

N-014

ネットワーク理論による疾病データの可視化
A visualization of KDB data by using network theory.

村館 靖之†
Yasuyuki Muradate

1. はじめに

地方自治体における財政難、特に国民健康保険事業からの財政赤字の問題は無視できない。そのため医療費適正化のための指針等[1][2]をもとにさまざまな地方自治体が地域の医療費適正化問題に取り組んでいる。地方自治体が医療費適正化や、地域医療の分析に使うためのシステムの一つが KDB (国保データベース) である。

本研究では、隣接行列の概念を用いて産業連関表の可視化手法を参考にし、ネットワーク理論を KDB (国保データベース) の掃出しレセプトデータに適用し、千葉市における疾病ネットワークを可視化する。

2. 方法と課題

KDB 掃出しレセプトデータは、列方向に被保険者証番号、性別、年齢、入院・外来区分、費用額、生活習慣ラベル、主病名から、第 6 番目に高い傷病名まで並んでいるデータである。KDB データの主病名と第二傷病名のリスト、および発生レセプト件数から閾値(100 件)を設けて 0・1 の要素のみの正方行列からなる隣接行列を作成し、疾病ネットワークを可視化する。樹状図の考えを用いて、疾病の近隣関係、クラスターを可視化する。

3. データと手法

H24.6 から H26.12 までの千葉市の KDB 掃出しレセプトデータ (約 515 万件) から H26 年の一年間に絞って分析する。手法はネットワーク理論である。ネットワーク理論の隣接行列の可視化手法は産業間ネットワーク、SNS の交友関係などさまざまなネットワークの可視化に用いられる。この手法を KDB の疾病ネットワークの可視化に応用を試みた。可視化にあたっては、統計ソフト R のグラフ描画機能パッケージである igraph を用いた。

4. 可視化

まずは社会保険表章用疾病分類表 (121 項目) の疾病の中で、正方行列作成の観点から、主病名だけに発生する「特定の処置 (歯の補てつを除く) 及び保健ケアのための保健サービスの利用者」の項目を除いた 120 疾病の可視化例を表示する。

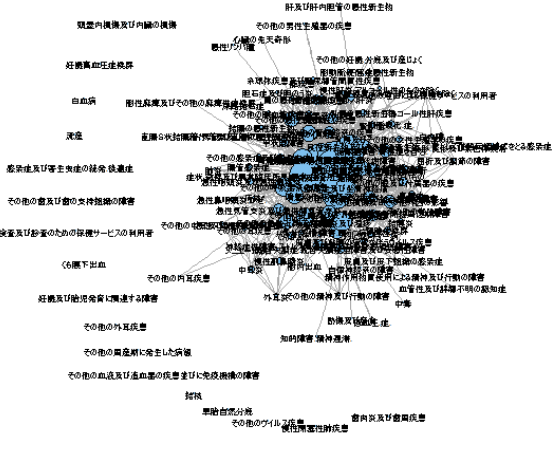
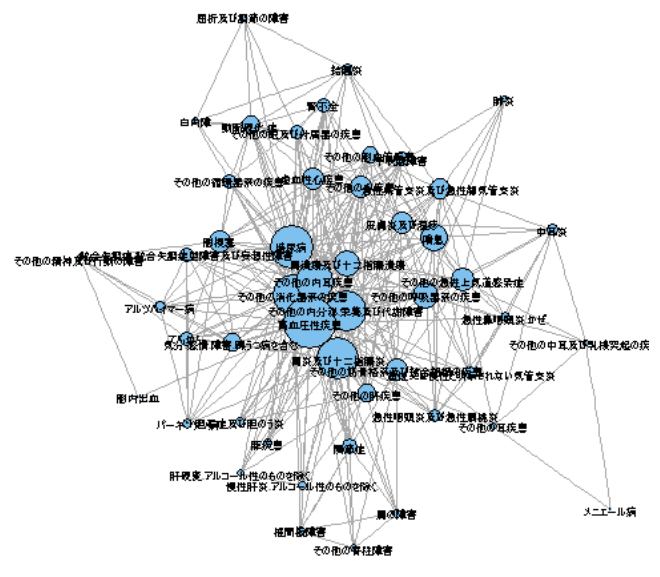


図 1 120 疾病ネットワークの可視化 (出所) KDB データをもとに筆者による計算結果

図 1 で 120 疾病ネットワークを可視化すると、孤立している疾患が多数存在することがわかる。またほぼ全ての中分類の疾患を可視化しているので、詰まっていて見にくい。分析上見やすくするため、疾病ネットワークの可視化にあたって、社会保険表章用疾病分類表 (121 項目) の疾病の内、出産、歯の疾患やウイルス性疾患、新生物などを除いた 49 疾患にしぼった。全疾患を可視化すると、出産や歯の疾患は、他の疾患とかかわりが薄く、孤立化し、周辺に島のように立地する。



