

PIOTR KULCZYCKI, KARINA DANIEL

## METODA WSPOMAGANIA STRATEGII MARKETINGOWEJ OPERATORA TELEFONII KOMÓRKOWEJ

### 1. WSTĘP

Rynek telefonii komórkowej jest nieustannie poddawany działaniom mającym na celu dostosowanie się do ulegających ciągłym zmianom wymagań klientów. Zarówno dynamiczny rozwój tego rynku, jak i poszerzający się zakres oferowanych usług, implikują coraz bardziej ekspansywną postawę operatorów wobec poszczególnych grup abonenckich. Ze względu na nasilającą się szczególnie w tej dziedzinie konkurencję ważne jest poszukiwanie narzędzi, które stanowiąc będą skuteczną pomoc w powiększaniu efektywnej aktywności firm operatorskich. Precyzyjne zdefiniowane potrzeb każdego klienta skutkuje zwiększeniem się szansy utrzymania go u danego operatora i uzyskania z tego tytułu możliwie jak największych zysków. Z kolei uniknięcie nadmiernego rozdrobnienia w charakteryzowaniu poszczególnych grup abonentów, a także utraty spójności w strategii postępowania wobec klientów, wymaga wprowadzenia w tym zakresie nowych rozwiązań o charakterze systemowym.

Powyższe szczególnie dotyczy rynku klientów korporacyjnych – grupy specyficznej i bardzo wymagającej. Kreowanie oferty wymaga w tym przypadku znacznej elastyczności w zakresie proponowanych usług oraz otwartości na indywidualne rozwiązania, które mogą stać się elementem budującym i utrwalającym współpracę, a także gwarantującym ją w dłuższym horyzoncie czasowym.

Przedmiotem prezentowanych tu badań jest zagadnienie wspomaganie decyzji dotyczącej określenia najkorzystniejszej strategii postępowania wobec konkretnego klienta korporacyjnego – poprzez odpowiednie zastosowanie zniżek – w celu maksymalizacji zysku operatora sieci telefonii komórkowej, przy jednoczesnym zapewnieniu satysfakcjonującej abonenta jakości usług. Procedura ta została stworzona z wykorzystaniem metodyki statystycznych estymatorów jądrowych, a jej postać oraz wyniki otrzymane na podstawie prowadzonych badań zweryfikowane zostały z użyciem danych rzeczywistych uzyskanych od jednego z polskich operatorów sieci GSM, w oparciu o wiedzę teoretyczną i doświadczenie wynikające z praktyki zawodowej związanej z rynkiem biznesowym.

Statystyczne estymatory jądrowe wykorzystano w prezentowanej pracy do zagadnień wykrywania elementów nietypowych (odosobnionych), analizy skupień (klasteryzacji) oraz klasyfikacji. W opracowanym algorytmie zastosowano także teorię zbiorów rozmytych, umożliwiającą przede wszystkim wykorzystanie wiedzy eksperta – znawcy specyfiki funkcjonowania rynku biznesowego abonentów sieci telefonii komórkowej.

Niniejszy artykuł jest podzielony na trzy zasadnicze części. Sekcję drugą stanowią preliminaria matematyczne – przedstawiono tu wybrane zagadnienia związane z estymatorami jądrowymi, a także ich zastosowaniami do wykrywania elementów odosobnionych, jak również w algorytmach analizy skupień oraz klasyfikacji. Sekcja trzecia stanowi główną część pracy – został tu przedstawiony i szczegółowo opisany algorytm wspomagania decyzji dotyczącej strategii marketingowej operatora sieci telefonii komórkowej. W poszczególnych podsekcjach kolejno zawarto opis struktury, analizę i przetwarzanie danych, sposób określania tablic ryzyka z wykorzystaniem liczb rozmytych typu L-R, a także wyznaczania funkcji celu oraz proces klasyfikacji klienta, co kompletuje prezentowaną tu formułę wspomagania decyzji. W sekcji czwartej przedstawiono wyniki otrzymane na podstawie badań przeprowadzonych na zbiorze rzeczywistych danych. W pełni potwierdziły one poprawność i użyteczność prezentowanej tu procedury, dostarczając także wniosków i komentarzy oraz ilustracyjnych interpretacji.

## 2. PRELIMINARIA MATEMATYCZNE

### 2.1. STATYSTYCZNE ESTYMATORY JĄDROWE

Estymatory jądrowe należą do grupy metod nieparametrycznych, w których nie formułuje się arbitralnie założeń dotyczących przynależności rozkładu badanej zmiennej do określonej klasy. Niech zatem dana będzie  $n$ -wymiarowa zmienna losowa, której rozkład posiada funkcję gęstości  $f$ . Jej estymator jądrowy  $\hat{f}: \mathbb{R}^n \rightarrow [0, \infty)$  wyznaczany jest na podstawie wartości  $m$ -elementowej próby losowej  $x_1, x_2, \dots, x_m$  i w swej podstawowej postaci dany jest wzorem

