

ポスター | 知識工学／医療データ解析

ポスター1

知識工学／医療データ解析

2019年11月22日(金) 14:50 ～ 16:20 ポスター会場2 (国際展示場 展示ホール8)

[2-P2-1-06] 介護施設入所者病歴に関するグラフマイニング

○串間 宗夫¹、山崎 友義¹、酒田 拓也¹、川野 光一¹、松尾 亮輔¹、小川 泰右¹、荒木 賢二¹、近藤 千博²（1. 宮崎大学医学部附属病院, 2. 宮崎市立田野病院）

キーワード：Medical Informatics, Graph Mining, Electronic Medical Record

【目的】介護ライフログの二次的利用法の開発が現場の介護職員から強く望まれている。本研究では、S介護老人保健施設入所者の20人について介護ライフログに記入された病歴を解析した。解析により介護施設入所者病歴関係の可視化図を示すことにより、多職種の医療従事者間で共有し活用することによって人材育成教育や医療従事者によって行われる作業等を効率化する手法を明らかにすることによって、実際の介護行為に反映させる。将来的には、要介護度に合わせて利用者の記録を半自動的に作成できる電子カルテの開発を目指す。

【方法】グラフマイニングとしての基礎的な分析および可視化のツールとして、Gephi 0.9.2 を用いた。可視化では Gephi に実装されている ForceAtlas2 というアルゴリズムを用い、重みの高いリンクの距離が短くなるようにノード配置の最適化を行った。

【結果】本研究解析結果として、介護施設入所者病歴関係可視化図を示し、ネットワークの一般統計としては、ノード数：90、辺の数：89、平均次数：0.989、平均重み時数：0.989、ネットワーク直径：1、グラフ密度：0.011、モジュラリティ：0.789 であった。この結果から、入所者一人ひとりに対して、多数の病歴があり、その病歴に対しては、多数の入所者が複数罹患されていた。入所者に多い病歴としては、心不全、不眠症、不眠症、廃用症候群、脳梗塞、糖尿病、骨粗鬆症、高血圧、となっている。この解析図を通して、感染症発見につながると考えられる。そのためには、データ数を増やす必要がある。

【まとめ】介護福祉施設入所者の病歴を可視化し、その可視化結果について考察し検討した。本研究は、宮崎大学の倫理指針に従っており、本研究の遂行は、宮崎大学倫理審査委員会の承認を得ている。宮崎大学承認番号：第 O-0383 号。

介護施設入所者病歴に関するグラフマイニング

串間 宗夫^{*1}、山崎 友義^{*1}、酒田 拓也^{*1}、川野 光一^{*1}、松尾 亮輔^{*1}、小川 泰右^{*1}、
近藤 千博^{*2}、荒木 賢二^{*1}、

^{*1} 宮崎大学医学部附属病院 病院 IR 部、

^{*2} 宮崎市立田野病院

Graph Mining on the Medical History of Nursing Home Residents

Muneo Kushima ^{*1}, Tomoyoshi Yamazaki ^{*1}, Takuya Sakata ^{*1}, Koichi Kawano ^{*1},

Ryosuke Matuo ^{*1}, Taisuke Ogawa ^{*1}, Kazuhiro Kondo ^{*2}, Kenji Araki ^{*1}

^{*1} Informatics Research, University of Miyazaki Hospital,

^{*2} Miyazaki City Tano Hospital

Purpose: In this study, we analyze and visualize medical histories which are written in the nursing care life log. The visualization enables medical workers to rationalize their work by sharing and exploiting the visualization results among the various types of medical workers.

Method: We employed graph mining by using Gephi 0.9.2 as a tool for the basic analysis and visualization.

Result: As a result, the visualization showed the relation of medical histories of nursing care center residents. The general statistics of the network are the number of nodes: 90, the number of edges: 89, the average degree: 0.989, the average weighted degree: 0.989, the network diameter: 1, graph density: 0.011, modularity: 0.789. The network indicated that there is a large number of medical histories for each resident, and each medical history corresponds to the various residents. The common medical histories of those residents are heart failure, insomnia, disuse syndrome, cerebral infarction, diabetes, osteoporosis, and hypertension. It is assumed that these analyses lead to the discovery of an infectious disease.

Summary: We visualized the medical histories of residents who are admitted to the nursing and welfare institution, and considered and discussed the visualization results.

Keywords: Medical information, Graph mining, Visualization, Electronic medical record.

