HOKKAIDO UNIVERSITY

Title	Network analysis of medical knowledge to investigate the structure of medical service : Toward a mathematical approach to medical service. [an abstract of dissertation and a summary of dissertation review]
Author(s)	日月, 裕
Degree Grantor	北海道大学
Degree Name	博士(理学)
Dissertation Number	乙第6960号
Issue Date	2015-06-30
Doc URL	https://hdl.handle.net/2115/59683
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Yutaka_Tachimori_abstract.pdf, 論文内容の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学 位 論 文 内 容 の 要 旨

博士（理 学） 氏 名 日 月 裕

学 位 論 文 題 名

Network analysis of medical knowledge to investigate the structure of medical service: Toward a mathematical approach to medical service.

(医療構造を調べるための医療知識のネットワーク分析: 医療に対する数学的アプローチを目指して)

医療知識は非常に複雑である。病名だけでも数万個を数え、それ以外に症状、検査結果など膨大な量の知識がある。医師はこのように大量、複雑な知識を扱いながら患者に対して適切な診断を行っている。医師はこのように多数の病名の中から如何にして適切な診断ができるのであろうか。一つの仮説として、医師が扱う医療知識そのものに診断を支えるための構造が含まれているということが考えられる。

医療知識は自然言語によって記述される場合が多い。その場合には、病気の多様な性質が他の病気との関連の中で自由に記述される。近年、このような自然言語に対して、ネットワーク分析の方法が開発されてきた。文章に記述されている単語同士を何らかの基準により結びつけネットワークを構成する。そのネットワーク構造を分析するのである。本研究では医学教科書の記述からネットワークを構成し、それがスモールワールド性、スケールフリー性、階層性を持つことを示した。さらに、実際の病院診療のデータベースが同じ性質を持つことを示し、医療知識の構造が実際の医療行動に影響している可能性を論じた。

医学知識として標準的教科書であるハリソンの「内科学」を採用した。まず、文章から医学的に意味のある一つあるいは数個の単語からなるひとまとまりを term と定義した。各 term は診断名(d)、自覚症状(s)、他覚的異常(f)、その他の医学専門用語(x)に分類した。例えば、胃癌を表す stomach cancer は二つの単語からなる診断名(d)に分類される term である。文章には term に含まれない冠詞、接続詞、動詞、副詞、形容詞なども含まれるが、これらの単語は分析の対象から外した。文章から抜き出した各 term をネットワークのノードとした。一つの文に含まれる全ての term(ノード)を枝によって互いに連結した。この操作を全ての文に行い、一つのネットワーク(医療知識ネットワーク：MKN) を構築した。

次に、病院情報システム内の病名データベースからネットワークを構築した。病名データベースは医師の日常の診療において、患者の病名をその都度、コンピュータに入力することにより作られたデータベースである。このデータベース内の各レコードには病名以外に、患者番号、科コード、病名コード(ICD コード)、医師コード、入力日などのデータが同時に記録されている。今回はこれらのデータを単純にネットワークのノードとし、同じレコード内のノード同士を枝で連結することによって病名データベースネットワーク(DDN)を構築した。

これらのネットワークに対して、以下の指標を計算した。(1)次数：各ノードに繋がる枝の数。(2)平均距離：任意のノード間をつなぐ枝の数の最小値の平均。(3)平均クラスター係数$\langle C\rangle$：各ノード vに対して k_vをその次数とすると、局所クラスター係数を$\mathrm{C_v} = \frac{(v\text{の近傍ノード間の枝の数})}{k_v(k_v-1)/2}$ と定義する。平均クラスター係

数は $\langle C \rangle = \frac{1}{n}\sum_v C_v$ (n: ノードの総数)と定義する。(3)次数分布 $p(k)$：ノードの次数の分布。

ネットワークがスモールワールド性を持つとは、小さな平均距離と大きなクラスター係数を持つことを意味し、スケールフリー性とは次数分布が$p(k) \backsim k^{-\alpha}$なる、べき指数 α の「べき分布」を示すことを意味する。さらに、階層構造についても調べた。

以上の指標を使って構築したネットワークを調べたところ、MKN、DDN のいずれもがスモールワールド性、スケールフリー性を持つことが分かった。しかも、DDN の平均クラスター係数は 0.83、べき指数は 2.084 であり、MKN のそれらはそれぞれ 0.86 と 2.045 といずれも非常に近い値であった。これらの事実は DDN と MKN とは似かよったネットワーク構造を持っていることを示唆する。さらに、これら二つのネットワークは共に階層構造の特徴も有していた。これらの事実は、数学的には今回作成したネットワークが従来のランダムグラフモデルでは表せないことを意味している。

医療知識がスモールワールド性を持つことは、全ての病名が少数の枝で繋がっていることを意味し、病名間の連想を行う際に有利であると考えられる。

一方、スケールフリー性と階層性は知識構造の生成に関係している。医療知識は常に変化している。診断名も常に新しいものが発生し、診断の定義もしばしば変更される。新しい語の生成は MKN に対しては新しいノードの追加を意味する。新しい語はすでに存在する医学用語と関連する。その際、すでに、多くの語と関連している語と関連する確率が高くなるであろう。このことは、MKN に新しいノードが追加される場合、すでにあるノードのうち次数の大きいものに繋がる可能性が高いことを意味する(優先選択原理)。これが、MKN がスケールフリー性を持つ理由である。また、新しい語が追加される場合には、すでに存在する医療知識の構造をあまり変化させないように追加されるであろう。このことが、MKN の中に階層構造が見られる理由であると思われる。MKN が成長するときの優先選択原理とすでにある構造の維持という二つの原理が MKN の構造を決めていると思われる。

最後に我々は実際の病名頻度の分布と知識ネットワークとの関係を調べた。MKN のノードのうち、診断名と分類されるノードの次数分布と病院の実際の病名頻度分布を比べた。その結果、どちらの分布もよく似た「べき分布」を示した。文章から作ったネットワークにおいてはノード(term)の次数と文章内でのその term の出現頻度が相関することが知られている。今回の結果はテキスト内によく出現する病名が実際の診療においても高頻度で出現することを意味している。医学テキストは常に新しい臨床経験を取り込んでいる。そのため、実際の臨床の中でしばしば現れる病名はテキストのなかでもしばしば扱われる可能性が高い。逆に、テキストにしばしば現れる病名は臨床の現場において医師によって想起される可能性が高くなるであろう。このような、医療知識と実際の診療の間の相互作用がこれら二つの間の構造を似たものにしていると考えられる。

医療知識や、病院データベースを使った医療行為のネットワーク分析は、世界で初めての試みである。このような分析は、医療の全体像を記述するために、今後、重要になると思われる。このネットワーク構造に対して、コミュニティ分析などの新しい分析をさらに進めることは、臨床応用のために、重要である。また。そのような分析によって、医療の全体像を明らかにすることが可能となるかもしれない。

しかし、スモールワールド性、スケールフリー性は従来の数学モデルであるランダムグラフモデルでは表現できない。そのため、新しい数学的枠組みが必要である。この枠組みの構築は医療界のみならず、数学界においても新しい分野の創造につながる可能性を秘めている。医療者と数学者の協働による、新しい数学分野が構築されることを心から願っている。