

上野 恵子 助教

京都大学大学院 医学研究科 社会疫学分野 様

フィールドワークにデータ解析を 組み合わせ、社会疫学の実相に迫る

■ Theme (目的)

社会疫学における複雑な課題の理解と解決。問題や課題を鮮明化し、より実効性の高い「現場に還元できる施策立案」。

■ Point (内容・効果)

- フィールドワークによる質的研究と**Alkano**のデータ解析による量的研究を組み合わせた混合研究。
- **Alkano**の二項ソフトクラスタリング機能（末尾コラム参照）を活用したデータ解析の有効性。

Interview

人々の健康格差を社会疫学の視点を交えながら研究し、解決策を探求している京都大学大学院の上野恵子助教。「現場に還元できる施策提案」を常に心がける中で、新たに取り入れた**Alkano**によるデータ解析の手法やその研究事例について伺った。

福祉や医療の現場に還元できる施策を

ご研究にデータ解析を取り入れた経緯を教えてください。

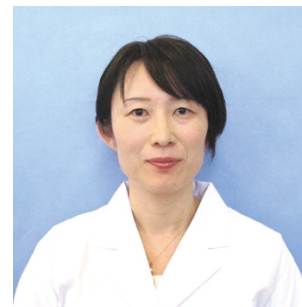
上野 社会疫学といって、学歴や所得、職業など個人や個人を取り巻く社会環境の違いで生じる健康格差に関する研究をしています。テーマは3つあり、1つめは生活保護受給者の健康・生活支援、2つめは救急車利用に関する問題、3つめは子どもと若者の社会的孤立・孤独の予防です。それぞれのテーマで対象となるデータを解析して状況や課題を整理し、その結果をもとに当事者の方々にお話を伺うことで現場に還元できる施策を提案しています。データ解析による量的研究とインタビューによる質的研究、そのどちらか一方だけでは探求できる範囲が限られてしまいます。ですから私は組み合わせることで、複雑な事象をより正しく理解し、課題に対応したいと考えています。

データ解析には**Alkano**の二項ソフトクラスタリング機能をお使いと聞きました。

上野 大学院時代に指導教官から、私たちの研究に役立つようなデータ解析手法としてPLSA（確率的潜在意味解析法）を教えてくださいました。当時、生活保護受給者に関する研究で対象者をグループ分けし、それぞれのグループごとの課題を見つけることができれば、福祉事務所が生活保護受給者に健康支援を提供する際に役立つのではないかというアイデアがあり、そのグループ分けの手法として指導教官とともに試してみたのが最初です。使い始めると、自分の研究スタイルに合っていると感じました。その後赴任したこの京都大学で、PLSAを改良した二項ソフトクラスタリング機能が実装されている**Alkano**を2022年に導入しました。

実際のクラスタリング事例を教えてください。

上野 私が学会発表した研究を例にご説明します。ある地方都市の5年間の約54,000件の救急搬送データをもとに、医療機関に救急搬送され受診後に帰宅した人たち（18～64歳の成人）を二項ソフトクラスタリングでグループ（セグメント）分けしました。その結果、セグメント1「怪我や神経系疾患で平日深夜に救急搬送された人たち」、同2「週末日中の外傷や火事による救急搬送困難事案*となった人たち」、同3「他の医療機関へ転院搬送された人たち」、同4「急病のため住宅から救急搬送された人たち」、同5「交通事故のため救急搬送された人たち」、同6「日中平日に労災、災害、精神疾患のために管轄外の医療機関に救急搬送



京都大学大学院
医学研究科 社会疫学分野
上野恵子 助教

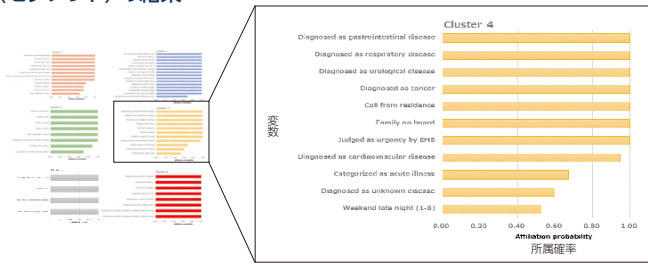
PROFILE

上野 恵子 助教

2003年、金沢大学医学部卒。2003～2004年横須賀米海軍病院インターン。2004～2009年、米国セントジョセフ・マーシー病院一般外科レジデント。国内救命救急センター勤務を経て、2018年、東京大学大学院 医学系研究科 公共健康医学専攻修了（専門職修士（公衆衛生学））。2022年、東京大学大学院 医学系研究科社会医学専攻修了（医学博士）。2022年より京都大学大学院 医学研究科社会疫学分野で研究に従事。

