

表現の特徴に着目したインシデントレポートからの知識抽出

Knowledge Extraction from Incident Reports Utilizing Sentence Features

寺井 政文[†]
Masafumi Terai原田 史子[‡]
Fumiko Harada島川 博光[‡]
Hiromitsu Shimakawa

1. はじめに

医療現場でのリスク管理として、重大事故に至る可能性はあるが、実際には事故には至らなかった潜在的事例であるインシデントレポートが収集されている。収集したインシデントレポートを分析することで、医療事故対策や改善に活用できる。日本医療機能評価機構が230の医療機関を対象に実施した調査 [1] では、3ヶ月間に62,000件ものインシデントレポートが収集された。現在の病院のインシデントは、電子データとして収集され計算機によって、より効果的な分析・対策が期待できる。

インシデントレポートの設問には、選択式項目であるメタデータ部と、インシデントの概要を自由に記述できるテキスト部がある。細かい状況までをインシデントレポートの選択式項目のみで表現できるわけではなく、テキスト部がなければ実際に発生したインシデントを的確に報告することはできない。一方、計算機でのテキスト部の記述は労力がかかり、書き手によって同様の内容であっても異なる表現が使われる。効果的なインシデント分析をするためには、十分な量のインシデントと、統一的に記述されたテキスト部が必要である。そのためには、インシデントレポートを簡潔かつ統一した表現で入力するための仕組みが求められている。

本論文では、項目の選択式によってテキスト部を記述する仕組みと、過去のインシデントデータから各項目を予測する手法を提案する。本手法では、テキスト部記述にかかる書き手の労力を減らし、表現が統一されることにより、分析に効果的なインシデントを得ることができる。

2. インシデントレポート構成要素

図1にインシデントレポートを入力するための設問を示す。図にあるように、設問は多岐に渡り、選択式入力のメタデータ部と、発生事例の概要が自由に記述されるテキスト部を持つ。メタデータ部には、「発生日」、「発生時間」などの発生時の環境状態に関する設問、「性別」、「経験年数」などの当事者の状態に関する設問、「年齢」、「疾患名」などの患者の状態についての設問がある。

テキスト部は、発生したインシデントに対して自由に記述することができ、発生したインシデントの概要を知る重要な設問である。しかし、計算機上で文章を入力する必要があるため、書き手の労力は大きく、同じ内容でも書き手によって異なる表現が使われる。これは、看護師が多忙であり、インシデント報告に当てる十分な時間がないことが原因と考えられる。

書き手には反省文というイメージがあるため、テキスト部は回りくどい表現や冗長な表現を用いて記述されることがある。例えば、「容器を落として割れてしまった」

メタデータ部	カテゴリ	転倒・転落
	重要度	エラーは生じたが、患者には未施行
	発生日	2010年5月12日(水)
	発生時間	15:20
	発生場所	病室
	当事者	患者
テキスト部	自由記述	39度の熱がでてる患者で左手に腫脹があり手を使いにくい状態の患者がトイレにいこうとして転倒された。熱があることや眼剤を飲んでることから移動の際はナースコールを押すよう前もって説明しておくべきであった。

図1: インシデントレポートの例

という文書と「容器を落としてしまった」という文書は、内容は同じであるが表現が異なる。さらに、「いつもはしていた」、「細心の注意を払っていた」など冗長な部分が多くなる傾向がある。これでは、計算機だけでなく、人が見て分析することも困難である。そのため、統一的にテキスト部を記述する仕組みが必要となる。

3. 自由記述表現の特徴に着目した知識抽出

3.1 項目の選択式によるテキスト部の記述

書き手の違いにより表現方法は異なるが、報告内容は同じであることが多い。計算機での文章入力に慣れていない看護師にとって、テキスト部の記述は大きな労力がかかる。そこで、図2のように、過去のインシデントデータから予測された項目を順次選択することによってテキスト部を入力する手法を提案する。本手法により、入力にかかる労力が減少でき、かつ記述内容が統一されることで、分析に十分なインシデントの量と質を得ることが期待できる。