1. 緒論

近年、膨大な医療データの中から得られる二次利用可能なデータの取捨選択や、その内容を素早く解析し理解するための環境が医療介護現場でも強く望まれている。また、医療現場における電子化が進み、電子データの利活用も広まってきた。

デジタルデータの利用によって物理的な量が減り、またデータの検索が素早く行えるなど、医療の効率化がなされている。しかし、その電子データの二次利用、特に介護福祉施設での介護記録の解析に基づく介護医療行為の改善、高度化という面では、まだ十分ではないというのが現状である。

介護記録の問題点として、他職種への連携・伝達や介護者間での被介護者状態把握のツールとして活用困難であり、更に、介護者の経験年数によって差があり、記述内容だけで介護要点の介護者間での共有が困難である。だが、何のために読むのか、を常に意識して書くことや、読み手が正しく理解し、ケアに活かすことができるように書くことが必要である。

2. 目的

電子カルテ医療記録としては、ライフログとして、人間の行い(Life)をデジタルデータとして記録(Log)に残すこと。個人に関する情報として、プライバシーは、それぞれ領域が異なる。基本属性として、個人、事業者:氏名、性別、生年月日、住所、職業など。事業者として、個人:ID、パスワードなど。行動情報として、医療情報(履歴)、医院診断結果(カルテ)、処方箋(量・頻度)、歯科医診断結果(通院カルテ)。個人情報として、特定の個人識別などから構成されている。

介護ライフログの二次的利用法の開発が現場の介護職員から強く望まれている[1, 2, 3]。介護ライフログは、被介護者の施設サービス利用内容記録と介護者の観察記録を統合した記録であり、他職種への連携・伝達や介護者

間での被介護者の状態を把握できるツールとして使用されている。これまで介護施設で大量に蓄積された詳細な医療介護処置の記録が履歴として記録されてきたが、その介護記録内容を綿密に解析して医療従事者に必要な知識の伝承を行うことまでは、十分に行われてこなかった。

医療情報の発達に伴い、医療記録自体も電子化が進みつつあるが、現時点では、ただ単に記録して残すというレベルにとどまっており、それ以上に活用されているとは言い難い。一方、介護業務の効果的運用と改善、介護業務従事者の教育・研修のため、介護ライフログの二次的利用法の開発が現場の介護職員から強く望まれているが、この課題を解決する具体的な検討は少ない。

今後の介護業務の向上には、前述の課題を解決し、電子的介護記録システムの開発が重要である。

3. 方法

本研究は、S 介護老人保健施設入所者の 20 人について介護ライフログに記入された病歴を解析した。

S 介護老人保健施設(T 病院併設)は、平成 27 年度から宮崎大学が指定管理者として運営しており、国立大学病院が、介護老人保健施設を管理しているのは全国初である。解析により介護施設入所者病歴関係の可視化図を示すことにより、多職種の医療従事者間で共有し活用することによって人材育成教育や医療従事者によって行われる作業等を効率化する手法を明らかにすることによって、実際の介護行為に反映させる。

将来的には、要介護度に合わせて利用者の記録を半自動的に作成できる電子カルテの開発を目指す。

3.1 原理手法

本研究では、ネットワーク分析の手法を用いている[4]。ネットワーク分析では、ネットワークに対して様々な処

理を行う。その際にネットワークは隣接行列によって表現される。隣接行列とはネットワークのノード間のエッジの有無を行列で表現したものである。 n 個のノードから構成されるネットワークは $n \times n$ の正方行列で表される。 n 次の隣接行列は以下のようになり、

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix} \quad \dots \dots \dots (1)$$

その成分は以下になる。

$$a_{ij} = \begin{cases} 1 \\ 0 \end{cases} \quad \dots \dots \dots (2)$$

(1: ノード i からノード j へのエッジが存在、0: ノード i からノード j へのエッジが存在しない)

ネットワーク構造を分析する際に、いくつかの重要な指標があり、代表的なものを以下に示す。

ネットワークにおける密度とは、グラフにおいて張ることのできるすべてのエッジの数と、実際に存在するエッジの数の比率である。

$$\text{density} = \frac{m}{N(N-1)} \quad \dots \dots \dots (3)$$

ノードは全てのノード間にエッジが存在する完全グラフなら最大値である 1 をとり、エッジが存在しない空グラフで最小値である 0 をとる。

次数とは、あるノードに接続するエッジの数である。ノード数 N のネットワークにおいて次数の最小値は 0、最大値は $N-1$ である。多くの場合、ノードごとに次数は異なる。そのため次数分布はネットワークの特徴を現わしているといえ、有用である。次数分布は次数 k のノードが占める割合を $p(k)$ とおく時、

$$\{p(k)\} = \{p(0), p(1), p(2), \dots, p(N-1)\} \quad \dots \dots (4)$$

と表せる。また、平均次数 $\langle k \rangle$ は以下で定義される。

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i \quad \dots \dots \dots (5)$$