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right), \quad (1)$$

przy czym dodatni współczynnik  $h$  określa się mianem parametru wygładzania, natomiast mierzalna funkcja  $K: \mathbb{R}^n \rightarrow [0, \infty)$ , która jest symetryczna względem zera (tj.  $K(x) = K(-x)$  dla każdego  $x \in \mathbb{R}^n$ ), posiada w tym punkcie słabe maksimum globalne (czyli  $K(0) \geq K(x)$  dla każdego  $x \in \mathbb{R}^n$ ) oraz spełnia warunek jednostkowej całki (tzn.  $\int_{\mathbb{R}^n} K(x) dx = 1$ ), nazywana jest jądrem. Wyboru postaci jądra  $K$  i wartości parametru wygładzania  $h$  dokonuje się najczęściej w oparciu o kryterium minimalizacji scałkowanego błędu średniokwadratowego. Okazuje się, iż postać jądra  $K$  nie wpływa w znaczącym stopniu na jakość statystyczną estymacji – stanowi to istotną praktyczną zaletę prezentowanej metody, gdyż pozwala zagwarantować cechy wymagane od estymatora (1) w konkretnym zastosowaniu, na przykład odpowiednią klasę regularności, przyjmowanie dodatnich wartości lub dogodność obliczeniową. W przeciwieństwie, istotne znaczenie ma właściwe wyznaczenie wartości parametru wygładzania  $h$  – istnieją wszakże dogodne algorytmy pozwalające obliczyć powyższą wartość na podstawie posiadanej próby losowej. W praktycznych zadaniach stosuje się ponadto dodatkowe

procedury polepszające własności estymatora (np. modyfikację parametru wygładzania lub transformację liniową) oraz dopasowujące jego własności do specyfiki rozważanego zagadnienia (m.in. ograniczenie nośnika). Estymatory jądrowe pozwalają również oszacować inne charakterystyki rozkładów zmiennych losowych, np. dystrybuantę lub wartość kwantyla. Szczegółowy opis powyższej metodyki można znaleźć w książce [11] oraz klasycznych monografiach [14, 15].

Statystyczne estymatory jądrowe stanowią dogodne narzędzie matematyczne przy rozwiązywaniu różnorodnych zadań z zakresu analizy i przetwarzania danych. W niniejszej pracy zostały one zastosowane do zagadnień wykrywania elementów odosobnionych, analizy skupień oraz klasyfikacji – zagadnienia te będą przedstawione poniżej w kolejnych podsekcjach.

## 2.2. WYKRYWANIE ELEMENTÓW ODOSONIONYCH

W wielu zagadnieniach analizy danych występuje zadanie wykrywania elementów nietypowych (odosobnionych), czyli takich, które w istotny sposób różnią się od ogółu populacji. Często umożliwia to eliminację tego typu elementów z posiadanego zbioru danych, co zwiększa jego homogeniczność (jednorodność), ułatwiając wnioskowanie, zwłaszcza w złożonych i nietypowych przypadkach. W praktyce, proces wykrywania elementów odosobnionych najczęściej jest dokonywany z wykorzystaniem procedur testowania hipotez statystycznych [1]. Poniżej zostanie przedstawiony test istotności, którego konstrukcja została oparta na metodyce estymatorów jądrowych.

Niech zatem dana będzie próba losowa  $x_1, x_2, \dots, x_m$  uznawana za wzorcową, czyli zawierająca możliwie jak najbardziej reprezentatywny zbiór elementów typowych. Niech ponadto  $\alpha \in (0,1)$  oznacza założony poziom istotności. Testowana jest hipoteza stanowiąca, że ustalony element  $\tilde{x} \in \mathbb{R}^n$  jest typowy, przeciwko hipotezie iż takim nie jest, czyli iż należy go uznać za odosobniony. Stosowaną tu statystykę  $S: \mathbb{R}^n \rightarrow [0, \infty)$  definiuje się wzorem

$$S(\tilde{x}) = \hat{f}(\tilde{x}), \quad (2)$$

przy czym  $\hat{f}$  oznacza jądrowy estymator gęstości otrzymany dla wzmiankowanej powyżej wzorcowej próby losowej  $x_1, x_2, \dots, x_m$ , podczas gdy zbiór odrzucenia (krytyczny) przyjmuje postać lewostronną  $(-\infty, a]$ , gdzie  $a$  stanowi jądrowy estymator kwantyla stopnia  $\alpha$  wyznaczony na podstawie próby  $\hat{f}(x_1), \hat{f}(x_2), \dots, \hat{f}(x_m)$ , przy założeniu iż nośnik zmiennej losowej jest ograniczony do zbioru liczb nieujemnych.

## 2.3. ANALIZA SKUPIEŃ

Celem analizy skupień jest dokonanie podziału zbioru danych – na przykład określonych w postaci próby losowej  $x_1, x_2, \dots, x_m$  na podgrupy (skupienia, klastry), z których każda zawiera „podobne” do siebie elementy, ale istotnie różniące się między poszczególnymi podgrupami [10]. W praktyce często pozwala to zdekomponować

duże zbiory danych o zróżnicowanych charakterystykach poszczególnych elementów, na podzbiory zawierające elementy o podobnych własnościach, co znacząco ułatwia lub wręcz umożliwia dalsze, szczegółowe wnioskowanie. W niniejszej podsekcji zostanie przedstawiona procedura analizy skupień oparta na statystycznych estymatorach jądrowych, wykorzystująca ideę metod gradientowych [5].

