

ASOCJACJE TREŚCI REKOMENDACJI BEZPIECZEŃSTWA

W artykule przedstawiono identyfikację i analizę powiązań (relacji), które mogą występować między składowymi treściami rekomendacji bezpieczeństwa i zdarzeniami kolejowymi. Do realizacji zadania wybrano metodę MBA (market basket analysis), a obliczenia przeprowadzono w programie STATISTICA. Przyjęto trzy warianty uzyskiwania rozwiązań, różniące się wartościami granicznymi (minimalnymi) współczynników jakości reguł oraz prawdziwością zdań i składowych reguł. Wyniki wraz z interpretacją przedstawiono w postaci tabel i sieci reguł asocjacji.

WSTĘP

W związku z implementacją do prawa krajowego przepisów dyrektywy o bezpieczeństwie kolei, we wszystkich państwach członkowskich Unii Europejskiej powołano tzw. krajowe organy dochodzeniowe (NIB, od ang. *National Investigation Body*). Efektem ich pracy są raporty sporządzane po wystąpieniu różnego rodzaju zdarzeń kolejowych. Istnieje prawny obowiązek informowania NIB o wszystkich wypadkach i incydentach, które miały miejsce na zasadniczej części sieci kolejowej w danym państwie członkowskim. Raporty są istotnym elementem polityki zarządzania bezpieczeństwem [SR-guide]. Zawierają one tzw. rekomendacje bezpieczeństwa, czyli nakazy zastosowania określonych działań (postępowań wobec ryzyka – por. prace np. [1, 2]). Wydawane są wszystkim instytucjom i podmiotom, których zaangażowanie może zapobiec występowaniu zdarzeń kolejowych (wypadków) w przyszłości.

Struktura rekomendacji jest bardzo różna. Zwykle jest to dłuższy tekst, poruszający kilka niezależnych kwestii (np. w zakresie systemów zarządzania bezpieczeństwem) lub lista doraźnych działań podejmowanych bezpośrednio po zdarzeniu. Możliwe jest jednak wskazanie pewnych prawidłowości znajdujących się w treściach rekomendacji. Przykładowo można stwierdzić, że procentowo najczęściej ogranicza się zakres stosowania rekomendacji po zdarzeniach nietypowych oraz zdarzeniach na przejazdach kolejowodrogowych, a więc w sytuacjach, gdy dużą rolę odgrywają warunki miejscowe (np. ukształtowanie i rodzaj drogi wiodącej przez przejazd). Możliwe jest także przyporządkowanie treści rekomendacji do pewnych grup tematycznych.

Przyjęto zatem następującą tezę badawczą: w danym obszarze systemu kolejowego treść rekomendacji bezpieczeństwa jest w znacznym stopniu powtarzalna i zależna od rodzaju zdarzenia kolejowego. Celem artykułu jest więc identyfikacja i analiza powiązań (relacji), które mogą występować między składowymi treściami rekomendacji bezpieczeństwa i zdarzeniami kolejowymi, do których odnoszą się te rekomendacje.

Analiza jest typowym zadaniem różnych projektów *data miningu*, jak również projektów tzw. *text miningu* (tj. podkategorii *data miningu*). Dobra skuteczność tych technik powoduje, że znajdują one zastosowanie w wielu obszarach działalności biznesowej oraz badawczej, od analizy preferencji klientów, poprzez wspomaganie zarządzania zasobami ludzkimi, bioinformatykę, immunologię, a kończąc na przykład, na badaniach historii rozwoju języka [3–6].

Realizację zadania przeprowadzono z wykorzystaniem metody analizy koszykowej (*market basket analysis*) i adekwatnego modułu dostępnego w programie STATISTICA. W szczególności analiza

polega na poszukiwaniu tzw. reguł asocjacyjnych, czyli pewnych modeli skojarzeniowych (znanych jako grupy modeli rozpoznawanych bez nauczyciela), które mają postać zdań warunkowych. W rozdziale 1 przedstawiono materiał źródłowy i przyjętą metodę analizy danych, w rozdziale 2 przedstawiono sposób rozwiązywania problemu, a w rozdziale 3 wyniki analizy.