ネットワークの次数分布は様々な形をとるが、多くのネットワークに見られるべき則がある。べき則とは、

$$p(k) \propto k^{-\gamma} \quad \dots \dots \dots (6)$$

で表され、 γ をべき指数と呼ぶ。

中心性とは、あるノードがネットワークの中で中心的な役割を果たしているかどうかを測定する指標であり、ノードごとの特徴を分析するのに用いる。以下では、特に記述がない限り、ノード数が N (頂点 $v_1 \sim$ 頂点 v_N) であり、隣接行列 A (成分は a_{ij}) の有効ネットワークとする。

次数中心性は最も基本的な中心性であり、あるノードの次数によって定義される。有効グラフではエッジの方向によって入次数 (indegree)、出次数 (outdegree) によって分けられるため、以下のように定義される。

$$\text{indegree}(i) = \sum_{j=1}^N a_{ij} \quad \dots \dots \dots (7)$$

$$\text{outdegree}(i) = \sum_{j=1}^N a_{ji} \quad \dots \dots \dots (8)$$

次数中心性は計算が平易であるが、あるノードの次数中心性は隣り合うノードの数のみによって定義されるため、ネットワーク全体の構造は反映していないといえる。

3.2 グラフマイニング

グラフマイニング[5, 6]としての基礎的な分析および可視化のツールとして、Gephi 0.9.2 を用いた。

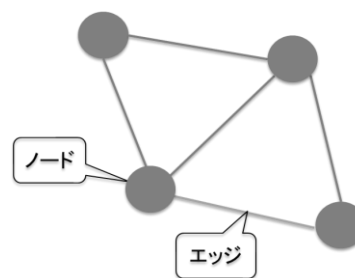


図 1 グラフ図例

図 1 に、グラフ図例を示す。グラフ構造は、ノードとノード間を結ぶエッジから構成されるデータ構造である。グラフ構造の分析結果を可視化は、ひとめで全体像を把握し、重要な情報の発見を容易にすることに役立つ。

可視化では Gephi [7, 8, 9, 10] に実装されている ForceAtlas2 というアルゴリズムを用い、重みの高いリンクの距離が短くなるようにノード配置の最適化を行った。Gephi は、大規模なネットワークグラフの可視化と分析のためのオープンソースソフトウェアである。3D レンダリングエンジンを使ってグラフをリアルタイムで表示し、グラフの探索をスピードアップできる。また、あらゆる種類のグラフを対象とした探索、分析、空間化、フィルタリング、クラスタリング、操作、エクスポートが可能である。更に、データ解析の支援を通じて、仮説を立て、パターンを直感的に把握し、データソーシング時に構造的な特異点や欠陥を分離し、グラフの探索や分析を行い、描画されたグラフと対話し、構造や形状、色などを操作しながら、隠された属性を引き出すことができる。これは CSV 形式などのいくつかの形式のファイルを読み込み、ノードとエッジで表されるグラフ構造を表示する。図 2 に、Gephi のスクリーンショット例を示す。

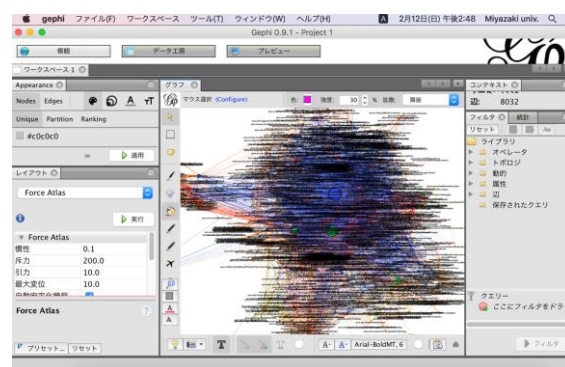


図 2 Gephi スクリーンショット例

4.結果

本研究解析結果として、介護施設入所者病歴関係可視化図を示し、表1に、グラフマイニングネットワークの一般統計解析数値を示す。

次数は、データの遷移先がいくつあるかを表す指標であり、これらは、データの多義性の程度を表す指標であると解釈できる。ネットワーク直径は、データ間で遷移するために必要な平均的な遷移数であり、小さい直径は、多くのデータが通過するハブの存在を示唆している。モジュラリティは、この値が小さいことで「コミュニティ」が明確に存在しないことを示唆しており、データの組み合わせは、分離しているわけではなく、比較的によく混ざっている。この結果から、入所者一人ひとりに対して、多数の病歴があり、その病歴に対しては、多数の入所者が複数罹患されていた。

