

公募シンポジウム

公募シンポジウム3

セキュアかつフェアなデータ流通プラットフォームとセキュリティ基盤技術

2019年11月23日(土) 14:40 ～ 16:40 C会場 (国際会議場 2階国際会議室)

[3-C-2-04] 多機関分散データの統合的利活用による生活安全の課題対象や特徴的な条件の発見への試み

○北村 光司¹、西田 佳史^{2,1}（1. 産業技術総合研究所、2. 東京工業大学）

キーワード：Multi-organizational Distributed Big Data, Injury Prevention, School Safety

子ども、高齢者、障がい者などの生活機能変化者は、日常生活の中で事故や不具合が生じるが、その課題を把握するためのデータは点在しており、個人情報保護などの観点から複数機関のデータを統合することは難しい。本研究では、セキュアに多機関分散データを統合的に利活用する基盤技術を、具体的な課題に取り組みながら、実証的に開発することを目的としている。

応用の1つとして、学校環境での安全について取り組みを行っている。複数の学校に分散する事故データを、セキュアに学校を特定せずに統合する仕組みを開発してきた。データを統合した後は、そのデータから課題の対象を発見したり、その課題をさらに焦点を絞って分析を進め、課題解決につなげる手法が必要である。本研究では、事故件数と重症度を考慮した可視化手法を開発した。また、着目すべき事故の条件を分析する手法として、指定した変数について集計やクロス集計を行い、発生確率を計算し、その発生確率分布間の違いをカルバック・ライブラー距離を類似度として算出し、他と異なる条件を見つけ出す手法を開発した。

これらの手法を実際の学校での事故データに適用することで、例えば、中学校の部活動の事故では、特に事故件数が多い部活は、バスケットボール、サッカー、バレーボール、野球、テニス、陸上球技であり、ヨット部や自転車競技部は、事故件数は少ないものの、他の部活よりも給付金額の平均値が高い、といったように課題の特徴を把握することができた。また、着目すべき事故の条件の分析については、例えば、幼稚園・保育所で事故発生時間帯の分布が、0～2歳、3～6歳という年齢群によって異なることが分かった。他には、小学校で起きた跳び箱での事故による負傷部位について、学年別の発生確率分布を元に分析を行うと、手・手指部の負傷割合が高く、その割合は低学年の方が高く、学年が上がるにつれて割合が低下する、という傾向が分かった。

多機関分散データの統合的利活用による 生活安全の課題対象や特徴的な条件の発見への試み

北村光司^{*1, *3}、西田佳史^{*2, *1, *3}

*1 産業技術総合研究所、*2 東京工業大学、

*3 科学技術振興機構 CREST

Application to the Finding Problems and Characteristic Conditions in Living Safety by Technology for Integratively Utilizing Multi- organizational Distributed Data

Koji Kitamura^{*1}, Yoshifumi Nishida^{*2, *1}

*1 National Institute of Advanced Industrial Science and Technology,

*2 Tokyo Institute of Technology,

Living safety technology is strongly needed for children, elderly and, persons with disabilities. Their ability changes day by day and their problems occur in everyday life. Since the combination of their ability, everyday environment, activities is diverse, we need to understand their problems based on actual data. But it is difficult because injury data, medical data, insurance data and so on, are distributed in multi-organization and it is difficult to share and integrate them due to privacy protection. To solve this issue, our project is developing technologies for integratively utilizing multi-organizational distributed big data based on security technology. The authors research on school safety based on developing technologies.

In this paper, the authors propose a method for determining the priority of injury prevention. The method is a visualization method based on the number of injury and severity. We also propose a method for understanding characteristic conditions. By using the method, we can find important conditions have different trends from others based on the probabilistic distribution of injury occurrence under a specified condition. The authors describe some examples of the results that we applied the proposed methods to actual injury data in schools.

