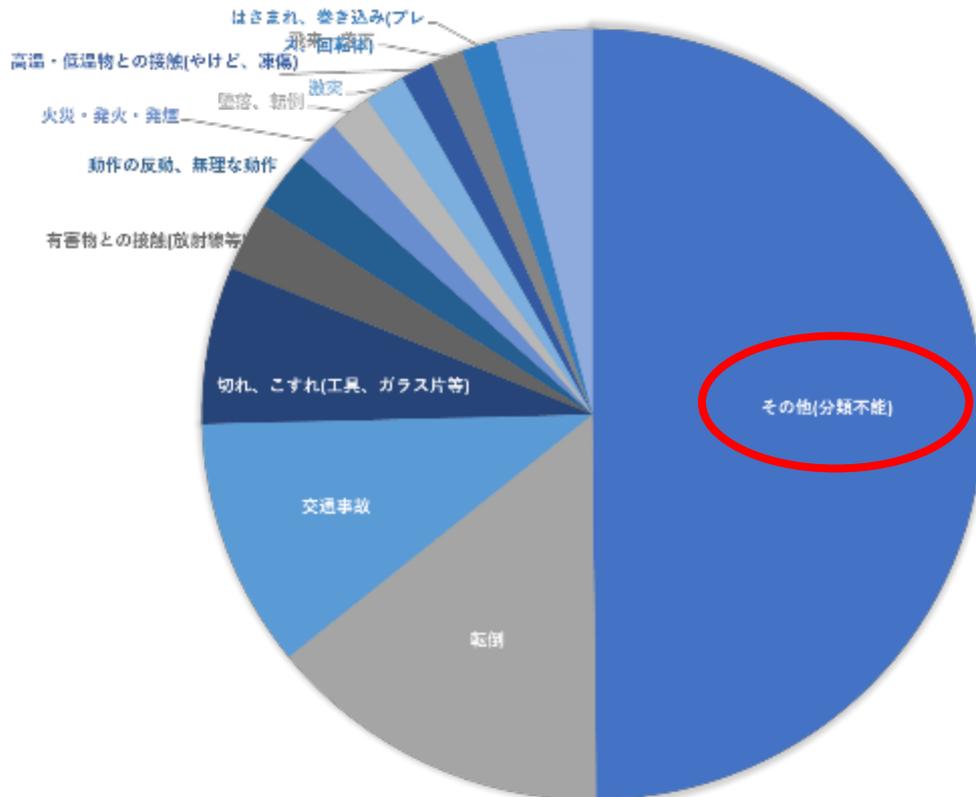
③事故分類のための構築学習と検証

事故分類を、負傷部位と傷の種類から分類できないかを考察する。

3.1結果 -事故分類「その他」の再分類-

事故分類を集計したところ、「その他(分類不能)」と報告されたものが、約半数を占めていた。

→事故連絡票が厚生労働省が指定している「事故の型」に基づいており、実際に発生した事故との分類に、差が生じたことが原因と考えられる。



自由記述部を自然言語処理し、使用されていた単語の頻度を確認することで、「その他」を再分類した。

図1 事故分類内訳(再分類前)

3.1結果 -事故分類「その他」の再分類-

[再分類方法]

“Text Mining Studio”を用いて、「その他」に分類されている自由記述部分に頻出する単語を確認ところ、「針刺し」に関連する言葉が多く見られた。「針刺し」との関連性は、“ことばネットワーク※1”と“原文検索※2”を併用した。また、“グルーピング機能※3”を用いて「針刺し」に関連する語句を含む情報を分類した。

