

## 看護記録のテキストマイニング

村松 洋<sup>†1</sup> 渡部 勇<sup>†1</sup>  
大崎 千恵子<sup>†2</sup> 小塚 和人<sup>†3</sup>

看護現場へフィードバックすることが有用な情報の抽出を主な目的として、看護の経過記録の分析用のテキストマイニングシステム Nursing Text Miner を開発した。本稿では、看護記録をマイニングする際の、一般分野のマイニングとは異なる要件と、そのための実装を報告する。看護記録の分析では、記述の内容だけでなく、どのように表現されているかを分析する必要がある。このため、汎用のマイニングツールでは通常、形態素解析結果を集計するが、本システムでは、形態素解析と正規表現によるパターンマッチングを併用した。実際の看護記録を分析した結果、時間軸での集計の必要性が高いこと等、看護記録の分析のうえで重要な集計方法を確認した。看護記録に特化したツールの開発により、ほかには事例のない、数年間に及ぶ看護記録の多様な分析が可能となった。分析結果の一部は、実際に現場にフィードバックを行い、記録の質の改善に有用であることが確認できた。

### Text Mining of Nursing Records

HIROSHI MURAMATSU,<sup>†1</sup> ISAMU WATANABE,<sup>†1</sup>  
CHIEKO OSAKI<sup>†2</sup> and KAZUTO KOZUKA<sup>†3</sup>

To extract information useful for feedback in nursing practice, a text mining system to analyze the nursing records, the Nursing Text Miner, was developed. In this paper, requirements for nursing record mining, that are different from those for mining in the general field and implementation of such requirements are reported. In the mining of nursing records, it is necessary to analyze not only the contents of descriptions, but also how they are expressed. Therefore, morphological analysis and pattern matching using regular expressions are used together in this system, whereas morphological analysis results are usually used in general-purpose mining tools. As a result of analyzing actual records, importance of some methods such as time series analysis were confirmed. By developing this system it has become possible to conduct various analyses of nursing records covering a few years. Portions of the analysis results were actually fed back to nursing practice and were confirmed to be useful for the improvement of the quality of records.

### 1. はじめに

テキスト・データの統計的解析は、多量のテキストがコンピュータに蓄積され始めた 1990 年代以降に活発となり、テキストマイニングと呼ばれて、様々な分野での適用が試みられている。分析対象のテキストは、Web 上のテキスト、新聞記事、特許情報、アンケート等、多岐にわたる。これらの分析対象のうち、特定の商品に対する Web 上の意見の分析（評判分析）や、特許分析では、その分析に特化したパッケージ・ソフトウェアやサービスが商品化されており、マイニング結果が定常業務の中に活用されている例は多い<sup>1)</sup>。しかし、テキストマイニング・ツールの使用を試みたものの、目に見える成果を得られない例も多く、「期待されているほどには使いこなされていない」との指摘もある。那須川<sup>2),3)</sup> は、テキストマイニングの適用を成功させる条件として、「分析の観点の定義が必要」であり、「テキストから抽出したデータを既成のデータマイニングツールで分析しても有用な結果につながらないことが多い」、「数百件の人手でも分析可能なテキストだと成果が出難い」と指摘している。少なからぬ工数が必要となることがある辞書整備のしかたも重要である。また、保田<sup>4)</sup> は、テキストマイニング・ツールは、分析目的に応じたカスタマイズが重要と指摘している。

これらの指摘は、テキストマイニングを活用するうえでは、今後も応用分野に即した利用方法の知識の蓄積と、それを反映した技術開発が必要なことを示唆している。上記の観点から本稿では、看護記録に着目した場合の、汎用テキストマイニング・ツールとはやや異なる機能や実装の事例を報告する。報告するプログラム Nursing Text Miner は、看護職の書く記録を統計解析し、現場にフィードバックする情報を抽出することを主な狙いとして開発したものである。

### 2. 看護記録とは

#### 2.1 看護記録の構成

看護記録とは、看護実践の一連の過程を記録したものである。記録の対象は、外来患者や

<sup>†1</sup> 富士通研究所ソフトウェア&ソリューション研究所  
Software and Solution Laboratories, Fujitsu Laboratories Ltd.

<sup>†2</sup> 昭和大学横浜市北部病院看護部  
Department of Nursing, Showa University Northern Yokohama Hospital

<sup>†3</sup> 昭和大学横浜市北部病院医療情報部  
Department of Medical Informatics, Showa University Northern Yokohama Hospital

訪問看護の場合もあるが、量的には入院患者が中心である。内容を大別すれば、A) 患者の基礎情報、B) 看護計画、C) 経過記録、および D) 看護サマリが基本的な要素である<sup>5)</sup>。インシデント・レポート(ヒヤリ・ハット報告)、高齢患者の転倒リスク評価シート等、個別の目的に応じて固有の書式を設定する記録も多種存在する。「電子カルテ」と称するシステムでは、通常、看護記録も電子的な記録となっている<sup>6)</sup>。

このうち、C) 経過記録は、患者の問題の経過や、治療・処置・ケア・看護実践を記載し、C1) テキストを中心とした叙史的な記録と、C2) 経過一覧表(フローシート)とが併用される。C2) 経過一覧表は、定常的な看護行為の実施状況や観察結果を記号や数字等で簡潔に記載するものであり、テキストマイニングの主な対象は、C1) 叙史的な記録となる。

## 2.2 記録の量

経過記録の特徴の1つは、その量の多さである。後述する昭和大学横浜市北部病院では、平均140字の記録が年間50万件作成されている。国内の病院全体では、病院勤務の看護職(看護師、准看護師、助産師、保健師)は80万人に及ぶ。このような多数の人がほぼ同じ書式で記録するテキストの種類は、あまり多くはない。

すなわち、テキストマイニング等を利用する効果的な分析手法が開発された場合、潜在的な適用対象はきわめて多いといえる。

## 2.3 記録の質への要請

看護記録のもう1つの特徴は、記録の質の向上が求められることである。適切な医療活動のうえでは、医師・看護職・その他の医療従事者の間で、正確・適切な情報が共有されることが必要なだけでなく、実施した看護が適切であったかを看護記録のうえで検証できることが求められる。訴訟等の医療上のトラブルの際は、看護記録の記載内容が論点となる場合もあり、リスクをとまなう処置への事前説明と合意の記録の有無も問題となる。また2000年より、患者・家族から記録の開示を請求された場合、開示することが原則となっており、開示した場合に患者・家族の信頼を損なう可能性のある記述は不適切とされる。