Keywords: Multi-organizational Distributed Big Data, Injury Prevention, School Safety

1. 緒論

子ども、高齢者、障がい者などは、日々生活機能が変わるため、日常生活の中で不具合が起きており、それらの不具合のうちの1つが事故である。そのため、生活機能が変わる人々の生活安全が求められている。日常生活の中で事故が起きるため、個別事例や不運にして起きた事故として扱われたり、そもそも包括的にデータを収集する仕組みがないために事故の実態が把握されていないという問題が起きている。そのため、データにもとづいて、科学的に課題を明らかにし、予防策を検討し、実際に予防策を実施し、その効果を検証する、という一連の取り組みを行う必要がある。しかしながら、生活安全を考える上で重要な事故データ、事故による傷害について医療機関を受診した際の医療データなどは、多機関に分散しており、事故の全数や事故が発生した原因から事故の結果生じた傷害まで、といった全体像を理解することが難しく、課題の把握や予防策の考案が難しい。そのような多機関に分散したデータを、統合的に活用することができれば、データにもとづいた課題把握から課題解決へとつなげることが可能となる。

しかし、機関をまたいだデータの共有や統合は、個人情報保護、プライバシー保護、情報漏えいなどの課題があるため、積極的には取り組まれないのが現状である。この現状が、データに基づいた生活安全を阻害している。JST CREST(ビッグデータ統合利活用促進のためのセキュリティ基盤技術の体系化)プロジェクトでは、セキュリティ基盤技術を活用した多機関分散データの利活用技術の開発を進めている。具体的にはセキュアに多機関のデータを統合的に活用する技術のベースに、実用用に活用する取り組みを行っている。著者らの研究グループでは、生活安全分野を対象に、データ保有機

関である医療機関等やデータ活用者である製品デザイン等の現場と連携し、実証的に多機関分散データの利活用技術を開発することを目的としている。これまでに、消防庁、日本スポーツ振興センター、複数の医療機関、保育所・小中学校などと連携し、分散された傷害関連データを統合的に活用する技術の開発を進めている。

具体的な応用の一つとして、学校安全を取り上げ、学校環境下で起きた事故を対象に技術開発と実証を進めている。これまでに、複数の小学校に分散する事故データを、学校を特定されないように情報を秘匿したまま、データを統合することで、事故情報を共有したり、重症となる事故要因を分析する手法を開発してきた¹⁾。本稿では、統合した多機関データを対象に、対策を取る優先度が高い対象を見つけ出すための可視化手法や、原因究明や着目点を把握につながる特徴的な条件を見つけ出す手法について報告する。

2. 対策の優先度検討のため可視化手法

学校環境下で発生する事故や病気についての保険業務を行っている日本スポーツ振興センターによると、毎年、幼稚園、保育園、小学校、中学校、高校で発生する保険給付件数は100万件以上発生しており、その90%以上は事故による負傷である。つまり、年間90万件以上の事故が発生していることとなり、これらの事故を一度に全て対策を取ることは困難であるため、何らかの優先順位をつけて、優先度が高いものから取り組んでいく必要がある。優先度の指標としては、重症度が高いもの、多発しているものが考えられる。また、何らかの分類で事故を分類して見たときに、個々の分類で比較したときに、重症度が高い事故が含まれる割合が高い、ということも重要な指標である。これらを考慮して、優先度を判断可能な

可視化手法として、事故件数と重症度の中央値のデータを元に個々の分類ごとに2軸マップ上にプロットし、さらに重症度のパーセンタイル値を円の大きさで表す可視化手法を開発した。図2に具体例を示して、詳細について述べる。図2は、高校で起きた事故を場所別に整理したマップで、重症度については保険給付額を採用し、横軸は給付金額の中央値、縦軸は件数の対数とした。また、場所ごとに給付金額の最大値を一番外側の円で、その内側に同心円状に給付金額の95, 75, 50, 25, 5パーセンタイル値に応じた半径の円で表している。つまり、一番外側の円の半径が大きいものは、最大給付金額が高いことを表し、内側のパーセンタイル値を表す円が大きいほど、給付金額が高い事例の割合が高いことを示す。このようにプロットすることで、右上のエリアにプロットされたも