表1 ネットワークの一般統計

項目名	数値
ノード数	90
辺の数	89
平均次数	0.989
平均重み次数	0.989
ネットワーク直径	1
グラフ密度	0.011
モジュラリティ	0.789

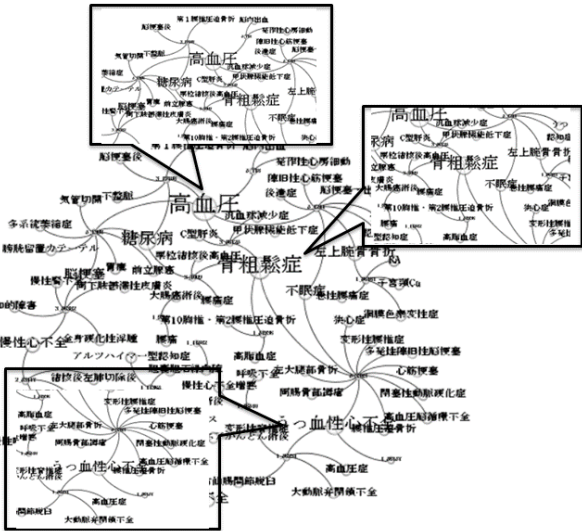


図3 介護施設入所者病歴関係可視化結果のスクリーンショット

入所者に多い病歴として頻出した単語は、心不全、不眠症、廃用症候群、脳梗塞、糖尿病、骨粗鬆症、高血圧、となっている。

図3に、介護施設入所者病歴関係可視化結果のスクリーンショットを示す。複数の点(ノード)とそれを結ぶ線(エッジ)によって構成されている。この図では、「クラスター」を見つけることができる。大きな集合を確認できる。今回は、ネットワークが小規模といえる。比較的に蜜につながりあっているところが見て取れる。ノードの色はデータのテーマ、ノードの大きさはHub 数値を表すように設定した。まず全データのグラフで

は、同一テーマ内の遷移が多いことを反映し、同じテーマのノードが固まる傾向が見られる。

図3の上部に位置している、高血圧、骨粗鬆症、糖尿病が、多数の入所者と関係していることを意味している。多数の人と満遍なく関係を持っていることを、意味グラフを用いることで、様々な物の関係を表すことができる。視覚的にパネルに表示することによって、介護士等の介護従事者が介護行為に関する事柄を視覚的に判断できるシステムとなり得る。

介護職員へのフィードバックの結果から、介護職員は、自分の担当被介護者について視点を持ちながら記録に臨み、その記録を介護に生かしている場面がうかがえた。

図4と図5に、サイズ分布と度数分布図を示す。この場合、12の可能なグループ化または相互接続されたノードのクラスターを確認している。また、データセット内のサイズ分布を示しており、モジュール化クラスとノード数のサイズが含まれている。平均次数が計算され、視覚化する前にグラフネットワークが理解できる。

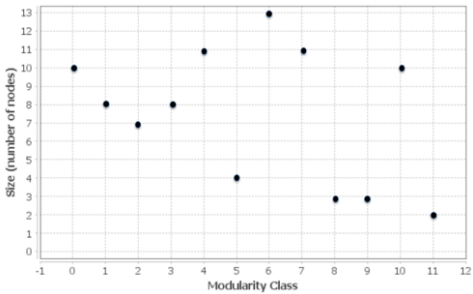


図4 サイズ分布(Modularity: 0.789, Modularity with resolution: 0.789, Number of Communities: 12)

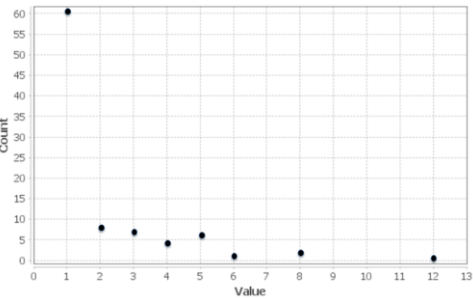


図5 度数分布

Source	Target	Type	Id	Label	Interval	Weight
5.4211	高血圧	無向	0			1.0
5.4211	不眠症	無向	1			1.0
5.4211	心不全	無向	2			1.0
5.4211	糖尿病	無向	3			1.0
5.4211	骨粗鬆症	無向	4			1.0
5.4211	左心不全	無向	5			1.0
5.4211	脳梗塞	無向	6			1.0
5.4211	脳梗塞・出血性脳梗塞	無向	7			1.0
5.4211	認知症	無向	8			1.0
5.4211	認知症・痴呆性認知症	無向	9			1.0
5.4211	痴呆性認知症	無向	10			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	11			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	12			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	13			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	14			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	15			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	16			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	17			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	18			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	19			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	20			1.0
5.4211	痴呆性認知症・痴呆性認知症	無向	21			1.0