※1 ことばネットワーク

Text Mining Studio中の分析機能の1つで、単語同士の係り受け関係をベクトルで表し、出現頻度を円の大きさに表したネットワーク図

※2 原文検索

Text Mining Studio中の機能の1つ。注視されている単語が元はどのような文章で使用されていたか、元となる原文を表示する機能。

※3 グルーピング機能

Text Mining Studio中の機能の1つ。テキスト中の単語および係り受け表現を抽出し、グループ分けを行う機能。

3.1結果 -事故分類「その他」の再分類-

再分類後も、「その他」の情報が全体の20%を占めていたことから、「その他」に対して同様の処理を二度行った。

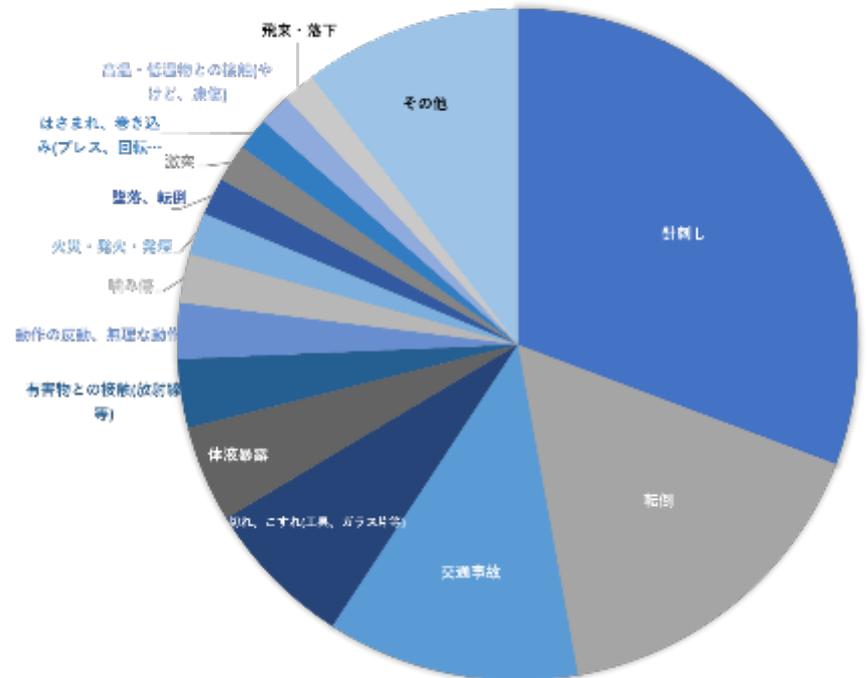
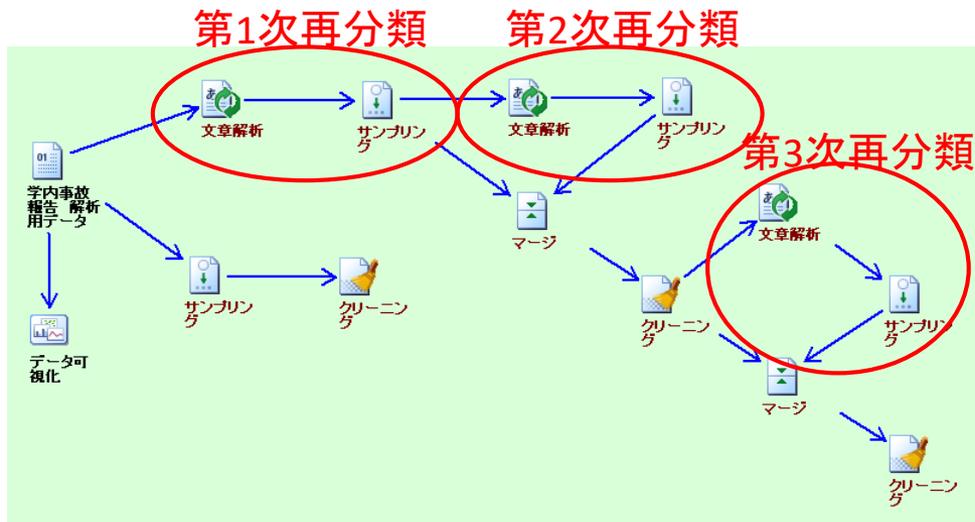


図2 事故分類内訳(再分類後)

