

死因究明におけるテキストマイニングの活用法に関する研究

Application of Text Mining for Death Analyses

谷藤 隆信† 山村 麻理子‡ 津田 和彦†
Takanobu TANIFUJI Mariko YAMAMURA Kazuhiko TSUDA

1. はじめに

近年、日本の死亡者数は年間約 100 万人で、その 1/6 に相当する約 15 万人が異状死である。異状死とは、未入院下での死亡など、死亡前に医師の管理下に置かれることなく死亡することを意味する。

異状死の場合、警察により検視と呼ばれる遺体調査と生前の調査が行われ通報書が作成される。その後、監察医制度のある地域は監察医により、ない地域は警察医により、遺体の調査が行われ最終的な死因を断定し検案調書が作成される。しかし、警察医の多くは町の臨床医であるため、死因調査の専門家でない。そのため、死因不明を表す“心不全”を用いる割合が、監察医制度のある地域は 15.3%、ない地域は 71.9%と大きな開きを生じている[3][4]。

この問題を改善するためには、警察医が死因を断定する際の援助となる情報が必要不可欠である。そこで本研究では、通報書と監察医が作成した検案調書の内容をマイニングすることで、その関連性を明らかにする。

2. 死因究明

監察医制度は、死体解剖保存法第8条で規定される死因を究明する制度である。しかし、日本で現在執行されているのは、わずかに東京 23 区など 5 都市のみである。この原因として、監察医制度の維持費用を地方負担とするなど、法制度の不備が原因とされる[1][2]。

このような状況により、多くの問題が生じている。例えば、1)先進諸国と比較して日本の死因統計が大きく偏っている点、2)91.8%の警察医が「死因の決定に何らかの不安がある」と回答している[8]など、警察医自身も死因調査に大きな負荷になっている点、3)埼玉、鳥取の連続不審死など、死因が明確になれば防げた可能性の高い事件の発生など、その影響は多方面に及んでいる。

しかし、監察医を育成するには多大な時間を要するため、早期の根本的な解決は困難である。それゆえ、警察医が決定する死因を監察医のレベルに近づけるためには、監察医が死因を決定する基準を明確化して警察医へ提供することは重要である。

3. 通報書と検案調書のマイニング

3.1 対象データ

データは、ベテラン監察医が担当した事例であり、病死の頻出な死因であることを抽出条件とした。期間は平

成 21 年。サンプリングは季節に影響されないように年間を通して行った。

データの基礎的属性を表 1 に示す。平均年齢は男性 72.3 歳、女性 81.4 歳である。

表 1. 対象データの基礎的属性

	男性	女性	計	
			件数	%
データ数	190	161	351	
死因内訳				
高血圧性心不全	16	14	30	8.5
急性心筋梗塞	20	10	30	8.5
虚血性心不全	21	14	35	10.0
急性大動脈解離	15	18	35	10.0
うっ血性心不全	19	17	36	10.0
クモ膜下出血	13	26	39	11.1
脳内出血	19	16	35	10.0
肺炎	20	17	37	10.5
アルコール性肝障害	33	3	36	10.3
老衰	14	26	40	11.4

3.2 テキストマイニング分析

通報書と検案調書をテキスト化したのち、数理システムの Text Mining Studio を用いて分析した。分析目的は、通報書（死体の生前情報）と検案調書（死体の外表情報）という 2 種類のテキストから、単語の分布の偏りを捉えて、死因ごとの特徴を抽出することである。

分析は、辞書登録→特徴語抽出→カテゴリ化→カテゴリ化による特徴語抽出の手順で行った。

特徴語抽出は、死因ごとに出現する単語の分布の偏りを捉えるものである。また、カテゴリ化による特徴語抽出は、単語や係り受けの意味的なまとまりを一つのカテゴリとしたうえで、死因ごとに出現するカテゴリの分布の偏りを捉えるものである。

3.2.1 辞書登録

警察官が記載する通報書は専門用語が少ないが、記載ミスがある。例えば、正しくは「大動脈解離」とするところを、実際にはない「大動脈瘤解離」と記載するなどである。監察医が記載する検案調書は専門用語が多く、ユーザ辞書への登録数が多い。また、医師の嗜好により記載する表現が異なるのが特徴である。例えば、「髄液が透明である」という意味を「髄液清澄」「髄液無色清明」「髄液透明」などと表記が分かれる。グルーピング辞書は、係り受けをひとつの単語として登録するものである。例えば、「心のう液検査（陰性）」というグルー

† 東京都監察医務院

‡ 筑波大学大学院ビジネス科学研究科

ピング語には「心のう」→「引く+できない」や「心臓血腫」→「ない」を登録した。

3.2.2 カテゴリ化

表記が異なる同義語と、係り受けで同義である表現をまとめて、意味的により大きな分類へ整理した。例えば、「救急搬送」というカテゴリには「119番」「119指令」という単語に加えて、係り受けの「意識+ない」→「通報」、「呼吸+ない」→「通報」、「救急車→呼ぶ」などをまとめた。

