



ANALIZA KOSZYKOWA I ANALIZA SEKWENCJI – WIELKI BRAT CZUWA

Mariusz Łapczyński, Uniwersytet Ekonomiczny w Krakowie

Reguły asocjacyjne

Skojarzenia/asocjacje (*associations*) to jeden z sześciu modeli *data mining*, zaliczany do grupy modeli rozpoznawanych bez nauczyciela lub inaczej do taksonomii bezwzorcowej. Modele skojarzeniowe przedstawiają współwystępowanie wartości/wariantów różnych zmiennych w danym przypadku. Modele skojarzeniowe (asocjacyjne) mają postać zdań warunkowych, w których pojawia się spójnik międzyzdaniowy: „jeżeli zdanie Z_1 , to zdanie Z_2 ”. Używając tego spójnika w mowie potocznej, przyjmuje się, że między zdaniami składowymi istnieje powiązanie rzeczowe lub formalne, tzn. pierwsze zdanie Z_1 implikuje drugie Z_2 . Z punktu widzenia logiki związku między poprzednikiem Z_1 a następnikiem Z_2 mogą mieć różnoraki charakter, jednak w przypadku badań rynkowych i marketingowych mowa o związkach przyczynowo-skutkowych i strukturalnych (tj. takich, które wynikają z rozmieszczenia przedmiotów w przestrzeni albo zdarzeń w czasie). Choć podobną postać zdania warunkowego mają reguły indukcyjne, to te należą jednak do narzędzi umożliwiających budowę modeli dyskryminacyjnych, czyli rozpoznawanych z nauczycielem (wzorcowych).

Reguła skojarzeniowa przyjmuje formę $A \rightarrow B$, gdzie A i B to zbiory atrybutów. Zapis ten oznacza, że jeśli w danym przypadku wariant zmiennej A przyjmuje 1 (prawda), to wariant zmiennej B również przyjmuje z określonym prawdopodobieństwem wartość 1 (prawda). Przykład takiej reguły może być następujący: „40% klientów, którzy kupili mrożoną pizzę, nabyło również keczup; 15% wszystkich klientów nabyło oba te produkty jednocześnie.” W tym przypadku zmienna A = klient kupił mrożoną pizzę, a zmienna B = klient kupił keczup; 40% to współczynnik *confidence*, a 15% to współczynnik *support*, czyli udział tego typu transakcji we wszystkich transakcjach w danej placówce handlowej.

Współczynnik *confidence* to inaczej prawdopodobieństwa warunkowe $P(B | A)$. Interpretując formalnie, trzeba by powiedzieć: „prawdopodobieństwo zdarzenia A obliczone przy założeniu, że zaszło zdarzenie B ”. Adaptując to na potrzeby analizy koszykowej, można powiedzieć: „prawdopodobieństwo zakupu produktu A obliczone przy założeniu, że zakupiono również produkt B ”. Współczynnik *support* to z kolei prawdopodobieństwo koniunkcji zdarzeń $P(AB)$, czyli prawdopodobieństwo łącznego zajścia dwu zdarzeń. Interpretacja tej miary jest znacznie łatwiejsza, chodzi bowiem o udział transakcji, w których kupiono



A i B jednocześnie, w zbiorze wszystkich transakcji. Problemem analityka jest znalezienie wszystkich reguł skojarzeniowych spełniających minimalną (przyjętą przez badacza lub menedżera) wartość *support* i minimalną wartość *confidence*.

Analiza koszykowa – *market basket analysis*

Typowe problemy decyzyjne menedżerów super- i hipermarketów związane są z wyborem produktów, jakie powinny znaleźć się na sali, projektowaniem gazetek reklamowych czy rozmieszczeniem asortymentu na półkach. Analiza transakcji z przeszłości jest powszechnie stosowanym podejściem poprawiającym jakość tych decyzji. Postęp w technologii kodów kreskowych umożliwił przechowywanie tzw. danych „koszykowych” – zbieranych w oparciu o pojedyncze transakcje. Dane „koszykowe” nie zawsze zawierają pozycje (produkty) nabyte razem w danej chwili. Mogą również zawierać produkty nabyte przez konsumenta w określonym przedziale czasowym. Przykładem mogą być miesięczne zakupy członków klubu książki lub sprzedaż odzieży przez firmy wysyłkowe³⁰.

Najpopularniejszą aplikacją reguł skojarzeniowych jest analiza koszykowa (*market basket analysis*). Zmienna (atrybut) nosi tutaj nazwę pozycji (*item*), a przypadek – transakcji. Tabela relacyjna to macierz, w kolumnach której znajdują się pozycje (produkty), a w wierszach transakcje. Celem analizy koszykowej jest:

1. Znalezienie wszystkich reguł, które w następniku zawierają konkretny produkt – ustalenie takich wzorców zakupów pozwoliłoby na optymalizację oferty w celu zwiększenia sprzedaży produktu z następnika;
2. Znalezienie wszystkich reguł, które w poprzedniku zawierają konkretny produkt – reguły te pozwoliłyby określić grupę produktów, których wielkość sprzedaży mogłaby ulec obniżeniu, gdyby sklep zaprzestał sprzedaży produktu z poprzednika;
3. Znalezienie wszystkich reguł wiążących pozycje (produkty) umieszczone na półce A i pozycje umieszczone na półce B – identyfikacja takich wzorców ułatwiłaby rozmieszczenie asortymentu w danej placówce handlowej oraz skuteczną promocję produktów.

