

令和 2 年 6 月 8 日現在

機関番号：16101

研究種目：基盤研究(C)（一般）

研究期間：2016～2019

課題番号：16K00372

研究課題名（和文）テキストからのリスク感性表現抽出とヒヤリハット潜在因子の横断的高速判定法の開発

研究課題名（英文）Extraction of the risk sensitivity from texts and development of the diagnosis of hiyari-hatto cases

研究代表者

森田 和宏（MORITA, Kazuhiro）

徳島大学・大学院社会産業理工学研究部（理工学域）・准教授

研究者番号：20325252

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：ハインリッヒの法則に基づき、多くの分野でヒヤリハット事例の報告が実施され、どのように事故を防ぐべきかの対策が講じられている。一方で、未然予防としてテキストからヒヤリハットを抽出する技術開発が必要である。本研究では、ヒヤリハットをリスク感性表現として抽出して事故の分類と関連付け、どの分野にも適用できる知識構造を構築する技術を確立し、テキストから分類判定の有用性を評価した。

研究成果の学術的意義や社会的意義

本研究の成果によって、危険感受性の低下やヒヤリハットを報告しないなど、対策に活かせない状況を解決するため、日々発生する業務報告などからリスク感性表現を検出することで、未然予防としてのヒヤリハットを判定できる新しい技術を確立した点、各分野で個別にヒヤリハット情報を抽出し、対策に結びつけていた人手の作業を自動化できるので、余った人手は安心安全対策の推進へと注力できる点に学術的、社会的意義がある。

研究成果の概要（英文）：Based on Heinrich's law, the report of hiyari-hatto cases are implemented in many fields, and it is devised how an accident should be prevented. On the other hand, the technical development which extract hiyari-hatto cases from texts as prevention before it happens is required. In this research, the hiyari-hatto case was extracted as risk sensitivity expression, and it related with the classification of an accident, and established the technology of building a knowledge contexture applicable to every field. The serviceability of the classification from texts was evaluated.

研究分野：感性情報処理，自然言語処理

キーワード：リスク感性 知識獲得

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

ハインリッヒの法則「1件の重大事故の陰には29件の重大事故となり得る軽傷事故があり、さらに300件の無災害ながらヒヤリまたはハッとした体験が隠されている」に基づき、医療、介護、交通、建築など多くの分野でヒヤリハット事例の報告が実施され、どのように事故を防ぐべきかの対策が講じられている。しかし、報告書作成の手間や、責任追及されることへの懸念から、ヒヤリハット情報を報告しない事例も存在し、有用な対策への見落としにつながったり、逆に有用な潜在因子を抽出するには対応しきれないほど大量の報告があったりするなど、人手に頼るだけでは課題が多い。

また近年、何がヒヤリハットか分からない、ヒヤリまたはハッとしないといった危険感受性の低下が問題となっており、危険に気づき事故を回避して安全行動をとるといった危険感受性を磨く教育が必要とされている。「床が濡れていたため滑って転びそうになった」では「床が濡れていた」が潜在因子であるが、「洗髪時に床を濡らしてしまった」から「転倒」の潜在因子に気付くことが危険感受性であり、体験型教育が有効といわれている。

一方、ヒヤリハットに関する研究動向は、事例の解析や要因の分析によって、対策を考察する研究が主であり、未然予防としてテキストからヒヤリハット潜在因子を抽出する技術開発が必要である。また、ヒヤリハット事例は医療、介護、交通、建築など多くの分野で独立して議論されるため、横断的に、どの分野にも適用できる体系化も必要である。

2. 研究の目的

(1) 膨大な報告から有効なヒヤリハット潜在因子(事故を引き起こす潜在的要因)をリスク感性表現として抽出し、危険因子(事故の型分類)と関連付ける知識構造を構築する。

(2) 医療、介護、交通、建築など各分野の既知のヒヤリハット事例を収集し、危険因子(転落、火災、倒壊、破裂など)の横断的な、どの分野にも適用できる体系を構築する。