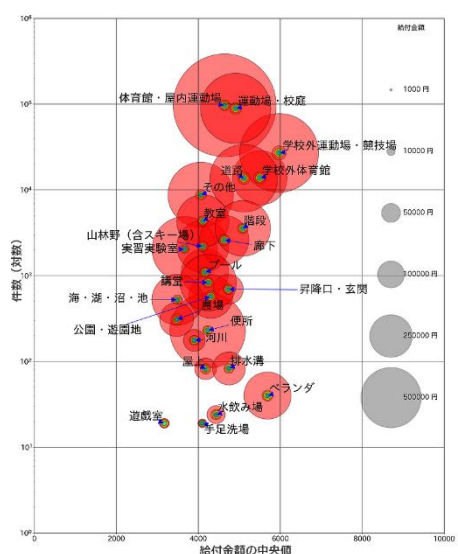


図2 高校における場所別事故のマップ

のは、件数も多く重症度も高いので、対策を取る優先度が高いと言える。上方のエリアにプロットされたものは、件数が多いので、件数を減らす対策が必要で、右側のエリアにプロットされたものは、重症度が高いので重症度を低減する対策が必要である。このように、それぞれの特徴に応じて、対策の方針を検討することも可能である。図2からは、体育館・屋内運動場や運動場・校庭は事故件数が特に多いこと、体育館・屋内運動場は最大給付金額が最も高いこと、学校外運動場・競技場やベランダは他の事故よりもやや重症度が高いことが分かる。

開発した可視化手法を、中学と高校の部活動での事故に適用したマップを図3、図1に示す。まず、図3の中学における部活動での事故のマップを見ると、給付金額の中央値が4千円付近に集中していることが分かる。最も件数が多いのはバスケットボール部で、サッカー・フットサル部、バレーボール部などが続く。特徴的なものは、ヨット部は件数が比較的少ないものの、給付金額の中央値が1万円近くと、他の多くの部活動の倍以上となっており、件数は少なくとも、事故が起きると重症度が高い事故になっていることが分かる。同様に、自転車競技部やレスリング部も、重症度が高い事故が起きていることが分かる。特に自転車競技部は、給付金額のパーセンタイル値を表す円が、他の部活動とは傾向が異なり、重症度が高い事故が占める割合が比較的高いことが分かる。次に図1の高校における部活動での事故のマップを見ると、給付金

額の中央値が5千円付近に集中しており、中学よりも全体的に重症度が高い事故が起きていることが分かる。最も事故件数が多いのはサッカー・フットサル部で、バスケットボール部、野球部が続く。このうち、野球部は、給付金額の中央値が約6千円となっており、重症度が高い事故が比較的多く起きていることが分かる。中学の部活動で件数は少ないものの、重症度が特に高かったヨット部は、給付金額の中央値が比較的高いが中学ほど顕著ではない。ヨット部に代わり、自転車競技部が、給付金額の中央値では最も高く、件数も増加していることから、対策が必要な課題であと考えられる。高校では、新たにスキー部が給付金額の中央値が高い部活として示されている。