Mimo swej przejrzystości i łatwości w interpretacji, reguły skojarzeniowe nie zawsze są przydatne do usprawnienia działań marketingowych firmy. Wyróżnia się trzy typy reguł asocjacyjnych:

1. Reguły użyteczne (*useful rules*) – najbardziej wartościowe z praktycznego punktu widzenia, odkrywają nieznanie wcześniej wzorce zakupów, np. „Jeżeli kupi 2 kg cukru, to kupi 4 butelki piwa”.
2. Reguły trywialne (*trivial rules*) – to klasyczne wzorce transakcji, znane każdemu pracownikowi z danej branży, np. „Jeżeli kupi węgiel drzewny, to kupi podpałkę do grilla”. Tu należy dodać, że reguły nie uznaje się za trywialne, jeśli analiza koszykowa

³⁰ Mamy wtedy do czynienia z innym bezwzorcowym modelem *data mining* – sekwencjami.



wykorzystana została do oceny skuteczności akcji promocyjnej, sprawdza się wtedy, czy oczekiwania dotyczące łącznego nabywania pewnych produktów, potwierdziły się.

3. Reguły niewytłumaczalne (*inexplicable rules*) – to faktyczne i nieoczekiwane wzorce zakupów, które nie przekładają się na działalność marketingową, np. „Jeżeli klient przychodzi w dzień otwarcia sklepu, to kupuje odświeżacz powietrza”.

Czasami do analizy włącza się tzw. wirtualne pozycje (*virtual items*), które zwykle są tożsame z okresem, w którym kupiono dany produkt, np. „Jeżeli kupił ziemię do kwiatów i jest maj, to kupi również rękawice ogrodowe”.

Handel elektroniczny

Większość studiów przypadków dotyczących aplikacji *data mining* w handlu elektronicznym jest skierowana do merchandiserów, którzy nazywani są tutaj internetowymi merchandiserami (*web merchandisers*). Łatwość wykorzystywania narzędzi *data mining* w handlu elektronicznym wynika z kilku udogodnień, jakie niesie ze sobą sposób gromadzenia danych transakcyjnych. Po pierwsze, zgromadzone dane zawierają sporo informacji o nabywcy oraz o nabywanym przez niego produkcie. Po drugie, ilość zgromadzonych danych jest na tyle duża, że dosyć prosto można wylosować odpowiednio liczną próbę testową – niezbędną do zbudowania rzetelnego modelu. Po trzecie, dzięki w pełni zautomatyzowanemu sposobowi gromadzenia danych otrzymuje się względnie rzetelne zbiory obserwacji. Po czwarte, bardzo łatwo wykazać menedżerom użyteczność modeli *data mining* oraz oszacować efektywność działań, np. poprzez obliczenie wskaźnika ROI. I wreszcie po piąte, wyniki analizy (wzorce i reguły) łatwo zintegrować z już istniejącymi w przedsiębiorstwie procesami. Trzeba przy tym uwzględnić czynniki polityczne i społeczne, jakie z pewnością pojawią się przy automatyzacji wykonywanych wcześniej „ręcznie” działań menedżerów.

W trakcie wydobywania reguł asocjacyjnych na potrzeby handlu elektronicznego można wykorzystać kilka ich rodzajów. Klasyfikacja spotykana w literaturze przedmiotu obejmuje:

- ◆ reguły dysocjacyjne/negatywne (*negative association rules*);
- ◆ cykliczne reguły asocjacyjne (*cyclic association rules*);
- ◆ sekwencje (*inter-transactional association rules*);
- ◆ reguły wartościowe (*ratio association rules*);
- ◆ reguły substytucyjne (*substitution association rules*).

W przypadku reguł dysocjacyjnych (jeśli A i nie B, to C) możliwe jest dynamiczne modyfikowanie wyglądu strony, tak aby internauta nie widział treści przez niego „nie lubianych”. W niektórych sytuacjach niezbędne jest uwzględnienie okresowej powtarzalności zakupów, co staje się możliwe dzięki regułom cyklicznym. Profilowanie strony odbywa się wówczas z uwzględnieniem elementu czasu. Internauta otrzymuje przekaz reklamowy



w okresie zwiększonej aktywności, kiedy – zgodnie z wynikami analizy – jego zakupy określonych produktów bądź usług rosą.

Reguły międzytransakcyjne (*inter-transactional association rules*) to inaczej sekwencje, czyli reguły asocjacyjne, w których bada się związek między zakupami dokonanymi w pewnym odstępie czasu. Reguły wartościowe są podobne do jakościowych reguł skojarzeniowych z tą różnicą, że oprócz informacji o nabywanych produktach zawierają też informacje o kwocie, jaką nabywca przeznaczył na zakup poszczególnych dóbr i usług. W efekcie, możliwe staje się prognozowanie wartości transakcji oraz zakupów konkretnych pozycji z oferty przedsiębiorstwa.

Reguły substytucyjne są oparte na powszechnie znanej koncepcji dóbr substytucyjnych. Po zestawieniu wstępnej listy zdań warunkowych konieczne jest jednak ich uzupełnienie wiedzą ekspercką. Za przykład może tu posłużyć księgarnia internetowa Amazon, gdzie nabywca, poszukując konkretnego tytułu, otrzymuje równocześnie listę substytutów, czyli książek o zbliżonej tematyce. Lista tytułów zastępczych obejmuje pozycje, które inni klienci nabywali razem z danym tytułem (bądź przeglądali ofertę przed zakupem) oraz pozycje, które są rekomendowane przez innych czytelników i klientów firmy.

Stosując reguły asocjacyjne, należy pamiętać, aby tuż po wygenerowaniu zestawu wzorców konieczne zweryfikować ich przydatność. W trakcie owej eksperckiej walidacji (*expert-filtering process*) następuje zazwyczaj znaczna redukcja liczby reguł (zdarza się, że nawet o 96%).