Przyjmuje się tu naturalne założenie, iż poszczególne skupienia odpowiadają modom (lokalnym maksimum) jądrowego estymatora gęstości  $\hat{f}$  określonego na podstawie rozważanej próby losowej  $x_1, x_2, \dots, x_m$ . W ramach wspomnianej procedury, poszczególne elementy są przemieszczane w kierunku zdefiniowanym przez gradient, według następującego algorytmu iteracyjnego:

$$x_j^0 = x_j \text{ dla } j = 1, 2, \dots, m \quad (4)$$

$$x_j^{k+1} = x_j^k + b \frac{\nabla \hat{f}(x_j^k)}{\hat{f}(x_j^k)} \text{ dla } j = 1, 2, \dots, m \text{ oraz } k = 1, 1, \dots, \quad (5)$$

przy czym  $b > 0$  (w praktyce rekomenduje się  $b = h^2/(n + 2)$ ), natomiast  $\nabla$  oznacza operator gradientu. Dzięki odpowiedniemu doborowi postaci jądra  $K$ , możliwe jest uzyskanie dogodnej, analitycznej formuły gradientu  $\nabla \hat{f}$ .

W wyniku wykonywania kolejnych kroków iteracyjnych, elementy próby losowej przemieszczają się sukcesywnie, coraz wyraźniej grupując w pewną liczbę skupień. Ich ostateczną postać określa się po wykonaniu ustalonego  $k^*$ -tego kroku<sup>1</sup>. W tym celu obliczany jest estymator jądrowy wzajemnych odległości elementów  $x_1^{k^*}, x_2^{k^*}, \dots, x_m^{k^*}$  (przy założeniu nieujemności nośnika zmiennej), po czym zostaje wyznaczona najmniejsza wartość, w której estymator ten osiąga pierwsze lokalne minimum, z pominięciem ewentualnego minimum w zerze. Do poszczególnych skupień zalicza się te elementy, których odległość jest nie większa niż tak wyznaczona wartość. Dzięki możliwości zmiany wartości parametru wygładzania  $h$ , możliwa staje się ingerencja w rząd wielkości liczby otrzymywanych skupień, aczkolwiek bez arbitralnych założeń dotyczących wartości tej liczby, co umożliwia dopasowanie jej do rzeczywistej struktury danych. Co więcej, ewentualne zmiany intensywności procedury modyfikacji parametru wygładzania, umożliwiają wpływ na proporcję liczby skupień wyznaczonych w obszarach zagęszczenia elementów posiadanej próby losowej do liczby skupień położonych w rejonach, gdzie są one rzadkie. Szczegółowy opis powyższego algorytmu analizy skupień zawarto w artykule [12].

#### 2.4. KLASYFIKACJA

Niech dana będzie liczba  $J \in \mathbb{N} \setminus \{0, 1\}$ . Załóżmy także, iż pozyskana z  $n$ -wymiarowej zmiennej losowej próba  $x_1, x_2, \dots, x_m$  została podzielona na  $J$  rozłącznych podzbiorów:

<sup>1</sup> Dla rozważanego w tej pracy zbioru danych, po przeprowadzeniu szczegółowych wstępnych badań, przyjęto  $k^* = 50$ . W pracy [11] znaleźć można ogólną formułę kryterium wyznaczania wartości  $k^*$ .

$$x_1^1, x_2^1, \dots, x_{m_1}^1 \quad (6)$$

$$x_1^2, x_2^2, \dots, x_{m_2}^2 \quad (7)$$

$$\vdots$$

$$x_1^J, x_2^J, \dots, x_{m_J}^J \quad (8)$$

gdzie  $m_1, m_2, \dots, m_j \in \mathbb{N} \setminus \{0\}$  oraz  $\sum_{j=1}^J m_j = m$ , reprezentujących wzorce wyróżnionych klas, przy czym wprowadzony dodatkowo górny indeks stanowi o przynależności elementu do wzorca danej klasy. Zadanie klasyfikacji polega na wskazaniu, do którego z nich należy przypisać ustalony element  $\tilde{x} \in \mathbb{R}^n$  [10].

Metodyka estymatorów jądrowych dostarcza naturalnego aparatu matematycznego do rozwiązania powyższego zagadnienia w optymalnym – w sensie minimum wartości oczekiwanej strat – ujęciu bayesowskim. Otóż niech  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_j$  oznaczają jądrowe estymatory gęstości określone dla prób odpowiednio (6)-(8). Jeżeli licznosci  $m_1, m_2, \dots, m_J$  są proporcjonalne do „częstości” pojawiania się elementów z poszczególnych klas, to rozważany element  $\tilde{x}$  należy zaliczyć do tej klasy, dla której wartość  $m_1 \hat{f}_1(\tilde{x}), m_2 \hat{f}_2(\tilde{x}), \dots, m_J \hat{f}_J(\tilde{x})$  jest największa.