3回の再分類を経て、「針刺し」、「体液曝露」を追加した。「その他」は約10%で、熱中症や体調不良などであった。

3.2結果 -事故情報の集計と因果の考察-

再分類後のデータを分析した結果を以下に示す。

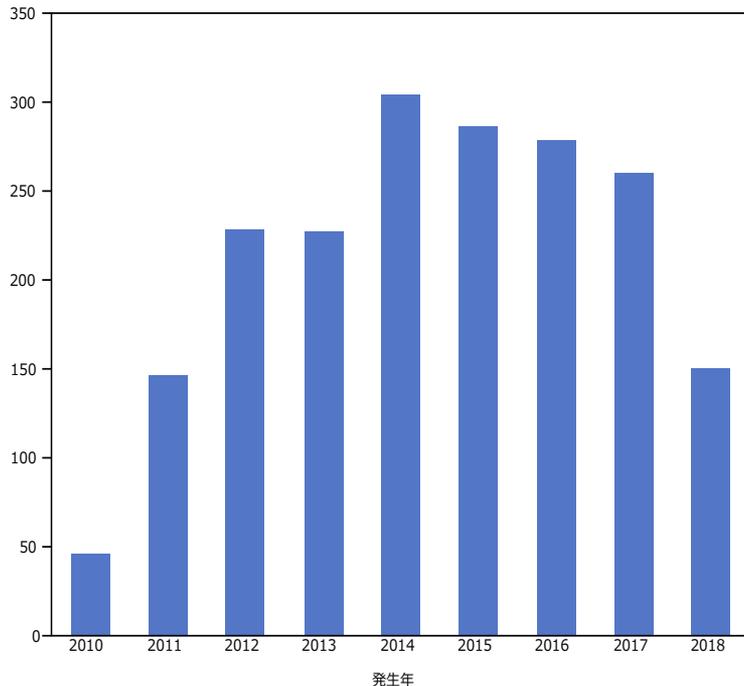


図3 報告件数の経年変化

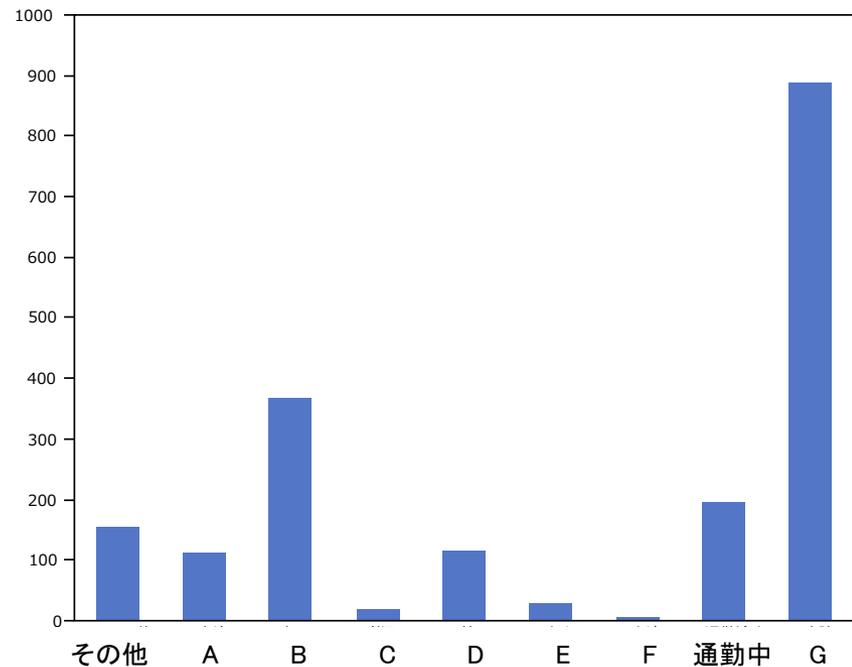


図4 場所ごとの報告件数の変化

図3より、2010年度から2014年度まで増加傾向を示していた。
図4より、場所「G」での事故が顕著に多いことが分かった。

3.2結果 -事故情報の集計と因果の考察-

月毎の事故の報告件数を図5(青が学生の件数)に、月毎の報告件数の割合を年度毎にまとめたものを表1に示す。

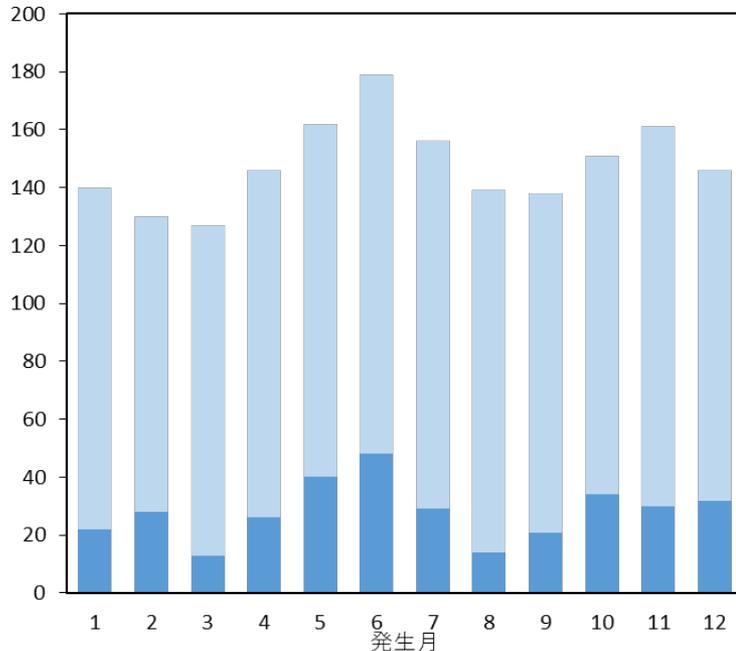


図5 月毎の報告件数の変化

表1 月毎の報告件数の経年変化(%)

	2010	2011	2012	2013	2014	2015	2016	2017
Jan.	0.0	6.2	6.1	6.6	10.2	7.7	8.6	9.6
Feb.	4.3	6.8	3.5	7.5	7.6	8.7	8.3	8.5
Mar.	0.0	9.6	5.3	6.6	5.9	7.0	8.3	9.6
Apr.	0.0	7.5	11.4	7.5	10.2	8.4	7.6	6.2
May	2.2	8.9	10.1	7.0	7.9	12.2	10.4	8.1
June	0.0	11.0	11.0	10.1	10.5	9.1	11.9	9.2
July	0.0	10.3	11.8	7.0	6.6	9.8	7.2	11.5
Aug.	4.3	7.5	9.2	7.0	9.9	5.2	9.0	7.3
Sept.	8.7	6.2	7.5	8.8	6.3	11.9	8.6	4.2
Oct.	19.6	9.6	6.6	12.3	6.3	7.0	7.9	9.2
Nov.	39.1	8.9	10.5	9.7	9.2	6.6	5.4	8.5
Dec.	21.7	7.5	7.0	9.7	9.5	6.3	6.8	8.1
number of data	46	145	228	227	304	286	278	260