医療機関の第三者評価を行う日本医療機能評価機構の評価項目では、記録について「質的点検(内容監査)が行われていること」と「評価結果がフィードバックされていること」を求めている。同評価機構の認定により患者の信頼を高めようとする病院にとっては、記録の質向上への取り組みが必要条件となっている。

このため、看護師向けの書籍や雑誌では、記録の書き方や監査の方法について多数の記事が書かれており、看護記録の書き方を主要なテーマとする専門誌も刊行されている。

記録の質への要求は、テキストマイニングの観点からは、ブログの評判分析や、コールセ

ンタのログの分析とは、異なった側面が看護記録の分析にはあることを意味する。ブログ分析等の多くのテキストマイニングの例では、記述されている対象が分析の主要な関心事項であり、記述のしかたを分析することは稀であるからである。これに対して、看護記録の分析では、記述の対象(患者の状態や看護行為)だけでなく、表現の正確性や適切性も重要な分析対象となる。

## 3. 電子化された看護記録の統計的分析の先行例

### 3.1 基礎的な分析

看護記録は、統計的な分析から有用な知見が得られる可能性を秘めた分野であり、従来から、紙をベースとした小規模な分析は、看護研究として広く行われている。しかし、電子化された記録を利用して、多量のデータを分析した事例は多くはない。

看護情報が電子化され始めた後、コード化されたデータの分析が1990年代より行われている。代表的な分析対象は、病院ごとに設定する各種のマスタファイルの各項目の利用状況である。マスタとしては、看護診断の名称とシステム内で使用するコードとの対応表が通常使用される。病院により、標準看護計画(主要な疾患別に用意された看護診断のリスト)、主要な看護行為の名称等もマスタに定義され、メニュー選択による入力に利用される。マスタの妥当性の検証を目的として、マスタの利用状況を把握することが、まず取り組まれた。また、看護の質の向上のうえで、利用状況の把握は有用と考えられた<sup>7)</sup>。

分析は、当初は1年程度の期間の平均値を確認するものであったが、データの蓄積にともない、数年間の時系列変化を調査し、着目すべき変化が検出されるようになった<sup>8)</sup>。

### 3.2 やや量の少ない記録のテキストマイニング

テキストマイニングの試みとしては、以下の事例が報告されている。アセスメント記録(患者の状況を全体的に評価し、看護計画立案の前提情報を整理する記録)の分析例では、単語の共起関係をデンドログラムで把握したり<sup>9)</sup>、使用される名詞を看護領域別に分類し、看護領域別の記載頻度を検討したりすることが試みられた<sup>10)</sup>。インシデント・レポートの分析では、係り受け解析を行って、頻度の高いインシデントの傾向を把握したり<sup>11)</sup>、さらに要約文を作ってマップ化することにより、インシデントの傾向を把握したりする試み<sup>12)</sup>や、類似事例の検索システム<sup>13)</sup>が試みられている。特定の疾患に対する看護の分析では、網膜剥離の患者について、看護計画内のテキストをキーワードで分類し、標準的な看護計画と、実際に立案された看護計画の相違を把握することが試みられている<sup>14)</sup>。

これらのテキストマイニング事例を概観すると、単語の共起関係や頻度の分析では、ある

程度の量のテキストが分析されているが、ただちに現場で有用な知識を得られたとはいえない。係り受け解析を行った事例は、いずれも、手作業でも分析可能な量のテキストでの実践にとどまっており、テキストマイニング手法の試行の段階といえる。分析には、主に茶釜や市販の汎用のテキストマイニング・ツールが使用されている。

### 3.3 経過記録のテキストマイニング

看護記録の中で量的な中心であり、質向上のうえでも重要な経過記録をテキストマイニングの手法で分析した例は、筆者らの報告<sup>15)-19)</sup>のほか、病院固有に開発した経過記録（フローシート内に一定のテキストを記録する方式）の機能の妥当性の検証<sup>20)</sup>、看護職の経験年数や特定の疾患について単語の共起関係を可視化する試み<sup>21);22)</sup>が行われている。海外では、経過記録の分析から医学的な知見を発掘するために、経過記録内の記述を、呼吸・血液循環・疼痛に関連した記述への分類<sup>23)</sup>、省略や略語の多いことを反映した構文解析ツールの開発<sup>24)</sup>、特定の患者の記録中の単語を分類し入院期間内の時系列変化を可視化することが試みられているが<sup>25)</sup>、現場へのフィードバックに役立っている事例は、調査した範囲では存在しない。

### 3.4 非臨床系テキストのマイニング

看護関係のテキストマイニングの例としては、このほかに非臨床系のテキスト、すなわち、患者や看護職等に対するアンケート等の分析が行われており<sup>26);27)</sup>、看護師国家試験問題を利用した形態素解析の精度向上や<sup>28)</sup>、看護関係の文献検索への応用<sup>29)</sup>が試みられている。従来報告されている非臨床系のテキスト分析例では、看護分野用の辞書整備を除くと、一般のテキストマイニングと異なる技術的要件の報告は見られない。

### 3.5 本報告のマイニングの位置

本稿で報告するのは、看護記録の中核的な部分である経過記録を分析対象とし、記録を作成した看護職へのフィードバックに有用な情報の抽出を目指した分析である。

他の事例の多くは、既存のツールの適用を主に行っているが、Nursing Text Miner では、集計機能等を新規に開発し、看護記録の分析に即した多様な集計や多量の記録の集計を可能としている。本報告の研究は、長期間の経過記録を網羅的に分析した、現時点では世界的にも唯一の事例である。

## 4. 昭和大学横浜市北部病院の看護記録

### 4.1 病院の概要

分析対象とする看護記録は昭和大学横浜市北部病院（以下、北部病院と略記）の記録であ

(F欄)			
2010/1/19 10:30	#02 慢性疼痛	D	「また、足が痛いです。伸ばしてもらえますか？」 下肢を自力では動かせず、疼痛がある。 体位変換。痛みのない体位を患者に聞きながら探す。
		A	
		R	「ありがとうございます。楽になりました。」
2010/1/19 15:00	.....	D	.....
		A	.....
		R	.....

本論文での  
1記録

図1 フォーカスチャーティングの形式  
Fig. 1 Format of focus charting records.

