

形式概念分析による交通事故データ分析について — 自転車事故の分析と形式概念分析のための属性抽出手法 —

○中村正樹 輿水樹 柴田大翔 小谷祥悟 榊原一紀 本吉達郎 星川圭介 (富山県立大学)

Traffic Accidents Analysis using Formal Concept Analysis : Bicycle Accidents and Improving Efficiency

*M. Nakamura, T. Tatsuki, D. Shibata, S. Kotani, K. Sakakibara, T. Motoyoshi and
K. Hoshikawa (Toyama Prefectural University)

概要— 情報通信技術の発展にともない、交通事故に関わる様々な情報がデータベース化されている。本研究では、過去に発生した交通事故データをもとに将来の交通事故を予測するため、形式概念分析に基づく交通事故データ分析手法を検討する。近年注目されている自転車事故に焦点をあてた分析を試みる。また、膨大なデータに形式概念分析を適用する際には、計算コストを考慮したデータの取捨選択が必要となる。交通事故データ分析のための効率的なデータ抽出手法について検討する。

キーワード: 交通事故予測, 形式概念分析, 自転車事故

1 はじめに

近年、交通人身事故の発生件数は全国的に減少傾向にあるが、今後の少子高齢化の状況を踏まえると高齢者や自転車事故などの交通弱者に関わる事故の防止が重要となってくる¹⁾。過去に発生した交通事故から得られるデータに基づく分析・予測に関する研究・開発は、近年活発に行われており、例えば、文献²⁾では、機械学習を用いた福岡市における交通事故予測手法が提案されている。我々は、先行研究^{3, 4)}において、富山県警察本部の協力のもと、過去数年に富山県内で発生した人身事故、物損事故の交通事故データの提供を受け、形式概念分析⁵⁾と呼ばれるデータ分析手法を用いて富山県内の交通事故を分析し、それに基づき機械学習モデルを作成し、死亡・重傷事故の発生予測手法を提案した。

本研究では、近年注目されている自転車に関わる交通事故に着目し、形式概念分析を用いた自転車事故の分析を試みる。文献⁶⁾では、アンケートに基づくヒヤリハット体験の分析から、交差点形状や自転車進行方向とヒヤリハット体験の関係が分析されている。文献⁷⁾では、交通事故データを分析することで、自転車が道路の左右どちら側を走行していたかを推測し、進行方向別の事故発生状況を分析している。本研究では、文献⁶⁾と同様に交差点形状と進行方向に着目し、過去に発生した交通事故データに基づき分析する。文献⁷⁾の方法で富山県内で発生した自転車交通事故を進行方向別に分類し、交通事故データと統合することで、形式概念分析の適用を試みる。

分析対象となる交通事故データは膨大であり、形式概念分析を直接適用すると、現実的な時間内で分析が終了しない。先行研究^{3, 4)}および本研究の自転車事故分析でも、交通事故データから主観的に事故分析に重要と考えられる属性を選択して、形式概念分析を適用している。形式概念分析の効率化のための手法は数多く提案されており、例えば文献⁸⁾では、分析対象の属性間に複数の距離を定義し、それらに基づきデータを分類し、それぞれの形式概念分析の結果を比較している。本研究では、文献⁸⁾と同様に複数の距離によるデータ

の分類を実データである交通事故データに適用し、事故予測に適したデータ抽出手法を検討する。

2 準備

2.1 交通事故データ

本研究では、富山県警察本部より提供された平成29年から令和3年の交通事故データを用いる(図1)。

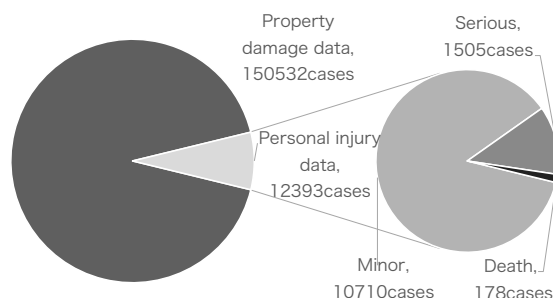


Fig. 1: Accident type and data size

162,925 件の交通事故は、人が損傷していない物損事故 150,532 件と人身事故 12,393 件に分けられる。人身事故には、重傷事故 1,505 件、死亡事故 178 件が含まれる。各事故データには事故の形態を示す事故類型や事故が発生した時間帯を示す昼夜などの事故内容が記されている。表1に人身事故データの一部を示す。実際のデータでは、100種類ほどの事故内容が数値と日本語で記されている。