### 3. ALGORYTM WSPOMAGANIA DECYZJI DOTYCZĄCEJ STRATEGII MARKETINGOWEJ

W poniższej części zostanie przedstawiony opis prezentowanej tu metody wspomaganie strategii marketingowej operatora telefonii komórkowej wobec klienta korporacyjnego. Sama procedura jest wieloetapowa i z tego powodu jej opis został podzielony na poszczególne podsekcje, w których zawarte będą kolejne fazy procesu analizy i przetwarzania danych oraz wspomaganie procesu decyzyjnego.

#### 3.1. STRUKTURA DANYCH

W praktyce istnieje rozległe spektrum wielkości charakteryzujących poszczególnych abonentów. Dzieje się tak przede wszystkim ze względu na ciągły rozwój rynku telefonii komórkowej, związany między innymi z wprowadzaniem nowych rozwiązań mobilnych w zakresie szeroko rozumianych usług głosowych i transmisji danych, mających na celu zaspokojenie zwiększających się potrzeb klientów. Przystępując do analizy należy spośród różnorodnych wielkości wytypować te, które możliwie najlepiej będą opisywać rozważane zagadnienie przy jednoczesnej gwarancji efektywnej ich aktualizacji. Warto również zaznaczyć, że nie wszystkie z badanych charakterystyk mogą być w praktyce identyfikowalne w równych odstępach czasu, część z nich ma charakter losowy, a niektóre sezonowy. Co więcej, niektóre z nich są tworzone i obserwowane w nieregularnych odstępach czasu, gdy przykładowo służą one weryfikacji skuteczności zastosowania jednorazowych działań operatora mających na celu polepszenie relacji z klientami lub zweryfikowanie słuszności wprowadzenia konkretnego produktu.

$$x_1^1, x_2^1, \dots, x_{m_1}^1 \quad (6)$$

$$x_1^2, x_2^2, \dots, x_{m_2}^2 \quad (7)$$

$$\vdots$$

$$x_1^J, x_2^J, \dots, x_{m_J}^J \quad (8)$$

gdzie  $m_1, m_2, \dots, m_j \in \mathbb{N} \setminus \{0\}$  oraz  $\sum_{j=1}^J m_j = m$ , reprezentujących wzorce wyróżnionych klas, przy czym wprowadzony dodatkowo górny indeks stanowi o przynależności elementu do wzorca danej klasy. Zadanie klasyfikacji polega na wskazaniu, do którego z nich należy przypisać ustalony element  $\tilde{x} \in \mathbb{R}^n$  [10].

Metodyka estymatorów jądrowych dostarcza naturalnego aparatu matematycznego do rozwiązania powyższego zagadnienia w optymalnym – w sensie minimum wartości oczekiwanej strat – ujęciu bayesowskim. Otóż niech  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_j$  oznaczają jądrowe estymatory gęstości określone dla prób odpowiednio (6)-(8). Jeżeli licznosci  $m_1, m_2, \dots, m_J$  są proporcjonalne do „częstości” pojawiania się elementów z poszczególnych klas, to rozważany element  $\tilde{x}$  należy zaliczyć do tej klasy, dla której wartość  $m_1 \hat{f}_1(\tilde{x}), m_2 \hat{f}_2(\tilde{x}), \dots, m_J \hat{f}_J(\tilde{x})$  jest największa.

### 3. ALGORYTM WSPOMAGANIA DECYZJI DOTYCZĄCEJ STRATEGII MARKETINGOWEJ

W poniższej części zostanie przedstawiony opis prezentowanej tu metody wspomaganie strategii marketingowej operatora telefonii komórkowej wobec klienta korporacyjnego. Sama procedura jest wieloetapowa i z tego powodu jej opis został podzielony na poszczególne podsekcje, w których zawarte będą kolejne fazy procesu analizy i przetwarzania danych oraz wspomaganie procesu decyzyjnego.

#### 3.1. STRUKTURA DANYCH

W praktyce istnieje rozległe spektrum wielkości charakteryzujących poszczególnych abonentów. Dzieje się tak przede wszystkim ze względu na ciągły rozwój rynku telefonii komórkowej, związany między innymi z wprowadzaniem nowych rozwiązań mobilnych w zakresie szeroko rozumianych usług głosowych i transmisji danych, mających na celu zaspokojenie zwiększających się potrzeb klientów. Przystępując do analizy należy spośród różnorodnych wielkości wytypować te, które możliwie najlepiej będą opisywać rozważane zagadnienie przy jednoczesnej gwarancji efektywnej ich aktualizacji. Warto również zaznaczyć, że nie wszystkie z badanych charakterystyk mogą być w praktyce identyfikowalne w równych odstępach czasu, część z nich ma charakter losowy, a niektóre sezonowy. Co więcej, niektóre z nich są tworzone i obserwowane w nieregularnych odstępach czasu, gdy przykładowo służą one weryfikacji skuteczności zastosowania jednorazowych działań operatora mających na celu polepszenie relacji z klientami lub zweryfikowanie słuszności wprowadzenia konkretnego produktu.

Ostatecznie, po szczegółowej analizie aspektów ekonomicznych rozważanego w niniejszej pracy zagadnienia przyjęte zostało, iż podstawową charakterystykę poszczególnych klientów stanowią trzy wielkości:

- średni miesięczny przychód przypadający na kartę SIM (ARPU),
- staż w sieci,
- liczba aktywnych kart SIM.

