

事故データベースを活用した高速道路における潜在的事故要因の抽出手法に関する基礎的検討

高田 翔太¹・澤田 英郎²・安 時亨³・大石 和弘⁴・米村 圭一郎⁵

¹非会員 西日本高速道路エンジニアリング関西 (株) (〒567-0032 大阪府茨木市西駅前町5番26号)
E-mail:sho_takada@w-e-kansai.co.jp

²正会員 西日本高速道路エンジニアリング関西 (株) (〒567-0032 大阪府茨木市西駅前町5番26号)
E-mail:h_sawada@w-e-kansai.co.jp

³正会員 西日本高速道路エンジニアリング関西 (株) (〒567-0032 大阪府茨木市西駅前町5番26号)
E-mail:s_an@w-e-kansai.co.jp

⁴非会員 西日本高速道路 (株) (〒530-0003 大阪市北区堂島1-6-20 堂島アバンザ18F)
E-mail:k.oishi.aa@w-nexco.co.jp

⁵正会員 片平エンジニアリング (株) (〒112-0002 東京都文京区小石川2-22-2)
E-mail:yonemura@katahira.co.jp

交通事故調書に交通状況や道路構造などの情報を付加した事故データベースを活用した事故分析が注目されている。本稿では、西日本高速道路・関西支社管内で発生した事故について事故データベースを構築し、従来の事故分析では見落とされてきた潜在的な事故要因の抽出についての基礎的検討を行う。データマイニング手法の一つであるアソシエーション分析を活用した分析の結果、従来から良く知られている知見が客観的なデータによって確認されるとともに、これまで認識が低かった事故要因が新たに見出される可能性が示唆された。

Key Words : *accident analysis, accident database, data mining, association analysis*

1. はじめに

NEXCO西日本では、従来の安全対策の知見に新たな視点からの対策を加えたものを「新交通安全対策」と位置付け、「交通安全対策アクションプラン」を策定した。その取り組みの一つに、事故データベース（以下、事故DBと略記する）を活用した事故要因分析がある。これは、交通事故調書に交通状況や道路構造などの情報を付加した事故DBを構築し、従来とは異なる視点から事故要因の抽出を試みるものである。

従来の事故分析では、事故件数や事故率に基づき抽出した事故多発箇所について詳細な要因分析を行い、効果的と思われる対策工を立案してきた。しかしこの方法では、事故多発箇所ほど事故は多くないが何らかの要因により事故発生の危険が高い箇所、すなわち潜在的な事故要因が存在する箇所を抽出することはできない。潜在的な事故要因を抽出するためには、事故件数や事故率のみならず、事故と道路条件、交通条件、環境条件との関連の強さに着目し、そうした各種条件の変化や特徴から事

故発生の危険性を導くような分析が必要である。

近年、事故と各種条件との関連性を把握する方法として、ビッグデータの解析手法であるデータマイニング¹⁾を活用して事故DBを分析する試みが注目されている²⁾³⁾。データマイニング手法の一つにアソシエーション分析があり、データベースに含まれる全項目間の関連の強さを計算することで、分析者の主観にとらわれることなく事故と各種条件の関連性を客観的に判断することができる。飯田ら³⁾は、阪神高速道路での事故を対象に、アソシエーション分析と階層グラフ構造を用いて、事故と関連性の強い当事者属性や道路構造、交通状況等を抽出する手法を提案している。しかしこの手法は、① 分析の結論部を衝突事故や追突事故といった特定の事故に固定するクラスアソシエーション分析を採用している点、② 事故と各条件との関連性の強さを判断する指標として各条件下での事故発生割合（confidence, 次章にて詳述する）を用いる点、において、事故の形態や事故の発生件数に係らず様々な事故と各種条件を全網羅的に分析できる手法とは言えない。

筆者らは、最終的に潜在的な事故要因を抽出することを目的とするため、結論部をあらかじめ固定しないアソシエーション分析を採用し、NEXCO西日本・関西支社管内で発生した事故を対象に、全事故形態と各種条件を総当たりで分析することとした。本稿は、これら分析方法の概要と結果について報告するものである。

2. 事故データベースの概要

NEXCO西日本・関西支社管内において平成18年～23年までの6年間に発生した20,530件の事故（飛来物・落下物事故を除く本線事故、NEXCO調べ）について、「事故調書データ」に「その他関連データ」を関連付けた事故DBを作成した。なお、その他関連データは事故調書データに記載されている事故発生日時・地点に基づいて関連付けを行っている。また、事故調書データ、その他関連データの中から重複する項目や有効データの少ない項目を削除し、最終的に36項目に集約した。（表-1）