表1における事故番号1の事故を例に挙げると、この事故は曇(天候:2)かつ昼(昼夜:12)のときに進行してきた車両が信号待ちなどで駐停車している前車の後部に衝突した事故(事故類型:23)であり、第一当事者には具体的危険がないとして注視を怠ったという人的要因(人的要因1:132)があるという事故である。このように、各事故内容を読み取ることで事故の状況を確認できる。ここで「第一当事者」とは、最初に交通事故に関与した車両等(列車を含む)の運転者又は歩行者のうち、当該交通事故における過失が重い者をいい、過失が同程度の場合には人身損傷程度が軽い者をいう。

Table 1: Personal injury data

Accident number	Accident type	Time	Weather	Human factor 1	...	Latitude	Longitude
1	23	12	2	132	...	xx.xx	yyy.yyy
2	46	12	2	114	...	xx.xx	yyy.yyy
3	23	12	2	107	...	xx.xx	yyy.yyy
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12,393	23	22	2	132	...	xx.xx	yyy.yyy

2.2 形式概念分析 (Formal Concept Analysis)

形式概念分析とは、束論に基づいて概念構造を分析するデータ分析手法である⁵⁾。自然現象などに出現する対象（オブジェクト）やオブジェクトが内包する性質（属性）の集合の対をコンテキスト表として記述し、表に存在する関係を形式概念として整理する。形式概念とは、共通の属性集合を持つオブジェクト集合である外延と、外延集合が共通に持つ属性集合である内包の組のことを指す。例えば、自然数のデータにおいて、集合の外延 $\{3, 5, 7\}$ と内包 $\{x \leq 10, \text{prime}(x), \text{odd}(x)\}$ (10以下の素数かつ奇数) の対が形式概念の例である。この形式概念をノードとしたコンセプトラティスという図や、コンセプトラティスをもとにした含意関係や関連というルールが抽出され、この結果をもとに属性間の関係を確認する。本研究では、形式概念分析ツールとして Concept Explorer⁴を使用する。

表2は生物とその特徴を表すコンテキスト表の例である。例えば、イヌのオブジェクトは四足・陸上・胎生・哺乳類の属性を持つことが読み取れる。

Table 2: A context table of animals

	Two-legs	Four-legs	Fil-let	Land	Water	Sky	Vivi-parous	Ovi-parous	Mam-mal
Dog		x		x			x		x
Cat		x		x			x		x
Crow	x					x		x	
Chicken	x			x				x	
Frog		x		x	x			x	
Killifish			x		x			x	
Platypus		x	x	x	x			x	x
Japanese giant salamander		x	x	x	x			x	
Whale			x		x		x		x

図2は表2から生成されたコンセプトラティスであり、含意関係のみを図示する。コンセプトラティスにおいてオブジェクトは上方に、属性は下方に向かって追加されるため、各ノードは追加されるオブジェクトや属性のみを表記する。形式概念の例として、イヌ・ネコ・クジラのオブジェクトと胎生・哺乳類の属性を持つ形式概念が存在する。

含意関係の例として、哺乳類と記述されたノードの下にはイヌ・ネコ・クジラ・カモノハシのオブジェクトが表示される。また、胎生と記述されたノードの下には哺乳類かつ胎生であるイヌ・ネコ・クジラのオブジェクトのみが表示される。そのため、「胎生ならば哺乳類」という含意関係が成り立ち、カモノハシというオブジェクトが原因で「哺乳類ならば胎生」という含意関係が成り立たないことが確認できる。

表3は抽出された含意関係や関連の例である。例として、ルール1は「胎生ならば哺乳類である」という含意関係、ルール4は「哺乳類である4種類の生物のうち、3種類が胎生である（確信度75%）」という関連である。