Przestrzenne reguły asocjacyjne

Następny obszar, w którym znalazły zastosowanie reguły skojarzeniowe, dotyczył danych ze spisu powszechnego ludności przeprowadzonego w Wielkiej Brytanii. Celem analizy było wsparcie decydentów zajmujących się polityką transportową miasta Manchester. Wykorzystano tu algorytm SPADA (*Spatial Pattern Discovery Algorithm*) odpowiedni dla generowania tzw. przestrzennych reguł skojarzeniowych (*spatial association rules*). Ten rodzaj reguł jest popularny szczególnie w analizie danych geograficznych, gdzie badane obiekty wykazują się często zmiennością przestrzenną. Badany obszar zawężono tutaj do dzielnicy Stockport złożonej z dwudziestu dwóch okręgów.

Odkrywanie przestrzennych reguł asocjacyjnych należy do grupy modeli opisowych (deskryptywnych), gdzie poszukuje się związku między zmiennymi charakteryzującymi dany obszar geograficzny a zmienną będącą przedmiotem zainteresowań badacza. W podejściu tym zmienną zależną³¹ był odsetek gospodarstw domowych dysponujących samochodem, zaś zmiennymi niezależnymi statystyki rynku pracy w tej dzielnicy. Zestaw zmiennych objaśniających obejmował m.in. liczbę zatrudnionych z podziałem na płeć w pełnym

³¹ Właściwie powinno się napisać „zmienna szczególnie interesująca dla badacza”, gdyż termin „zmienna zależna” sugeruje, że mamy do czynienia z modelem predykcyjnym, podczas gdy jest to tylko zmienna wybrana *a priori* jako element następnika.



i niepełnym wymiarze czasu pracy, liczbę samozatrudnionych wg płci, stopę bezrobocia. Autorzy pracy wykazali, że przestrzenne reguły asocjacyjne są skuteczne w odkrywaniu wzorców i zależności w przestrzennych bazach danych charakteryzujących się wysokim stopniem skorelowania zmiennych.

Innym problemem badawczym, jaki udało się rozwiązać dzięki przestrzennym regułom skojarzeniowym, była próba znalezienia związku między charakterystykami dzielnic w Helsinkach a liczbą wypadków samochodowych w tych dzielnicach. Dane o wypadkach pochodziły z baz danych helsińskiego pogotowia ratunkowego i straży pożarnej, natomiast charakterystyki rejonów, w których owe wypadki miały miejsce – z bazy danych rady miejskiej w Helsinkach. Zmienne uwzględnione w analizie to m.in.: bary i restauracje, przedszkola, drogi główne, drogi poboczne, alejki, parki i cmentarze, cieki wodne i inne. Przed rozpoczęciem wyszukiwania reguł ustalono minimalną wartość wsparcia dla reguły i współczynnika ufności na poziomie 0%, co spowodowało, że wynikiem analizy był pełny zestaw zdań warunkowych. Zestaw ów został następnie zredukowany poprzez zdefiniowanie elementów poprzednika i następnika – w każdym zdaniu miała występować zmienna „wypadek”.

Przykładowa reguła brzmiała: „jeśli restauracja albo bar, to wypadek samochodowy (1,7%; 40%)”. Pierwsza wartość w nawiasie oznacza wsparcie dla reguły, które interpretujemy w następujący sposób: 1,7% wszystkich wypadków miało miejsce wyłącznie w pobliżu barów albo restauracji. Druga wartość to współczynnik ufności, który w tym przypadku interpretujemy następująco: w pobliżu 40% helsińskich barów i restauracji miał w badanym okresie miejsce wypadek samochodowy.

Reguły asocjacyjne w marketingu bezpośrednim

Kolejny obszar, w którym wykorzystano reguły asocjacyjne, dotyczył marketingu bezpośredniego, a ściślej rzecz ujmując, ustalenia efektywnej bazy adresowej dla fundacji zajmującej się zbieraniem datków na rzecz weteranów wojennych w Stanach Zjednoczonych. Celem analizy było zestawienie listy adresatów, którzy odpowiedzą na ofertę fundacji, przy czym głównym kryterium owej efektywności była wysokość darowizny z uwzględnieniem kosztów korespondencji. W analizach tego typu wykorzystuje się zazwyczaj narzędzia wzorcowe (modele rozpoznawane z nauczycielem), takie jak drzewa klasyfikacyjne czy sieci neuronowe, jednak autorzy zdecydowali się skorzystać z reguł skojarzeniowych, definiując *a priori* elementy następnika reguły. Były to: zakodowana binarnie zmienna „fakt odpowiedzi na prośbę fundacji” oraz zdyskretyzowana zmienna „wysokość darowizny”.