また、事故DBには名義尺度・比例尺度データが混在しているが、アソシエーション分析を実施するためにはこれらをすべて名義尺度に変換する必要がある。そこで、各データの度数分布や交通工学的視点に基づいて名義尺度への変換を行った。

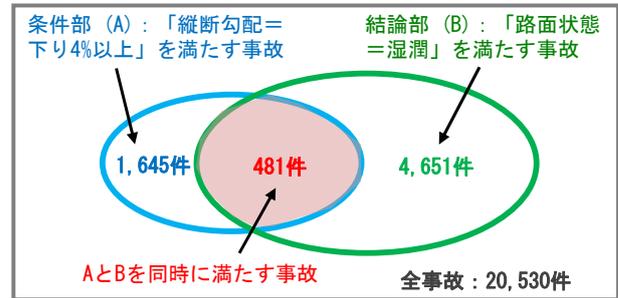
表-1 事故データベースの構成要素

＜事故データベース＞		
事故調書データ		その他関連データ
1 年次	道路構造データ	20 対面通行
2 月		21 道路構造
3 通常期／繁忙期		22 車線数
4 平日／休日		23 曲線半径
5 路面状態		24 縦断勾配
6 明暗	舗装データ	25 舗装種別
7 事故対象物		26 高機能舗装経過年
8 事故類型		27 わだち掘れ量
9 死傷の有無	交通データ	28 年平均日交通量
10 第1当事者車種		29 年平均大型車混入率
11 第1当事者速度		30 事故時交通量
12 第1当事者車線		31 事故時大型車
13 第1当事者行動		32 事故時速度
14 違反種別		33 渋滞状況
15 意図・原因	34 渋滞巻き込まれ時間	
16 行動操作	その他データ	35 走行時間
17 利用頻度		36 15分間雨量
18 性別		
19 年齢		

3. アソシエーション分析の概要

アソシエーション分析は、データベースに含まれる項目間の関係を「『項目x = a』のとき、『項目y = b』の割合が高い」といったルール（条件の組合せ）の形で抽出する手法である。ここで、「項目x = a」の部分で「条件部」、 「項目y = b」を「結論部」とよび、「A（条件部） ⇒ B（結論部）」の形で表現する。アソシエーション分析により計算される指標はsupport（支持度）、confidence（確信度）、lift（改善度）の三つである（図1）。分析の際には、これらの指標について分析者が閾値の条件を指定し、有用なルールを抽出する。

部） ⇒ B（結論部）」の形で表現する。アソシエーション分析により計算される指標はsupport（支持度）、confidence（確信度）、lift（改善度）の三つである（図1）。分析の際には、これらの指標について分析者が閾値の条件を指定し、有用なルールを抽出する。



$$\text{support}(A \Rightarrow B) = \frac{A \text{ と } B \text{ を同時に満たす事故}}{\text{全事故}} = \frac{481}{20530} = 0.023$$

⇒ 「4%以上の下り勾配における湿潤事故」は全事故の2.3%

$$\text{confidence}(A \Rightarrow B) = \frac{A \text{ と } B \text{ を同時に満たす事故}}{A \text{ を満たす事故}} = \frac{481}{1645} = 0.29$$

⇒ 4%以上の下り勾配における事故のうち、湿潤事故は29%

$$\text{lift}(A \Rightarrow B) = \frac{\text{confidence}(A \Rightarrow B)}{\text{support}(B)} = \frac{481}{1645} \div \frac{4651}{20530} = 1.29$$

⇒ 「4%以上の下り勾配での事故に占める湿潤事故の割合」は、「全事故に占める湿潤事故の割合」より1.29倍高い

図-1 アソシエーション分析より算出される3指標のイメージ

4. 分析の概要

(1) 分析の視点

既往研究³⁾では、アソシエーション分析の三つの指標のうち、「ある条件（条件部）の中で特定の事故（結論部）がどの程度含まれるか」を表すconfidenceに着目し、supportやliftの下限値に一定の条件を設けたうえでconfidenceが高いルールを抽出している。

しかし、本稿の目的に照らし合わせると、「ある条件部が課されたときに特定の結論部の割合がどの程度高まるか」を表すliftに着目するのが望ましいと考えられる。すなわち、事故全体から見ればその割合は高くないが、ある条件（条件部）が課されることで割合が高まるような形態の事故（結論部）に着目することで、潜在的な事故要因を抽出できると期待される。