1. MATERIAŁY I METODY

1.1. Analiza koszykowa

Analiza koszykowa jest najpopularniejszą aplikacją wspomnianych wcześniej reguł asocjacyjnych (skojarzeniowych) [4]. Służy do znajdowania w dużym zestawie danych ukrytych zależności w postaci reguł o schemacie implikacji: *Jeżeli A to B*. Część A jest nazywana poprzednikiem reguły, część B – następnikiem reguły. Zarówno następnik jak i poprzednik reguł mogą być pojedynczym zdaniem lub związkiem zdań (składać się z kilku zdań). W analizie koszykowej zdania łączy się według schematu koniunkcji. Analiza powstała w dziedzinie marketingu, a pierwotnym jej zastosowaniem była badanie danych transakcyjnych pochodzących z supermarketów [3, 6].

Miary jakości reguł asocjacji

Miary jakości reguł asocjacji opierają się na określonych prawdopodobieństwach występowania cech, zmiennych lub zdarzeń sformułowanych w zdaniach i składowych reguł tj. zdarzenia określonego w poprzedniku reguły (np. zdarzenie A), zdarzenia określonego w następniku (np. zdarzenie B), oraz zdarzenia będącego kombinacją dwóch poprzednich. Na tej podstawie można określić trzy następujące współczynniki uznawane za miary jakości reguł asocjacyjnych:

1. Współczynnik *W* – *wsparcia* (lub *pokrycia*), wyrażony poprzez prawdopodobieństwo koniunkcji niezależnych zdarzeń A i B (poprzednika i następnika).
2. Współczynnik *P* – *pewności* (lub *zaufania*), wyrażony poprzez warunkowe prawdopodobieństwo zdarzenia B, pod warunkiem, że zaszło zdarzenie A.
3. Współczynnik *K* – *korelacji* jako miary, której wartość obliczana jest na podstawie wartości wsparcia dzielonej przez pierwiastek kwadratowy z iloczynu wsparć liczonych osobno dla zdarzenia A i dla zdarzenia B.

Zdania składowe reguł (lub w tym przypadku implikacji) zbudowane są z nazw zmiennych i odpowiednich znaczników nazywanych inaczej – *kodami dychotomii*. Wartości tych kodów (0 lub 1) decydują o prawdziwości poszczególnych zdań, z których zbudowana jest reguła.

1.2. Materiał badawczy

Analizie poddano łącznie 185 rekomendacji, których treści przedstawiono w pracy [7]. Rekomendacje przyporządkowywano do siedmiu następujących grup tematycznych:

1. Przenoszenie odpowiedzialności za wybór rozwiązania na podmiot.
2. Nakazywanie zwiększenia nadzoru.
3. Zmiana zakresu i problematyki szkoleń.
4. Nakazywanie uaktualnienia dokumentów.
5. Bezpośrednie polecenia, o treści innej niż w grupach 1-4.
6. Rekomendacje o charakterze wyłącznie informacyjnym.
7. Zakres stosowania rekomendacji.

Grupy zostały określone na podstawie wcześniejszej analizy treści rekomendacji.

2. SPOSÓB ROZWIĄZANIA

2.1. Przygotowanie danych źródłowych

Podstawą uzyskania rozwiązań jest zdefiniowanie zmiennych (jakościowych, zwykle o wartościach lingwistycznych) nazywanych zmiennymi wielokrotnych odpowiedzi lub zakodowanie informacji zmiennymi binarnymi w postaci tzw. wielokrotnych dychotomii. Tego rodzaju zmienne występują najczęściej przy analizie danych pochodzących z badań opinii publicznej, gdy na postawione pytanie jest możliwych wiele różnych odpowiedzi [8]. Różnica w zmiennych polega na sposobie kodowania informacji. W zmiennych typu „wielokrotne dychotomie”, dopuszczalne są jedynie wartości binarne.

