

BADANIE ATRAKCYJNOŚCI OFERTY DOSTĘPU DO INTERNETU ZA POMOCĄ ANALIZY GRADACYJNEJ

Tomasz Ząbkowski, Wiesław Szczesny

Katedra Informatyki

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: tomasz_zabkowski@sggw.pl, wieslaw_szczesny@sggw.pl

Streszczenie: Celem badań była analiza atrakcyjności oferty dostępu do mobilnego Internetu wśród klientów. W pracy wykorzystano podejście gradacyjne do wizualnej prezentacji danych oraz dokonania oceny podobieństwa pakietów pod względem ceny w różnych grupach wiekowych klientów. Pozwoliło to sformułować wnioski, że atrakcyjność oferty dostępu do mobilnego Internetu różni się w zależności od grupy wiekowej klientów oraz kategorii dochodowej. Jednocześnie badanie pozwoliło zidentyfikować segmenty nabywców o podobnych cechach.

Słowa kluczowe: analiza gradacyjna, Internet, mapy nadreprezentacji

CEL BADAŃ

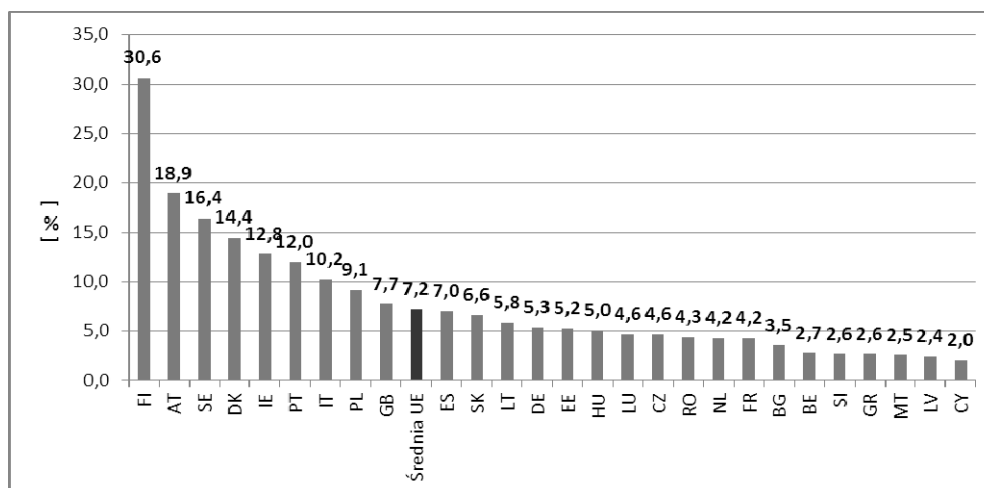
Celem prezentowanych badań była ocena atrakcyjności oferty dostępu do internetu mobilnego. W opracowaniu wykorzystano podejście gradacyjne do zebranych danych empirycznych, co umożliwiło wizualną prezentację danych oraz porównanie i ocenę podobieństwa ofert pod względem ceny w różnych grupach wiekowych oraz różnych kategoriach dochodowych. Wykorzystanie analizy gradacyjnej w pracy wynika z przesłanek o dużym znaczeniu regularności w danych. Występowanie zależności między zmiennymi i obiektami połączone z regularnością może być pomocne w opisie i segmentacji danych. Odkrycie pewnych jednorodnych podgrup klientów w danych empirycznych, mających podobne charakterystyki może zostać wykorzystane do tworzenia odpowiednio dobranych strategii marketingowych w celu budowania trwałych relacji z klientem (CRM).

OPIS PROBLEMU I DANE

Rynek dostępu do Internetu w Polsce rozwija się bardzo dynamicznie, obejmując swoim zasięgiem coraz szersze grono odbiorców. Bezpośrednim tego odzwierciedleniem jest rosnąca wartość rynku i duża liczba klientów. Według danych Urzędu Komunikacji Elektronicznej [UKE 2010] w 2010 roku blisko 6,1 mln osób posiadało stacjonarny i około 2,8 mln osób dostęp za pomocą modemów 2G/3G do Internetu. Usługi mobilnego dostępu przez modemy 2G/3G w urządzeniach przenośnych stanowiły najbardziej dynamicznie rozwijający się segment rynku, co powodowało, że była to druga (zaraz po dostępie xDSL) najczęściej wybierana forma dostępu do sieci. W dużej mierze był to efekt działalności operatora Play, który w 2010 roku, o blisko 86% zwiększył bazę klientów i tym samym osiągnął prawie 22% liczby użytkowników ogółem. Upowszechnianie się dostępu modemami 2G/3G do Internetu wynikało m.in. (a) z dostrzeżonych przez klientów zalet mobilności usługi, (b) wzrostu pokrycia siecią 3G terytorium kraju, (c) zachodzących procesów modernizacyjnych umożliwiających korzystanie z większych prędkości przesyłowych.