Przeprowadzone badania potwierdziły, iż poszerzanie powyższej listy o dalsze wielkości praktycznie nie polepsza otrzymywanych wyników, natomiast zwiększa czas i złożoność obliczeń.

W dalszej części pomocniczo rozważana będzie również wartość średniego abonamentu danego klienta (średnia arytmetyczna wszystkich abonamentów jego konta). Iloczyn liczby kart SIM oraz tak wyznaczonego średniego abonamentu stanowi główną część stałych przychodów, jakie dany klient generuje dla operatora. Pozostały przychód uzależniony jest od czasu połączeń wykonywanych poza pakietem uwzględnionym w kwocie abonamentu. Jest on związany z rodzajem usług, z których korzysta klient, dlatego też coraz częściej uwzględnia się nie tylko opłaty za połączenia głosowe, ale także za transmisję danych oraz inne usługi bezpośrednio z nią związane.

Uzupełnienie dla prowadzonych analiz stanowią będą także pomocnicze wielkości o charakterze ilościowym i jakościowym, które w podlegającym badaniom zbiorze klientów wyróżniają jednostki prestiżowe, a także firmy transportowe oraz typu ochrona-monitoring, jak również przedsiębiorstwa małe, średnie, duże oraz bardzo duże. Zostaną one wykorzystane w procesie eliminacji nietypowych elementów i mało licznych skupień.

Niech zatem charakterystyka poszczególnych klientów dana będzie w postaci 3-wymiarowego wektora:

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} \text{ dla } i = 1, 2, \dots, m, \quad (9)$$

gdzie  $x_{i,1}$  stanowi średni miesięczny przychód na kartę SIM  $i$ -tego klienta,  $x_{i,2}$  – jego staż w sieci,  $x_{i,3}$  – liczbę kart SIM tegoż klienta, podczas gdy  $m$  stanowi liczbę posiadanej bazy danych. Dodatkowo przez  $y_{i,1}$  oznaczona zostanie wielkość jego średniego abonamentu, natomiast zmienne  $y_{i,2}$ ,  $y_{i,3}$ ,  $y_{i,4}$  określono następująco:

$y_{i,2}$  – przynależność do jednostek prestiżowych charakteryzuje się za pomocą zmiennej skategoryzowanej o następujących wartościach: 0 – klient nieprestiżowy, 1 – klient istotny dla operatora z punktu widzenia strategii firmy, zwany dalej prestiżowym, 2 – klient wyjątkowo prestiżowy (kluczowy);

$y_{i,3}$  – wielkość przedsiębiorstwa charakteryzowana jest przez zmienną skategoryzowaną o wartościach: 1 – małe przedsiębiorstwo, 2 – średnie przedsiębiorstwo, 3 – duże przedsiębiorstwo, 4 – bardzo duże przedsiębiorstwo;

$y_{i,4}$  – dla firm transportowych oraz typu ochrona-monitoring jako jej wartość przypisuje się odpowiednio 1 oraz  $-1$ , natomiast 0 dla jednostek o innym profilu działalności.

Powyższe wielkości zostaną użyte do scharakteryzowania poszczególnych klientów korporacyjnych w procesie analizy i przetwarzania danych przedstawionym w dalszej części artykułu.

### 3.2. ANALIZA I PRZETWARZANIE DANYCH

We wstępnej fazie analizy, ze zbioru danych eliminowane są jednostki nietypowe (elementy odosobnione), zgodnie z procedurą przedstawioną w podsekcji 2.2. Zastrzeżone zostaje jednak, iż nie można usunąć tu jednostek kluczowych oraz wszystkich jednostek z grupy transportowych lub typu ochrona-monitoring. Niniejszym zwiększona zostaje jednolitość struktury danych, a efekt ten otrzymuje się poprzez pominięcie jedyne tych elementów, które miałyby pomniejsze znaczenie w dalszej procedurze.

Po selektywnym wyeliminowaniu elementów nietypowych, dokonuje się analizy skupień zbioru danych, z użyciem algorytmu opisanego w podsekcji 2.3. W rezultacie uzyskuje się podział zbioru danych, reprezentujących poszczególnych klientów korporacyjnych, na podgrupy elementów o zbliżonej charakterystyce. Następnie dokonuje się ponownej, nieco odmiennej eliminacji elementów nietypowych, realizowanej poprzez usunięcie skupień o małej liczności, aczkolwiek z ponownym zastrzeżeniem pozostawienia skupień zawierających jednostki kluczowe, a ponadto skupień składających się co najmniej w połowie z klientów prestiżowych, jak również pozostawienia co najmniej jednego skupienia z firmą transportową oraz jednego skupienia z jednostką typu ochrona-monitoring.

W konsekwencji przeprowadzonego powyżej procesu analizy i przetwarzania danych, ich zbiór pomniejszony jest o elementy wnoszące pomniejszą informację przedmiotową, a pozostałą jego część grupuje się w skupienia zawierające elementy o zbliżonej – z punktu widzenia rozważanego zadania – charakterystyce.