⁴Concept Explorer, <http://conexp.sourceforge.net>

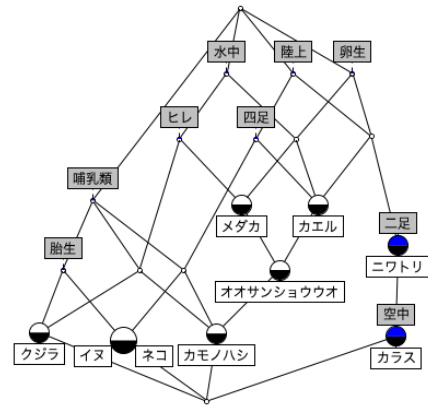


Fig. 2: The concept lattice of the animals

Table 3: Examples of implications and association rules of animals

Number	Prerequisite	Confidence	Conclusion
1	Viviparous	100% (3/3)	Mammal
2	Land, Water	100% (3/3)	Four-legs, Oviparous
3	Water	80% (4/5)	Oviparous
4	Mammal	75% (3/4)	Viviparous
⋮	⋮	⋮	⋮

3 自転車事故分析

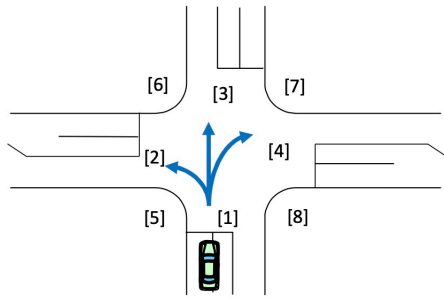
本節では、近年注目されている自転車に関連する交通事故の事故要因の解明のため、自転車に関わる人身事故に対象をしばって形式概念分析を試みる。

3.1 コンテキスト表の作成

交通事故データを形式概念分析へ適用するためにデータの2値化を行い、コンテキスト表を作成する。本研究で対象とする事故オブジェクトは、交差点で発生した自動車対自転車の人身事故である。事故件数は、平成29年から令和2年の689件である。形式概念分析は、属性数が増えることで計算時間が指数関数的に増える。そのため本分析では、表4の47属性のみを用いる。表中の括弧内の数字は属性の種類数である（例えば、天候は晴、曇、雨、雪の4種）。文献⁶⁾では交差点形状と自転車事故の関連が示唆されていたため、本研究では信号機、車道幅員、中央分離帯、歩車道区分の属性に着目する。車道幅員は、第一当事者側が大（13.0m以上）、第二当事者側が小（5.4m以下）のような3×3属性である。自転車進行方向は、文献⁷⁾に方法にしたがって、交通事故データから推測する。元の交通事故データでは、交差点を図3のように番号付けし、第一当事者の進行方向が左折の場合[1][2]と記録される。対向から直進してきた第二当事者との事故の場合、第二当事者の進行方向[3][1]と記録される。この際、第一当事者が自動車、第二当事者が自転車の場合、事故（接触）が起きていることから、（自動車が逆走していない限り）自転車は右側通行していたと推察される。このような推察から自転車進行方向（左右）の属性をデータに追加する。なお、直進同士など、組み合わせによって自転車進行方向が不明な場合も存在する。

Table 4: Selected forty-seven attributes

重傷 (1), 天候 (4), 路面状態 (2)
信号機 (6), 車道幅員 (9), 中央分離帯 (4), 歩車道区分 (4)
環境要因 (2), 自転車進行方向 (2), 当事者 2(3)
日 (2), 時間 (4), 季節 (4)