† 東京大学大学院情報学環

図2 49疾病ネットワークの可視化
出所) KDBデータをもとに筆者による計算結果

図2では、49の疾病のネットワークを可視化した。中心には、ネットワークの次数が最大な「高血圧性疾患」が位置している。さらに糖尿病も中心に近い位置に来ている。ネットワークの中心には、次数の大きい疾病、つまり他の疾病と併発しがちな疾病が配置されている。一方、ネットワークの周縁には、他の疾病と併発しにくい疾病が配置されている。

可視化にあたって、円の大きさはネットワークの次数に比例するように設定して描画した。

表1 49疾病のネットワークの次数(ランキング上位5位)
出所) KDBデータをもとに筆者による計算結果

ランク	疾病名	次数
1	高血圧性疾患	41
2	糖尿病	34
3	胃炎及び十二指腸炎	33
4	その他の内分泌、栄養及び代謝障害	32
4	その他の消化器系の疾患	32

表1では、49疾病のネットワークの次数(手の本数)を上位5位までならべた。高血圧性疾患の次には、糖尿病が来ている。糖尿病は千葉市がターゲットとしている疾患である。

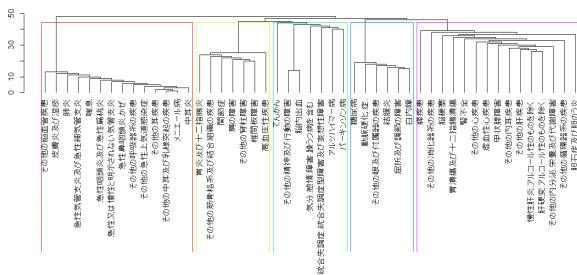


図3 49疾病のデンドログラム
出所) KDBデータをもとに筆者による計算結果

図3では、49疾病のデンドログラム(樹状図)を可視化した。全体で5のクラスターに分類できた。第一のクラスターは呼吸器系と耳に関する疾患、つまり耳鼻咽喉科系の疾患であった。第二のクラスターは高血圧性疾患を含んでいる。第三のクラスターは脳神経系の疾患が中心である。第四のクラスターは目と糖尿病である。第五のクラスターは、腎不全を含む内臓系の疾患である。機械はクラスターに分類する際、基本的に医学的知識なしに、自動的にネットワークの近縁関係をもとに分類する。しかし、機械学習による分類結果は医学的な解釈も可能な形になっている。

5. 考察

ネットワークの次数(手の本数)が最大なものは高血圧性疾患であった。つまり高血圧は、「万病のもと」である。高血圧症が、ネットワークのへそに来ている。この手法では因果関係は特定できず、相関関係を可視化するだけだが、このような結果が得られた。疾病のデンドログラムは市民の健康相談でも、たとえば白内障や動脈硬化と、糖尿病を併発している方が多いんですよといったコンサルが可能になる。

疾病のネットワークのへそが高血圧疾患であるという結果は、疾病の相関関係を分析して得られた事実である。一方で、千葉市の政策担当者の認識としては、糖尿病から腎不全にいたる経路、疾病の相関を可視化してほしいというニーズがある。デンドログラムでは、本来近隣関係にあるはずの腎疾患と糖尿病が距離が開いていた。機械学習による自動的な分類手法は、場合によっては人間の実感と離れた結果を出力する。分析者による可視化結果の解釈が重要である。

可視化にあたっては、閾値を増加させると、ネットワーク図は疎になる。閾値を減少させると、ネットワークは密になるという関係がある。そもそもどのような疾病を中分類からピックアップするかという問題と、閾値をどのように設定するか、どのような可視化パターンが見やすいかはトレードオフの関係にある。

6. 結論

疾病ネットワークの可視化の結果、ネットワークの次数が最大なものは高血圧性疾患であった。

今回の結果は、疾病ネットワークを、主病名と第二傷病名のリストをもとに、KDBの掃出しレセプトデータから可視化することによって得られた。

KDBの掃出しレセプトデータをもとに、産業連関表の内生部門に対応する表を作成し、隣接行列に変換することで、産業連関表の可視化手法と同様の手法を医療系データにも適用した点が、今回の研究の新要素である。

今後の課題としては時間を軸に見た病気の遷移(因果関係の可視化)に迫る必要がある。経済分野の産業連関表のように医療費の視点を取り込んだ疾病のネットワーク表を作成できるようになることが望ましい。現在の研究では、レセプトの発生件数をベースに可視化しているが、できればレセプトの疾病ごとの金額をもとに可視化できると、医療経済面の視点から分析が可能になる。

どのような年齢層で、どの性別で、どのような疾患が発生するか、など疾病データの可視化研究に対する要望は尽きない。

また研究者の側には、市のデータを分析する以上、政策立案に役立つ成果を出すことが求められる。

参考文献

- [1]鎌形喜代美(2014)「保健事業推進のための国保データベース(KDB)システムの活用」保健医療科学 63(5),463-466.
[2]東京都保険者協議会医療費分析部会(2012)『医療費の分析とその活用-医療費適正化に向けて-』http://www.tokyo-kokuhoren.or.jp/tokyo_conference/insurance_conference/pdf/guid_iryoubunseki_katuyou.pdf
(最終訪問 2015/05/19)