### 3.3. OKREŚLENIE TABLIC RYZYKA

Podstawę przedstawionej w niniejszej podsekcji procedury wspomagania decyzji podejmowanej wobec firm o charakterystyce właściwej dla poszczególnych, wyznaczonych uprzednio skupień, stanowić będzie wiedza eksperta skutkująca zdefiniowaniem wprowadzonych dla potrzeb dalszej części procedury tabel nazywanych poniżej tablicami ocen rezygnacji.

Na podstawie praktyki tworzenia ofert dla abonentów przyjęto dwie zmienne decyzyjne, których wartości stanowią kolejno o wielkości udzielonego upustu na abonament (oznaczana jako  $d_a$ ) oraz wielkości rabatu za minutę połączenia (oznaczana przez  $d_m$ ). Zmienne te zostały wybrane ze względu na fakt, iż stanowią one w każdym procesie negocjacyjnym prowadzonym między operatorem a klientem, podstawowe narzędzie służące tworzeniu nowej i modyfikowaniu obowiązującej oferty, która ma regulować współpracę w przyszłym okresie. Zakłada się, że wynegocjowany poziom upustów i rabatów obowiązywać będzie przez cały okres umowy między firmą i operatorem, czyli przez 24 miesiące.



W dalszej części pracy przyjęto uzasadnioną praktyką dyskretyzację wartości tych zmiennych do następujących „poziomów”:

- możliwy upust na abonament  $d_a \in \{0\%, 5\%, 10\%, 15\%, 50\%, 99\%\}$ ,
- rabat za minutę połączenia  $d_m \in \{0\%, 3\%, 6\%, 12\%, 25\%, 50\%\}$ .

Dla każdej kombinacji powyższych „poziomów”, ekspert powinien dokonać oceny ryzyka rezygnacji klienta z usług danego operatora. Biorąc pod uwagę przyjęte wartości zniżek, zadaniem eksperta jest przygotowanie dla każdego z otrzymanych skupień, 36 ocen ryzyka. W przypadku eksperta znającego specyfikę badanego rynku, a także zależności w nim występujące, możliwa jest specyfikacja występujących prawidłowości, co istotnie ułatwia określenie takich tablic. Co więcej, dla nieodległych skupień powyższe tablice będą relatywnie podobne. W celu usprawnienia powyższego procesu ekspert może wyznaczyć pewną ilość tablic wzorcowych, które będą stanowić element wyjściowy do dalszych prac, polegających jedynie na ich odpowiednich modyfikacjach. Podstawę do tworzenia tablic wzorcowych powinny stanowić przede wszystkim jednostki kluczowe i prestiżowe oraz firmy transportowe i typu ochrona-monitoring. Dodatkowo ekspert może wykonać modele tablic przygotowane dla klasycznych przedsiębiorstw małych, średnich, dużych i bardzo dużych. Dzięki takiemu ujęciu każda tworzona tablica będzie stanowić połączenie informacji zawartej w odpowiednich tablicach wzorcowych, co korzystnie wpłynie na jakość i czas jego pracy.

Ponieważ praktyka wskazuje, iż tego typu oceny formułuje się w sposób werbalny na podstawie nieprecyzyjnych przesłanek intuicyjnych, to wynik powinien zostać określony w postaci liczby rozmytej. Do powyższego celu użyte będą liczby typu L-R z uwagi na ich prostotę oraz wyrazistą interpretację. Liczba rozmyta  $\mathcal{A}$  jest typu L-R w przypadku, jeśli jej funkcja przynależności dana jest w postaci:

$$\mu_{(w,\alpha,\beta)}(x) = \begin{cases} L\left(\frac{x-w}{\alpha}\right), & \text{gdy } x \leq w \\ R\left(\frac{x-w}{\beta}\right), & \text{gdy } x > w \end{cases}, \quad (10)$$

gdzie  $w \in \mathbb{R}$ ,  $\alpha, \beta > 0$ , a ponadto  $L: (-\infty, 0] \rightarrow [0, 1]$  jest funkcją niemalejącą oraz spełniająca warunek  $L(0) = 1$ , i analogicznie,  $R: [0, \infty) \rightarrow [0, 1]$  jest funkcją nierosnącą oraz  $R(0) = 1$ . Parametry  $\alpha$  oraz  $\beta$  decydują zatem o lewo- i prawostronnym „rozmyciu” powyższej liczby względem wartości  $w$ . Ponadto przyjmuje się dla wartości indeksów  $\alpha = 0$  albo  $\beta = 0$ :

$$\mu_{(w,0,\beta)}(x) = \begin{cases} 0, & \text{gdy } x < w \\ R\left(\frac{x-w}{\beta}\right), & \text{gdy } x \geq w \end{cases} \quad \text{dla } \beta > 0 \quad (11)$$

$$\mu_{(w,\alpha,0)}(x) = \begin{cases} L\left(\frac{x-w}{\alpha}\right), & \text{gdy } x \leq w \\ 0 & \text{gdy } x > w \end{cases} \quad \text{dla } \alpha > 0. \quad (12)$$

Po dokonywanym na wstępie arbitralnym ustaleniu postaci funkcji  $L$  oraz  $R$ , liczbę rozmytą typu L-R definiują trzy parametry  $w$ ,  $\alpha$  oraz  $\beta$ , a zatem można ją zapisać

w postaci  $\mathcal{A} = (w, \alpha, \beta)$  i w konsekwencji proces jej identyfikacji wymaga określenia jedynie tych trzech wartości.