Porównując Polskę z krajami Unii Europejskiej pod względem wielkości wskaźnika dostępności mobilnego dostępu do Internetu, Polska, na podstawie danych z Digital Agenda Scoreboard [UKE 2010], zajęła ósme miejsce spośród państw Unii Europejskiej (por. rys. 1), przewyższając średnią unijną o 1,9 pp.

Rys. 1. Penetracja mobilnego szerokopasmowego dostępu do Internetu w UE na 100 mieszkańców (w %)



Źródło: opracowanie własne na podstawie [UKE 2010]

Dane wykorzystane w analizie stanowił zbiór 2821 klientów, którzy zdecydowali się w 2010 roku na zakup pakietu dostępu do mobilnego Internetu u operatorów świadczących usługi telefonii (por. tabela 1).

Na podstawie dostępnych danych, wydzielono sześć kategorii wiekowych: klienci w wieku do lat 20, 21-25 lat, 26-35 lat, 36-45 lat, 46-60 lat oraz powyżej 60 lat. Klienci mieli do wyboru w ofercie operatora pięć pakietów dostępu do Internetu w cenie: 39 zł, 49 zł, 69 zł, 89 zł oraz 129 zł.

Dodatkowo została wyodrębniona kategoria dochodowa, która dla A1 oznacza, że dany klient nabył pakiet o wartości wyższej niż średnia wartość jego dotychczasowych rachunków z ostatnich 6 miesięcy. A2 przedstawia sytuację odwrotną tzn., że dany klient nabył pakiet o wartości niższej niż średnia wartość jego dotychczasowych rachunków z ostatnich 6 miesięcy. Przykładowo, dla pierwszej grupy wiekowej (do lat 20), 233 klientów nabyło pakiet internetowy w cenie 39 zł, co było wartością wyższą niż średnia wartość dotychczasowych rachunków z ostatnich 6 miesięcy. Z kolei 41 klientów w tej grupie wiekowej nabyło pakiet za 39 zł, co stanowiło wartość niższą niż średnia wartość dotychczasowych rachunków z ostatnich 6 miesięcy. Innymi słowy w kategorii A1 mamy do czynienia z klientami, którzy płacą średnio niższe rachunki niż wartość nabywanego pakietu internetowego, natomiast w kategorii A2 mamy klientów, którzy płacą średnio wyższe rachunki niż wartość nabywanego pakietu.

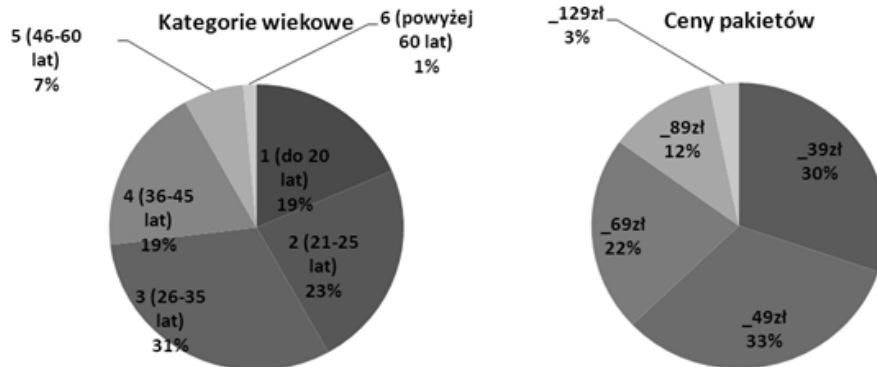
Tabela 1. Dane empiryczne wykorzystane w analizie

Grupa wiekowa	Kategoria	Wartość pakietu					Suma	Suma
		39zł	49zł	69zł	89zł	129zł		
1 (do 20 lat)	A1	233	92	38	16	2	381	523
	A2	41	37	46	17	1	142	
2 (21-25 lat)	A1	106	107	61	10	1	285	659
	A2	79	122	109	47	17	374	
3 (26-35 lat)	A1	78	160	155	138	47	578	879
	A2	34	115	95	45	12	301	
4 (36-45 lat)	A1	121	124	79	46	10	380	533
	A2	46	77	23	6	1	153	
5 (46-60 lat)	A1	68	55	11	8	0	142	186
	A2	23	17	4	0	0	44	
6 (powyżej 60 lat)	A1	21	12	1	0	0	34	41
	A2	2	3	2	0	0	7	
Suma		852	921	624	333	91	2821	2821

Źródło: opracowanie własne

W zamieszczonych w tabeli danych zwraca uwagę to, że w strukturze wiekowej najbardziej liczna jest trzecia grupa wiekowa (26-35 lat). Natomiast najmniej licznie reprezentowana jest grupa klientów powyżej 60 roku życia, gdyż stanowi ona jedynie 1%. W przypadku pakietów internetowych najbardziej popularne są pakiety najtańsze, w cenie 39 zł i 49 zł i stanowią one odpowiednio: 30% i 33% spośród wszystkich nabywanych pakietów (por. rys. 2).