された人たち」の6つのセグメントに分けられました。
上記のセグメントの特徴は、セグメントに所属する変数をもとに決めます。セグメント4であれば、「消化器疾患」「住

医療機関を受診後帰宅となった成人の救急車利用者のグループ分け(セグメント)の結果



2023年4月「第12回アジア救急医学会総会(フィリピン)」発表資料を元に再構成

居からの救急要請」「急性疾患」などです。変数がセグメントに所属する度合いは所属確率といい、0から1の数値で表されます。今回の研究では、所属確率が0.5以上の変数をそのセグメントを特徴づける変数と定義しています。なお、所属確率が低い変数は他のセグメントにも所属している可能性があります。

この結果から、【セグメント4:急病のため住宅から救急搬送された人たち】で医療機関に救急搬送され受診後に帰宅した人は、「消化器系、呼吸器系、泌尿器系やがんといった疾患をもつ人」で、さらに「家族が救急車に同乗した」「週末・深夜の救急搬送」という情報もこのセグメントには含まれていることがわかります。

* 救急搬送困難事案:救急車が現場に到着してから出発までに要した時間が30分以上の症例を指す。

複雑な現象がこの解析で整理される。だから解決のヒントも浮かんでくる

この解析手法を使う意義は、どんなところにありますか。

上野 研究に向かい合うとき、私は「この現象はおそらくこうなっているのだろう」と仮説を立てます。仮説をデータから二項ソフトクラスタリングを活用し検証しているといえます。データに基づくグループ化によって、対象となる人たちが抱えているであろう問題やその原因が、より鮮明に見えてくるからです。

解析結果として出てくるグループは、変数や解析条件を変えれば違う結果になります。何回か試していくうちに自分の知識や経験に照らし合わせて、「そうだよな」と納得できる結果が得られます。そうして得られた結果は現実世界に近い状況を表現していて、他の研究者との意見交換や、対象者への支援策を検討するのに役立ちます。私が研究対象としている「社会」は答えのない世界です。多くの人がさまざまに考え動いている。そういう複雑な現状を研究者の視点から整理し、納得感のある形で表現するのに、この解析手法はとても有効だと思っています。

Alkanoを導入して良かった点は何か。

上野 Rでも二項ソフトクラスタリングと同様の解析ができると聞いて、試してみたことがあります。ただ私は統計解析の専門家ではないのでRを扱うのは難しく、二項ソフトクラスタリングの適切なコマンドを見つけて解析することができませんでした。**Alkano**を導入して良かったことは、NTTデータ数理システムのきめ細かなサポートを受けられたことです。ソフトウェアの使い方に始まり、データの整形やフォーマットといった初歩的なことから、解析の仕方まで分からないことへのアドバイスまでいただくことができました。そのおかげで納得いく結果を短時間で得られ、他の研究業務に集中することができました。

使い勝手の面では、データ入力の際に手持ちのExcelファイルのデータを使えるという簡単な操作が気に入っています。データの整形といった前処理が必要なく、データを登録後、アイコンで必要な処理を行うことができます。パラメーター設定もクリック操作で可能なのがいいですね。

決定木など、他の解析手法からのアプローチもお考えはどうですか。

上野 研究を通じて現場に還元できる施策を立てたいと、常にも思っています。そのためにはこれまでとは違うアプローチも必要ではないか、そうすれば課題解決への近道が見つかるのではないかと、という思いで二項ソフトクラスタリングを使い始めました。ですから他の解析手法にも関心があります。いま使ってみようと思っているのは**Alkano**の決定木です。決定木では、ある目的変数と関連する複数の条件を可視化できるので、例えば救急搬送されて重症と判断される人たちはどのような条件が揃っているのかといった検証ができると思っています。

さらに、**BayoLinkS**でベイジアンネットワークにもトライしたいと考えています。救急車利用者とその関連要因の因果関係を分かりやすく整理し、消防署の方々と結果をディスカッションすることで、持続的な救急医療体制の構築に向けた方策が見つかるのではないかと期待しています。

二項ソフトクラスタリング

一般的に「PLSA (確率的潜在意味解析法)」と呼ばれる解析手法。個人に対する行動・属性、といった二項 (ペア) のデータをクラスタリングして内容把握を行う際に特に有用である。NTTデータ数理システムではPLSAに独自の計算方法を採用しており、二項ソフトクラスタリングという機能名で**Alkano**に付加している。

個人ID付きPOSデータの二項ソフトクラスタリング計算例

顧客・商品ごとの購入点数行列に対して、顧客・商品を入れ替えて、同時に買われている組をクラスタとして抽出したものを。