(2) 分析の方法

上記を踏まえ、本稿における分析の方法を図2のように定めた。なお、条件部に含まれる項目が多すぎると結果の解釈・分析が煩雑になることに加え、supportの値が小さくなりすぎる（条件部と結論部を同時に満たす事故が少なくなりすぎる）ため、最大3項目までとした。

本分析によって潜在的な事故要因を抽出するイメージを図3に示す。図中の例では、「路面状態＝湿潤」および「違反種別＝操作不適」という2条件を「4%以上の上

り勾配における事故」と関連が強い要因とみなし (④), 「路面状態=湿潤」を潜在的な事故要因と判断 (⑤) している。

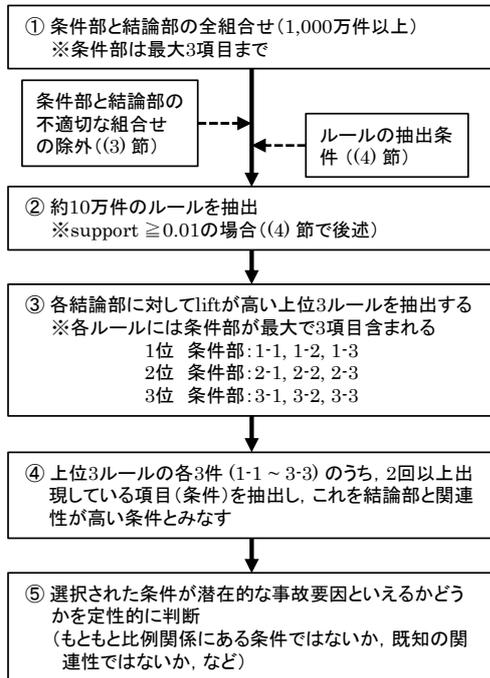


図-2 分析の流れ

結論部	事故件数	support	confidence	lift	条件部		
					項目1	項目2	項目3
縦断勾配=上り4%以上	234	0.01	0.15	3.25	路面湿潤	操作不適	高機能舗装
	241	0.01	0.13	2.87	路面湿潤	操作不適	死傷者なし
	284	0.01	0.13	2.85	路面湿潤	操作不適	

図-3 潜在的な事故要因の抽出 (手順③~⑤) イメージ

(3) 条件部と結論部の不適切な組合せの除外

2章で構築した事故DBには, 「15分間雨量」と「路面状態」, 「渋滞巻き込まれ時間」と「渋滞状況」など, 内容が重複していると考えられる項目が多く存在している。こうした項目が存在すると, 例えば「『15分間雨量=0mm』⇒『路面状態=乾燥』」といった極めて自明なルールが上位3ルールとして抽出されてしまう。そこで, 条件部または結論部に含まれる項目としてふさわしくないものは分析の対象外とした。

また, 条件部および結論部の項目に「その他」, 「不明」等を含むルールは解釈が困難なため, これらについても分析対象から除外した。

(4) ルールの抽出条件

ルール抽出の条件として表2に示す3条件を設定した。条件3で0.1と0.01の2条件を設定したのは,

- ① 事故発生割合10%以上の比較的発生しやすい形態の事故 (事故件数 ≥ 2,053件)
- ② 事故発生割合1%程度の比較的発生しにくい形態の事故 (事故件数 ≥ 205件)

のそれぞれについて考察を行うためである。

表-2 ルールの抽出条件

条件	抽出条件
条件1	liftの下限値は1.2とする 理由: 割合の増加が一定 (20%) 以上のルールのみを抽出
条件2	confidenceの下限値は0.1とする 理由: confidence (条件部に占める結論部の事故の割合) が小さすぎる場合, 「条件部が結論部の主要な要因である」とは考えにくい
条件3	supportの下限値は0.1または0.01とする 理由: support (条件部と結論部を同時に満たす事故の割合) が小さすぎる場合, 実務的な観点から有用な知見とはいえない

5. 分析結果と考察

(1) 事故発生割合10%以上 (support ≥ 0.1) の結果

support ≥ 0.1の条件では, これまで認識が低かったような潜在的な関連性は抽出されなかったが, 従来から知られている関連性がいくつか抽出された。(表-3)

つまり, アソシエーション分析において事故発生件数 (割合: support) に一定の閾値を設けることで, 基本的な事故要因を抽出することができ, またその関連性の強さが定量的指標 (support, confidence, lift) によって具体的に示されることが分かる。

