

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 7 日現在

機関番号：24506

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25463332

研究課題名(和文) 看工連携による看護ケアテキスト評価・分析支援システムの開発

研究課題名(英文) Development of Computer Aided Nursing-care Text Classification System

研究代表者

新居 学(Nii, Manabu)

兵庫県立大学・工学(系)研究科(研究院)・助教

研究者番号：80336833

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：看護ケアの質向上のため、看護質評価基準に沿った質問に対して回答された看護過程の自由記述文章(看護ケアテキスト)を、専門家の評価と同様に自動評価することを目指して研究開発を行ってきた。看護ケアテキスト数値化プログラムを開発し、日本語自由記述文章を数値で表現された特徴ベクトルへ変換し、機械学習アルゴリズムにより自動分類を行った。単語の係り受けなどの文構造を考慮した特徴ベクトル化により分類精度向上を図った。決定木構造を導入したサポートベクトルマシンやWord2Vecと呼ばれる単語の関係性を多次元空間へ配置する手法を導入して、分類性能の向上や分類過程の可視化を行った。

研究成果の概要(英文)：For improving nursing-care quality, a nursing-care text evaluation system has been developed. The nursing-care text evaluation system consists of a feature vector extraction system and machine learning based classification. The nursing-care texts are freestyle Japanese texts which are described by nurses. Nurses report their own nursing-care processes which were actually done by the nurses. The feature vector extraction system analyzes dependency relations between words and represents such relations by a matrix type expression. Support vector machines with binary decision tree are used for visualization of the evaluation process of the nursing-care texts. The word2vec is a tool for representing words by high-dimensional vectors. Using word2vec, the classification performance of the nursing-care texts was improved.

研究分野：計算知能工学

キーワード：看護ケア テキスト分類 コンピュータ支援システム ソフトコンピューティング 機械学習

1. 研究開始当初の背景

看護ケアの質向上を目的として Web アプリケーションにより収集された看護ケアテキスト（看護過程を記載した自由記述文章）を自動評価・分析支援するシステムの構築を行っており、自動分類システムの性能向上や評価過程の可視化を目指して研究を開始した。

2. 研究の目的

看護ケアテキスト分類システムの性能向上を図るためには、テキストデータから適切な特徴量を抽出し数値ベクトル化する必要がある。本研究では、文の構造を係り受け関係として捉え、係り受け関係を反映した特徴ベクトル生成を目指す。これまでの研究では係り受け関係を一次元の特徴ベクトルに反映させるために一部の情報が欠落するという問題を持っていた。本研究ではこれを行列形式にすることで解決する。

また、分類性能の向上と分類過程の可視化のために Support Vector Machine using Binary Decision Tree を利用する。この手法はこれまでの研究で用いてきた SVM と決定木を融合した方法で決定木構造を解析することでデータの分類過程を可視化することができる。

3. 研究の方法

(1) 看護ケアテキストは看護の質評価基準に基づき設定された質問に対して、看護師が実際に行った、もしくは類似の事例に関する経験について回答したテキストである。看護ケア専門家はこの回答を精査することにより4段階に評価を行う。

(2) 自動分類システムは上記(1)の看護ケア専門家の評価と同様の評価を行うことを目指す。自動分類システムの構築には、上記(1)の専門家による評価結果を利用する。すなわち、機械学習アルゴリズムを教師あり学習の枠組みで構築する。ただし、実際に使用するシステムでは、継続的に教師データを利用できるわけではない。

(3) Support Vector Machine(SVM)はカーネル関数を用いてデータを高次元空間へ写像し、データを2クラスに分類する手法である。サポートベクターと呼ばれるデータを選択して利用することにより効果的にクラス分類を行うことができる。本研究では看護ケアテキストの自動分類に SVM に基づくシステムを使用する。Support Vector Machine using Binary Decision Tree(SVM-BDT)は上記 SVM を決定木構造に配置しデータを分類する手法である。データをよく分類できるクラスとそれ以外のクラスに2分割することを繰り返し二分決定木を構築する。各ノードでは SVM が配置されており学習が行われる。