Fig. 3: Estimation of bicycle directions ⁷⁾

3.2 形式概念分析の適用

前節のコンテキスト表で、信頼度 80%以上、すなわち前提部の条件を満たす事故のうち 80%以上の事故が結論部の条件を満たすアソシエーションルールを形式概念分析ツールで出力した結果、23,557 件のルールが出力された。そのうち、結論部の属性を自転車通行方向: 右, あるいは, 自転車通行方向: 左を含むルールを抽出した。その結果それぞれ 154 件, 455 件のルールが得られた。得られたルールの一部を示す。

$$\langle 7 \rangle \left(\begin{array}{l} \text{路面状態: 乾燥} \\ \wedge \text{信号: 点灯 (3 灯式)} \\ \wedge \text{幅員: 中・大} \\ \wedge \text{中央分離帯: なし} \\ \wedge \text{歩車道区分: 緑石} \\ \wedge \text{日: 平日} \end{array} \right) \Rightarrow \langle 6 \rangle (\text{左側通行})$$

このルールは、左辺の 6 属性をすべて満たす事故 7 件のうち、6 件 (86%) が右辺の属性 (左側通行の自転車) を満たすことを表す。

図 4, 5 に、結論部が左側通行および右側通行のルール的前提部に出現する属性の出現数を示す。左右の進行方向における違いに注目すると、左側通行のルールには中央分離帯なしを前提部を持つルールが多く含まれるのに対し、右側通行のルールには中央分離帯や中央線 (ペイント) を前提部を持つルールが多く含まれることがわかる。

3.3 形式概念分析に基づく事故要因分析

形式概念分析は、入力データにおける注目すべき属性間に関する関係に対する気づきを与えるツールとみなすことができる。3.2 節の結果は、交差点における自転車事故において、中央分離帯と自転車進行方向に何らかの関係があるのではないかと示唆を与えるが、その裏付けのために実際のデータを分析する必要がある。

表 5, 図 6, 7 は、交通事故データにおける進行方向と中央分離帯の有無, 中央線 (ペイント) の関係である。自転車事故全体では左側通行に対して右側通行の割合が高いのに対して、中央分離帯なしの自転車事故では、半数近くが進行方向不明ではあるが、不明を除くと左側通行に比べて右側通行の割合が低くなっている。一方で、中央分離帯および中央分離帯 (ペイント)

Table 5: Relationship between bicycle direction and central reservation

	中央分離帯なし	中央分離帯, 中央線 (ペイント)
左側通行	79	167
右側通行	93	96
不明	165	78

の自転車事故の場合、右側通行の割合が全体の半数近くまで高くなっている。この結果から、例えば、自転車の進行方向に着目した交通パトロールや交通安全マップの作成など、交通事故防止に向けた取り組みへの支援が考えられる。自転車の進行方向は、例えば、高等学校の付近では朝と夕方の登下校時に自転車の進行方向が予測できるため、時間帯をしぼることで交差点ごとにどちらの進行方向の自転車がよくなるか予測可能である。

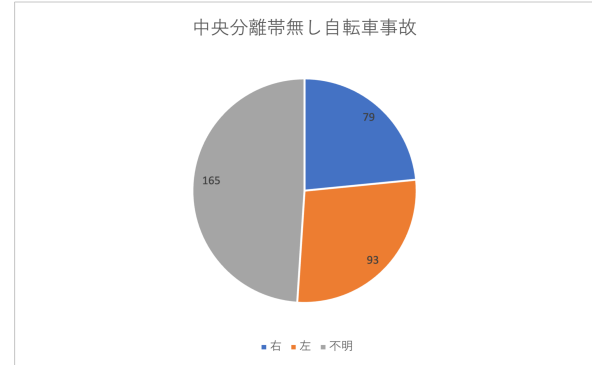


Fig. 6: Relationship between bicycle directions without central reservations

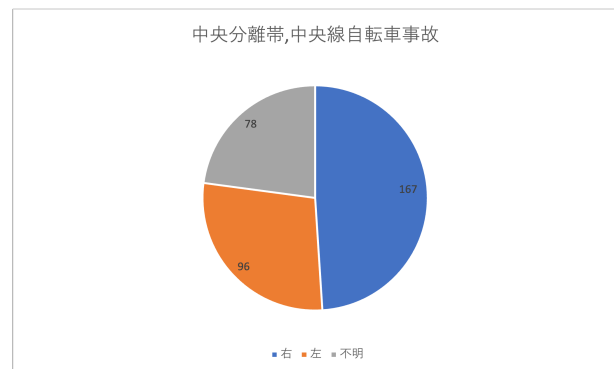


Fig. 7: Relationship between bicycle directions with central reservations or painted center lines