W ramach analiz przygotowano jedną zmienną wielokrotnych odpowiedzi, która dotyczy rodzaju zdarzenia kolejowego (zmn. „Zd_kolejowe”) oraz siedem zmiennych binarnych (dychotomicznych), które odpowiadają grupom tematycznym treści w rekomendacjach bezpieczeństwa. Na rysunku 1 przedstawiono przykład sformułowania zmiennych oraz ich przykładowe wartości. Zmienna dotycząca rodzaju zdarzeń kolejowych przyjmuje cztery wartości lingwistyczne: *zderzenie*, *wykolejenie*, *przejazd*, *inne*. Zamienna dotycząca zakresu stosowania rekomendacji („Zakr_stosown”) została potraktowana jako zmienna dychotomiczna. W związku z tym, gdy rekomendację skierowano do jednego podmiotu lub jego jednostki organizacyjnej, jej treść została zakwalifikowana do grupy z ograniczonym zakresem stosowania, a wartość zmiennej przyjęła „1”. W przeciwnym wypadku, gdy rekomendacja dotyczyła wszystkich podmiotów, dla których miała ona sens, uznawano, że zakres jej stosowania nie jest ograniczony, a wartość zmiennej przyjęła „0”.

Dane: Rekomendacje (8 zmn. * 185 prz.)					
	1	2	3	4	5
	Zd_kolejowe	Przenos_odp	Zwieksz_nadz	Szkolenia	Aktual_dok
1	zderzenie	0	0	0	0
2	zderzenie	0	0	0	0
3	zderzenie	0	0	0	0
4	zderzenie	0	1	0	0
5	zderzenie	0	1	0	0
6	zderzenie	0	1	0	0
7	zderzenie	0	0	0	0
8	zderzenie	0	0	0	1
9	zderzenie	1	0	0	0
10	zderzenie	0	0	0	0
11	przejazd	0	0	0	0
12	przejazd	0	0	0	0
13	wykolejenie	0	0	0	0

Rys. 1. Przykład zdefiniowania danych źródłowych (w postaci zmiennej wielokrotnych odpowiedzi i wielokrotnych dychotomii) oraz ich przykładowe wartości w modelu identyfikacji reguł asocjacyjnych w rekomendacjach bezpieczeństwa

Założono badanie dwóch rodzajów relacji:

- a) między składowymi treści rekomendacji,
- b) między treściami rekomendacji a zdarzeniami kolejowymi.

Odwzorowano to w modelu komputerowym poprzez generowanie reguł implikacji z odpowiednio sformułowanymi składowymi (zdaniem) poprzednika i następnika. Przyjęto trzy warianty uzyskiwania rozwiązań, różniące się wartościami granicznymi (minimalnymi) współczynników jakości reguł oraz prawdziwością zdań i składowych reguł. Współczynniki ustalono na poziomach przedstawionych w tabeli 1.

Tab. 1. Przyjęte graniczne wartości miar jakości reguł asocjacyjnych

Lp.	Nazwa miary jakości reguł	Wartości współczynników (miar) jakości reguł		
		Wariant 1	Wariant 2	Wariant 3
1	Wsparcie (W)	0,50	0,1	0,1
2	Pewność (P)	0,50	0,2	0,2
3	Korelacja (K)	0,50	0,2	0,5

2.2. Warianty uzyskiwania rozwiązań

Wariant 1 wykorzystano w identyfikacji wszystkich możliwych relacji występujących między zmiennymi. Wygenerowano zatem możliwe kombinacje zdań w regułach asocjacyjnych, ale ograniczono zakres generowania reguł za pomocą parametrów podanych w tabeli 1. Liczba reguł (lub implikacji), nawet przy nieznacznie zawężonym zakresie poszukiwań (przykładowo, wartości parametrów odpowiednio: $W = 0,01$, $P = 0,01$, $K = 0,01$) wyniosłaby 116186, co utrudniałoby ich dalszą interpretację.