図5から、月毎の報告件数は二峰性を示し、調査大学の学期等と関連している可能性が考えられた。学生からの報告は約17.5%を占めていた。

表1から、6月はどの年度も、期待値の8.3%を毎年上回っていた。

3.2結果 -事故情報の集計と因果の考察-

図6より、「転倒」に関連する形態素とその関連性が俯瞰できたことから、自由記述部における[転倒]という単語と共存する形態素と、その頻度を図7に示す。

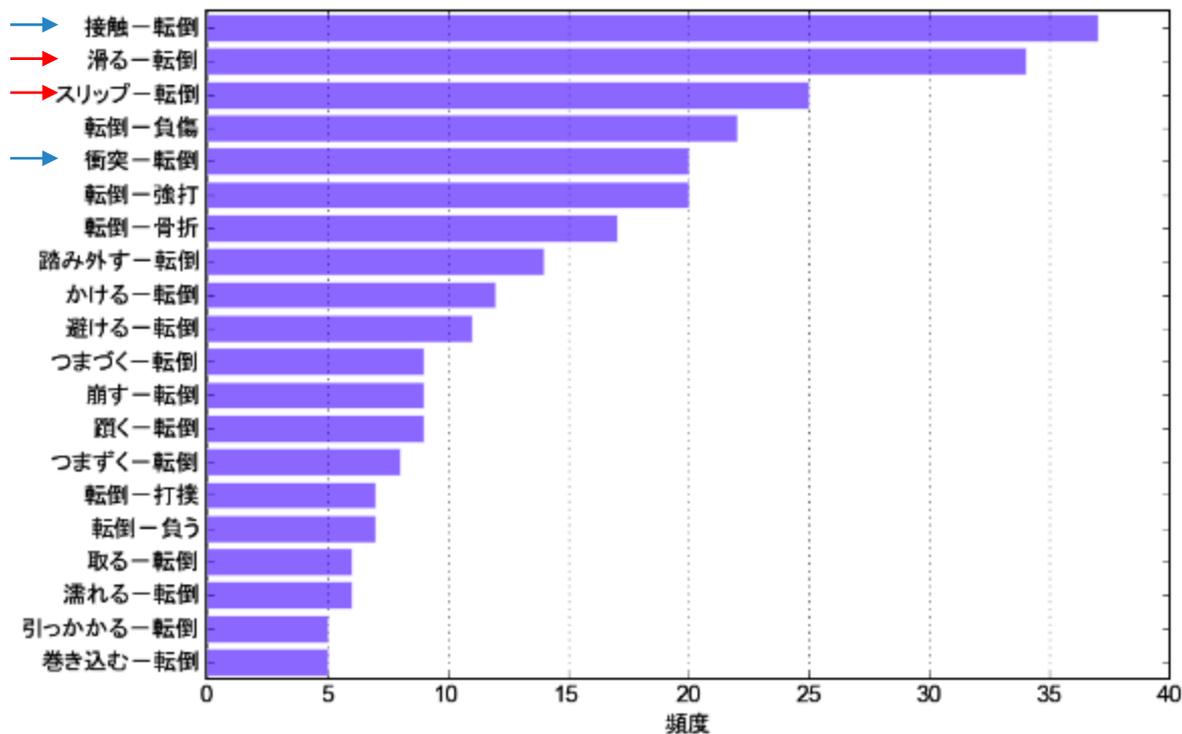


図7 [転倒]と共存する形態素の件数

[転倒]を含む事故は434件あり、その中で[接触-転倒]を含むものは37件、[滑る-転倒]を含むものは34件あった。

青矢印で指すものは[接触-転倒]にまとめられ、赤矢印で指すものは[滑る-転倒]にまとめることができる。