4 形式概念分析のための属性抽出

交通事故データの属性を形式概念分析のコンテキスト表のために二値化すると、属性数は 1,821 となる。このような多数の属性を含むコンテキスト表に対する形式概念分析は、計算コストが膨大になり、現実的な時間内に計算が終了しないか、メモリ不足で分析ができない。図 8 は、交通事故データからオブジェクトをランダムで 1,000 件、属性をランダムに 10 から 200 属性抽出したコンテキスト表に対する形式概念分析の計算時間である。計算時間が指数関数的に増加するのが確認できる¹⁾。

先行研究^{3, 4)} および前節の自転車事故分析では、交通事故データから主観的に事故分析に重要と考えられる属性を選択している。そのため、選択されなかった属性が関係する事故要因が見逃されている可能性がある。

本節では、交通事故データの形式概念分析のための属性抽出方法を検討する。文献⁸⁾では、分析対象の属性間に複数の距離を定義し、それらに基づき属性を階

¹⁾実行環境: CPU 3.2Ghz Intel Xeon W, メモリ 96GB

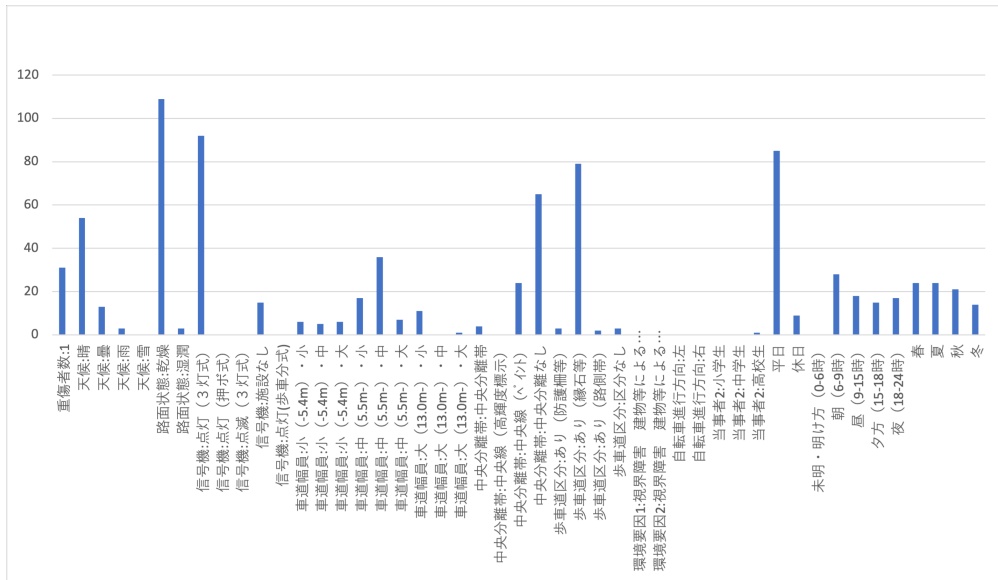


Fig. 4: Number of attributes in premises of rules with left direction in consequent