Rys. 2. Udział procentowy według kategorii wiekowych oraz cen pakietów



Źródło: opracowanie własne

GRADACYJNA ANALIZA DANYCH

Gradacyjna analiza danych jest zaliczana do technik wielowymiarowej eksploracji danych, określanych także jako Data Mining. Istotną zaletą analizy gradacyjnej jest jednoczesna integracja obszarów z pogranicza teorii pomiaru, statystyki, taksonomii, wielowymiarowej analizy danych (także wizualizacji) oraz zastosowanie podejścia umożliwiającego jednolite traktowanie danych różnego typu. Szerszy jej opis można znaleźć m.in. w [Szczesny 2002], [Kowalczyk i in. 2004]. Ważną cechą analizy gradacyjnej jest jej ukierunkowanie na wizualizację, celem lepszego zrozumienia i poznania całościowej struktury danych i występujących w nich współzależności.

Wykorzystanie analizy gradacyjnej w pracy wynika z przesłanek o dużym znaczeniu regularności w danych, mogących mieć znaczenie z punktu widzenia badań rynku i kierowania odpowiedniej oferty marketingowej (pakietów dostępu do Internetu) do odpowiednich grup klientów.

Podstawowym zagadnieniem w ramach metody gradacyjnej analizy danych jest badanie zróżnicowania (koncentracji) rozkładów poprzez dążenie do takiego uporządkowania macierzy danych, aby osiągnąć maksymalny kontrast pomiędzy skrajnymi wierszami i kolumnami [Borkowski i Szczesny 2005], [Pleszczyńska i in. 2006]. Dane w postaci tabeli zostają poddane jednoczesnemu przestawieniu zarówno wierszy jak i kolumn, tak, aby w tabeli wynikowej cecha porządkująca wiersze była jak najsilniej dodatnio zależna od cechy porządkującej kolumny.

Celem przybliżenia zagadnienia zostanie przedstawiony przykład, w którym mamy dwa rozkłady: $p = (p_1, \dots, p_k)$, $s = (s_1, \dots, s_k)$ określone na tym samym zbiorze. W prezentowanym przykładzie posłużymy się danymi dotyczącymi pakietów o wartości 39 zł i 69 zł (por. tabela 2). Na ich podstawie wyznaczona została krzywa koncentracji rozkładu s względem rozkładu p . Jest to krzywa

łamana, wpisana w kwadrat jednostkowy, łącząca kolejno punkty $(0,0), (p_1, s_1), (p_1 + p_2, s_1 + s_2), \dots, (p_1 + \dots + p_k, s_1 + \dots + s_k), (1,1)$. Nachylenie odcinków krzywej wyznacza iloraz s_i/p_i , który może przebiegać dowolnie według naturalnego porządku (por. kol. 6 w tabeli 2) lub w sposób uporządkowany rosnąco (por. kol. 7 w tabeli 2).

Tabela 2. Tabela licznosci według grup wiekowych oraz pakietów internetowych o wartości 39 zł i 69 zł

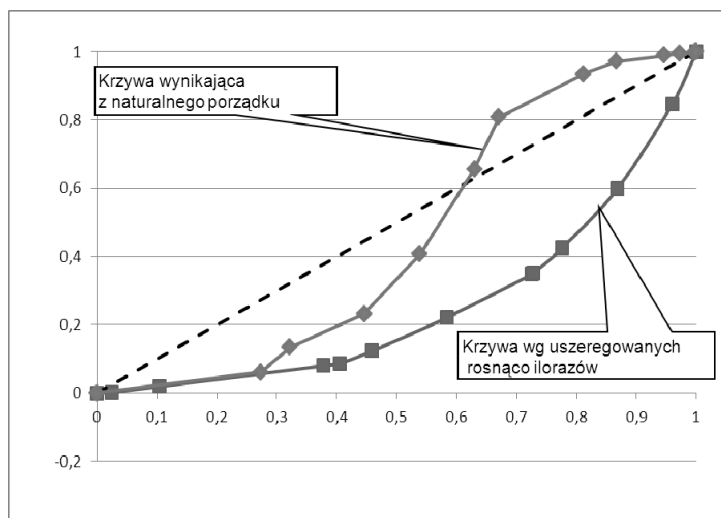
Grupa wiekowa	Pakiet		Struktura		Iloraz s_i / p_i	Kolejność ilorazu	Punkty krzywej	
	39	69	p_i	s_i			Fp_i	Fs_i
1	2	3	4	5	6	7	8	9
1_A1	233	38	0,273	0,061	0,223	3	0	0
1_A2	41	46	0,048	0,074	1,532	9	0,025	0,002
2_A1	106	61	0,124	0,098	0,786	6	0,104	0,019
2_A2	79	109	0,093	0,175	1,884	10	0,378	0,080
3_A1	78	155	0,092	0,248	2,713	11	0,405	0,087
3_A2	34	95	0,040	0,152	3,815	12	0,459	0,123
4_A1	121	79	0,142	0,127	0,891	7	0,583	0,221
4_A2	46	23	0,054	0,037	0,683	5	0,725	0,348
5_A1	68	11	0,080	0,018	0,221	2	0,728	0,351
5_A2	23	4	0,027	0,006	0,237	4	0,776	0,425
6_A1	21	1	0,025	0,002	0,065	1	0,869	0,599
6_A2	2	2	0,002	0,003	1,365	8	0,960	0,848
Razem	852	624					1	1