(3) 既存の収集済みコーパスを用いて、リスク感性表現を抽出し、危険因子体系から「転落」などの危険因子を判定する手法を開発し、実験により性能を評価する。

3. 研究の方法

(1) ヒヤリハット潜在因子をリスク感性表現として抽出し、危険因子と関連付ける知識構造の構築

収集されたヒヤリハット事例「踏み台を踏み外しそうだった」から意味共起「踏み台 + 踏み外す」を抽出し、危険因子「転落」を分類名として定義することで意味共起による抽出知識を構築する。抽出には、意味共起知識による文脈解析の研究成果を利用する。次に、抽出した意味共起に対しリスク感性表現となり得る表現を検出する照合ルールを作成する。最後に、リスク感性表現と危険因子の関連づけには、印象知識構築法の研究成果を利用して、語彙レベルで相互検索可能な知識構造を構築する。

(2) 各分野の既知のヒヤリハット事例を収集し、危険因子の横断的な、どの分野にも適用できる体系の構築

ヒヤリハット事例の収集はWEB上からおこなう。意味共起の抽出作業において、分類名称は作業時に「装置を使用する場合の危険因子」、「乗り物移動の場合の危険因子」などのように、それぞれの状況に応じた分類候補として作成しておく。その際、分野のその区別なく、先入観無しに一括で抽出作業を実施する。これにより、横断的(普遍的)な体系構築を効率化する。分類が曖昧となり分類場所がわからない場合は、曖昧性を考慮した分類の研究成果を利用し、最も近い複数の分類に対応させて、曖昧性を残して全体候補を構築しておき、意味共起の分類体系構築の研究成果を利用して、意味共起の包含関係による分類を決定することで絞り込む。この際、意味共起が非常に少ない分類が生じる場合があるので、最終的に人手で振り分けることで構築する。

(3) 収集済みコーパスを用いて、リスク感性表現を抽出し、危険因子体系から「転落」などの危険因子を判定する手法の開発と、実験による性能評価

ヒヤリハット潜在因子から危険因子の判定手法の開発は、共起情報格納型トライ構造の研究成果を利用して実装する。開発した判定手法の評価には、既存の収集済みコーパスに対して提案手法を適用し、精度を評価する。ここで、知識構築は人手に頼るところが多く、検出漏れが生じることが想定されるので、不足する言語表現を補うため、感性表現コーパスを利用する。また、「転倒」などの表現においては「転ぶ」「倒れる」などの類似表現が存在するので、これらは類義語辞書を拡張することで精度向上を図る。

4. 研究成果

(1) ヒヤリハット潜在因子をリスク感性表現として抽出し、危険因子と関連付ける知識構造の構築

例えばヒヤリハット事例「午前8時30分頃、給油が終わった車を送り出そうとして、車道に出たところ、走行中の車に撥ねられそうになった」(職場のあんぜんサイト:ヒヤリ・ハット事例)の場合、「車に撥ねられそう」の部分に注目すると、交通事故につながると考えられるので、意味共起「車 + 撥ねる」として抽出し、危険因子「交通事故」を分類名として仮定義することで

意味共起による抽出知識を構築した。次に、意味共起「車 + 撥ねる」を検出する照合ルールは、字面のままルール化すると「バス + 撥ねる」は検出できないので、多属性照合技術を利用して「[乗り物] + 撥ねる」のように要素をまとめながら作成した。最後に分類名 < 交通事故 > との関連付けにより知識構造を構築した。構築した照合ルールと分類の例を表 1 に示す。要素をまとめることにより、少ないルールで多くの表現を検出できるようになり、メンテナンス性にも優れる知識構造となっている。

表 1 照合ルールと分類の例

分類	照合ルール
転倒	[身体] + [滑る]
転倒	[身体] + [転倒]
接触	[乗り物] + [挟まれ巻き込まれ表現]
薬剤	[薬剤関連] + [気づき]
療養上の世話	[崩れ] + [医療用具]