同義語が異なる件数としてカウントされてしまうことは、今後の課題と言える。

3.2結果 -事故情報の集計と因果の考察-

図7で多かった[滑る-転倒]と[接触-転倒]に関して、各単語を含む事故に絞り、どのような形態素が高頻度に含まれるを確認した。

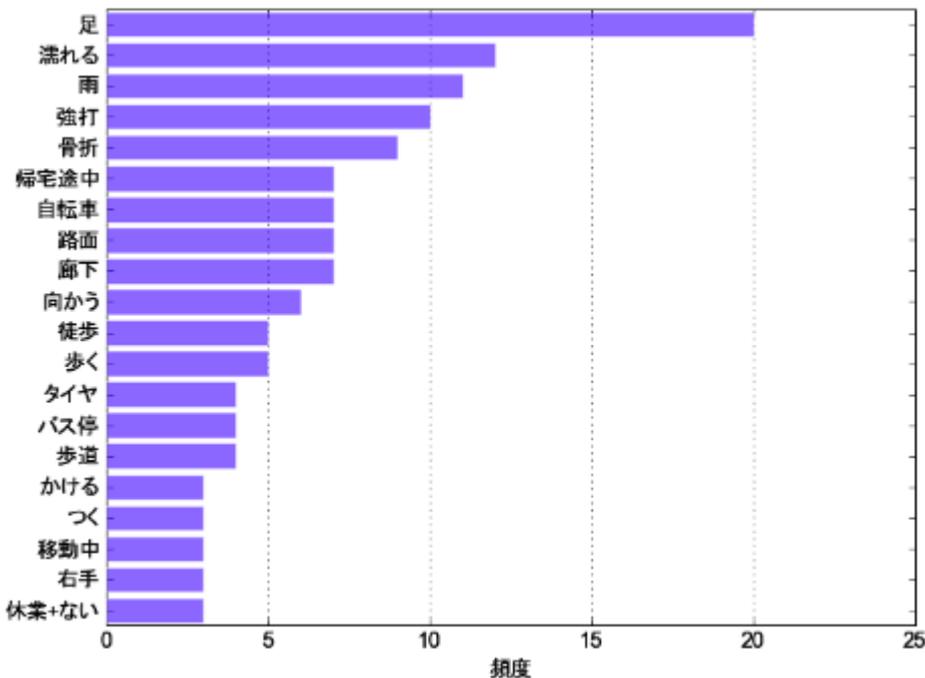


図8 [滑る]と[転倒]を含む形態素数

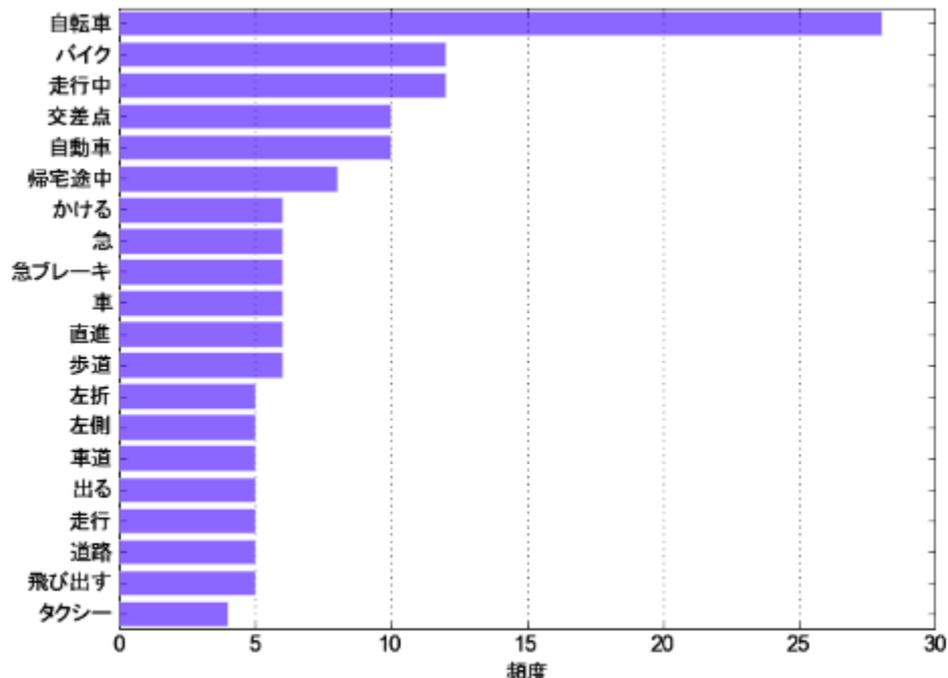


図9 [接触]と[転倒]を含む形態素数

図8からは、[滑る]が関わる転倒には[足]や[濡れる]、[雨]などの形態素が頻出することが分かった。図9からは、[接触]が関わる転倒には[自転車]や[バイク]、[走行中]などの形態素が使われており、特に[自転車]は28件と突出していた。