Źródło: opracowanie własne

Krzywa koncentracji wynikająca z naturalnego uporządkowania może przebiegać zarówno nad, jak i pod przekątną. Jako miernik koncentracji ar wykorzystuje się podwójną różnicę między dwoma polami zawartymi między przekątną a krzywą. Dokonując szeregowania grup wiekowych z przykładu w tabeli 2, tak aby kolejne ilorazy były rosnące (biorąc pod uwagę kolejność z kol. 7), w wyniku zostaje utworzona krzywa zróżnicowania leżąca pod krzywą zróżnicowania wyznaczoną przy dowolnym innym uporządkowaniu. Krzywą tą nazywa się maksymalną krzywą zróżnicowania (koncentracji) rozkładu s względem rozkładu p , oraz odpowiadający jej wskaźnik koncentracji nazywa się wskaźnikiem maksymalnej koncentracji z oznaczeniem ar_{max} [Borkowski i Szczesny 2005], [Pleszczyńska i in. 2006].

Ułożenie krzywej koncentracji może być dwojaki (por. rys. 3). W pierwszym przypadku, gdy dane reprezentują naturalny porządek, krzywa koncentracji przebiega zarówno nad i pod przekątną o współrzędnych $(0,0)$ oraz $(1,1)$. Natomiast w drugim, gdy dane są uszeregowane według niemalejących ilorazów krzywa przebiega pod przekątną.

Rys. 3. Krzywe koncentracji



Źródło: opracowanie własne

W rozważanym przykładzie przebieg krzywej koncentracji wskazuje, że klienci w ramach grup wiekowych w sposób odmienny byli zainteresowani rozważanymi pakietami dostępu do Internetu. Wskazuje na to maksymalna krzywa różnicowania, która jest znacznie oddalona od przekątnej.

W analizie gradacyjnej rozwiązanie określonego problemu dla danych w postaci tabeli (o nieujemnych wartościach) oznacza znalezienie takiego uporządkowania kolumn i wierszy, aby wskaźnik zależności gradacyjnej osiągnął maksimum globalne. Do tego celu wykorzystuje się współczynnik korelacji rho-Spearmana, oznaczany jako ρ^* oraz τ Kendalla [Szczesny 2002]. W celu wyznaczenia maksimum lokalnego dla współczynnika ρ^* należy wprowadzić pojęcie rozkładu gradacyjnego dla danej tabeli kontyngencji, oznaczone jako $P = \{p_{ij} : i = 1, \dots, m, j = 1, \dots, k\}$, gdzie $p_{ij} = n_{ij}/n$. Rozkład gradacyjny, oznaczony jako P^* jest rozkładem ciągłym z gęstością opisaną w kwadracie jednostkowym. Zatem jeśli P jest rozkładem pary zmiennych losowych, to P^* jest rozkładem łącznym tych zmiennych otrzymanym w wyniku przekształcenia gradacyjnego. Kwadrat jednostkowy $\langle 0,1 \rangle$ dzieli się na $m \times k$ prostokątów, o wymiarach $p_{i+} \times p_{+j}$:

$$p_{i+} = \sum_{j=1}^k p_{ij}, p_{+j} = \sum_{i=1}^m p_{ij} \quad (1)$$

gdzie $i = 1, \dots, m, j = 1, \dots, k$.