本研究では、(1) から (3) を繰り返し、ユーザはテキストを記述するものとする。

- (1) フレームの選択
- (2) 主要素の選択
- (3) フック要素の選択

(1) テキスト部を構成する骨組みに当たる部分である。何が何をどうしたといった、発生事例の文型を選択する箇所であり、この文型を本研究ではフレームと定義する。

(2) 選択したフレームに対応して、(1) で選択した文型に当てはめる、主語・目的語・述語に対応する要素を選択する。この要素を本研究では主要素と定義する。

(3) (2) で選択された単語に紐付けられている、補語や修飾語に当たる要素であり、本研究ではこの要素をフック要素と定義する。

[†]立命館大学大学院理工学研究科
[‡]立命館大学情報理工学部

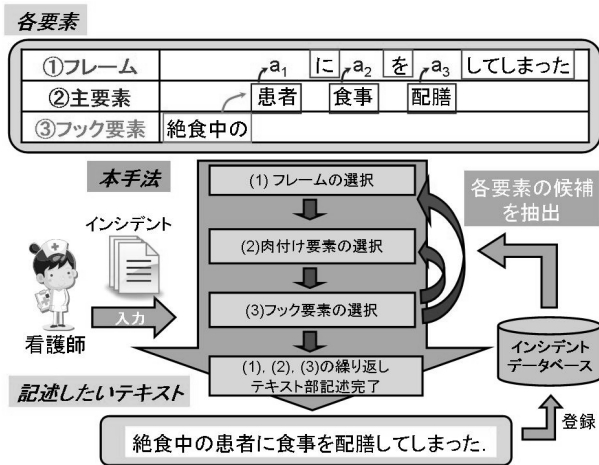


図 2: 項目の選択式によるテキスト部の記述

主要素とフック要素は、過去のインシデントのテキスト部のデータをマイニングすることで予測できる。

3.2 フレーム抽出

発生するインシデントには似通ったものが多く、報告されるインシデントのテキスト部には、特徴ある表現が用いられるので、いくつかに分類することができる。各フレームに対応する主要素 a_n とする。例えば、書き手がインシデントの発端となった場合は、「 a_1 に a_2 を a_3 してしまった」や「 a_1 を a_2 し忘れた」などといった文型で頻繁に記述される。一方、書き手がインシデントの発生には関わっていないが、インシデントを発見する場合もある。そのような場合には、「 a_1 が a_2 になっていた」や「 a_1 が a_2 してあった」などといった文型で記述される。

書き手がインシデントに遭遇するかたちによって、いくらかの文型のカテゴリがあり、そのカテゴリには特徴ある表現が用いられる。文型に沿って分類することで、記述したい内容に関連する各要素を予測できる。そこで、この文型に基づいてテキスト部の文型を組み立てる骨組みであるフレームを抽出し、それらをユーザに提示する。

3.3 主要素抽出

文型の種類によって分類された、各フレームの主要素 a_n を抽出する必要がある。例えば、「 a_1 が a_2 してあった」というフレームの場合、先頭の「 a_1 」に対応する主要素は、「カルテ」、「注射」などが挙げられ、2番目の「 a_2 」に対応する主要素には、「二重登録」、「間違えて用意」などが挙げられる。フレームごとに、記述される主要素は限定されることが多いため、予測し提示することができる。

過去のインシデントデータから、各フレームにおいて各要素の出現データによって主要素を予測する。具体的には、分類されたインシデントに文節区切りを適応し、区切られた文節を形態素解析することで、主要素を抽出する。形態素解析した文節の中でも、ある形態素同士が高い共起を示す場合は一つの要素としてみなす。抽出した主要素は、各フレームでの出現頻度、フレームを構成する要素との距離からランク付けする。算出されたランクの高いものから順にユーザに提示される。また、提示

された要素の中に記述したい内容に沿った要素がない場合は、そのつどユーザが入力することとする。

3.4 フック要素抽出

フレームと主要素を選択するだけでは、テキスト部を十分に記述することはできない。主要素を修飾する要素や、発生時間や発生要因などに相当する要素であるフック要素を記述する必要がある。これらの要素の多くは、主語・目的語・述語を修飾していることが多い。そのため、それらを構成する要素である主要素との関連性から抽出することができると考えられる。そこで、主要素との修飾関係に着目することでフック要素を抽出する。まず、過去の全インシデントの形態素の中から、主要素と一致するものを検出する。次に、検出した形態素と修飾関係にある形態素をフック要素として抽出する。さらに、抽出されたフック要素は、主要素との距離、修飾の有無から関連性が算出される。このように、ある主要素を選択したさい、その主要素に対して関連性の高いフック要素が順次にユーザに提示される。また、提示された要素の中に記述したい内容に沿った要素がない場合は、そのつどユーザが入力することとする。

4. 関連研究

これまでも、インシデントレポートを対象として、メタデータ部とテキスト部の語句の情報から知識を抽出する手法 [2] や、テキスト部のシーケンス構造に着目し、知識を抽出する手法 [3] が提案されている。これらの手法では、語句間の位置関係や出現関係から記述されたインシデントの分析を試みている。そのため、人手による、インシデントレポート分析を支援できる。しかし、記述されたインシデントの分析を目的としているため、書き手にとってのテキスト部記述の労力削減や、内容の統一化はできない。本手法では、項目の選択化によってテキスト部を記述するため、書き手の労力を削減し、内容の統一化を図ることができる。

5. おわりに

本論文では、効果的な分析のためのインシデントレポート入力支援を目的とした、項目の選択式によってテキスト部を記述する仕組みを提案した。

今後は、本論文で述べた内容を実現するテキスト部記述システムを構築し、有効性の検証を行う予定である。

参考文献

- [1] 財団法人日本医療機能評価機構, “医療事故情報収集等事業第20回報告書,” http://www.medsafe.jp/pdf/report_20, 2010
- [2] 岡部貴博, 吉川大弘, 古橋武, “メタデータと語句の共起情報を利用したインシデントレポート解析システムの提案,” 知能と情報 (日本知能情報ファジィ学会誌) vol.18, No.5, pp.689-700, Oct.2006
- [3] 河崎佑一, 宮本定明, “ファジィ近傍に基づく内積空間とカーネル法によるテキストデータ解析,” 知能と情報 (日本知能情報ファジィ学会誌) vol.21, No.4, pp.461-469, Feb.2009