4 評価

4.1 基本情報

本評価に用いたテキストデータの基本情報を表2に示す。

表2. 基本情報

	総文数	述べ単語数	単語種別数
通報書	1,889	21,842	5,021
検案調書	4,534	17,638	6,300

4.2 辞書登録

辞書登録数を表3に示す。グルーピング辞書は、係り受けを登録したものである。また、辞書登録前後による名詞数とグルーピング語数の変化を表4に示す。

表3. 辞書登録数

	ユーザ辞書	類義語辞書	分割辞書	グルーピング辞書
通報書・検案調書	48	86	45	327

表4. 辞書登録前後の名詞数とグルーピング語数の変化

	辞書登録前		辞書登録後	
	名詞数	名詞数	グルーピング語数	
通報書	17,641	14,894	2,426	
検案調書	14,684	12,571	1,564	

4.3 特徴語抽出

通報書の分析結果の一部を表5に示す。指標値は、その属性について特徴的であると判断できる特徴具合を表すものである。補完類似度から計算される値で大きいほど特徴的である。

表5. 特徴語抽出の例

死因	特徴語	属性頻度	全体頻度	指標値
高血圧性心不全	高血圧	15	80	28.92
急性心筋梗塞	胸痛	7	12	21.4
虚血性心不全	入浴	10	28	24.73
急性大動脈解離	起床	6	22	13.68
うっ血性心不全	ベッド	11	55	15.84
クモ膜下出血	倒れる	18	91	26.17
脳内出血	いびき	7	19	17.22
肺炎	熱	8	14	20.96
アルコール性肝障害	酒	13	24	34.99
老衰	食事	21	86	35.11

4.4 カテゴリ化

カテゴリ数は、通報書と検案調書で47となった。代表的なカテゴリと、そこに含まれる単語と係り受けの数を括弧内に示す。

救急・警察通報(4)、救命処置(5)、意識低レベル(5)、体調不良(5)、心のう内穿刺(陽性)(2)などである。

4.5 カテゴリ化による特徴語抽出

検案調書の分析結果の一部を表6に示す。死因に特徴的な身体所見・病歴・背景などが浮かび上がる。また、穿刺など遺体検査の結果から、死因を導き出していることがわかる。なお、指標値は特徴語抽出と同様に計算される。

表6. カテゴリ化による特徴語抽出の例

死因	カテゴリ化による特徴語	属性頻度	全体頻度	指標値
高血圧性心不全	循環器系疾患	24	106	50.29
急性心筋梗塞	胸痛・拡散痛	8	19	22.30
虚血性心不全	後頭下穿刺(陰性)	15	101	21.82
急性大動脈解離	異液吸引(陽性)	17	23	17.00
うっ血性心不全	下肢浮腫(陽性)	14	71	20.21
クモ膜下出血	後頭下穿刺(陽性)	9	31	16.66
脳内出血	後頭下穿刺(陽性)	17	31	46.15
肺炎	脳梗塞	7	28	16.93
アルコール性肝障害	胸腔穿刺(陰性)	8	38	17.20
老衰	子	16	93	21.21

5 おわりに

監察医が死因を決定する基準を探索するために、実際の記録である通報書と検案調書をテキストマイニング分析した。生前情報から教科書的に死因を決定できる事例もあるが、全てがそうではない。生前情報が乏しいのが異状死の特徴だからである。分析結果から、2種類のテキストにある単語やカテゴリは死因により分布が偏っていた。これは、監察医が死因を決定するために一定の判断基準や経験則を持つことを意味する。

本報告により、死因決定に苦慮する警察医に対して、死因を導く判断基準を提供し、決定した死因の信頼性を向上させることができると考える。

今後は、対象とする死因分野を拡張し、分析作業を継続する予定である。

参考文献

- [1] 福永龍繁. 死亡診断・死体検案システムの現状と問題点. 科学 2004;1298-1304
- [2] 勝又義直. 日本における検視等の手続きの問題点. 現代医学 2004;51 巻-2:499-501
- [3] 福永龍繁, 上野易弘, 山本健二他. 内因性急死の調査研究-兵庫県監察医業務区域とその他の区域における異状死体の検案結果の比較(1986-88年). 神緑会学術誌 1990;6:16-21
- [4] 溝井康彦, 柳田純一, 佐藤喜宣他. 日本の死因調査の現状と問題点. 日本医事新報 1994;3639:28-34
- [5] 福永龍繁, 上野易弘, 中川加奈子他. 兵庫県下における異状死体の検案結果(1986年)-監察医業務区域とその他の区域の比較-. 日法医誌 1988;42(4.5):431-442
- [6] 早川陸, 武市尚子, 矢島大介他. 千葉県における警察嘱託医業務の実態調査. 千葉県医師会雑誌 2007;第59巻-2:34-38