る。北部病院は、急性期医療を中心とした地域中核病院であり、病床数は661（以下、2009年現在）である。経過記録を書く看護職は、約500名である。19の診療科と、17の病棟により運営されている。病棟は、施設内の物理的な区画でもあるが、看護職のチームの運営単位である。入院患者は、特定の診療科と特定の病棟が担当する。特定の病棟に着目すると、多くは1~3の診療科の患者であるが、それ以外の患者も若干存在する。

北部病院は2001年4月に開設され、開設と同時に電子カルテ（HOPE/EGMAIN-EX、富士通製）の運用を開始した。電子的なカルテを保存して原本とする運用は1999年に認められ、以後電子カルテの導入が行われていったが、北部病院は最も早い時期に導入した病院の1つである、このため長期間の看護記録の変化を分析可能である。

### 4.2 経過記録の記録形式

経過記録の形式は、1) 経時記録、2) SOAP、3) フォーカスチャーティング等がある。北部病院の場合、フォーカスチャーティングで記述することを原則としている。フォーカスチャーティングは、D（Data：患者の訴えや客観的な情報）、A（Action：実施した看護行為）、R（Response：Actionに対する患者の反応）の3種に構造化し、DAR全体の標題をフォーカス欄（F欄）に記載する形式である<sup>30)</sup>。フォーカスチャーティング記録の例を図1に示す。

以下では、FDAR（F欄とD、A、Rの行）の1つのまとまりを「記録」として扱い、記録の件数を分析する。

各記録に付与されるコード化データには、患者ID、利用者（看護職）ID、日時、患者を担当する診療科、患者を担当する病棟、関連する看護問題の番号と看護診断コード、評価記録フラグ、中間サマリフラグ等がある。看護問題は、患者ごとに数個設定され、一定期間ごとに解決したか否かの評価が行われる。評価の結果の記録や、長期間入院している患者の状況を要約した中間サマリ記録は、入力時の指定によりフラグが設定される。経過記録の一部

をメニュー選択(テンプレート)で入力する病院もあるが、北部病院の場合、上記のコード以外はすべてフリーテキストとして入力している。

## 5. テキストマイニングの観点から見た経過記録の特徴

### 5.1 正規表現によるパターンマッチングの必要性

経過記録の分析のうえでは、記述の内容だけでなく、表現も分析対象となる。その1つは、患者への記録開示の観点から不適切とされる表現の集計である。不適切表現の例は、看護記録についての36の文献から300種を収集した。不適切表現の一部は、「ボケ」、「頑固」(患者の状態や性格に関する否定的な表現)、「いまいち」、「まずまず」(客観性に乏しく誤解を招きやすい表現)等、単語であるが、「～のように見える」(あいまいな表現)、「転倒していたところを発見」(転倒そのものを目撃したのではなく、推定にすぎないため、状況説明が不適切とされる表現)等、フレーズも分析対象となる。フレーズの集計では、形態素解析で単語を切り出すよりも、正規表現によるパターンマッチングが効率的である。

また、記録の読みやすさの観点から、表現への推奨事項もある。フォーカス欄では、「フォーカス欄のみを縦読みしただけで、患者の状況の概要が把握できる」ことが望まれ、たとえば、入院時には、単に「入院」と記載するのではなく、「独歩で入院」、「検査目的で入院」等、状況が把握しやすい表現が推奨される。「入院」が1単語のみかどうかの分析も、パターンマッチングを使用するのが効率的である。

### 5.2 引用表現の扱い

経過記録では患者・家族の発言をそのまま記録することが推奨される。このため、経過記録内には、患者・家族の発言の引用は多い。北部病院の場合、経過記録の中の57%の記録にカギ括弧による引用があり、記録全体の平均文字数は139字、引用がある場合の引用の平均文字数は41字である(2008年)。引用部分の文字数は記録全体の17%を占める。

記録を分析するうえでは、引用内と引用外を区別する必要がある。たとえば、「痛みはない」といった表現は、患者自身の言葉か、看護職の観察による推定なのかで異なる意味を持つ。「ボケ」という表現は、引用外であれば不適切表現の可能性が高いが、患者自身の発言の一部なら不適切表現とはいえない。

フォーカスチャージングでは、引用は「…」のようにカギ括弧で区切って記載するのが原則とされる。このためNursing Text Minerでは、カギ括弧外とカギ括弧内とに分離して集計処理を行う。ただし、「～とのこと」等カギ括弧なしの引用表現も存在し、例外的ではあるが、カギ括弧がなく、R行全体が引用の場合もある。このようなカギ括弧なしの引用表

現は、引用表現のうち、約5%と推定され、通常は無視できる。

### 5.3 FDARの扱い

フォーカスチャージングの場合、たとえば「F欄の中の単語を別扱いしたい」、あるいは「D行の中の単語とA行の中の単語の関係を知りたい」等、FDARを区別した集計が必要となる。このため、Nursing Text Minerでは、集計はFDAR別の集計を可能としている。

### 5.4 定型文

看護記録では、記録作業の効率化や標準化のために、定型文を用意したり、既存の記録の一部をコピーして再利用したりすることが行われる。分析項目によっては、この影響は無視できないため、確認が必要となる場合もある。この確認のうえでは、KWIC形式の出力が有用である。

## 6. 集計方法

様々な分析試行を行って確認した、経過記録を分析する際の主要な分析の軸は、1) 記録を書く看護職、2) 記録日時、3) 記録対象の患者の3点である。このほか、看護問題の種類等も分析軸となる場合がある。

### 6.1 看護職の属性

記録を書いた看護職を分類するうえで基本となるのは、看護職の所属する病棟である。病棟ごとに記録の特徴が異なることと、結果を現場にフィードバックする際には、病棟へのフィードバックとなるためである。

看護職の所属する病棟別の集計を正確に行うためには、看護職がどの病棟に所属するかの情報が必要となるが、北部病院の電子カルテシステムの経過記録内には、看護職の所属する病棟の情報はない。Nursing Text Minerでは、簡便のため、通常は記録対象の患者を担当する病棟の情報を使用している。ある病棟が担当する患者に着目すると、その記録を書く看護職は、担当する病棟の看護職だけでなく、手術室や検査室等の看護師も含まれている。このため、患者を担当する病棟をもとに集計すると、若干の病棟外の看護職の記録も含まれるが、その影響は通常10%程度であり、無視できることが多い。

看護職を分類するもう1つの属性は、経験年数やスキルレベルである。北部病院では、看護職のスキルレベルは、4段階のラダーとして管理されている。各看護職がどのラダーに属するかは、直近のデータは把握できるものの、過去の時点での情報は、保存していない。このため、過去のデータも含めた分析のためには、代替方法として入職年を使用して分析を

行っている。北部病院の場合、入職年は各看護職の ID コードより、例外はあるものの、判定が可能である。特定の年に入職した看護職は、新人だけでなく、他病院からの転職者が数割含まれる。しかし、8.4 節に示すとおり、入職年で分類することにより経験年数により変化する特徴と変化しない特徴を統計的に識別できる。