Realizacja badań wymagała ponadto przyjęcia odpowiednich wartości logicznych dotyczących zdań implikacji to jest tzw. kodów dychotomii. Ustalają one prawdziwość zdań reguł, a w przypadku prowadzonych analiz – występowanie/niewystępowanie określonej treści rekomendacji lub zdarzenia kolejowego.

W wariacie 1 wartości kodów ustawiono na „0-1”, co pozwala generować wszystkie możliwe reguły i badać dowolne związki między zmiennymi. Możliwe jest także ustawienie kodów dychotomii tylko na wartość „1”. Powoduje to generowanie reguł „badających” wyłącznie pojawianie się w rekomendacjach pewnych cech lub treści. W ten sposób można poddać np. licznosci występowania treści lub licznosci występowania zdarzeń kolejowych. Omówione podejście wykorzystano w generowaniu wstępnych wyników badań i wyników według wariantu 2.

Przyjęte w wariacie 1 wartości minimalnego wsparcia reguł ($K = 0,50$) oraz minimalnej pewności reguł ($P = 0,50$) oznaczają odpowiednio, że prawdopodobieństwo wystąpienia danej reguły asocjacji w zbiorze wszystkich reguł wynosi 0,5, a prawdopodobieństwo następnika reguł (warunkowe) jest także nie mniejsze niż 0,5.

Jak wskazano, badania zgodnie z wariantem 1 ujmują sytuacje niewystępowania określonych treści w rekomendacjach. Zostało to zasygnalizowane znakiem „= 0” w zapisie generowanych reguł (patrz np. rys. 2). Interpretacja reguł w takiej postaci jest stosunkowo trudna. Dlatego w kolejnym etapie badań (wariant 2) zdecydowano się zastosować inne podejście. Poszukiwane są reguły dotyczące wyłącznie występowania określonych cech lub treści w rekomendacjach. W zapisie tak wygenerowanych reguł jest to sygnalizowane poprzez znak „= 1” (patrz np. tab. 4).

W wariacie 3 zachowano koncepcję badań z wariantu 2, ale zmieniono parametry generowania reguł. Założono, że współczynnik korelacji (K) nie może być niższy niż 0,5 (tab. 1).

Maksymalną licznosci poprzedników reguł i maksymalną licznosci następników przyjęto równą 8, niezależnie od wariantu.

3. WYNIKI ANALIZY

We wstępnym etapie analiz wyznaczono licznosci występowania odpowiednich składowych reguł asocjacyjnych (tab. 3). Wybrano składowe dotyczące obecności w rekomendacjach treści z określonej grupy tematycznej oraz składowe dotyczące rodzajów zdarzeń

kolejowych. Uzyskano 18 kombinacji struktur składowych reguł asocjacyjnych tj. struktur, które mogą być zarówno poprzednikiem lub następnikiem każdej z reguł. Wyniki przedstawiono w tabeli 3. Można zauważyć, że najczęściej występującym zdarzeniem kolejowym był wypadek na przejeździe kolejowym (nazwa składowej: „przejazd”), a treść rekomendacji bezpieczeństwa kwalifikowano do grupy z ograniczonym zakresem stosowania (nazwa składowej: „Ogr_stosown”).

Tab. 3. Wybrane składowe reguł asocjacyjnych wraz z wartościami liczności ich występowania

Lp.	Składowa reguł asocjacyjnych (poprzednik lub następnik reguł)	Liczność
1	Ogr_stosown == 1,	120
2	przejazd	75
3	Bezp_polecenie == 1,	74
4	zderzenie	58
5	Bezp_polecenie == 1,, Ogr_stosown == 1,	58
6	Ogr_stosown == 1,, przejazd	53
7	wykolejenie	44
8	Przenos_odp == 1,	40
9	Ogr_stosown == 1,, zderzenie	37
10	Bezp_polecenie == 1,, przejazd	34
11	Bezp_polecenie == 1,, Ogr_stosown == 1,, przejazd	32
12	Zwiesz_nadz == 1,	30
13	Szkolenia == 1,	28
14	Przenos_odp == 1,, Ogr_stosown == 1,	23
15	Bezp_polecenie == 1,, zderzenie	23
16	Ogr_stosown == 1,, wykolejenie	23