(4) 係り受けによる特徴ベクトルの生成の

ためには、まず、日本語の文章を形態素解析および係り受け解析を行う。

4. 研究成果

(1) 係り受け関係による特徴ベクトルはこれまで、単語の係り元から係り先への関係を直接利用して順序関係として一次元に並べ特徴ベクトルとしていた。看護ケアテキスト p の特徴ベクトル x_p は以下の式のように表現する。

$$x_p = e_p + r_p,$$

$$e_p = (e_1, e_2, \dots, e_{N_{TL}}),$$

$$r_p = (r_1, r_2, \dots, r_{N_{TL}}),$$

$$e_k = \begin{cases} 1, & e_k \text{ に対応する単語が存在する場合} \\ 0, & \text{それ以外} \end{cases},$$

$$r_k = \begin{cases} d/N_{TL}, & \text{係り先単語が存在する場合} \\ 0, & \text{それ以外} \end{cases},$$

ここで、 N_{TL} はテキストに含まれている全単語数である。この表現法では特徴ベクトル x_p は単語の存在を示す特徴ベクトル e_p と係り受けを表現する特徴ベクトル r_p の和で表現されることになる。そのため、看護ケアテキストに出現する単語それぞれに対して一種類のみの係り受け関係が表現されることになる。一部の語は複数の係り受け関係を持つことが有り、これを表現できていなかった。そこで、行列表形式で係り受け関係を表現することとした。行列表形式では、単語数の二乗の要素数を持つ行列を用いて係り受け関係を表現する。単語 t_i が単語 t_j に係る関係を持つとする。この時、特徴行列 X_p は以下のように表現できる。

$$X_p = (x_{ij})_{i,j=1}^{N_{TL}},$$

$$X_p = \begin{cases} 1, & t_i \rightarrow t_j \text{ が存在する, } i \neq j \\ 1, & t_i \text{ が文の末尾である, } i = j, \\ 0, & t_i \rightarrow t_j \text{ が存在しない} \end{cases}$$

このように係り受け関係の存在を行列表形式で表現することにより情報の欠落がなく特徴ベクトルとして表現できる。機械学習システムに入力する際には、すべての行を一行に並べることにより一次元のベクトルとする。

(2) 看護ケアテキストのみに含まれる単語だけを利用する方法では単語の情報不足を避けることができない。そこで蓄積されている看護ケアテキスト以外の文章集合から単語情報を抽出し特徴ベクトル生成に利用する方法を考案した。大量の一般のテキストデータから抽出した単語情報を用いることで学習用の

看護ケアテキストに含まれない単語であっても特徴として利用することができる。また、看護ケアテキストには看護分野の専門用語以外の一般の単語が頻繁に出現する。これは看護師が患者に関する情報を収集した結果であり、より良い看護のためには患者の理解は不可欠であるため、広く一般の単語を考慮することが必要である。そこで、本研究では単語を任意の次元の空間に配置しそれぞれの語の関係性を表現できる word2vec と呼ばれる手法を利用した。この手法は大量のテキストデータを入力として学習を行い、各単語を任意の次元の数値ベクトルとして表現できる。

(3) 2007 年および 2008 年に収集された看護ケアテキストデータを学習用とし、2009 年に収集された看護ケアテキストデータを評価用データとして分類を行った。看護ケアテキストデータは毎年収集され、専門家により評価・リコメンデーションの返却が行われている。本研究に利用できるよう整備されたデータは上述のものであり、これらよりも新しいデータは現在整備中である。これらの 3 年分のテキスト集合は 12 種類の質問項目ごとに分けられた回答テキスト集合である。このデ

表 1 SVM による分類結果

質問項目	全テキスト数	係り受け	係り受け (行列)
P123	827	502	521
P131	769	495	447
P132	638	413	427
P212	718	365	344
P213	735	394	413
P221	768	438	427
P222	703	563	570
P322	604	447	439
P411	566	458	471
P423	636	548	544
P425	690	442	473
P431	659	548	544
合計	8313	5613	5620

表 2 SVM による分類結果

質問項目	全テキスト数	単語クラスタ	word2vec
P123	827	543	512
P131	769	515	530
P132	638	409	399
P212	718	366	361
P213	735	375	392
P221	768	421	431
P222	703	564	565
P322	604	430	422
P411	566	475	481
P423	636	581	582
P425	690	460	487
P431	659	545	533
合計	8313	5684	5695