Pole prostokąta o wymiarach i, j jest równe $p_{i+} p_{+j}$. Funkcja gęstości rozkładu P^* jest stała w każdym prostokącie i równa jest ilorazowi $p_{ij}/p_{i+} p_{+j}$. Wartość tego ilorazu interpretuje się jako stopień nadreprezentacji w stosunku do

rozkładu jednostajnego. Im iloraz ten jest większy od 1, tym bardziej nadreprezentowana jest zawartość umieszczona w komórce p_{ij} w stosunku do tego, co wynikałoby z rozkładu proporcjonalnego. W przypadku wartości ilorazu poniżej 1, mamy do czynienia z niedoreprezentacją. Współczynniki regresji dla wierszy oraz kolumn, oznaczone jako $\text{regr}_{r|c}^*$ i $\text{regr}_{c|r}^*$, oraz współczynnik korelacji ρ^* (rho) dla P^* , nazywane są odpowiednio, funkcjami regresji gradacyjnej oraz współczynnik korelacji gradacyjnej dla rozkładu P . Miary zależności ρ^* rho-Spearmana oraz współczynnik korelacji τ Kendalla dla kolumn wyrażają się następującymi równaniami:

$$\rho^*(P) = 3 \sum_{t=2}^m \sum_{s=1}^{t-1} (S_t + S_{t-1} - S_s - S_{s-1})(p_{t+} p_{s+}) ar(t:s) \quad (2)$$

oraz

$$\tau(P) = 2 \sum_{t=2}^m \sum_{s=1}^{t-1} [(p_{t+} p_{s+}) ar(t:s)] \quad (3)$$

gdzie w powyższych równaniach

$$ar(t:s) = \sum_{i=1}^k \sum_{j=t+1}^k [(p_{si} p_{tj} - p_{sj} p_{ti}) / (p_{i+} p_{s+})] \quad (4)$$

oznacza wskaźnik koncentracji dla kolumny t względem kolumny s , oraz

$$S_u = \sum_{i=1}^u p_{i+}, \text{ dla } u = 1, \dots, k. \quad (5)$$

W sposób analogiczny przeprowadza się obliczenia dla wierszy.

Wartość ρ^* zależy od kolejności wierszy i kolumn w tablicy o wyrazach p_{ij} . Liczba jednoczesnych przestawień wierszy i kolumn w małych tabelach jest skończona, stąd można znaleźć parę lub więcej par ustawień, przy których ρ^* osiąga maksimum. Ustawienia tabeli, przy których $\rho^* = \rho^*_{\max}$ są interesującym źródłem informacji o danych w niej zawartych i często daje się je interpretować [Książyk i in. 2005]. Jest to możliwe zwłaszcza wtedy, gdy tablica po maksymalizacji ρ^* staje się na tyle regularna, że wszystkie wskaźniki $ar(t:s)$ dla kolumn i i wierszy są równe odpowiadającym im $ar_{\max}(t:s)$.

W praktyce trudno oczekiwać znalezienia takich zależności, aby wszystkie wskaźniki $ar(t:s)$ oraz $ar_{\max}(t:s)$ były równe. W szczególności, dla tabel o dużych m i k , przeszukiwanie po wszystkich kombinacjach wierszy i kolumn jest procesem długotrwałym, dlatego do znalezienia maksimum globalnego wykorzystywane są metody symulacyjne typu Monte Carlo. W tym celu poszukuje się w pętli, rozpoczynając od losowego uporządkowania wierszy i kolumn tabeli, takiego ich ustawienia, dla którego dane ρ^* osiąga maksimum lokalne. W postępowaniu tym wykorzystywany jest fakt, że jeśli któraś z regresji wierszy lub kolumn nie jest niemalejąca, to przestawiając odpowiednio wiersze lub kolumny zgodnie z wartościami regresji, otrzymujemy rozkład o większej wartości

ρ^* . Rozkład ten, dla obydwu rodzajów regresji ustawionych niemalejąco, uzyskiwany jest po skończonej liczbie przestawień. Najwyższa wartość ρ^* wybrana ze zbioru maksimum lokalnych jest z reguły równa maksimum globalnemu.

BADANIE ATRAKCYJNOŚCI OFERTY DOSTĘPU DO INTERNETU

Dostępny zbiór danych został poddany analizie gradacyjnej w pakiecie GradeStat opracowanym w Instytucie Podstaw Informatyki PAN. W pierwszym kroku, na podstawie danych zawartych w tabeli 1 zostały wyliczone wskaźniki nadreprezentacji dla każdej komórki leżącej na przecięciu grup wiekowych oraz cen pakietów. Przykładowo, dla kolumny z pakietem o wartości 39 zł i grupy wiekowej 1 oraz kategorii A1, wartość wskaźnika wynosi $2,025=233/(852 \times 381/2821)$. Wyliczenia zawiera tabela 3.