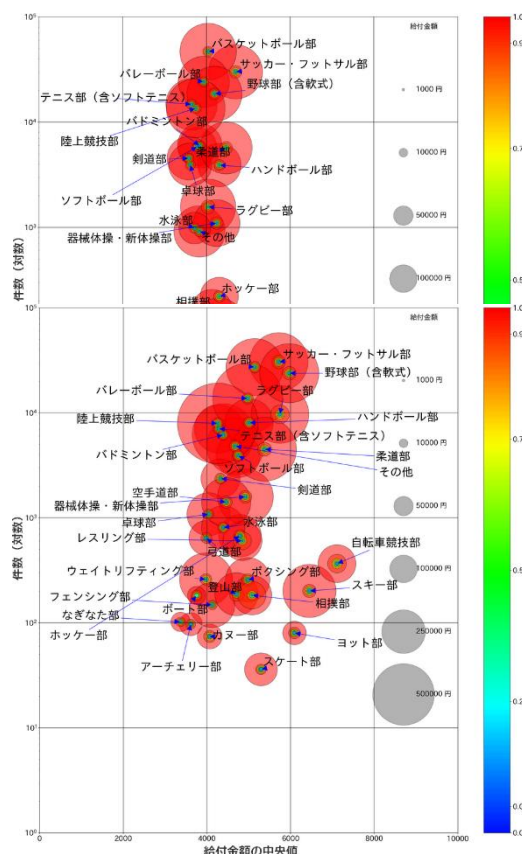


図1 高校における部活動の事故マップ

このように複数の指標を統合して可視化することで、課題の発見や考察に有用である。このマップを時系列に作成すれば、どのように事故の特徴が変化したのかを把握したり、対策を講じた対象であれば、重症度を下げることができているのか、件数を減らすことができているのか、といった評価を視覚的に行うこともできる。

3. 着目すべき事故条件の分析手法

事故の原因を究明したり、対策を考案するには、事故の実態を漠然と捉えるだけでなく、例えば、特に重症度が高くなる条件や、他とは傾向が異なる条件といったように、着目すべき視点を決め、その視点で詳細に分析する必要がある。着目すべき視点を、人手で事故データを見ながら検討することも考えられるが、たまたま目についた事例に偏ってしまったり、経験や勘に頼ってしまい、課題を適切に把握することが難しい。

そこで、指定した対象の事故について、網羅的に条件を比較し、特徴的な条件を見つけ出す手法を開発した。具体的には、対象となる事故データについて、指定した変数で集計やクロス集計を行い、各条件の発生確率を算出し、その発生確率分布間の違いを、Jensen-Shannon ダイバージェンス距離を類似度として算出し、他と異なる特徴を示す条件を見つけ出す手法を開発した。図5に具体例を示して、詳細について述べる。図5は、幼稚園と保育園での事故データを対象に、年齢群(0～2歳、3～6歳、全体)ごとに、事故が発生する時間帯の確率分布を示している。横軸が時間で、縦軸が確率を示している。このとき、年齢群の確率分布同士で Jensen-Shannon ダイバージェンス距離を算出して、類似度を計算する。これによって、他の条件とは異なった傾向を示す条件を見つけ出すことができる。同様の計算を、複数の変数を条件として繰り返すことで、特徴的な条件を網羅的に見つけ出すことが可能である。また、視覚的にも確認できるように、算出した Jensen-Shannon ダイバージェンス距離を元に、多次元尺度構成法を用いて、2次元マップ上に可視化したものを図4に示す。これにより、0～2歳と3～6歳では事故発生時間帯に傾向の違いがあることが分かる。0～2歳は昼過ぎの時間帯では事故発生確率が低くなるが、3～6歳では事故が発生していることが分かる。これは当然のことで、0～2歳は保育園児であるため、午睡の時間帯には事故が発生しにくく、3～6歳は幼稚園児も含まれ、幼稚園では基本的には午睡の時間が無いため、事故の発生確率が高くなっているのである。このように網羅的に特徴的な条件を見つけ出した上で、分析者が背景も含めて検討することで、分析者の経験や勘にとらわれず、意味のある特徴を把握することが可能である。