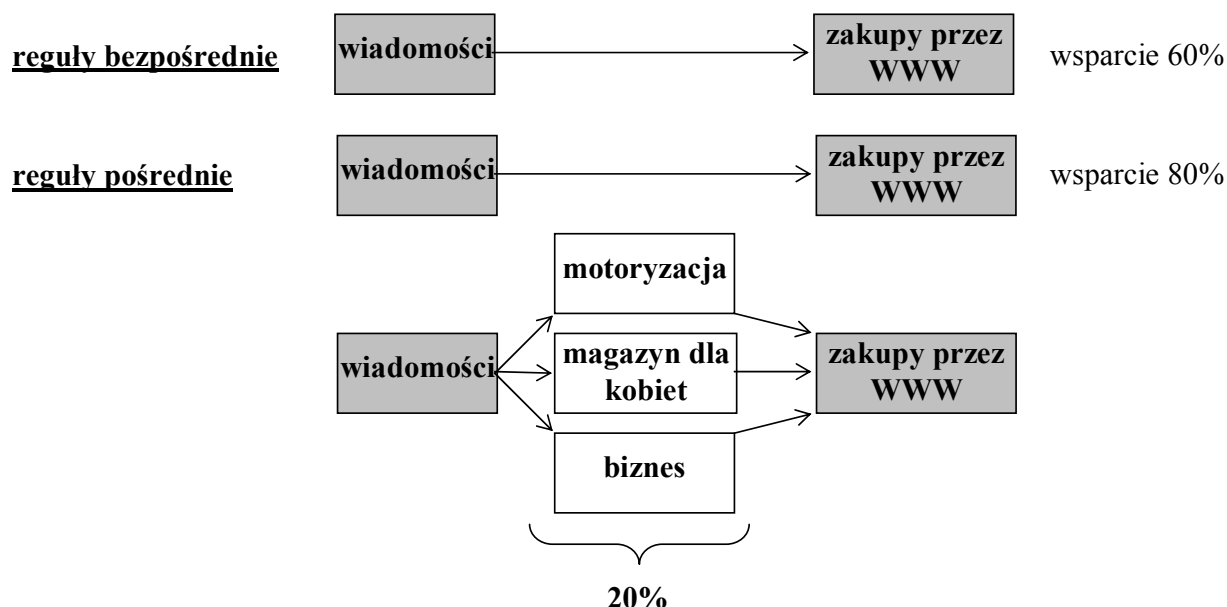
Innowacyjnym elementem tej analizy było wprowadzenie tzw. zogniskowanych reguł skojarzeniowych (*focused association rules*). Ich celem było zredukowanie problemu niezrównoważonych klas, bowiem odsetek donatorów w bazie adresowej fundacji wynosił zaledwie 5%. Pozostałe 95% stanowiły osoby, które nie odpowiedziały na prośbę fundacji. Zadaniem badacza było podwójne ustalenie wartości wsparcia dla reguły (*support*): najpierw minimalnego wsparcia dla reguł wśród tych adresatów, którzy wsparli fundację, a potem maksymalnego wsparcia dla reguł wśród tych adresatów, którzy nie pomogli



fundacji. Innymi słowy, poszukiwano wzorców, które z dużą częstotliwością występują w grupie donatorów i jednocześnie nie są zbyt częste w grupie nie-donatorów. Taka też była kolejność procedury analitycznej. Algorytm wyszukiwał reguły z uwzględnieniem minimalnego wsparcia dla mniej licznego podzbioru, a następnie usuwał z wygenerowanego zestawu te reguły, które zbyt często pojawiały się w drugim – bardziej liczny podzbiorze (reguły, dla których współczynnik *support* był wyższy od przyjętego na wstępie poziomu). Minimalna wartość wskaźnika wsparcia była zwykle równa 1%, zaś maksymalna wartość wsparcia była równa odsetkowi klasy „odpowiedział na prośbę fundacji” w całym zbiorze uczącym.

Analiza sekwencji

Analiza sekwencji to odmiana analizy asocjacyjnej, w której między elementami z poprzednika a elementami z następnika występuje „przerwa czasowa”. Oznacza to, że produkty pojawiające się w regule zostały nabyte podczas co najmniej dwóch transakcji. W odróżnieniu od „tradycyjnej” analizy koszykowej, znaczenie ma tutaj kolejność występowania pozycji w poszczególnych koszykach. Popularnym obszarem, w którym budowane są modele sekwencyjne, jest Internet, a w szczególności analiza strumienia stron (*web clickstream analysis*). Wyniki analizy pozwalają poznać kolejność odwiedzanych stron internetowych (kolejność uruchamianych aplikacji) i stworzyć wzorce zachowań w obrębie jednej lub kilku witryn WWW.

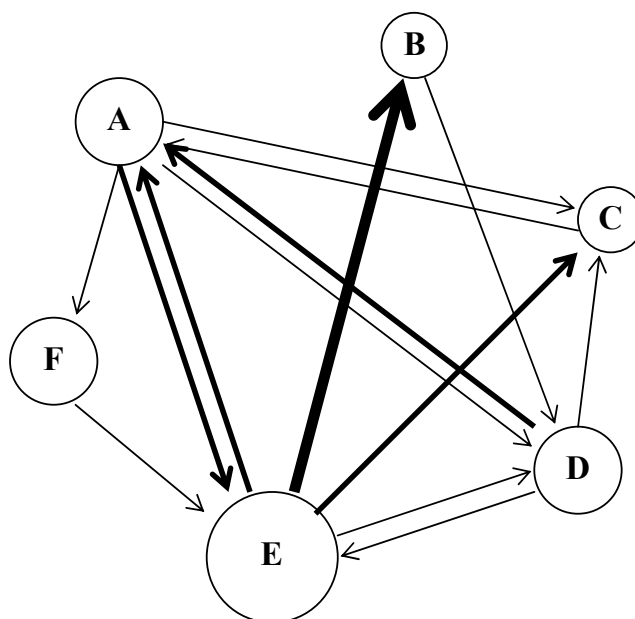


Rys. 1. Przykład sekwencji pośrednich i bezpośrednich.