Tabela 3. Wskaźniki nadreprezentacji

Grupa wiekowa	Kategoria	Wartość pakietu				
		39 zł	49 zł	69 zł	89 zł	129 zł
1 (do 20 lat)	A1	2,025	0,740	0,451	0,356	0,163
	A2	0,956	0,798	1,464	1,014	0,218
2 (21-25 lat)	A1	1,231	1,150	0,968	0,297	0,109
	A2	0,699	0,999	1,318	1,065	1,409
3 (26-35 lat)	A1	0,447	0,848	1,212	2,023	2,521
	A2	0,374	1,170	1,427	1,266	1,236
4 (36-45 lat)	A1	1,054	0,999	0,940	1,025	0,816
	A2	0,995	1,541	0,680	0,332	0,203
5 (46-60 lat)	A1	1,586	1,186	0,350	0,477	0,000
	A2	1,731	1,183	0,411	0,000	0,000
6 (powyżej 60 lat)	A1	2,045	1,081	0,133	0,000	0,000
	A2	0,946	1,313	1,292	0,000	0,000

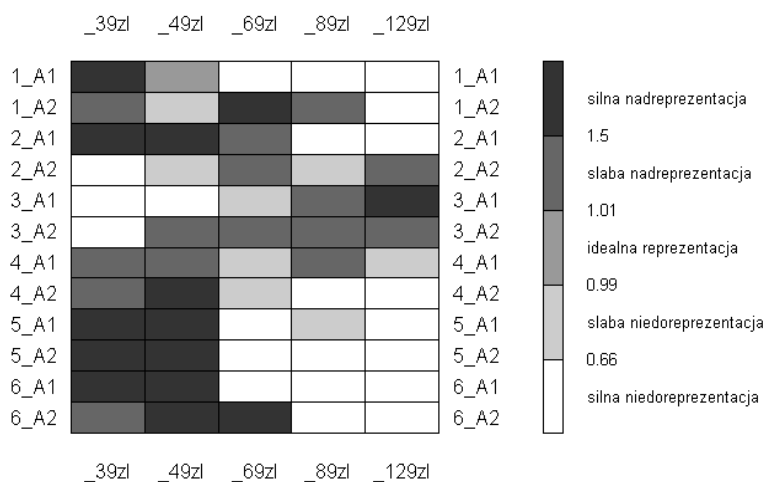
Źródło: obliczenia własne

Dla wyznaczonych w tabeli 3 wskaźników nadreprezentacji można stworzyć mapę obrazującą stopień reprezentacji danych. W tym celu, za pomocą kilku odcieni szarości można określić obszary niedoreprezentacji, idealnej reprezentacji oraz nadreprezentacji w danych. W rozpatrywanym przykładzie obszary te oznaczono w następujący sposób:

- wartość poniżej 0,66 oznacza silną niedoreprezentację;
- wartość w przedziale 0,66-0,99 oznacza słabą niedoreprezentację;
- wartość w przedziale 0,99-1,01 oznacza idealną reprezentację;
- wartość w przedziale 1,01-1,5 oznacza słabą nadreprezentację;
- wartość powyżej 1,5 oznacza silną nadreprezentację.

Reprezentację graficzną danych z tabeli 3 przedstawia rysunek 4.

Rysunek 4. Mapa nadreprezentacji dla danych z tabeli 3



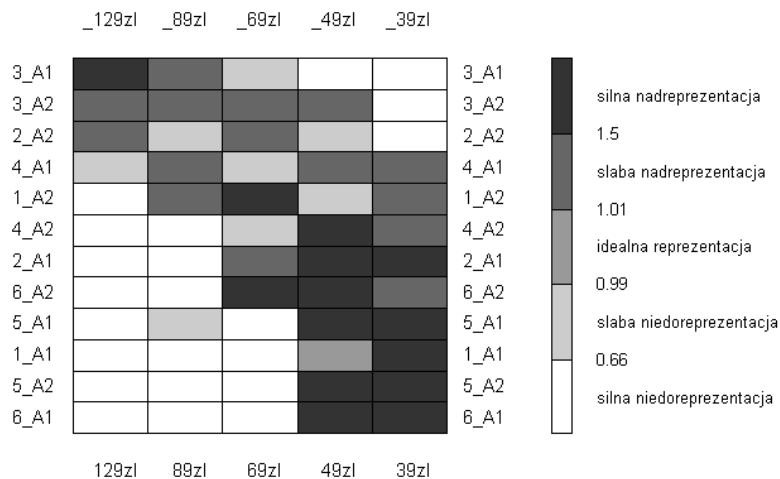
Źródło: opracowanie własne

Na rys. 4 widzimy wyraźnie, że pewne pakiety są bardzo licznie reprezentowane w określonych grupach wiekowych, a inne wprost przeciwnie. Rozważmy np. pakiety najtańsze, w cenie 39 zł i 49 zł, które są najczęściej wybierane przez klientów najmłodszych oraz tych powyżej 45 roku życia. Z kolei pakiety najdroższe są najpopularniejsze wśród klientów w trzeciej kategorii wiekowej (36-45 lat).