また、分析項目によっては、厳密に特定の病棟に所属する看護職のみの集計が必要となる場合もある。病棟に所属する看護職も、直近は把握可能であるが、過去のデータを分析するうえでは、残された記録の状況から推定することとなる。このため、特定の看護職が該当期間にどの病棟が担当する患者の記録を何件書いているかを集計することにより、看護職の所属する病棟を把握することが必要となる。

## 6.2 記録日時

時間軸による集計では、時刻分布や曜日による差を集計する場合もあるが、重要なのは、年度単位の集計と、年月単位の集計である。半期ごと、4 半期ごと、暦年ごとの集計を行う場合もある。

経過記録の内容は、数年間を比較すると、様々な項目で変化している。変化の背景には、インフォームド・コンセントや、医療安全に対する社会的な関心の高まり、記録開示の原則化等があり、院内での看護の質の向上や効率化の努力が反映した変化も多い。病棟別に年度単位の変化を集計すれば、各病棟で変化のしかたに差があることが確認できる。たとえば、特定の不適切表現の減少の状況の病棟差を知ることができ、現場の運営に参考となる数値が得られる。また、直近の記録の集計だけでは、病棟間の差が、病棟による患者プロファイルの差か病棟間の看護業務・記録のスタイルの差か不明であっても、長期の時系列データを見ると、看護業務・記録スタイルの差によるものであることが判明する例も多い。

年月単位の集計は、変化が見られる指標が、病棟単位での意識的な看護業務・記録の変更か否かを確認するのに有用である。意識的な変更があった場合には、特定の月を境に大きな変動が見られる。

## 6.3 患者属性

基本的な患者属性は、担当する診療科である。経過記録中の単語は、疾患や治療方法により異なる。特定の診療科の患者が複数の病棟により担当されている場合、病棟間の差が患者プロファイルの差に起因する差か、看護業務・記録スタイルの差に起因する差かを識別するには、特定の診療科が担当する患者について病棟比較を行うことで判別できる場合がある。

患者の疾患の種類や、年齢、性別、入院日・手術日からの相対日数等を集計の軸とする場合もある。これらのデータは、経過記録内には存在せず、他のファイルのデータと突き合わ

せて集計対象とする患者 ID を設定する。

## 6.4 出力形式

効果的な分析のためには、看護職の属性と年度/年月のマトリクスを、多数の指標（多数の表現の頻度や共起関係等）について、一括して出力することや、特定の条件での高頻度語や共起語を時系列で出力できることが重要となる。テキストに付加されたコード情報でテキストを分類し、分類別のテキストの特徴を出力する機能は、汎用のテキストマイニング・ツールの基本的な機能であるが、複数の条件で絞り込んだ記録を看護職の ID 別に集計する等、やや複雑な条件で分類・集計することが、経過記録の分析の場合、特に必要である。

## 7. Nursing Text Miner の実装

### 7.1 概要

#### 7.1.1 ファイル構成

Nursing Text Miner での主なファイルと処理の流れを図 2 に示す。

原テキストファイルは、電子カルテシステムから抽出された看護記録のコピーであり、XML

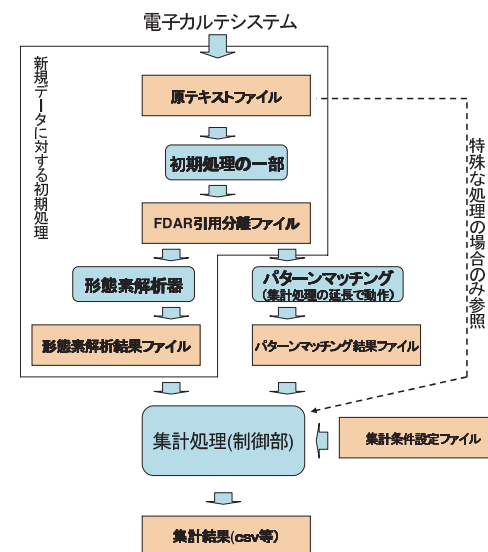


図 2 Nursing Text Miner の主なファイルと処理の流れ  
Fig. 2 Files and processing flow in the Nursing Text Miner.

でタグ付けされ、時系列に記録が並んでいる。新しい原テキストを Nursing Text Miner に取り込む際は、初期処理として、FDAR 引用分離ファイルと形態素解析結果ファイルに新規データを追記する。FDAR 引用分離ファイルでは、各記録は、F 引用内、F 引用外、D 引用外、D 引用内、… とテキストを 8 種のセグメントに分割して格納する。XML タグは区切り記号に置き換え、フィールドの位置のみで該当するデータを参照可能とする。すなわち、FDAR 引用分離ファイルは後続の処理を容易にするために作成するファイルであり、データの意味は、原テキストファイルとほぼ同一である。形態素解析結果ファイルは、FDAR 引用分離ファイルを入力として作成し、各記録のセグメント別に出現した単語のリストを格納する。

パターンマッチング結果ファイルは、集計条件設定ファイルで指定された複数の正規表現が、各記録のセグメントにマッチしたか否かを格納する。多くの集計は、直近の 1~2 年を対象とするので、集計対象の期間のデータのみを対象とする。前回と類似した条件で異なる集計を行う場合には、パターンマッチング結果ファイルの内容は、前回作成した内容の再利用または、追記を行う。追記にくい条件（年月単位の集計範囲が以前と異なり、かつ、新たな正規表現の指定が追加された等）の場合は、前回のパターンマッチング結果ファイルの内容を破棄して、再作成する。

各記録のコード化データは、FDAR 引用分離ファイルとパターンマッチング結果ファイルにも格納する。

### 7.1.2 集計処理

集計条件は集計条件設定ファイルに指定し、結果は主に csv ファイルに出力する。集計条件としては、病棟・年月・看護職 ID 等による集計範囲の絞り込みと、X 軸 Y 軸に取り上げる項目の選択とがある。これにより、クロス集計表、指定条件での高頻度語のリスト、該当する記録を書いた看護職の ID のリスト等の出力を指定する。

分析する語を正規表現で指定する集計や、コード化データのみを集計では、パターンマッチング結果ファイルのみを参照する。

形態素解析結果ファイルは、特定の表現を指定せずに、高頻度語の一覧を出力する際と、共起語の可視化グラフ作成の際に使用する。

引用内外の分割前のテキストが必要な、特定記録の全文出力と KWIC 作成には、原テキストファイルを参照する。

形態素解析には、富士通研究所の実験用の既存の分析器<sup>31)</sup>を使用し、初期処理・集計条件の設定機能・集計機能は Perl を用いて新規に作成した。パターンマッチングは Perl の機