表-3 分析結果 (support ≥ 0.1)

結論部	事故件数	support	confidence	lift	条件部		
					項目1	項目2	項目3
運転者属性に関するルール							
車種1当=乗用車	2,273	0.11	0.73	1.37	昼光	追越車線	休日
	2,239	0.11	0.72	1.36	路面乾燥	追越車線	休日
	2,573	0.13	0.72	1.36	追越車線	死傷の有無=死傷者なし	休日
車種1当=大型貨物	2,222	0.11	0.33	1.71	通常期	事故対象物=第2当事者等	平日
	2,398	0.12	0.31	1.60	事故対象物=第2当事者等	平日	平日
	2,164	0.11	0.28	1.48	通常期	路面乾燥	平日
性別=男性	2,094	0.10	0.87	2.35	通常期	操作不適	急ハンドル
	2,360	0.11	0.87	2.35	操作不適	急ハンドル	
	2,105	0.10	0.87	2.35	走行中	負傷事故	
追突事故に関するルール							
車種1当=追越車線	2,519	0.12	0.63	1.36	昼光	車種1当=乗用車	追突
	2,331	0.11	0.63	1.34	車種1当=乗用車	追突	高機能舗装
	2,228	0.11	0.62	1.33	昼光	車種1当=乗用車	わだち掘れ量=0.0~8.0
事故類型=追突	2,568	0.13	0.98	2.19	わき見運転	低速走行	事故対象物=第2当事者等
	2,222	0.11	0.98	2.19	路面乾燥	わき見運転	低速走行
	2,370	0.12	0.98	2.19	わき見運転	低速走行	死傷者なし

※着色部: 結論部と関連性が高いと思われる条件部の項目

(2) 事故発生割合1%以上 (support ≥ 0.01) の結果

support ≥ 0.01の条件では, support ≥ 0.1の場合と同様に, 既知の関連性が抽出されているが, 一方で急勾配, トンネル, 運転者属性に関するルールについて, 潜在的な事故要因と思われる条件部の項目 (黄色塗潰し) もいくつか抽出されている。(表-4)

つまり、「『4%以上の上り勾配』が湿潤事故と関連が強い」など、必ずしも自明でない事故と各種条件との関連性がいくつか抽出されていることから、アソシエーション分析が潜在的な事故要因の抽出に有用である可能性が示唆されたと言える。

性を全網羅的に分析した。その結果、従来から周知の知見に加えて、従来の事故分析では認識が低かった潜在的な事故要因を見出せる可能性を示すことができた。

表-4 分析結果 (support ≥ 0.01)