W następnym kroku surowe dane w postaci wskaźników nadreprezentacji zostały poddane w pakiecie GradeStat właściwej analizie polegającej na wyznaczeniu wartości współczynników regresji, zarówno dla wierszy oraz kolumn, celem reprezentacji poziomu korelacji gradacyjnej pomiędzy wymiarem kolumn, a wymiarem wierszy. W tym celu algorytm poszukuje w pętli, rozpoczynając od losowego uporządkowania wierszy i kolumn tabeli, takiego ustawienia wierszy i kolumn, dla którego dane ρ^* osiąga maksimum lokalne. W postępowaniu tym wykorzystywany jest fakt, że jeśli któraś z regresji wierszy lub kolumn nie jest niemalejąca, to przestawiając odpowiednio wiersze lub kolumny zgodnie z wartościami regresji, otrzymujemy rozkład o większej wartości współczynnika ρ^* . Najwyższa wartość ρ^* wybrana ze zbioru maksimum lokalnych jest z reguły równa maksimum globalnemu, co często uzyskiwane jest przy niewielkiej liczbie powtórzeń algorytmu.

Wynikiem ustawienia są duże wartości wskaźnika nadreprezentacji w lewym górnym rogu oraz prawym dolnym rogu tabeli. Z kolei małe wartości wskaźnika nadreprezentacji znajdują się w lewym dolnym oraz prawym górnym rogu tabeli. Wizualizację końcowego ustawienia dla rozważanych danych prezentuje mapa nadreprezentacji na rys. 5.

Rysunek 5. Mapa nadreprezentacji po przekształceniach



Źródło: opracowanie własne

Na podstawie wizualizacji z rys. 5, można zauważyć pewien związek pomiędzy grupą wiekową, a zakupem określonego pakietu internetowego. Najdroższe pakiety, za 89 zł i 129 zł, są wybierane przez klientów w trzeciej grupie wiekowej (26-35 lat), z kolei pakiety najtańsze, w cenie 39 zł i 49 zł są nabywane przez klientów najmłodszych (z pierwszej grupy wiekowej) oraz klientów najstarszych (z piątej oraz szóstej kategorii wiekowej). Można również zauważyć, że pakiety w cenie 39 zł i 49 zł są także często nabywane przez klientów z drugiej kategorii wiekowej, a dokładnie tych, którzy płacą niskie rachunki (kategoria A1). Z kolei pakiet w cenie 49 zł jest popularny (silna nadreprezentacja) wśród klientów z czwartej grupy wiekowej płacących wyższe rachunki (kat. A2).

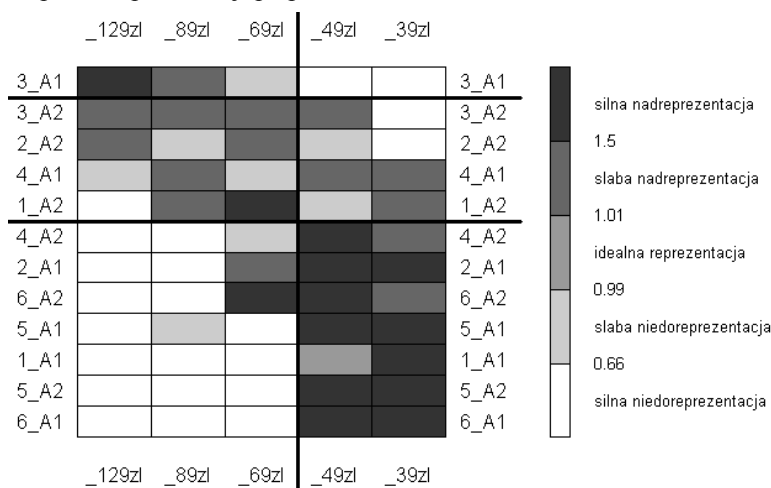
Kolejny krok w badaniu polegał na pogrupowaniu badanej zbiorowości klientów na jednorodne grupy nabywców. W obecnych, bardzo konkurencyjnych, warunkach rynkowych segmentacja jest częstym podejściem, za którym stoi szereg przesłanek m.in. [Kotler 1999]:

- w otoczeniu przedsiębiorstwa zachodzą ciągle zmiany, co powoduje, że rynek nie ma charakteru homogenicznego, a postępowanie nabywców na rynku staje się coraz bardziej różnorodne;
- z ogółu nabywców danego produktu można wyróżnić podgrupy o podobnych cechach i zachowaniach;
- podgrupy te są mniej liczne i jednocześnie bardziej homogeniczne niż grupa jako całość;
- działanie na rzecz mniejszej i podobnej podgrupy nabywców w odróżnieniu od dużych, zróżnicowanych grup nabywców jest łatwiejsze i bardziej skuteczne.