Wystąpiły także złożone składowe reguł (zawierające więcej niż jeden element) czego przykładem jest składowa sformułowana: „Bezp_polecenie == 1,, Ogr_stosown == 1,”. Oznacza ona jednocześnie wystąpienie w rekomendacji treści z grupy „Bezpośrednie polecenia...” i treści dotyczącej ograniczonego zakresu stosowania rekomendacji. Zgodnie z tabelą 3, taki przypadek wystąpił 58 razy. Poniżej przedstawiono wyniki badań według trzech przyjętych wariantów oraz interpretację tych wyników.

Wariant 1

Na rysunku 2 przedstawiono przykład 15 „najsilniejszych” reguł asocjacyjnych. Wyodrębniono je zakładając parametry generowania według wariantu 1 (tab. 1). W tym przypadku liczba możliwych reguł wyniosła 138, gdyż w odróżnieniu od badań we wstępnym etapie (badań liczności) ujmowały one także sytuacje niewystępowania określonych treści w rekomendacjach (znacznik „== 0” w składowych reguł wnioskowania). Jak wynika z rysunku 2, tego typu reguły uzyskiwały najwyższe wartości miar jakości.

Podsumowanie reguł asocjacji (Rekomendacje)						
Min. wsparcie = 50,0%, Min. zaufanie = 50,0%, Min. korelacja = 50,0%						
Maks. liczność poprzednika = 8, Maks. liczność następnika = 8						
	Poprzednik	==>	Następnik	Wsparcie%	Zaufanie(%)	Korelacja(%)
44	Aktual_dok == 0,	==>	Rekomend_inf == 0,	84,32432	94,54545	91,54324
60	Rekomend_inf == 0,	==>	Aktual_dok == 0,	84,32432	88,63636	91,54324
29	Szkolenia == 0,	==>	Rekomend_inf == 0,	80,00000	94,26752	89,03393
59	Rekomend_inf == 0,	==>	Szkolenia == 0,	80,00000	84,09091	89,03393
116	Zwiesz_nadz == 0,	==>	Rekomend_inf == 0,	79,45946	94,83871	89,00105
58	Rekomend_inf == 0,	==>	Zwiesz_nadz == 0,	79,45946	83,52273	89,00105
4	Przenos_odp == 0,	==>	Rekomend_inf == 0,	74,05405	94,48276	85,75910
28	Szkolenia == 0,	==>	Aktual_dok == 0,	74,05405	87,26115	85,11944
42	Aktual_dok == 0,	==>	Szkolenia == 0,	74,05405	83,03030	85,11944
57	Rekomend_inf == 0,	==>	Przenos_odp == 0,	74,05405	77,84091	85,75910
15	Zwiesz_nadz == 0,	==>	Aktual_dok == 0,	72,97297	87,09677	84,41623
41	Aktual_dok == 0,	==>	Zwiesz_nadz == 0,	72,97297	81,81818	84,41623
114	Szkolenia == 0,, Aktual_dok == 0,	==>	Rekomend_inf == 0,	69,18919	93,43066	82,43153
119	Szkolenia == 0,, Rekomend_inf == 0,	==>	Aktual_dok == 0,	69,18919	86,48649	81,91004
14	Zwiesz_nadz == 0,	==>	Szkolenia == 0,	69,18919	82,58065	82,05299

Rys. 2. Wybrane reguły asocjacyjne treści rekomendacji bezpieczeństwa charakteryzujące się największymi wartościami miar jakości

Na podstawie przedstawionych wyników można sformułować ogólny wniosek, że rekomendacje zwykle ujmowały kilka nakazów. Jeżeli treść (nakaz) z jednej grupy tematycznej nie występowała w relacji, to zwykle nie występowała też treść relacji z innej grupy. Przykładowo, brak zapisu o konieczności aktualizacji dokumentów

(reguła nr 44, zmienna „Aktual_dok”) bardzo często pociąga za sobą brak rekomendacji o charakterze informacyjnym.