Uwzględniając sposoby określania lojalności klienta korporacyjnego, wynikające ze stosowanej do takich ocen praktyki, przyjęto istnienie siedmiu stanów:

- 0 – praktycznie nie istnieje możliwość rezygnacji;
- 1 – ryzyko odejścia jest nikłe, uzależnione jedynie od czynników losowych;
- 2 – istnieje małe ryzyko rezygnacji;
- 3 – istnieje średnie ryzyko odejścia;
- 4 – istnieje duże ryzyko odejścia;
- 5 – jest omal pewne, że klient zrezygnuje z usług;
- 6 – praktycznie jest pewne, że klient zrezygnuje z usług.

W tej sytuacji można uściślić, iż  $w, \alpha, \beta \in \{0, 1, \dots, 6\}$ , przy czym powinno być spełnione  $w - \alpha \geq 0$  oraz  $w - \beta \leq 6$ .

Na przykładowej tablicy ocen rezygnacji, otrzymanej w trakcie numerycznej weryfikacji prezentowanej tu metody (zamieszczonej w podsekcji 4.3 jako tab. 1), można zauważyć, iż stosunkowo małym rozmyciem charakteryzują się te oceny eksperta, które dotyczą skrajnie małych lub skrajnie dużych zniżek, co świadczy w takich przypadkach o większej precyzyjności oceny pozytywnej albo negatywnej reakcji klienta. Jednakże im większe jest zróżnicowanie w proponowanych poziomach upustów i rabatów, czyli w prawym-górnym i lewym-dolnym obszarze tablicy, tym oceny eksperta stają się nieco mniej precyzyjne. Z powyższych obserwacji wynikają zauważalne prawidłowości przydatne przy określaniu tablic ocen rezygnacji.

Schemat postępowania w stosunku do każdego z typów klientów jest inny, wyszczególnione jednostki swoje decyzje o zmianie operatora podejmują na podstawie odmiennych przesłanek. Niektóre z nich – przede wszystkim jednostki kluczowe i prestiżowe lub bardzo duże przedsiębiorstwa – w sposób bardziej zdecydowany są w stanie podjąć decyzję o rezygnacji z usług firmy świadczącej usługi mobilne w momencie, kiedy ich często zbyt wysokie oczekiwania nie zostają w pełni spełnione przez operatora. Prestiż i wielkość klienta stanowią bowiem dla niego główny argument w procesie prowadzenia negocjacji z operatorem, tym samym formułowane oczekiwania są bardziej rygorystyczne, niż ma to miejsce w przypadku innych klientów. Wobec mniejszych klientów korporacyjnych ekspertowi jest trudniej precyzyjnie określić reakcje na zaproponowane im kombinacje upustów i rabatów. W tym przypadku pozytywna decyzja o pozostaniu w sieci zarówno przy zaproponowaniu niskiej, jak i wysokiej kombinacji zniżek, jest o wiele trudniejsza do ustalenia.

Opracowane tablice stanowią podstawę przy wyznaczaniu funkcji celu, której zadaniem jest wskazanie kombinacji upustu i rabatu przynoszących najmniejsze potencjalne straty operatorowi, uwzględniając iż klient może pozostać w sieci i z uwagi na otrzymane zniżki stać się mniej zyskownym lub – z drugiej strony – zrezygnować z usług danego operatora.

#### 3.4. WYZNACZANIE FUNKCJI CELU

Na podstawie ekonomicznej analizy zagadnienia, w oparciu o elementy teorii preferencji rozmytych [4], przyjęta została następująca postać funkcji celu  $\mathcal{G}$

$$\begin{aligned}
 \mathcal{P}(d_a, d_m; x_i, y_{i,1}, w, \alpha, \beta) = & \frac{\int_{-\infty}^{\infty} x \mu_{(w, \alpha, \beta)}(x) dx}{\int_{-\infty}^{\infty} \mu_{(w, \alpha, \beta)}(x) dx} \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (y_{i,1} + b) + \\
 & + \left( 6 - \frac{\int_{-\infty}^{\infty} x \mu_{(w, \alpha, \beta)}(x) dx}{\int_{-\infty}^{\infty} \mu_{(w, \alpha, \beta)}(x) dx} \right) \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (d_a \cdot y_{i,1} + d_m \cdot b),
 \end{aligned} \tag{13}$$

gdzie  $i_1, i_2, \dots, i_I$  oznaczają indeksy tych elementów zbioru danych, które zostały zaliczone do rozważanego skupienia,  $I$  oznacza jego licznosc,  $b$  jest iloczynem średniej stawki za minutę połączenia oraz średniego czasu połączeń generowanego poza pakietem objętym opłatą abonamentową. W przypadku funkcji przynależności danej dla liczb rozmytych typu L-R wzorami (10)-(12), i typowych postaci funkcji  $L$  oraz  $R$ , jej wartość będzie można wyrazić za pomocą dogodnego wzoru analitycznego.