転倒は接触型とスリップ型に大別でき、雨や雪などで足元が悪い時と、自転車を避けたりぶつかったりした時がそれぞれの要因として考えられた。

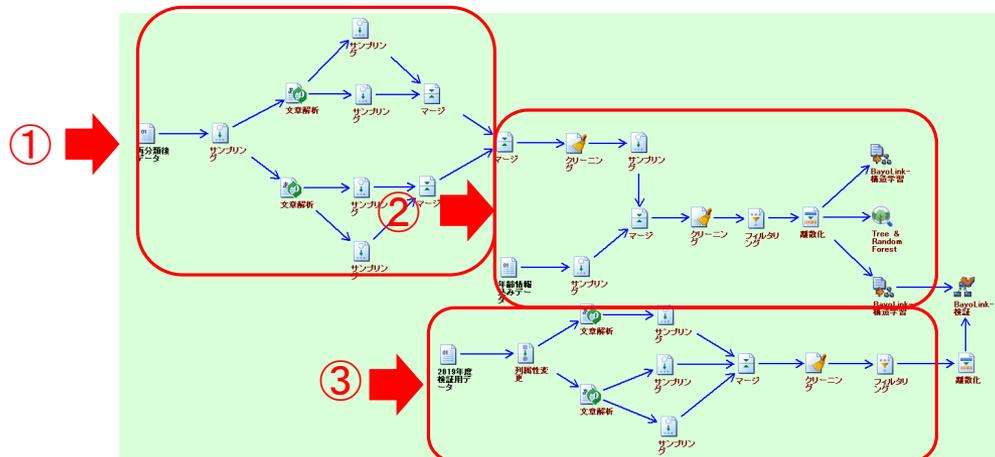
3.3分析 -事故分類のための構築学習と検証-

①学習用データの処理に関して

TMSのグルーピング機能を用いて分類し、それぞれ以下のグループとした。

負傷部位: 指、眼、頭、腕、脚、体幹

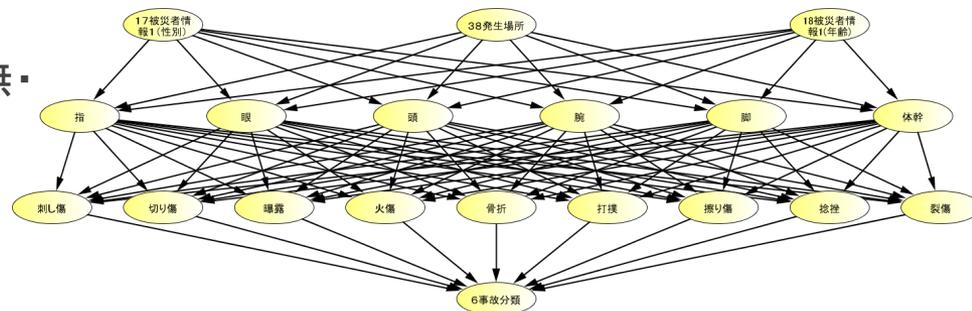
傷の種類: 刺し傷、切り傷、曝露、火傷、骨折、打撲、擦り傷、捻挫、裂傷



②学習データを用いた構造学習

目的変数: 事故分類

親ノード: 各負傷部位の有無・各傷の種類の有無・性別・発生場所・年齢



③検証用データ: 2018年度～2019年度(260件)

3.3結果 -事故分類のための構築学習と検証-

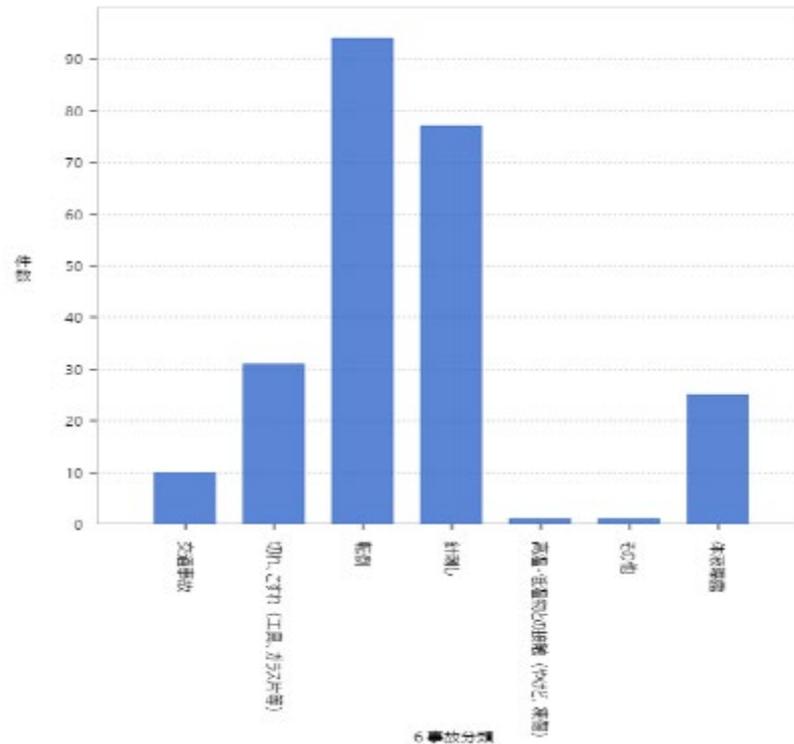


図10 検証結果

「切れ、こすれ」は77%を示した。

10件未満のものは検証に反映されずに「転倒」に含まれていた。

4. 考察

● 事故分類に関して

厚生労働省が定めた事故の型に該当しづらい事故が多く、「その他」として分類されていた。また、「転倒」と「交通事故」は、個人によって選択が異なる可能性があり、分類項目を検討しなおす必要があった。

● 再分類に関して

自由記述部から頻度の高い単語群をグルーピングする手法は、再分類に効果的であった。また、傷の種類に着目するのであれば、構造学習も有効と言えた。

● 事故の俯瞰に関して

車道から歩道に移動する際の自転車の転倒など、分類だけでは認識することが難しい傾向を“ことばネットワーク”から得ることができた。

転倒に関する分析では接触型とスリップ型に分類できることが明らかとなった。それぞれの要因についても形態素の発生頻度や種類からある程度予想でき、最終的に元の事故情報をたどることで、要因を特定することができた。

5. 今後の課題

● 事故分類

過去に生じた事故の種類に基づき、分類を再検討する必要がある。

● 同義語の扱い

テキストマイニングでは[滑る]と[スリップ]など、同義語が別単語としてカウントされる時があった。また、医学用語などの、漢字が続く語句は分かち書きしきれない部分があった。