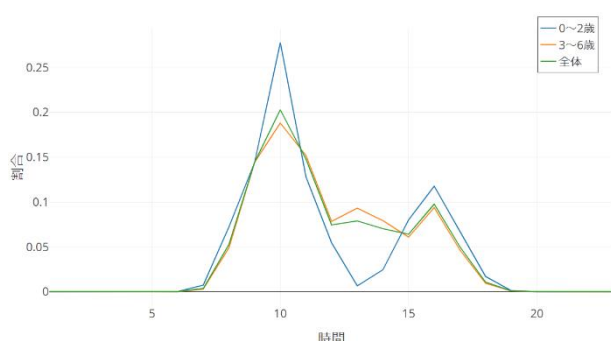


図6 年齢群別の事故発生時間帯の確率分布

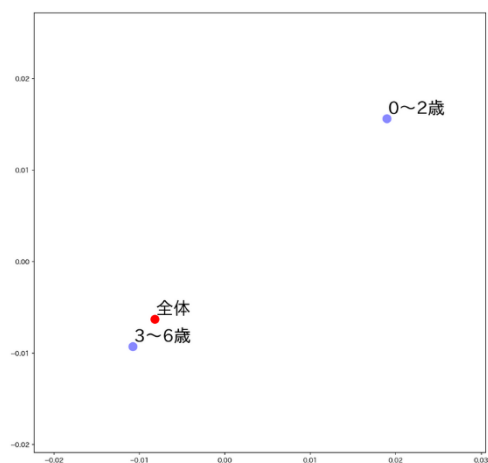


図5 事故発生時間帯の確率分布の年齢群間距離を元にした多次元尺度構成法マップ

開発した手法を、幼稚園と保育園で起きた遊具による事故について、傷害の種類の発生確率分布を対象に適用した結果を図6に示す。他の傾向と異なる遊具として、回旋塔、登り棒、固定タイヤ、雲てい、シーソー、砂場などが見られる。砂場については他の遊具よりも骨折の割合が低く、回旋塔については他の遊具よりも裂創や挫創の割合が高い。登り棒と固定タイヤは遊具としては特徴が異なるが、傷害としてはよく似た傾向を持ち、他の遊具よりも捻挫が多いのが特徴である。シーソーは、他の遊具よりも脱臼の割合が特に高い。雲ていについては、他の遊具は、打撲・挫傷が最も多く、2番目に骨折が多いが、雲ていは骨折が最も多く、2番目が打撲・挫傷である特徴を持つ。このように各遊具の特徴を普段とは異なる観点で把握できると、ある遊具について既に対策があれば、特徴が類似している他の遊具に同様の対策を適用できないか検討する、といったことも可能となる。

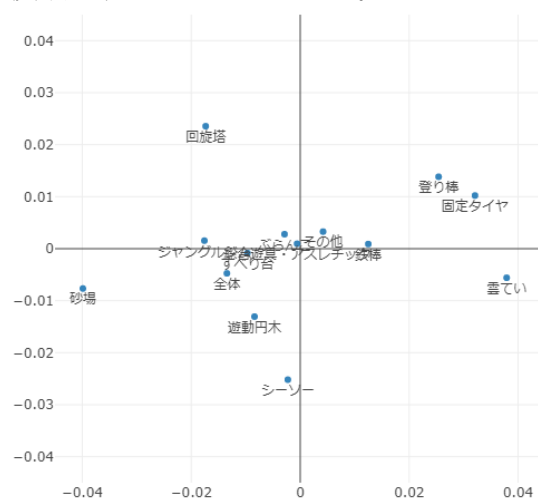


図4 傷害の種類の確率分布の遊具間距離を元にした多次元尺度構成法マップ

4. 結論

本稿では、セキュリティ基盤技術を活用した多機関分散データの利活用技術の応用として、学校安全を目的にした傷害データの分析について、対策を取る優先度が高い対象を見つけ出すための可視化手法や、原因究明や着目点の把握につながる特徴的な条件を見つけ出す手法について報告した。可視化手法については、優先度に関連する重症度と件数を合わせて検討可能な可視化手法を提案し、部活動の事故データに適用した事例を紹介した。特徴的な条件を見つけ出す手法については、指定した変数について、事故発生確率分布を元に他とは異なる傾向を示す条件を見つけ出す手法を提案し、幼稚園・保育園での事故データに起用した事例を紹介した。今後、これまでに開発してきた手法やシステムを統合し、データの入力から分析までを一貫して行えるシステムとして開発し、継続的な現場での活用を目指す。

謝辞

本研究は、JST, CREST, JPMJCR1404 の支援を受けたものである。

参考文献

- 1) 北村光司, 西田佳史. 多機関分散データの統合的利活用による生活安全の課題把握と課題解決への試み. 第38回医療情報学連合大会予稿集, 2018.