結論部	事故件数	support	confidence	lift	条件部		
					項目1	項目2	項目3
急勾配に関するルール							
路面状態 =湿潤	260	0.01	0.70	3.10	通常期	操作不適	上り4%以上
	234	0.01	0.70	3.09	操作不適	高機能舗装	上り4%以上
	284	0.01	0.69	3.04	操作不適	上り4%以上	
道路構造 =土工	242	0.01	0.95	1.68	死傷者なし	高機能経過年 =15年以上	上り4%以上
	241	0.01	0.95	1.68	走行中	高機能経過年 =15年以上	下り4%以上
	273	0.01	0.94	1.68	高機能経過年 =15年以上	上り4%以上	
縦断勾配= 上り4%以上	234	0.01	0.15	3.25	路面湿潤	操作不適	高機能舗装
	241	0.01	0.13	2.87	路面湿潤	操作不適	死傷者なし
	284	0.01	0.13	2.85	路面湿潤	操作不適	
縦断勾配= 下り3~4%	207	0.01	0.16	2.20	高機能経過年 =8~10年	わだち掘れ量 =0.0~8.0	
	207	0.01	0.16	2.20	高機能経過年 =8~10年	高機能舗装	わだち掘れ量 =0.0~8.0
	208	0.01	0.15	2.09	死傷者なし	高機能経過年 =8~10年	土工部
縦断勾配= 下り4%以上	210	0.01	0.26	3.23	事故対象物 =第2当事者等	高機能経過年 =8~10年	休日
	226	0.01	0.25	3.06	死傷者なし	高機能経過年 =8~10年	休日
	252	0.01	0.23	2.91	高機能経過年 =8~10年	休日	
トンネルに関するルール							
曲線半径= 左クロソイド	256	0.01	0.72	4.03	追突	高機能舗装	TN部
	257	0.01	0.72	3.99	追越車線	高機能舗装	TN部
	267	0.01	0.68	3.77	事故対象物 =第2当事者等	高機能舗装	TN部
車線1当 =追越車線	257	0.01	0.92	1.97	高機能舗装	TN部	左クロソイド部
	258	0.01	0.90	1.93	追突	TN部	左クロソイド部
	211	0.01	0.88	1.89	車種1当 =乗用車	TN部	左クロソイド部
明暗=昼光	251	0.01	0.96	1.36	V1当=20~60km	追越車線	TN部
	269	0.01	0.96	1.35	V1当=20~60km	わき見運転	TN部
	295	0.01	0.95	1.35	V1当=20~60km	追突	TN部
運転者属性に関するルール							
車種1当 =小型貨物	355	0.02	0.26	2.21	路面湿潤	操作不適	平日
	352	0.02	0.25	2.11	路面湿潤	追越車線	平日
	480	0.02	0.25	2.10	路面湿潤	昼光	平日
明暗=未明	237	0.01	0.54	1.85	車種1当 =大型貨物	V1当=90~100km	事故対象物 =第2当事者等
	215	0.01	0.53	1.83	12月	第1走行車線	土工部
	214	0.01	0.52	1.79	車種1当 =大型貨物	V1当=90~100km	走行中
違反種別 =後退不適	245	0.01	0.70	8.44	車種1当 =大型貨物	右車線変更時	接触
	263	0.01	0.69	8.37	車種1当 =大型貨物	右車線変更時	事故対象物 =第2当事者等
	224	0.01	0.69	8.26	路面乾燥	車種1当 =大型貨物	右車線変更時
車種1当 =乗用車	215	0.01	0.84	1.58	わき見運転	休日	高齢層
	226	0.01	0.82	1.55	繁忙期	わき見運転	TN部
	262	0.01	0.82	1.55	事故対象物 =第2当事者等	休日	高齢層

6. おわりに

(1) まとめ

本稿では、NEXCO西日本・関西支社管内で発生した事故を対象に、潜在的な事故要因を抽出する手法として、データマイニング手法の一つであるアソシエーション分析を用いて、事故と道路、交通、環境の各種条件の関連

(2) 今後の課題

a) 分析対象データの整理方法の改良

アソシエーション分析は、名義尺度のデータしか扱わないため、比例尺度データについては名義尺度への変換、つまりデータの Kategorizatsion が結果に大きく影響する。また、supportやconfidence等の事故発生割合を表す指標はデータの有効数が影響するため、もともと有効データの少ない項目はルールの抽出条件で除外されている可能性があるが、こうした項目の中に有用な知見が含まれていたことも考えられる。今後は、分析対象データの特徴を踏まえたうえで、事故と各種条件の関連性を偏りなく均等な条件で分析するためのデータ整理方法を検討し、事故DBを構築していく必要がある。

b) 新たな評価指標の検討

本稿において関連性の強さを判定する指標として採用したliftの値が表すのは、「ある条件部が課されたときに特定の結論部の割合がどの程度高まるか」であり、事故件数の大小を直接反映していないという問題がある。そのため、liftに着目した分析によって潜在的な事故要因を抽出できたとしても、その要因に対応する事故が少なければ「もっとも効率良く事故を減らすことのできる対策」の提案につながるとは限らない。これについては、アソシエーション分析の三つの評価指標と事故件数・事故率との組み合わせによる指標など、より客観的で総合的な評価指標について検討していく必要がある。

参考文献

- 1) 豊田秀樹：データマイニング入門 -R で学ぶ最新データ解析-, pp.9-24, 東京図書, 2008.
- 2) 小澤友記子, 兒玉崇, 大藤武彦：阪神高速道路の事故要因分析と今後の事故削減に向けた課題, 第30回交通工学研究発表会論文集, CD-ROM, 2010.
- 3) 飯田克弘, 小柳航：クラスアソシエーション分析を用いた都市高速道路における事故要因の階層的可視化, 第32回交通工学研究発表会論文集, CD-ROM, 2012.

(2014. 8. 1 受付)

A BASIC STUDY ON ANALYTICAL METHOD TO DERIVE POTENTIAL ACCIDENT FACTOR IN EXPRESSWAY FROM ACCIDENT DATABASE

Shota TAKADA, Hideo SAWADA, Sihyoung AN, Kazuhiro OISHI and Keiichiro YONEMURA