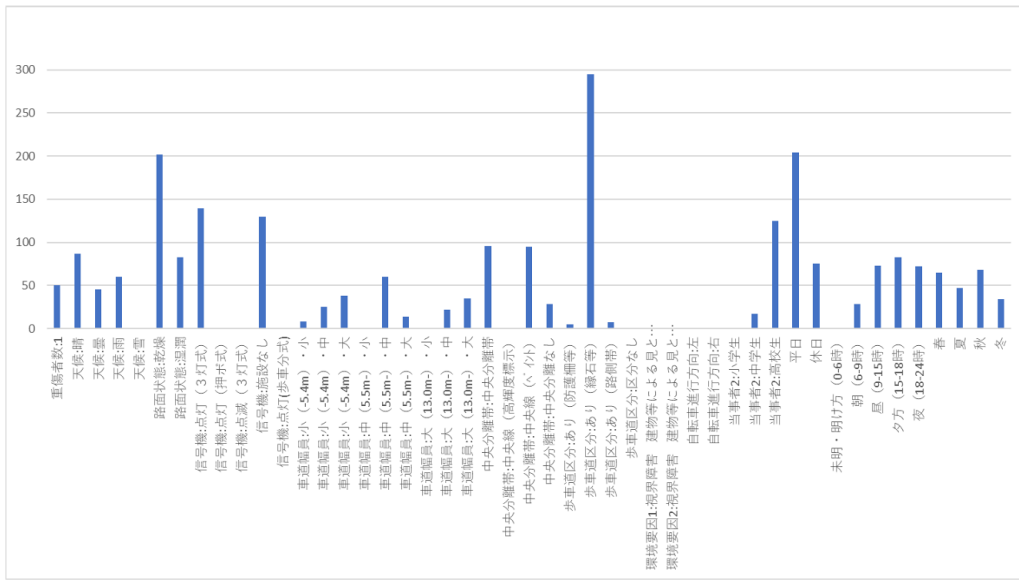


Fig. 5: Number of attributes in premises of rules with right direction in consequent

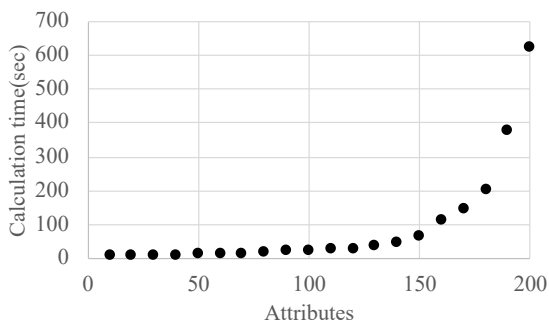


Fig. 8: Calculation time of FCA

層型クラスタリングで分類し、それぞれの分類に対する形式概念分析について考察している。階層型クラスタリングの長所の1つは、目的に応じて柔軟に分類数を決められる点であり、本研究の適切な属性数の分類を得る目的に適している。そこで、本研究では、文献⁸⁾と同様に、複数の距離に基づくデータ分類を、実デー

タである交通事故データに適用し、事故予測に適した属性抽出手法を検討する。

4.1 属性抽出手法

属性間の距離として、文献⁸⁾で用いられた関連度(標本相関係数の絶対値)に加え、ユークリッド距離、マンハッタン距離、コサイン類似度について検討する。提案手法では、各属性同士の距離を算出して、その距離から距離行列を作成し、階層型クラスタリングを用いて属性を分割する。階層型クラスタリングでは、データのグループ間の距離も必要となる。グループ間の距離として、最長距離法、最短距離法、ワード法、群平均法、重心法、メディアン法を検討する。

提案手法では、まず、分析対象となる交通事故データに対して、形式概念分析を適用するためにコンテキスト表を作成する。次に、分析対象となる属性を1つ決める。各属性同士の距離を求めて、距離行列を作成する。その際、元データで1つの属性だったコンテキスト表の属性群については、距離を0とする。例えば、

天候の5属性(晴れ,曇り,雨,雪,霧)の互いの距離は0とする.階層型クラスタリングを適用し,形式概念分析が適用可能な属性数になるまで分割する.本手法では,得られた分類それぞれに対し,最初に定めた分析対象の属性の列を加えたコンテキスト表に対して,形式概念分析を適用する.

4.2 距離の検討

交通事故データ分析において,複数のデータおよびグループ間の距離のどの組み合わせが適切か検討するため,先行研究³⁾で用いられた10,422オブジェクト,表6の82属性のコンテキスト表に対して,提案手法を適用する.