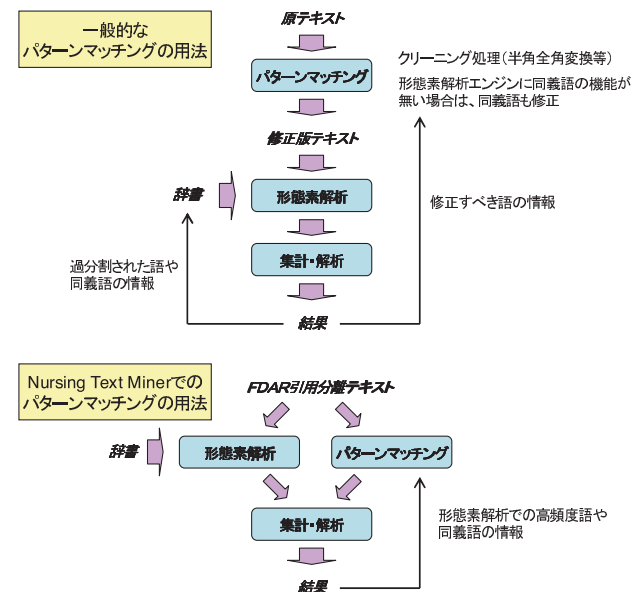


図 3 一般的なマイニングと Nursing Text Miner でのパターンマッチングの使用法の相違  
Fig. 3 Pattern matching in the Nursing Text Miner and conventional mining tools.

能を利用している。

### 7.2 パターンマッチングの使用方式

一般的なテキストマイニングと Nursing Text Miner でのパターンマッチングの使用法の違いを図 3 に示す。

テキストマイニングでの、パターンマッチングの用法は、多くは前処理であり<sup>32)</sup>、分析自体は形態素解析の結果を使用する機会が多い。手軽な方法として、正規表現や KWIC 等のみで分析を行い、形態素解析を使用しない手法も紹介されている<sup>33)</sup>。Nursing Text Miner では両者を併用し、着目すべき単語を抽出するための高頻度語のリストアップには、形態素解析を使用し、特定の単語やフレーズの使用頻度等の集計には、主にパターンマッチングを使用している。

### 7.3 パターンマッチング併用の利点

パターンマッチングを主要な集計手段とした分析の利点は、上記のフレーズの分析が容易であることにとどまらず、辞書整備の負担が小さくなる点にもある。精度の良い形態素解析

を行ううえでは、分析対象のテキストに即した辞書が必要となる場合が多い。しかし、辞書整備は多くの時間を要する場合もあり、辞書を修正すると予期しなかった副作用が集計結果に現れることもある。

未定義語と複合語の扱いは、マイニングのうえで問題となる場合が多いが、Nursing Text Miner の形態素解析では、複合語の可能性のある語については、分解した語と複合語の両方を出力している<sup>31)</sup>。たとえば、辞書に「看護」と「看護婦」しかなくても、「看護師」という表現が多ければ「看護師」と「看護」の両方を出力する。これにより、形態素解析のみによるマイニングで問題となることの多い、単語の過分割回避のための、多数の語の辞書登録を不要としている。

また、結果の解釈を容易にするために、同義語を辞書で定義する場合もあるが、経過記録を分析する場合、画一的に同義語を定義しにくい。たとえば、「医師」と「Dr」は同義語であり、記述の内容を分析するには等価として扱いたい。「患者・家族から見て分かり難い不適切な略語が使用されていないか」との観点から分析する際には、区別して扱うことが必要である。このように分析目的に応じて同義語の扱いが異なる場合、辞書を更新せずに、検索パターンとする正規表現を使い分ける方が効率的である。

主要な集計手段としてパターンマッチングを用いた結果、本事例では、汎用の辞書を使用し、現在まで形態素解析の精度を上げるための辞書更新は行っていない。

#### 7.4 係り受け解析との関係

精度の高い形態素解析は、係り受け解析を行ううえでは必要性が高い。評判分析やコールセンタのログ分析の例では、係り受け解析を行って単語の対を把握することは重要である。看護記録の場合、インシデント・レポートの分析では、係り受け解析の必要性は高い。「点滴」の係り先が「抜ける」か「遅れる」かを区別することは、分析の主要な部分である。しかし、経過記録の分析では、副次的な役割にとどまる。すなわち、表現の頻度分析の範囲内で検討すべき事項が多数あるため、係り受け解析の必要性は高いとはいえない。「点滴」等の語の使用頻度の病棟別や経年的な変化の特徴を分析することがまず必要となるからである。

本事例では、今後係り受け解析を利用する可能性はあるものの、単語間の関係は、正規表現の共起関係等で分析している。

## 8. 分析例

### 8.1 記録内容の規定要因

経過記録の内容を規定する要因としては図4のとおり4種のものがある。患者の疼痛を

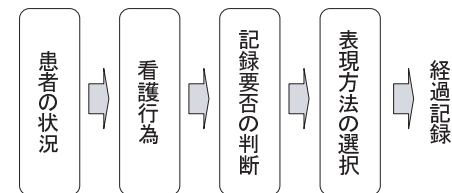


図4 経過記録中の単語頻度の規定要因

Fig. 4 Influential factors of word frequency in nursing narratives.

例とすると、経過記録に「疼痛」が現れるか否かは、まず患者に疼痛があるかに依存する。看護職は患者の表情等の観察により疼痛の有無を判断したり、患者に話しかけて疼痛の有無を確認したりする。疼痛があっても気づかない場合もありうる。疼痛があると認識した場合、記録すべき情報と考えるか否かは、看護職の判断である。記録する場合に、「疼痛あり」と書くか、「～が痛い。」と患者発言の形で書くかは、表現方法の選択である。

以下では、各要因に対応した分析例を示す。

### 8.2 基礎項目

基礎的な統計量として、記録の件数や文字数、R行のある割合、評価レコードの割合、参照している看護問題の内容等がある。

これらは主に表現方法の分析である。これらの量は、病棟間の比較、年月による変化を見ると、着目すべき差や、変化が分かる場合がある。年度単位の表示の例を図5に、月単位の表示の例を図6に示す。

図5, 6に、例示した「看護計画に対応した記録の割合」と「Rのある割合」は、いずれも経過記録として大きいことが望ましいとされており、この指標が他の病棟と比べてどのような値かは、各病棟にとって参考となる情報の1つである。図6は、2008年4月ごろに2病棟で急激な変化が見られ、新年度に記録改善に向けた意識的な努力が行われたことが推察される例である。

### 8.3 不適切表現

表現方法の分析の次の例は、不適切表現の使用状況の集計である。5.1節で述べたとおり不適切表現の例は文献から収集することができる。収集した不適切表現の使用状況を、パターンマッチングにより、病棟別、年月別、個人別等の集計ができる。この情報は、各病棟にフィードバックすることで、削減に有効であることが実証された<sup>17)-19)</sup>。不適切表現として文献に例示されている表現の中には、「～と思われる」等、どのような使用法が実際に不