Wariant 2

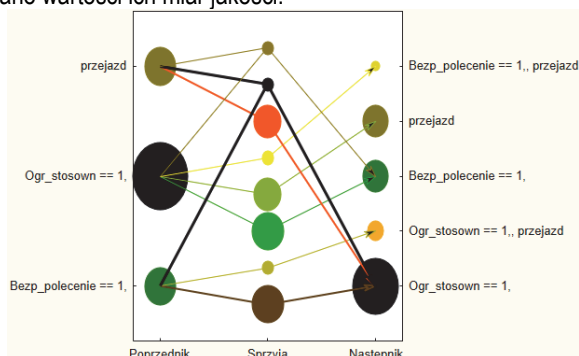
W ramach wariantu uzyskano 19 reguł asocjacyjnych. Wynika z nich, że najczęstszą sytuacją jest pojawianie się treści rekomendacji będących bezpośrednim poleceniem (reguła 6) łącznie z ograniczonym zakresem ich stosowania (Ogr_stosow) w przypadku zdarzeń na przejazdach kolejowo-drogowych. Nieznacznie większe jest przy tym prawdopodobieństwo wystąpienia treści rekomendacji o ograniczonym zakresie stosowania, gdy rekomendacja jest bezpośrednim poleceniem (31,08%), niż odwrotnie (30,83).

Tab. 4. Reguły asocjacji treści rekomendacji bezpieczeństwa wraz z wartościami ich miar jakości

Lp.	Poprzednik	==>	Następnik	Wartość miary jakości reguł [%]		
				W	P	K
1	Przenos_odp == 1,	==>	Ogr_stosown == 1,	11,89	73,33	36,67
2	Zwiesz_nadz == 1,	==>	Ogr_stosown == 1,	31,35	78,38	61,55
3	Bezp_polecenie == 1,	==>	Ogr_stosown == 1,	12,43	31,08	35,11
4	Bezp_polecenie == 1,	==>	zderzenie	18,38	45,95	45,64
5	Bezp_polecenie == 1,	==>	przejazd	17,30	43,24	51,10
6	Bezp_polecenie == 1,	==>	Ogr_stosown == 1, przejazd	31,35	48,33	61,55
7	Ogr_stosown == 1,	==>	Bezp_polecenie == 1,	20,00	30,83	44,35
8	Ogr_stosown == 1,	==>	zderzenie	28,65	44,17	55,87
9	Ogr_stosown == 1,	==>	przejazd	17,30	26,67	50,10
10	Ogr_stosown == 1,	==>	Bezp_polecenie == 1, przejazd	12,43	39,66	35,11
11	zderzenie	==>	Bezp_polecenie == 1,	20,00	63,79	44,35
12	zderzenie	==>	Ogr_stosown == 1,	18,38	45,33	45,64
13	przejazd	==>	Bezp_polecenie == 1,	28,65	70,67	55,87
14	przejazd	==>	Ogr_stosown == 1,	17,30	42,67	48,52
15	przejazd	==>	Bezp_polecenie == 1, Ogr_stosown == 1,	12,43	52,27	31,65
16	wykolejenie	==>	Ogr_stosown == 1,	17,30	55,17	48,52
17	Bezp_polecenie == 1, Ogr_stosown == 1,	==>	przejazd	17,30	94,12	50,10
18	Bezp_polecenie == 1, przejazd	==>	Ogr_stosown == 1,	17,30	60,38	51,10
19	Ogr_stosown == 1, przejazd	==>	Bezp_polecenie == 1,	11,89	73,33	36,67

Wariant 3

Liczba reguł odpowiadająca przyjętym parametrom badań pozwoliła uzyskać 8 reguł asocjacyjnych. Na rysunku 3 przedstawiono je w sposób graficzny w postaci sieci reguł asocjacji, a w tabeli 5 podano wartości ich miar jakości.