Podstawowe miary wykorzystywane w analizie sekwencji są takie same jak w analizie asocjacji – wsparcie dla reguły (*support*) oraz współczynnik ufności (*confidence*). Pewną innowacją jest tutaj podział reguł na pośrednie (*indirect rules*) i bezpośrednie (*direct*

rules)³². Różnicę pomiędzy nimi przedstawiono na rysunku (rys. 1). Wsparcie dla reguł bezpośrednich (tu: 60%) jest zwykle niższe od wsparcia dla reguł pośrednich (tu: 80%), co wynika z kolejności nabywanych produktów (tu: odwiedzanych stron na witrynie internetowej). W klasycznej analizie sekwencji ważne było jedynie to, czy oba produkty znajdują się w jednym koszyku. W przypadku sekwencji, B może pojawić się bezpośrednio po A ($A \rightarrow B$), bądź też pośrednio po A, np. $A \rightarrow C \rightarrow B$, $A \rightarrow D \rightarrow B$. Różnica między wartościami wsparcia dla obu rodzajów reguł jest wynikiem pojawienia się innego zdarzenia (tu: wizyty na innej stronie) między zdarzeniem A i B. Te brakujące 20% przypada zatem na strony pośrednie, czyli „motoryzację”, „magazyn dla kobiet” i „biznes”.

Kolejnym etapem procedury może być zamiana zestawu reguł sekwencji na wykres analizy połączeń (*link analysis*). Po określeniu minimalnej wartości wsparcia dokonywana jest selekcja reguł, które stają się następnie podstawą do utworzenia wykresu składającego się z węzłów (*nodes*) i połączeń (*links*) (rys. 2). Grubość strzałki łączącej węzły (połączenia) symbolizuje liczbę obserwacji, które spełniają daną regułę. Jest to podstawowa miara występująca w analizie połączeń, którą można nazwać przelicznikiem reguły (*count of sequence*). Wielkość węzła jest zwykle zależna od tzw. miary centralności (*centrality measure*), która z kolei zależy od miary centralności pierwszego rzędu (*first-order*) – informującej o liczbie połączeń „wychodzących” oraz od miary centralności drugiego rzędu (*second-order*) – informującej o liczbie połączeń „przychodzących”.



Rys. 2. Przykład wykresu analizy połączeń.

³² P. Giuidici, *Applied Data Mining. Statistical Methods for Business and Industry*, John Wiley & Sons 2003, s. 240-241.



Przykład zastosowania reguł asocjacyjnych w analizie danych ankietowych

Reguły asocjacyjne można również zastosować w analizie danych ankietowych w odniesieniu do pytań, których odpowiedzi znajdują się na nominalnym lub przedziałowym/ilorazowym poziomie pomiaru. Poniższy przykład dotyczy związków między wartościami, jakimi kierują się w życiu potencjalni nabywcy samochodów, a preferowanymi przez nich atrybutami pojazdów.

Lista wartości i analiza struktury korzyści

Uznaje się, że wartości mogą wyjaśniać zachowanie nabywców w równym lub nawet wyższym stopniu niż takie konstrukty, jak: postawy, zaangażowanie w zakup, atrybuty produktu czy style życia. Pierwsze próby wyjaśnienia zachowań nabywczych za pomocą wartości pojawiły się już na początku lat 60. ubiegłego stulecia. Współcześnie wartości wykorzystywane są m.in. w popularnej metodzie segmentacji rynku – VALS (*values and life styles*). Jeśli chodzi o rynek motoryzacyjny, to wymienić można m.in. prace D.E. Vinsona i J.M. Munsona (segmentacja nabywców samochodów z zastosowaniem skali RVS)³³, R.E. Pittsa i A.G. Woodside'a (wskazano na ścisły związek wartości z preferencjami i intencjami zakupu)³⁴ oraz L. Bozinoffa i R. Cohena (dowodzono, że istnieje interakcja między wartościami a sytuacjami zakupowymi)³⁵.

Jedną z pierwszych skal wartości była wprowadzona przez M. Rokeacha lista 18 wartości instrumentalnych i 18 wartości ostatecznych (*Rokeach Value Survey*, w skrócie *RVS*)³⁶. Zadaniem respondenta było uporządkowanie 18 wartości od najważniejszej do najmniej ważnej. Przeciwnicy tej metody twierdzili, że mimo iż kwestionariusz ze skalą Rokeacha jest długi, to i tak nie wyczerpuje wszystkich możliwych wartości. Drugie ograniczenie tej metody wiązało się z zadaniem, jakie stawiano respondentom. Dowiedziono, że człowiek potrafi skutecznie uporządkować około 7 pozycji (+/- 2)³⁷, zatem rzetelny ranking 18 wartości stawał się kwestią problematyczną. Trzecią wadą tej metody był czas przeznaczony na wypełnienie kwestionariusza – badania wykazywały, że respondent potrzebował od 10 do 20 minut na porangowanie obu zestawów. Ostatni zarzut dotyczył związku między niektórymi wartościami ostatecznymi (np. pokój na świecie, bezpieczeństwo narodowe) a sytuacjami zakupowymi. Wątpliwości budziła konieczność umieszczenia w ankiecie pełnej listy wartości.

³³ D.E. Vinson, J.M. Munson, *Personal values: an approach to market segmentation*, [w:] K.L. Bernhardt (red.) "Marketing: 1877-1976 and beyond", AMA Chicago 1976, s. 313-317.

³⁴ R.E. Pitts, A.G. Woodside, *Personal value influences on consumer product class and brand preferences*, "Journal of Social Psychology" 119/1983, s. 37-53.

³⁵ L. Bozinoff, R. Cohen, *The effects of personal values and usage situations on product attribute importance*, [w:] B.J. Walker et al. (red.) "An assessment of marketing thought and practice", AMA Chicago 1982, s. 25-29.