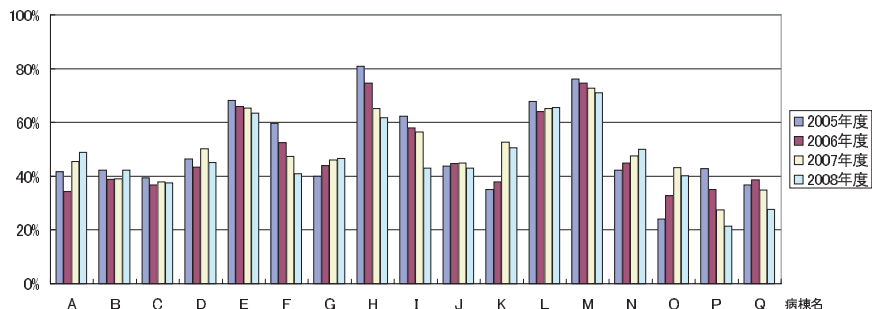


図5 病棟別年単位時系列の出力例：看護計画に対応する記録の割合  
注：匿名化のため病棟の順序は本来の並びを入れ替えてある

Fig. 5 Ward vs. year series output sample: Ratio of nursing problem related records.

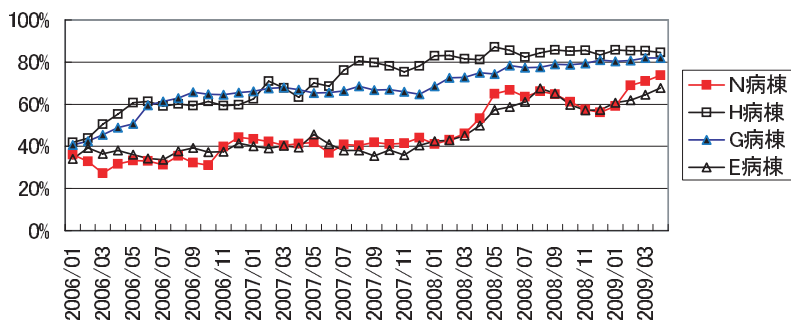


図6 病棟別月単位時系列の出力例：R行のある記録の割合

Fig. 6 Ward vs. month series output sample: Ratio of record having R-lines.

適切に議論を要する例もあるが、多数の用例を収集して分析することも可能となる。

#### 8.4 単語の使用頻度と共起関係

経過記録の中の単語の出現頻度は、疾患や治療方法、担当する病棟、記録を作成した時期等、多数の要因に影響される。患者の属性が同じ記録を集計すると、看護行為・記録要否の判断・表現方法の選択の影響を知ることができる。単語使用頻度の経年的な変化は顕著であり、その背景は6.2節に記したとおりである。

北部病院での結果では、インフォームド・コンセントや、安全対策の重要性が高まった影響を受け、「確認」等の単語頻度は大きく増加している。また、看護職や、医師・看護職によ

順位	2004年度		2006年度		2008年度	
	確認	45,185	確認	55,341	確認	67,964
1	説明	13,111	説明	16,565	説明	25,114
2	本人	9,200	話	10,921	話	15,686
3	話	8,758	医師	10,204	入院	13,229
4	内服	8,249	内服	9,432	医師	12,584
5	状態	8,129	本人	9,350	本人	12,337
6	本日	8,126	本日	9,163	内服	12,288
7	施行	7,888	施行	8,969	本日	10,733
8	医師	7,731	状態	8,871	経過	10,615
9	経過	7,573	経過	8,583	施行	10,376
10	伝える	6,871	観察	8,206	状態	10,158

表1 共起語の変化の集計例：「確認」の共起語

Table 1 Annual change of concurrent words with “CONFIRM”.

るカンファレンスの内容を記載する頻度も増加している。あいまいな記録を避けるという観点から「様子」は大きく減少した。これらは、病棟別・年月別に集計することで状況を把握でき、病棟ごとの状況の差は、フィードバックする意味のある情報となる。

変化の内容をより詳しく知るためには、該当する単語と記録内に共起する単語を、年月別や病棟別に集計することが有効である。表1に、「確認」の共起語の長期的変化を集計した例を示す。この例では、共起語で増加しているのは、「説明」や「話」であり、増加した確認対象は、機器やデータではなく、患者の理解であることが分かる。

また使用単語の頻度を看護職別に集計すれば、個人別の特徴を抽出することができ、平均との差の大きい人や重要な語が顕著に少ない人の情報は、フィードバックや目視による記録監査に有用な情報となる。

看護職を入職年度別にグループ化して単語頻度の変化を集計すると、経験年数によりどのような変化があるか観察できる。北部病院の場合、最も顕著な変化をしている語は「医師 | Dr | 先生」であった。結果を図7に示す。新人1年目は、看護職が直接医師へ報告したり、指示を受けたりすることはなく、リーダーを経由する。このため、「医師」の頻度は少なくなる。グラフからは、医師とのコンタクトがベテランと同様になるには、数年かかることが読み取れる。「医師 | Dr | 先生」の共起語について、図7と同様な図を描くと、経験年数による増加が最も大きいのは、「医師 | Dr | 先生」&「説明」であることも分かる。個人別に見て、経験が長くても「医師」の頻度が少ないことはフィードバック情報となりうる。

このほか、単語の頻度を用いた分析としては、D行の単語とA行の単語との関係、特定の看護診断名とそれに関係した記録中の単語の関係等がある。これらは、看護職の判断が適切に行われているか等の検討の材料となる。

#### 8.5 患者属性と単語頻度の関係

疾患別や、年齢・性別等の患者プロフィール別に集計することは、看護上の新たな知見を



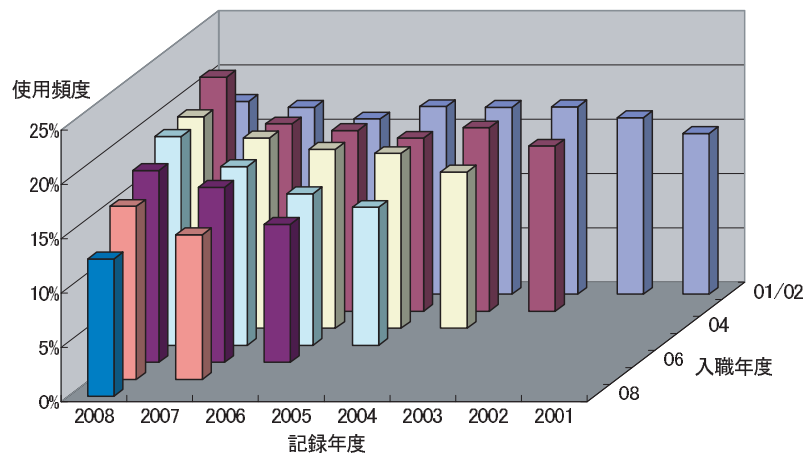


図 7 経験年数に依存する単語の例:「医師 | Dr | 先生」の頻度 (引用外)  
Fig. 7 Frequency of nursing experience dependent word "DOCTOR".