Rys. 3. Wyniki analizy rekomendacji bezpieczeństwa w postaci sieci reguł asocjacji (wsp.: wsparcie = 01; pewność = 0,2; korelacja = 0,5)

Na wykresie (rys. 3) wsparcie dla poprzednika i następnika każdej ze znalezionych reguł asocjacji ilustrowane jest przez wielkość i kolor odpowiedniego koła. Kolor i grubość linii natomiast, wskazują poziom zaufania (warunkowe prawdopodobieństwo wystąpienia następnika, o ile wystąpi poprzednik). Wielkość środko-

wych kół (nad etykietą „Sprzyja”) określa współwystępowanie odpowiednich par poprzednika i następnika. Z graficznego podsumowania analizy, można odczytać, że stosunkowo dużą wartością wsparcia charakteryzuje się zmienna Ogr_stosow (skojarzona z bezpośrednim poleceniem i zdarzeniem wypadku na przejeździe). Oznacza to, że bardzo często występuje ograniczenie zakresu stosowania rekomendacji. Według tabeli 5, ograniczenie pojawia się bardzo często, gdy rekomendacja zawiera bezpośrednie polecenia (reg.1, $P = 78\%$).

Tab. 5. Reguły asocjacji treści rekomendacji bezpieczeństwa o korelacjach powyżej 50% wraz z wartościami ich miar jakości

Lp.	Poprzednik ==>	Następnik	Wartość miary jakości reguł [%]		
			W	P	K
1	Bezp_polecenie == 1, ==>	Ogr_stosown == 1,	31,35	78,38	61,55
2	Bezp_polecenie == 1, ==>	Ogr_stosown == 1, przejazd	17,30	43,24	51,10
3	Ogr_stosown == 1, ==>	Bezp_polecenie == 1,	31,35	48,33	61,55
4	Ogr_stosown == 1, ==>	przejazd	28,65	44,17	55,87
5	Ogr_stosown == 1, ==>	Bezp_polecenie == 1, przejazd	17,30	26,67	50,10
6	Przejazd ==>	Ogr_stosown == 1,	28,65	70,67	55,87
7	Bezp_polecenie == 1, przejazd ==>	Ogr_stosown == 1,	17,30	94,12	50,10
8	Ogr_stosown == 1, przejazd ==>	Bezp_polecenie == 1,	17,30	60,38	51,10

Istotnym wynikiem badań jest prawdopodobieństwo pojawienia się treści rekomendacji o ograniczonym zakresie stosowania, w związku z wypadkiem na przejeździe kolejowo-drogowym. Wynosi ono aż 70,67 % (tab. 4 – reg. 6). Łącznie taka reguła stanowiła 28,65% wszystkich reguł.

PODSUMOWANIE

Analiza reguł asocjacji pozwala zidentyfikować wiele ukrytych zależności występujących w badanych obszarach (zjawiskach, procesach itp.). W porównaniu z wynikami otrzymanymi na przykład z użyciem tabelach licznosci/częstości, można zauważyć, że w niektórych przypadkach zdania (i związane z nimi zmienne) mają wysokie częstości, ale nie znajdują się w regułach asocjacji. Tabela częstości wykazuje na przykład, że większość rekomendacji zawiera treści z jednej grupy tematycznej jednak może się zdarzyć, że żadna z tych treści nie będzie wyraźnie skojarzona z inną grupą (nie będzie miała wysokiego zaufania ani korelacji w regule asocjacji). Z drugiej strony, względnie rzadkie występowanie pewnych treści rekomendacji bezpieczeństwa może dawać wyraźne asocjacje np. ze zdarzeniem kolejowym.