³⁶ Szerzej w: A. Sagan, *Badania marketingowe. Podstawowe kierunki*, Wydawnictwo AE, Kraków 2004, s. 154 oraz M. Rokeach, *The Nature of Human Values*, The Free Press, NY 1973.

³⁷ L.B. Peterson, M.J. Peterson, *Short-term retention of individual verbal items*, "Journal of Experimental Psychology", 58/1959, s. 193-198.



Alternatywą dla metody RVS jest tzw. lista wartości (*List of Values*, w skrócie *LOV*), która składa się tylko z dziewięciu pozycji:

- ◆ POCZUCIE PRZYNALEŻNOŚCI (do rodziny, do grupy przyjaciół, do organizacji i stowarzyszeń),
- ◆ ŻYCIE PEŁNE WRAŻEŃ (poszukiwanie silnych wrażeń i emocji, udział w niezwykłych wydarzeniach, „wzrost poziomu adrenaliny”),
- ◆ PRZYJEMNOŚĆ I RADOŚĆ ŻYCIA (potrzeba zabawy, zapewniania sobie przyjemności materialnych i niematerialnych),
- ◆ SERDECZNE STOSUNKI Z INNYMI (dobre relacje z najbliższymi, przyjacielskie kontakty ze znajomymi),
- ◆ SAMOREALIZACJA (spełnianie marzeń, planów życiowych, realizacja własnych celów),
- ◆ BYCIE POWAŻANYM (zdobycie szacunku innych, dbanie o własny wizerunek, zabieganie o pozytywne opinie najbliższych, dbanie o dobre imię),
- ◆ POCZUCIE SPEŁNIENIA (satysfakcja ze zrealizowanych działań, spełnienie w roli ojca (matki) / męża (żony), satysfakcja z pracy),
- ◆ BEZPIECZEŃSTWO (bezpieczeństwo swoje i najbliższych, bezpieczeństwo finansowe, ochrona zdrowia, bezpieczna starość),
- ◆ SZACUNEK DLA SAMEGO SIEBIE (postępowanie zgodne z własnym sumieniem, przestrzeganie ustalonych przez siebie zasad)³⁸.

W przeciwieństwie do RVS w metodzie LOV proszono respondentów, aby z listy wartości wybrali tylko 2 najważniejsze, którymi kierują się w codziennym życiu.

Do oceny atrakcyjności (stopnia pożądania) cech samochodów osobowych wykorzystano analizę struktury korzyści. Analiza struktury korzyści (*Benefit Structure Analysis*, w skrócie BSA) jest narzędziem zaprojektowanym przez J.H. Myersa³⁹, służącym do identyfikacji cech produktu najbardziej pożądanych przez konsumentów. Analiza struktury korzyści to technika będąca alternatywą dla popularnej analizy *conjoint*, skali SIMALTO czy domu jakości. Pierwotnie wykorzystywana była do ustalenia morfologii produktu, identyfikacji podstawowych jego charakterystyk, określeniu korzyści osiąganych przez konsumentów oraz stworzenia profilu produktu o unikatowych cechach. W pierwszych pracach Myersa BSA było zestawem około 75-100 korzyści, jakie respondenci osiągają z użytkowania produktu, oraz zestawem około 50-75 cech (fizycznych charakterystyk) tegoż produktu. Zadanie respondenta polegało na określeniu stopnia pożądania danej korzyści (cechy) oraz na określeniu stopnia niedoboru tej samej korzyści (cechy) w produkcie przez niego użytkowanym. Jest to sposób postępowania znany z techniki SERVQUAL, w której badany ma ocenić istniejący obiekt oraz, korzystając z tej samej skali, zaznaczyć jego ocenę idealną.

³⁸ S.E. Beatty i in., *Alternative Measurement Approaches to Consumer Values: The List of Values and the Rokeach Value Survey*, „Psychology & Marketing”, Vol. 2 No. 3 Fall 1985, s. 181-200.

³⁹ J.H. Myers, *Benefit Structure Analysis: A New Tool for Product Planning*, „Journal of Marketing” No 40 (Oct 1976), s. 23-32.

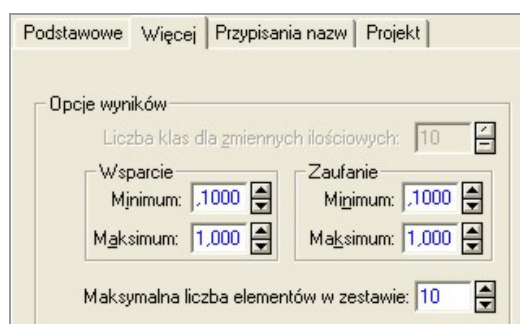


Kolejne aplikacje BSA przeprowadzone przez J.H. Myersa⁴⁰ dotyczyły głównie produktów (usług) częstego zakupu, np. szamponów do włosów, detergentów, napojów, usług telekomunikacyjnych itp. Mimo pozornej prostoty tych dóbr, kwestionariusze ankietowe liczyły przeciętnie po około 75-125 atrybutów (czyli dwa razy tyle skal), choć zdarzały się przypadki, że respondenci analizowali nawet 289 cech (!). Ankietowanymi byli respondenci z panelu otrzymujący kwestionariusz drogą pocztową. W instrukcjach kwestionariuszowych zaznaczono, że po przeczytaniu dwóch stron ankiety, należy zrobić krótką przerwę. Mimo tych utrudnień, zwracano uwagę, że BSA jest narzędziem, które może być przeznaczone do badania oczekiwań konsumentów w zakresie atrybutów, które jeszcze nie istnieją.