生み出す可能性がある。一例として、患者の男女別に単語頻度を比較した結果では、「不安」に着目すると、男性患者に比較して女性患者では 1.43 倍頻度が高く、特に精神神経科では 1.67 倍の差がある、あるいは、癌患者は 60 代が多いが癌患者の「不安」の記述は 40 代前後がピークである等の結果が得られている<sup>16)</sup>。異なった病棟が担当する疾患の間では、病棟の記録スタイルの差により、有意な比較は難しいが、同じ病棟内で異なるプロファイルの患者間の比較では、有意な差を検出できる。

特定の疾患に着目する等、専門性の高い分析を行ううえでは、単語の共起関係をグラフ化することは、分析を進めるうえで手がかりとなる場合もある。Nursing Text Miner での共起関係の可視化(骨格となる強い単語間関係のみを表示したスケルトンマップ)の例を図 8 に示す<sup>31)</sup>。

特定の疾患の分析で、たとえば、手術後 3 日以内等に対象を絞り込むと該当する記録の件数は少なくなり、目視でも分析可能な量の記録となる。すなわち、必ずしもテキストマイニング手法が有効な領域とはいえなくなる。テキストマイニングが効果的となるのは、患者数の多い疾患について入院期間全体の特徴を抽出したり、多数の疾患について同じ分析方法で集計したりする等、多量のテキストから情報を抽出することである。こうした手法は、今後看護のうえで有用な知見を得られる可能性を秘めていると思われる。

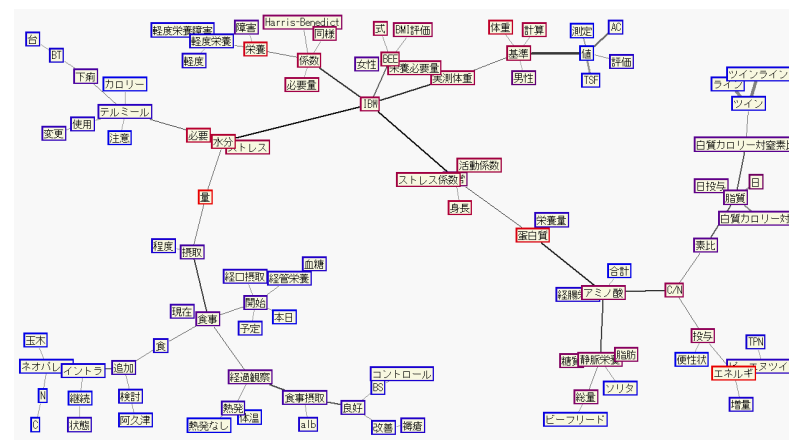


図 8 スケルトンマップによる共起関係の可視化の例: 栄養サポートチーム・カンファレンス記録の高頻度語  
Fig. 8 Visualization of word concurrency using skeleton map.

## 9. 結 語

Nursing Text Miner は、多数の分析を試みることにより、経過記録の分析を行ううえで重要なことが判明した機能を実装したものである。汎用のテキストマイニングツールではカバーできない点が少ない。汎用ツールの適用にとどまらずに、ある程度の機能の作りこみを行うことにより、看護現場に有用な情報を抽出する基本的な仕組みを構築できた。

従来、看護記録の監査等で目視によってチェックしていた項目の一部は、明らかに、コンピュータ処理の方が、効率的で網羅的なチェックが行える。マイニングの結果多量に出力されるデータを、現場で使いやすい情報としていくうえでは、有用な指標となる項目についてさらに知見を積み重ねる必要がある。

試行錯誤的に得られた結果の一部、たとえば、単語使用頻度が数年間の間に大きく変化していることは、分析の初期には予想していなかったことである。また、患者属性に関する特徴を分析するうえでは、看護職の記録スタイルの差としてどのようなものがありうるかを知っておくことが重要であることも分かってきた。これらの知見は、今後の看護記録の分析で、有用である。