Do wyznaczenia segmentów posłużyła dostępna w GradeStat procedura analizy skupień. Proces wyszukiwania skupień dla danych był przeprowadzony

według zaimplementowanego algorytmu krzywa korelacji/rozkład brzegowy dla zmiennych [Ciok 2004], [Kowalczyk i in. 2004]. Efektem tej analizy jest rys. 6, który prezentuje sześć najbardziej jednorodnych segmentów ze względu na cenę pakietu oraz kategorię wiekową i dochodową.

Rysunek 6. Mapa nadreprezentacji po przekształceniach



Źródło: opracowanie własne

Wyróżnione na rys. 6 segmenty mogą stanowić podstawę do różnicowania oferty rynkowej dla określonych grup klientów. Dzięki czemu można zaproponować klientom te pakiety, z których będą najbardziej zadowoleni. Ponadto, w marketingu bezpośrednim, jest to niezwykle cenna informacja, dzięki której łatwiej jest przeprowadzać kampanie akwizycyjne. Przykładowo, biorąc pod uwagę wiek klienta i dotychczasową informację o wydatkach (kategoria dochodowa), łatwiej jest zaproponować określony pakiet, który wpisuje się w charakterystykę danego segmentu. Trudno oczekiwać, że klient z piątej lub szóstej grupy wiekowej będzie zainteresowany pakietem dostępu do Internetu w najwyższej cenie. Działanie akwizycyjne prowadzone zgodnie z przyjętą mapą segmentów, pozwolą zaoszczędzić czas oraz środki, co w efekcie może przełożyć się na większy zwrot nakładów z kampanii.

PODSUMOWANIE

Otrzymane w badaniu rezultaty pozwalają na sformułowanie trzech zasadniczych wniosków.

1. Podejście gradacyjne, poprzez identyfikację zależności między zmiennymi i obiektami, jest pomocnym narzędziem do opisu danych wielowymiarowych. Dużą zaletą jest możliwość wizualizacji, przez co porównywanie i ocena podobieństwa w danych jest łatwiejsza.

2. Atrakcyjność oferty dostępu do mobilnego internetu różni się w zależności od grupy wiekowej klientów oraz ich kategorii dochodowej. Można zauważyć, że pakiety najtańsze wybierane są przez klientów najmłodszych i najstarszych. Z kolei pakiety najdroższe są nabywane przez klientów z grupy wiekowej 26-35 lat..
3. Wyznaczone w badaniu segmenty nabywców o podobnych cechach, mogą być podstawą do różnicowania oferty rynkowej dla określonych grup klientów. W efekcie, tak zidentyfikowane grupy mogą zostać wykorzystane do tworzenia odpowiednio dobranych strategii marketingowych w obszarze akwizycji nowych klientów.

BIBLIOGRAFIA

- Borkowski B., Szczesny W. (2005) Metody wizualizacji danych wielowymiarowych jako narzędzie syntezy informacji, *Roczniki Naukowe SERIA*, Tom VII, zeszyt 5, s. 11-15.
- Ciok A. (2004) On the number of clusters – a grade approach, IPI PAN, Warszawa.
- Kotler P. (1999) *Marketing. Analiza, planowanie, wdrażanie i kontrola*, Felberg, Warszawa.
- Kowalczyk T., Pleszczyńska E., Ruland F. eds. (2004) *Grade Models and Methods of Data Analysis. With applications for the Analysis of Data Population, Studies in Fuzziness and Soft Computing*, vol. 151, Springer Verlag, Berlin - Heidelberg - New York.
- Książek J., Matyja O., Pleszczyńska E., Wiech M. (2005) *Analiza danych medycznych i demograficznych przy użyciu programu GradeStat*. Instytut Podstaw Informatyki PAN i Instytut Pomnik – Centrum Zdrowia Dziecka, Warszawa.
- Pleszczyńska E., Jarochońska E., Szczesny W. (2006) Wielowymiarowa analiza danych oparta na modelach gradacyjnych z implementacją w programie GradeStat. Materiały VI Krajowej Konferencji Inżynieria Wiedzy i Systemy Ekspertowe, Wrocław.
- Szczesny W. (2002) Grade correspondence analysis applied to contingency tables and questionnaire data, *Intelligent Data Analysis*, vol. 6(1), IOS Press, Amsterdam, s. 17-51.
- UKE (2010) Raport o stanie rynku telekomunikacyjnego w Polsce w 2010 r., dostępny na stronie www.uke.gov.pl

GRADE ANALYSIS TO ASSESS THE ATTRACTIVENESS OF INTERNET PACKAGES

Abstract: The aim of the study was to analyze the attractiveness of the mobile internet offer. In this work the grade analysis approach was used to visualize the data and to assess the similarity of internet packages in terms of price and customers age groups. This allowed formulating conclusions that the attractiveness of mobile internet packages varies depending on the customers' age and income category. At the same time research has identified segments of customers with similar characteristics.

Keywords: grade analysis, internet, overrepresentation