(2) 各分野の既知のヒヤリハット事例を収集し、危険因子の横断的な、どの分野にも適用できる体系の構築

ヒヤリハット事例のデータ収集は Web 上から行ったが、過去の研究では、教育や学習の一環としたヒヤリハット事例の研究もあり、外部で公開されているヒヤリハット事例を取り上げている場合もあったため、まずはそれらの公開事例から始め、その後、収集量を増やすために SNS なども利用した。決定した危険因子(分類)と事例数(一部)を表 2 に示す。

名称としては特定の分野を表すものも存在するが、これらを分野に分けることはしておらず、同等に扱う形で体系を構築している。これにより、例えば < 転倒 > などは、どの分野にも起こりうる事例であるし、< 薬剤 > であれば、医療現場だけでなく工場などの現場でも適用できるようになっている。つまり、一つの体系で分野横断的に適用できることは、各業種におけるヒヤリハット潜在因子の補完が可能となり、早期対策への貢献が期待できる。

表 2 危険因子と事例数の一部

危険因子(分類)	事例数
転倒	72
接触	77
薬剤	266
療養上の世話	118

(3) コーパスを用いて、リスク感性表現を抽出し、危険因子体系から < 転落 > などの危険因子を判定する手法の開発と、実験による性能評価

ヒヤリハット潜在因子から危険因子の判定手法の開発は、共起情報格納型トライ構造の研究成果を利用して実装した。先に述べた多属性照合技術を利用した照合ルールによって、潜在因子となる意味共起を検出すると同時に関連付いた危険因子である分類名を取得することで判定する。

この判定手法の評価には、コーパスに対して提案手法を適用し、潜在因子の検出精度と危険因子の判定精度を評価した。図 1 に精度を示す。判定精度が高いことから、潜在因子と危険因子の関連付けに大きな間違いはないと考えられるが、精度の不足部分は、照合ルール作成にて要素のまとめすぎによって別の分類名となっているもので、重複部分を見直すことで改善できる。

また、データ 3 において検出精度が低くなっている。これは検出漏れによるが、主な原因は潜在因子となる意味共起に固有名詞が含まれるものであった。固有名詞は量も多く知識構築も人手に頼るところが多くなるが、辞書への拡張で補え、精度も改善できる。

最後に誤検出の調査のためにヒヤリハットを含まないコーパスにて評価を行ったが、1 件も検出しなかったため、現状で誤検出されることはないと考えられる。以上より、判定手法の有効性は示せたと言える。

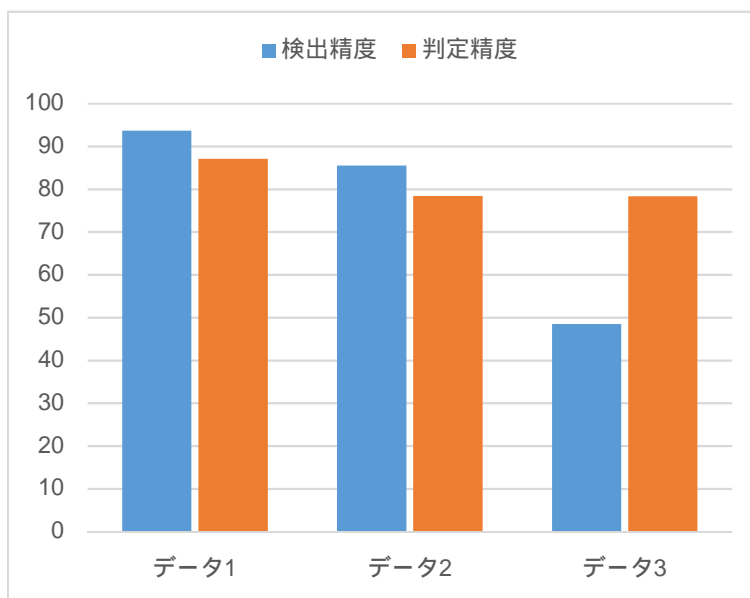


図 1 検出・判定精度の結果

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Kazuhiro Morita, Kenzaburo Yashiro, Masao Fuketa
2. 発表標題 A classification method of the incident by extraction from text
3. 学会等名 5th International Conference on Recent Trends in Computer Science and Electronics (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----