また、疾患固有の看護上の知見等を効率的に見出ししていくには、分析方法を開発してい

くことが、今後必要である。

### 参考文献

- 1) 松井くにお, 渡部 勇, 内野寛治: ナレッジマネジメントにおけるテキストマイニング, 情報処理, Vol.47, No.8, pp.893-899 (2006).
- 2) 那須川哲哉: テキストマイニングを使う技術・作る技術, 東京電機大学出版局 (2006).
- 3) 那須川哲哉: テキストマイニングの普及に向けて—研究を実用化につなぐ課題への取組, 人工知能学会誌, Vol.24, No.2, pp.275-282 (2009).
- 4) 保田明夫: テキスト・マイニングの概要, 電気学会論文誌 C, Vol.125, No.5, pp.682-689 (2005).
- 5) 日本看護協会: 看護記録および診療情報の取り扱いに関する指針, 日本看護協会出版会 (2005).
- 6) 安野洋一, 根元大介: ナースのための電子カルテ導入・活用ガイド, 学習研究社 (2005).
- 7) 葛西圭子, 依田安代, 佐藤美幸, 益田亜佐子, 山田由美, 山崎治敏, 林 浩平: 看護情報システムマスター使用状況, 医療情報学, Vol.27 (Suppl.), pp.471-473 (2007).
- 8) 角谷千恵美, 村松由美子: 患者看護情報システムにおける過去3年間の診断ラベルの使用頻度の変化, 看護きろく, Vol.15, No.8, pp.31-35 (2005).
- 9) 遠藤 晃, 佐藤ひとみ, 高橋久美子, 寺下貴美, 櫛引一則, 鈴木研禰, 小笠原克彦, 桜井恒太郎: テキストマイニングを用いた自由記載入力型看護記録の分析, 医療情報学, Vol.24 (Suppl.), pp.716-717 (2004).
- 10) 渡邊仁美, 桑田成規, 歌野由紀, 池上秀和: 患者アセスメントデータベースに登録された患者全体像と患者目標のテキスト分析, 医療情報学, Vol.24 (Suppl.), pp.1112-1113 (2004).
- 11) 大倉博文, 小野千恵子, 丸山ひさみ, 神田博仁, 鈴木敏郎, 丸山康孝, 坂田信裕, 村瀬澄夫: テキストマイニングによる注射・点滴に関する作業過程の検討, 医療情報学, Vol.24 (Suppl.), pp.638-639 (2004).
- 12) 大谷芳弘, 川中普晴, 吉川大弘, 山本皓二, 篠木 剛, 鶴岡信治: 自由形式インシデント報告を対象としたキーワードマップの作成とその検討, 医療情報学, Vol.25 (Suppl.), pp.676-679 (2005).
- 13) 竹村匡正, 廣瀬昌博, 岡本和也, 糸 直人, 黒田知宏, 吉原博幸: インシデント・アクシデントレポート自動分析システムの構築, 医療情報学, Vol.29 (Suppl.), pp.766-777 (2009).
- 14) 佐藤ひとみ, 寺下貴美, 石井英樹, 遠藤 晃, 桜井恒太郎: テキストマイニングを用いた自由記載型看護計画の分析—標準看護計画との比較, 医療情報学, Vol.27 (Suppl.), pp.849-852 (2007).
- 15) 渡部 勇, 村松 洋, 小塚和人, 大崎千恵子: テキストマイニング技術を用いた看護記録分析システムの開発, 医療情報学, Vol.27 (Suppl.), pp.853-854 (2007).
- 16) 小塚和人, 大崎千恵子, 渡部 勇, 村松 洋: 看護記録の自然言語処理による分析と検討, 医療情報学, Vol.27 (Suppl.), pp.855-856 (2007).
- 17) 大崎千恵子, 小塚和人, 村松 洋, 渡部 勇: テキストマイニングによる看護記録の監査, 医療情報学, Vol.27 (Suppl.), pp.857-858 (2007).
- 18) 大崎千恵子, 小塚和人, 村松 洋, 渡部 勇: テキストマイニングを利用した看護記録内容の経年的な比較, 医療情報学, Vol.28 (Suppl.), pp.935-937 (2008).
- 19) 大崎千恵子, 小塚和人, 村松 洋, 渡部 勇: テキストマイニングを用いた看護記録の記述改善に向けた分析, 医療情報学, Vol.29 (Suppl.), pp.414-415 (2009).
- 20) 鳥越千秋, 向窪世知子, 宇都由美子, 窪田美行: 叙時的経時記録の実現を目指す看護記録システムの改造とその評価, 医療情報学, Vol.29 (Suppl.), pp.410-412 (2009).
- 21) 串間宗夫, 荒木賢二, 鈴木斎王, 荒木早苗, 田村宏樹, 淡野公一, 外山貴子, 石塚興彦, 池田 満: マイニング技法を活用した電子カルテ (IZANAMI) のネットワーク可視化, 信学技報, Vol.108, No.388 CAS2008-97, pp.187-192 (2009).
- 22) 串間宗夫, 荒木賢二, 鈴木斎王, 荒木早苗: 肺癌入院患者看護記録のネットワーク可視化, 医療情報学, Vol.29 (Suppl.), pp.963-969 (2009).
- 23) Hiissa, M., et al.: Towards automated classification of intensive care nursing narratives, *Int. J. Med. Inform.*, Vol.76, No.S3, pp.S362-S368 (2007).
- 24) Laippala, V., et al.: Towards automated processing of clinical Finnish: Sublanguage analysis and a rule-based parser, *Int. J. Med. Inform.*, Vol.78, No.11, pp.836-849 (2009).
- 25) Fang, S., Lwin, M. and Ebright, P.: Visualization of Unstructured Text Sequence of Nursing Narratives, *Proc. Symp. Appl. Comp.* '06, pp.240-244 (2006).
- 26) 服部兼敏, 鷺田万帆: 学際的技術としてのテキストマイニング—その意義と看護における可能性, 看護研究, Vol.41, No.3, pp.239-248 (2008).
- 27) 藤井美和, 小杉孝司, 李 政元: 福祉・心理・看護のテキストマイニング入門, 中央法規出版 (2005).
- 28) 相良かおる, 小野正子, 穴井めぐみ, 青山和子, 工藤二朗, 小田正枝: 看護領域における形態素解析実験と看護専門用語辞書, 第4回看護情報研究会論文集, pp.52-55 (2003).
- 29) 鷺田万帆, 服部兼敏: 看護におけるテキストマイニングとその活用事例, 看護研究, Vol.41, No.3, pp.249-259 (2008).
- 30) 川上千栄子: フォーカスチャータリング活用術 (改訂4版), メディカ出版 (2008).
- 31) 渡部 勇, 三末和男: 単語の連想解析によるテキストマイニング, 情報処理学会第55回情報学基礎研究会資料, No.55-8, pp.57-64 (1999).
- 32) 金 明哲: テキストデータの統計科学入門, 岩波書店 (2009).
- 33) 佐良木昌, 新田義彦: 正規表現とテキストマイニング, 明石書店 (2003).

(平成 22 年 2 月 26 日受付)

(平成 22 年 7 月 7 日採録)

(担当編集委員 伊川 洋平)



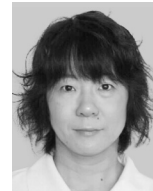
村松 洋 (正会員)

1947 年生。1969 年東京大学理学部物理学科卒業。1971 年東京大学理学系大学院物理課程修士課程修了。1972 年富士通入社。システム性能評価、オペレーティングシステムの開発等に従事。1994 年富士通研究所。現在は医療 IT システムの研究に従事。日本医療情報学会、研究・技術計画学会、ACM 各会員。



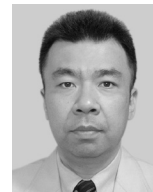
渡部 勇 (正会員)

1961 年生。1985 年慶應義塾大学理工学部機械工学科卒業。1987 年東京工業大学大学院制御工学専攻修士課程修了。同年富士通国際情報社会科学研究所入所。1990 年富士通研究所。テキストマイニング、特許検索・分析等の研究に従事。人工知能学会会員。



大崎千恵子

1964 年生。1985 年横浜市立大学医学部附属看護専門学校卒業。同年昭和大学藤が丘病院入職。急性期病棟等で勤務。2000 年昭和大学横浜市北部病院へ異動。以後、看護部門における医療情報システム専従。2004 年明星大学人文学部心理学科卒業。2007 年日本大学大学院総合社会情報研究科修士課程修了。日本医療情報学会会員。



小塚 和人

1960 年生。1988 年昭和大学医学部卒業。同年昭和大学産婦人科学教室。2001 年から昭和大学横浜市北部病院医療情報部。日本医療情報学会評議員。