Table 6: Selected eighty-two attributes

昼夜 (6), 天候 (5), 路面状態 (5)
信号機 (7), 地形 (3), 事故類型 (33)
年齢 1(6), 年齢 2(7), 人身損傷程度 1(7), 人身損傷程度 2(7)

階層型クラスタリングの結果から,分類が適切にされているかどうかを確認する.図9は,クラスタリングの結果の一部の樹形図である.図9(上)のように樹形図が階段状にある現象は鎖効果と呼ばれ,複数属性を含む分類が得られにくい.図9(下)のようなバランスのよい分類が望ましい.4.1節の距離に対する階層型クラスタリングの結果を確認したところ,複数属性を含む分類が2つ以上あり,階層の反転が起きていない組み合わせは,相対係数の絶対値-群平均法,ユークリッド距離-最長距離法,ユークリッド距離-ワード法,マンハッタン距離-最長距離法,マンハッタン距離-ワード法,cos類似度-最短距離法,cos類似度-群平均法の7つの組み合わせであった.

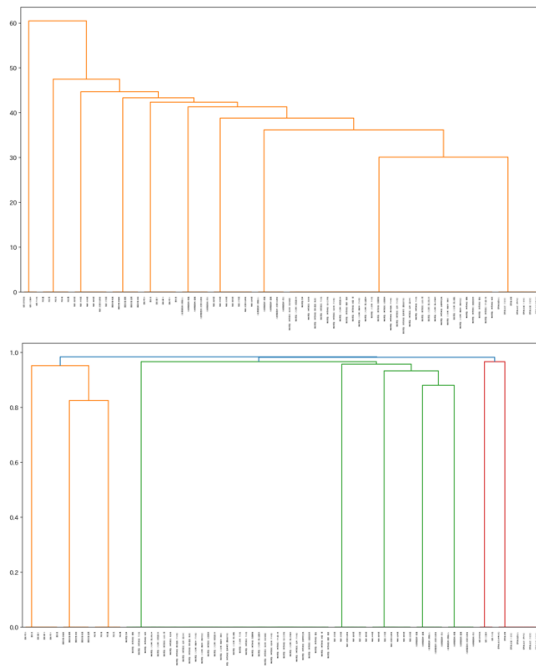


Fig. 9: Results of classification

先行研究⁴⁾では,形式概念分析で取得されたルールの前提部の属性を用いて,結論部の属性を分類する機械学習モデルを作成する交通事故予測手法が提案されている.そこで,本手法の各距離の評価として,形式概念分析の結果のルールのうちで,分析対象の属性を結

論部に持つルールを抽出し,それらの前提部に現れる属性を説明変数とし,結論部を目的変数とするニューラルネットワークによる機械学習モデルを作成する.平成29年から令和2年のデータを教師とする.作成された機械学習モデルに対し,令和3年のデータで検証する.また,比較対象として全属性を説明変数とした機械学習モデルも作成する.ニューラルネットワークの各種パラメータは先行研究⁴⁾に従い,詳細は省略する.分析対象の属性を死亡・重傷事故および高齢者とし,それぞれ5回作成した機械学習モデルの検証結果におけるF1 Score(適合率と再現率の調和平均)の分布を図10に示す.横軸に距離の組み合わせ,縦軸にF1 Scoreを取る.横軸に現れない組み合わせは,結論部に分析対象が現れるルールが存在しなかった分類である.死亡・重傷事故の予測(図10(上))では,cos類似度以外についてF1 Scoreが30前後となった.高齢者事故の予測(図10(下))では,ユークリッド距離(マンハッタン距離)-最長距離法のF1 Scoreが良い結果であった.

4.3 提案手法の適用

提案手法による全交通事故データ(659属性,10422オブジェクト)の分類を試みる.分析対象を死亡・重傷事故として,ユークリッド距離-最長距離法を用いた階層型クラスタリングを適用すると,表7の分類が得られる.

Table 7: 属性のグループ分けの結果

分類1	分類2	分類3	分類4	分類5	分類6	分類7
指定速度 1(7)	指定速度 2(7)	署コード (16)	人的要因 1(8)	人的要因 2(9)	当事者 1(14)	市町村 (15)
天候 (4)	車道幅員 (16)	月 (12)	加害部位 2(16)	地形 (3)	当事者 2(21)	認知速度 2(10)
性別 2(3)	違反 1(27)	曜日 (7)	—	—	事故類型 (28)	—
道路形状 (7)	—	通行目的 1(17)	—	—	行動類型 1(17)	—
—	—	通行目的 2(18)	—	—	中央分離帯 (7)	—
—	—	認知速度 1(18)	—	—	—	—
—	—	年齢 1(6)	—	—	—	—
—	—	年齢 2(7)	—	—	—	—