ータを用いて分類評価を行って提案した特徴ベクトルによる分類性能の変化を調査した。

表 1 および 2 は SVM を用いて分類を行った結果である。これらの表から word2vec を用いた特徴ベクトルを利用した場合には、看護ケアテキスト集合全体としては正しく分類できるテキスト数が最大となっていることがわかる。個別の質問項目ごとに見ていくと必ずしもいずれか一つの特徴ベクトル表現が良いというわけではないこともわかる。それぞれの質問項目では看護ケアの質評価のポイントが異なることが原因であると考えられる。たとえば、痛みを持つ患者への対応を評価するためには、痛みの状況について情報収集しなければならない。そして得られた痛みの様子を考慮して適切な看護をする必要がある。看護師は痛みの情報収集、痛みの状況、具体的対処を述べなくてはならない。これらの内容が記述されていれば良い評価となる。このようなテキストを評価するためには、キーワードが含まれていれば良いというわけではなく、単語間の関係性が必要となる。これは単に係り受けのような文構造のみならず、単語間の意味的な関係性も必要である。この実験結果が示すように本研究で考案した特徴ベクトル表現は、分類性能向上に貢献することもある一方で全ての質問項目に対して一様に性能向上を果たせるわけではなかった。質問項目ごとの評価のポイントを考慮して、特徴ベクトル表現を使い分けることができれば、看護ケアテキスト集合全体を良い性能で分類することができる。このことは、本研究を実施中に明らかになった点であり、今後はこのことを念頭に置いた自動分類システム構築を行っていく予定である。

表 3 SVM-BDT による分類結果

質問項目	係り受け	係り受け (行列)	単語クラスタ
P123	509	550	524
P131	508	470	504
P132	408	429	406
P212	356	345	352
P213	418	401	407
P221	415	415	403
P222	565	575	558
P322	435	440	434
P411	458	471	475
P423	548	555	578
P425	474	486	382
P431	548	544	545
合計	5642	5681	5568

表 3 は SVM-BDT を用いて分類した結果である。この結果からわかるように SVM を用いた場合と傾向が異なる。単語クラスタ特徴ベクトルとの相性が悪いという結果であった。このことは一部の看護ケアテキストデータのクラス分布が著しく偏っていること等が原因であると考えられる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

- ① Manabu Nii, Kazunobu Takahama, Shota Miyake, Atsuko Uchinuno, Reiko Sakashita, Rule Representation for Nursing-care Process Evaluation Using Decision Tree Techniques Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), 査読有り, Vol.18 No.6, pp.918-925, 2014
- ② Manabu Nii, Yoshinori Hirohata, Atsuko Uchinuno, Reiko Sakashita, Dependency relation based feature vectors for classifying nursing-care texts, International Journal of Intelligent Computing in Medical Sciences & Image Processing, 査読有り, Volume 5, Issue 1, pp.57-66, 2013, doi:10.1080/1931308X.2013.799319

[学会発表] (計 5 件)

- ① Manabu Nii, A Matrix-based Feature Vector Definition and a SVM-BDT-based Classification System for Classifying Nursing-care Texts, IEEE International Conference on Fuzzy Systems, PID15271, 2-5 Aug. 2015, Istanbul(Turkey)
- ② Manabu Nii, Soft Class Decision for Nursing-care text classification using a K-Nearest Neighbor based System, 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp.1825-1830, 6-11 Jul. 2014, Beijing(China)
- ③ Manabu Nii, A Directed Graph based Feature Definition for Classifying Nursing-care Texts, 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp.3712-3716, 5-8 Oct. 2014, San Diego, CA(USA)
- ④ Manabu Nii, Consideration about Utilizing Text Architecture for Making Feature Vectors in Classifying Nursing-Care Texts, 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1817-1821, 13-16 Oct. 2013, Manchester(UK)
- ⑤ Manabu Nii, Classification Rule

Representation for Nursing-care Process Evaluation, The 14th International Symposium on Advanced Intelligent Systems, F1c-1(10pages, CD-ROM), 13-16 Nov. 2013, Daejeon(Korea)

6. 研究組織

(1) 研究代表者

新居 学 (NII, Manabu)

兵庫県立大学・大学院工学研究科・助教

研究者番号：80336833