図6 リンク次数解析のスクリーンショット

図6に、リンク次数解析のスクリーンショットを示す。Gephiのネットワーク解析機能を使うと、他の知識と関連つけられて

いないページを容易に抽出することができる。

5.考察

- (1)入所者一人ひとりに対して、多数の病歴があり、その病歴に対しては、多数の入所者が複数罹患されていた。
- (2)入所者に多い病歴としては、高血圧、脳梗塞、骨粗鬆症、廃用症候群、糖尿病、うつ血性心不全、不眠症となっている。
- (3)視覚的にパネルに表示することによって、介護士等の介護従事者が介護行為に関する事柄を視覚的に判断できるシステムとなり得る。
- (4)介護職員へのフィードバックの結果から、介護職員は、自分の担当被介護者について視点を持ちながら記録に臨み、その記録を介護に生かしている場面がうかがえた。
- (5)被介護者のさまざまな情報を共有するために役立てるといふ作業を行っていない傾向が読み取れる。
- (6)多忙を極め、忙しい業務に携わっている介護士の仕事の性格から、現在の施設内の記録は、介護士が空き時間にメモや申し渡しの側面をもっており、一般的な記録文章としての性格をもったものではない。
- (7)分析を考えた記入がされていない資料からの検討である等の問題点が挙げられる。
- (8)今後は、本研究で明らかにした解析結果を更に積み重ね、記録に関するデータベースを構築したい。

6.結論

介護ライフログとして記録されている介護福祉施設入所者の病歴に関して、グラフマイニングを用いて可視化し、その可視化結果について考察し検討した。

将来的には、要介護度に合わせて利用者の記録を半自動的に作成できる電子カルテの開発を目指す。

謝辞

本研究は JSPS 科研費 JP18K11530、公益財団法人大同生命厚生事業団地域保健福祉研究助成を受けたものである。

本研究は、宮崎大学の倫理指針に従っており、本研究の遂行は、宮崎大学倫理審査委員会の承認を得ている。宮崎大学承認番号:第 O-0383 号。

関係者各位の協力に感謝する。

参考文献

- 1) 齋秀娟、神部智司、笠原幸子、清水由香、岡田進一、白澤政和、介護支援専門員が認識する時間を要する業務内容とそれに関連する心理状態に関する探索的研究—テキストマイニングに基づいた自由記述文からのキーワード抽出—、生活科学研究誌・Vol.12 (2013)、2013。
- 2) 福田賢一郎、中島正人、福原知宏、渡辺健太郎、西村拓一、本村陽一、介護現場における情報収集と利用、2015 年度人工知能学会全国大会(第 29 回)JSAI2015、1K5-NFC-05b-2、2015。
- 3) 柴田邦臣、服部哲、松本早野香、介護に関する記録を利用者の生きがいに結びつける「福祉生活の思い出日記」サイトの構築、情報処理学会研究報告、Vol.2011-GN-79 No.3、2011。
- 4) 進藤暁俊、村上進亮、山富二郎、銅供給構造の変遷の分析 - 企業ネットワークからのアプローチ - , Journal of MMIJ, p.559 - 566, 2015.
- 5) 田邊哲哉、大島純、廣田雅春、石川博、横山昌平、グラフ型データベースを用いたアクティブラーニングにおける会話分析システムの提案、第 9 回データ工学と情報マネジメントに関するフォーラム DEIM Forum 2017 F4-3, 2017.
- 6) 飯田恭弘、岸本康成、藤原靖宏、塩川浩昭、鬼塚真、大規模グラフ構造データからのコミュニティ抽出と重要度計算—高速化

への取組みと応用—, 人工知能, 29 巻 5 号, 2014 年。

- 7) Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.
- 8) M. Thangaraj, S. Amutha, Description of GNP (Gephi, NodeXL, Pajek) Social Network Analysis Tools, International Journal Science and Research (IJSR) ISSN (Online): 2319-7064, 2015.
- 9) “Gephi”. <https://gephi.org/>
- 10) Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, Using Computer Games Techniques for Improving Graph Visualization Efficiency, IEEE-VGTC Symposium on Visualization, 2010.