上記の各分類に形式概念分析を適用した結果,信頼度80%以上のルール351,943件が得られた.一方で,分類4や5のような属性数が少ない分類からは信頼度80%以上のルールを抽出できなかった.一例として,「中央分離帯のない道路で等速で直進中に発生した自転車単独事故11件は必ず死亡・重傷事故」というルールが得られた.全データを統計分析したところ,自転車単独の死亡・重傷事故は,死亡・重傷事故全体と比較して,中央分離帯のない道路を等速で直進中に発生する割合が高いことが確かめられた.全データに対する提案手法の適用により,死亡・重傷事故に対する複数属性が関係する事故要因の分析が可能なが確かめられた.

5 まとめ

過去に発生した交通事故のデータに対する形式概念分析による交通事故分析を検討した.自転車事故の分析では,交通事故データから自転車進行方向などの情報を推察し,得られたデータセットに対する形式概念分析の適用結果から,自転車進行方向と事故の関係を分析することで,中央分離帯の種類と自転車進行方向に関する傾向を分析した.

形式概念分析のための属性抽出手法として,属性および属性分類間の距離に基づく階層型クラスタリングによる属性分類手法を提案した.一例であるが,小さなデータセットに対する死亡・重傷事故および高齢者事故の分析に提案手法を適用し,ユークリッド距離(マンハッタン距離)-最長距離法による分類が,機械学習に



Fig. 10: Results of machine learning

基づく交通事故予測に効果的であるという結論を得た。

今後は、交通事故予測システムの作成に向けて、自転車事故を含む様々な交通事故に対し、提案した属性抽出手法を適用し、効果を確認する。

謝辞

本研究に取り組むにあたって富山県警察本部に貴重なデータを御提供頂くとともに、有益な議論および御提案を頂きました。心より感謝申し上げます。本研究の一部は、公益財団法人富山第一銀行奨学財団の令和4年度研究助成金を受けて行ったものです。

参考文献

- 富山県警察本部, 令和2年版交通事故白書及び交通死亡事故発生マップ, <https://police.pref.toyama.jp/6115/toukei/koutsuujiko/jikohakusho/kj202104.html>, (参照:Jan.18, 2023)
- 鳥居和史, 水野由裕, 遠山一則, 清水茂樹, 向後颯太: 構造化データと衛星画像のマルチモーダル AI モデルによる交通事故発生確率の予測手法の開発, 自動車技術会論文集, **53-2**, 404/409 (2022)
- 小谷祥悟, 浅沼佑汰, 中村正樹, 榎原一紀, 本吉達郎, 星川圭介: 交通事故データに対する形式概念分析による事故予測システムの構築の検討, 信学技報, MSS2021-56, **121-443**, 7/12 (2022)
- 小谷祥悟, 中村正樹, 榎原一紀, 本吉達郎, 星川圭介: 形式概念分析と機械学習による交通事故予測システムの検討, 計測自動制御学会 システム・情報部門学術講演会 2022, SS04-02, 364/369 (2022)
- 鈴木治, 室伏俊明: 形式概念分析—入門・支援ソフト・応用一, 知能と情報, **19-2**, 103/142 (2007)
- 草野優太, 亀谷友紀, 山中英生: 高齢者・高校生の自転車による交差点ヒヤリハットの比較分析, 土木学会論文集 D3 (土木計画学), **67-5**, 1251/1257 (2011)
- 萩田賢司, 横関俊也: 自転車走行空間における自転車通行方向別の交通量と事故特性の比較, 土木学会論文集 D3 (土木計画学) **75-5**, 493/506 (2019)
- 深谷有吾, 石樽隼人, 武藤敦子, 森山甲一, 犬塚信博: 属性間の関連度を用いた分解による概念束の単純化, 情報処理学会論文誌数理モデル化と応用 (TOM), **12-3**, 18/25 (2019)