Wskazaną użyteczność analizy reguł asocjacji wykorzystano w niniejszym artykule do identyfikacji i analizy powiązań (relacji), które mogą występować między składowymi treści rekomendacji bezpieczeństwa i zdarzeniami kolejowymi. Do realizacji zadania wybrano metodę MBA (*market basket analysis*), a obliczenia przeprowadzono w programie STATISTICA. Przyjęto trzy warianty użycia rozwiązań, różniące się wartościami granicznymi (minimalnymi) współczynników jakości reguł oraz prawdziwością zdań i składowych reguł. Wyniki badań potwierdziły przypuszczenia, że w danym obszarze systemu kolejowego treść rekomendacji bezpieczeństwa jest w znacznym stopniu powtarzalna i zależna od rodzaju zdarzenia kolejowego. Przykłady przedstawiono w interpretacji wyników analiz. Szczególnie silnie powiązanie wykazują relacje między grupami tematycznymi ograniczenie stosowania rekomendacji i bezpośrednie polecenie a zdarzeniem wypadku na przejeździe kolejowo-drogowym. Istotne jest także silne powiązanie rzadziej występujących treści tj. przenoszenia odpowiedzialności za

wybór rozwiązania na podmiot i nakazów zwiększenia nadzoru z ograniczeniem zakresu stosowania rekomendacji.

BIBLIOGRAFIA

- Kadziński A., Kobaszyńska-Twardowska A., Gill A., The concept of method and models for risk management of hazards generated at railway crossings, in: Proceedings of 20th International Scientific Conference. Transport Means (5-7 October 2016, Juodkrante, Lithuania), Lithuania 2016, s. 297–302.
- Kadziński A., Studium wybranych aspektów niezawodności systemów oraz obiektów pojazdów szynowych. Monografia, Poznań 2013.
- Paszyta A., Przykład badania wzorców zachowań klientów za pomocą analizy koszykowej. StatSoft Polska Sp. z o.o., Kraków 2005.
- Łapczyński M., Analiza koszykowa i analiza sekwencji-wielki brat czuwa. Mat. Konf. z Seminarium Zastosowania Statystyki i Data Mining, Statsoft, Warszawa 2009.
- Suriyah M., Karthikeyan S., Karky M., Ganapathy V., Lyric basket: Market basket analysis of Tamil words in lyrics. International Journal of Pure and Applied Mathematics, nr 8 Special Issue (115), 2017, s. 15–21.
- Aguinis H., Forcum L.E., Joo H., Using market basket analysis in management research. Journal of Management, nr 7(39), 2013, s. 1799–1824.
- Smoczyński P., Kadziński A., Gill A., Recommendations of Polish NIB. Mendeley Data, (1), Poznań 2017.
- StatSoft, Internetowy podręcznik statystyki. 2011.
- Plachecka M., Efekty działań na rzecz poprawy bezpieczeństwa w transporcie publicznym w Polsce, Autobusy – Technika, Eksploatacja, Systemy -Transportowe 2016, nr 10.
- Brożyna E., Czynniki ludzki a bezpieczeństwo w ruchu drogowym. Autobusy – Technika, Eksploatacja, Systemy - Transportowe 2017, nr 7-8.

Association rules of the sentences of safety recommendations

The article presents the identification and analysis of relationships that may occur between the safety recommendations and rail events. The MBA method was used, and the calculations were performed in the STATISTICA. Three variants of solutions have been developed, differing in the limits of the quality factors of the association rules. Results and interpretations are presented in the form of tables and a network of association rules.

Autorzy:

dr inż. **Adrian Gill** – Politechnika Poznańska, Instytut Silników Spalinowych i Transportu, Zakład Pojazdów Szynowych, e-mail: adrian.gill@put.poznan.pl, tel. +48 616652017.

mgr inż. **Piotr Smoczyński** – Politechnika Poznańska, Instytut Silników Spalinowych i Transportu, Zakład Pojazdów Szynowych, e-mail: piotr.smoczynski@put.poznan.pl, tel. +48 616652841

dr hab. inż. **Adam Kadziński** – Politechnika Poznańska, Instytut Silników Spalinowych i Transportu, Zakład Pojazdów Szynowych, e-mail: adam.kadzinski@put.poznan.pl, tel. + 48 616652267.

dr inż. **Anna Kobaszyńska-Twardowska** – Politechnika Poznańska, Instytut Silników Spalinowych i Transportu, Zakład Pojazdów Szynowych, tel. +48 616654511. e-mail: anna.kobaszynska-twardowska@put.poznan.pl,