Bez względu na liczbę analizowanych cech należy się upewnić, że lista zawiera zmienne należące do następujących wymiarów: a) konkretne fizyczne cechy produktu, b) korzyści z użytkowania produktu, c) wizerunek produktu, d) wizerunek producenta oraz e) osobiste cele i wartości.

Reguły asocjacyjne

Jak wcześniej wspomniano, podstawowymi miarami jakości wykorzystywanymi w poszukiwaniu reguł asocjacyjnych są współczynnik *support* (tu: Wsparcie) oraz współczynnik *confidence* (tu: Zaufanie). Obie miary ustalono tu na poziomie 0,1 (rys. 3)



Rys. 3. Okno wyboru opcji dla analizy asocjacji.

Na rys. 4 przedstawiono kilka najczęściej pojawiających się reguł. I tak osoby z wykształceniem średnim kierują się w swoim życiu poczuciem bezpieczeństwa. Dotyczy to 24% badanych (wsparcie = 23,81974), z zastrzeżeniem, że przeciętnie co druga osoba ze średnim wykształceniem kładzie nacisk na tę właśnie kategorię ze skali LOV (zaufanie = 51,62791). Obecność drugiej podobnej reguły (LOV bezpieczeństwo → wykształcenie średnie) wynika z innej wartości współczynnika *confidence*. W pierwszej można było powiedzieć, że 52% osób ze średnim wykształceniem przywiązuje wagę do bezpieczeństwa, natomiast w drugim przypadku można powiedzieć, że 48% osób, które zwracają uwagę na bezpieczeństwo, to osoby ze średnim wykształceniem. Warto dodać, że współczynnik wsparcia jest dla obu reguł taki sam.

⁴⁰ J.H. Myers, *Segmentation and positioning for strategic marketing decisions*, AMA Chicago 1996, s. 297.



Podsumowanie reguł asocjacji (j_ANKIETA preferencje)
 Min: wsparcie= 10,0%, zaufanie = 10,0%
 Maks. liczność zestawu = 10

Poprzednik	==>	Następnik	Wsparcie%	Zaufanie(%)	Przyrost
WYKSZTAŁCENIE (średnie)	==>	LOV bezpieczeństwo	23,81974	51,62791	1,037009
LOV bezpieczeństwo	==>	WYKSZTAŁCENIE (średnie)	23,81974	47,84483	1,037009
WYKSZTAŁCENIE (wyższe)	==>	LOV bezpieczeństwo	19,52790	48,14815	0,967114
LOV bezpieczeństwo	==>	WYKSZTAŁCENIE (wyższe)	19,52790	39,22414	0,967114
LOV bezpieczeństwo	==>	faza-3	18,24034	36,63793	0,975616
faza-3	==>	LOV bezpieczeństwo	18,24034	48,57143	0,975616
WYKSZTAŁCENIE (średnie)	==>	faza-3	16,73820	36,27907	0,966060
faza-3	==>	WYKSZTAŁCENIE (średnie)	16,73820	44,57143	0,966060

Rys. 4. Przykładowe wyniki analizy dla całego zbioru obserwacji.

Program *STATISTICA* umożliwia wyszukiwanie reguł z wybranymi przez badacza elementami poprzednika bądź następnika. Na rys.5 przedstawiono reguły, których poprzednikiem jest kategoria „bezpieczeństwo” ze skali LOV.

Podsumowanie reguł asocjacji dla wybranych elementów (j_ANKIETA preferencje)
 Min: wsparcie= 10,0%, zaufanie = 10,0%
 Maks. liczność zestawu = 10

Poprzednik	==>	Następnik	Wsparcie%	Zaufanie(%)	Przyrost
LOV bezpieczeństwo	==>	PLEĆ (kobieta)	15,23605	30,60345	1,198421
LOV bezpieczeństwo	==>	BSA koszty eksploatacji	14,80687	29,74138	1,154957
LOV bezpieczeństwo	==>	WYKSZTAŁCENIE (średnie)	23,81974	47,84483	1,037009
LOV bezpieczeństwo	==>	BSA funkcjonalność	14,59227	29,31034	1,034744
LOV bezpieczeństwo	==>	BSA zabezpieczenia	13,09013	26,29310	1,029629
LOV bezpieczeństwo	==>	BSA bezpieczeństwo	13,51931	27,15517	1,004310

Rys. 5. Lista reguł z wybranym przez badacza elementem poprzednika reguły.

Można zaobserwować, że wśród osób, dla których poczucie bezpieczeństwa jest priorytetem:

- ◆ 31% to kobiety,
- ◆ 30% przywiązuje wagę do niskich kosztów eksploatacji samochodu,
- ◆ 48% to osoby ze średnim wykształceniem,
- ◆ 29% kładzie nacisk na funkcjonalność samochodu,
- ◆ 26% zwraca uwagę na zabezpieczenia samochodu przed kradzieżą,
- ◆ 27% dba o ogólne bezpieczeństwo nabytego pojazdu (poduszki powietrzne, kurtyny powietrzne, kontrolowane strefy zgniotu, ABS itp.).

Analiza asocjacyjna z uwzględnieniem płci respondentów

W niektórych sytuacjach warto przeprowadzić analizę asocjacji w wyodrębnionych segmentach rynku bądź w inaczej określonych względnie jednorodnych grupach respondentów/nabywców. Na dwóch kolejnych rysunkach zestawiono reguły skojarzeniowe w grupie mężczyzn (rys. 6) i w grupie kobiet (rys. 7). W każdym zestawie elementem poprzednika jest kategoria „poczucie bezpieczeństwa” ze skali LOV, zaś elementami następnika są zagregowane atrybuty samochodu osobowego ze skali BSA.



Podsumowanie reguł asocjacji dla wybranych elementów (j_ANKIETA preferencje)					
Min: wsparcie= 10,0%, zaufanie = 10,0%					
Maks. liczność zestawu = 10					
Warunek uwzględniania: v17=0					
Poprzednik	==>	Następnik	Wsparcie%	Zaufanie(%)	Przyrost
LOV bezpieczeństwo	==>	BSA bezpieczeństwo	13,25648	28,57143	1,022091
LOV bezpieczeństwo	==>	BSA koszty eksploatacji	12,96830	27,95031	1,154614
LOV bezpieczeństwo	==>	BSA funkcjonalność	12,96830	27,95031	1,010287
LOV bezpieczeństwo	==>	BSA wygląd	12,10375	26,08696	0,862112
LOV bezpieczeństwo	==>	BSA zabezpieczenia	11,52738	24,84472	1,026324
LOV bezpieczeństwo	==>	BSA komfort	11,23919	24,22360	0,884799
LOV bezpieczeństwo	==>	BSA osiągi techniczne	10,66282	22,98137	0,857477

Rys. 6. Zestaw reguł asocjacyjnych w grupie mężczyzn (poprzednikiem jest kategoria „bezpieczeństwo” ze skali LOV).

Podsumowanie reguł asocjacji dla wybranych elementów (j_ANKIETA preferencje)					
Min: wsparcie= 10,0%, zaufanie = 10,0%					
Maks. liczność zestawu = 10					
Warunek uwzględniania: v17=1					
Poprzednik	==>	Następnik	Wsparcie%	Zaufanie(%)	Przyrost
LOV bezpieczeństwo	==>	BSA koszty eksploatacji	20,16807	33,8028	1,117371
LOV bezpieczeństwo	==>	BSA funkcjonalność	19,32773	32,3944	1,070814
LOV bezpieczeństwo	==>	BSA zabezpieczenia	17,64706	29,5775	1,005634
LOV bezpieczeństwo	==>	BSA bezpieczeństwo	14,28571	23,9437	0,982516
LOV bezpieczeństwo	==>	BSA komfort	13,44538	22,5352	1,072676

Rys. 7. Zestaw reguł asocjacyjnych w grupie kobiet (poprzednikiem jest kategoria „bezpieczeństwo” ze skali LOV).

Mężczyźni, którzy na liście wartości wybrali „poczucie bezpieczeństwa”, podczas zakupu samochodu zwracają uwagę na cechy decydujące o ogólnym bezpieczeństwie pojazdu (29%), na koszty eksploatacji samochodu (28%) oraz na jego funkcjonalność (28%). W grupie kobiet, które również wskazały kategorię „bezpieczeństwo” z LOV najważniejsze są koszty eksploatacji pojazdu (34%), funkcjonalność auta (32%) oraz zabezpieczenia przed kradzieżą (30%). Ogólnie rzecz biorąc, warto tutaj zwrócić uwagę, że wartość „poczucie bezpieczeństwa” o wiele rzadziej łączy się z atrybutami samochodu w grupie mężczyzn niż w grupie kobiet. Świadczą o tym różnice w wartościach współczynnika *support* w obu tabelach.

Literatura

1. Agrawal R., Imielinski T., Swami A., *Mining association rules between sets of items in large databases*. In Proceedings of the ACM SIGMOD Conference on Management of Data, Washington D.C., May 1993, s. 207-216.
2. Berry M.J.A., Linoff G., *Data Mining Techniques For Marketing, Sales and Customer Support*. John Wiley & Sons, New York 1997.
3. Cengiz I., *Mining association rules*. Bilkent University, Department of Computer Engineering and Information Sciences, Ankara, Turkey, 1997.
4. Giuidici P., *Applied Data Mining. Statistical Methods for Business and Industry*, John Wiley & Sons 2003.



5. Han J., Fu Y., *Discovery of Multiple-Level Association Rules from Large Databases*, Proceedings of 21th International Conference on Very Large Data Bases, Zurich, Switzerland, September 1995.
6. Han J., Kamber M., *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco 2001.
7. Karasova V., Krisp J.M., Virrantaus K., *Application of Spatial Association Rules for Improvement of a Risk Model for Fire and Rescue Services*, [w:] Hauska, H., Tveite, H., (red.), 10th Scandinavian Research Conference on Geographical Information Science (ScanGIS2005), 13-15 June, Stockholm, Sweden, s. 183-193.
8. Kohavi R., Provost F., *Applications of Data Mining to Electronic Commerce*, „Data Mining and Knowledge Discovery”, 5(1/2)/2001, s. 5-10.
9. Malerba D., Esposito F., Lisi F.A., *Mining Spatial Association Rules in Census Data*, Proceedings of the Joint Conference on „New Techniques and Technologies for Statistics” and „Exchange of Technology and Know-how”, 2001, s. 541-550.
10. Natarajan R., Shekar B., *Interestingness of association rules in data mining: Issue relevant to e-commerce*, „Sadhana” Vol. 30, Parts 2 & 3, April/June 2005, s. 291–309.
11. Srikant R., Agrawal R., *Mining quantitative association rules in large relational tables*. In Proceedings of the ACM SIGMOD Conference on Management of Data, June 1996.
12. Wang K., Zhou S., Yang Q., Yeung J.M.S., *Mining Customer Value: From Association Rules to Direct Marketing*, Proceedings from 19th International Conference on Data Engineering, 5-8 March 2003, s. 738-765.