Równanie (13) określające funkcję celu złożone jest z dwóch składników. Pierwszy z nich przedstawia iloczyn średniej rozmytej oceny odejścia oraz straty wynikłej z rezygnacji z usług, natomiast drugi stanowi iloczyn dopełnienia do średniej rozmytej oceny odejścia i straty powstałej w wyniku przyznania zniżek. Wielkości  $y_{i,1} + b$  oraz  $d_a \cdot y_{i,1} + d_m \cdot b$  reprezentują wartości potencjalnych strat w przypadkach – odpowiednio – nieprzyznania zniżek i rezygnacji z usług, a także zmniejszenia przychodów operatora wynikłych z przyznania rabatów i upustów.

Parametr  $b$  wprowadzony we wzorze (13) jest stały w relatywnie długich okresach czasu i obecnie jego wartość wynosi 24 PLN, jednakże ulega ona okresowej aktualizacji. Wartość tego parametru stanowi iloczyn średniej stawki za minutę połączenia przyjętej jako 0,40 PLN oraz 60 minut czasu połączeń. Ze względu na dynamiczny rozwój rynku telefonii mobilnej, jak również panującą na nim silną konkurencję, skutkującą sukcesywnym obniżaniem stawek za minutę połączenia zarówno u operatorów mobilnych, jak i stacjonarnych, powyższa stawka powinna być uaktualniana co 3-6 miesięcy, aczkolwiek w przypadku utrzymywania się stabilnej sytuacji, akceptowalne może być wydłużenie tego czasokresu do 12 miesięcy.

Tablice ocen rezygnacji określone są odrębnie dla kolejnych skupień. Jak podkreślano wcześniej, wyznaczenie dla każdego z nich 36 ocen dla poszczególnych wariantów rabatów i upustów, w praktyce nie stanowi dla eksperta nadmiernej trudności, biorąc pod uwagę dostępny czas przygotowania takiej informacji oraz istnienie zauważalnej systematyki wartości tych ocen. Następnie dla każdego ze skupień wyznaczane jest minimum funkcji celu, co pozwala wskazać, który wariant wartości upustu na abonament oraz rabatu za minutę połączenia wydaje się najkorzystniejszy z punktu widzenia spodziewanego zysku operatora. Ze względu na możliwość uelastycznienia indywidualnych negocjacji możliwe jest także ewentualne wskazanie tych zniżek, dla których wartość funkcji celu jest niewiele większa od minimalnej. W procesie negocjacji niezmiernie bowiem ważne jest wykorzystanie alternatywnych upustów i rabatów, które

z punktu widzenia klienta mogą wydawać się znacznie korzystniejsze, a operatorowi przynoszą zysk niewiele mniejszy.

W kolejnej podsekcji zostanie przedstawiona procedura klasyfikacji konkretnego klienta do jednego z otrzymanych skupień, co w połączeniu z powyższymi wynikami wyczerpuje prezentowaną tu procedurę.

### 3.5. KLASYFIKACJA KLIENTA

W wyniku przeprowadzonej dotychczas realizacji kolejnych etapów analizy danych dokonano dekompozycji elementów bazy klientów korporacyjnych na rozłączne grupy i ustalono dla każdej z nich odpowiednią strategię wspomagającą proces decyzyjny. Aby skorzystać z powyższych wyników w przypadku negocjacji prowadzonych z konkretnym abonentem, konieczne było przypisanie go do jednej z wyróżnionych w ten sposób grup. Celowi temu służy klasyfikacja klienta przedstawiona w poniższym podrozdziale.

Rozważany klient korporacyjny, z którym prowadzone są negocjacje dotyczące udzielenia mu zniżek, scharakteryzowany jest za pomocą trzech wielkości, w nawiązaniu do wzoru (9) zapisanych jako:

$$\tilde{x} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}, \quad (14)$$

gdzie  $\tilde{x}_1$  oznacza jego średni miesięczny przychód na kartę SIM,  $\tilde{x}_2$  – staż w sieci, natomiast  $\tilde{x}_3$  – liczbę kart SIM tegoż klienta.

W tym przypadku – jeśli prowadzone są re negocjacje warunków umowy – dane te mogą dotyczyć dotychczasowej historii abonenta w konkretnej sieci. Mogą one także stanowić informacje historyczne pochodzące od konkurencji, gdy została podjęta próba przejścia klienta. Istnieje również możliwość nawiązania współpracy z całkowicie nowym użytkownikiem usług telefonii mobilnej – dane potrzebne do przeprowadzenia jego klasyfikacji oparte zostały na oszacowaniach wynikających ze znajomości rynku oraz na planowanej strukturze usług świadczonych temu klientowi. Wielkości charakteryzujące każdego klienta przyjęte do opracowanej analizy są dostępne wprost, na przykład na podstawie danych zawartych w miesięcznych fakturach za korzystanie z usług telekomunikacyjnych. Nie wymagają one dodatkowych przekształceń, jak również specjalnych opracowań. W przypadku wszystkich operatorów potrzebne do klasyfikacji informacje przyjmują formę właściwie identycznych raportów. Dodatkowe, nie zawarte w fakturach dane, na przykład informacja o stażu w danej sieci mogą być łatwo uzupełnione z dokumentacji operatora lub – w przypadku próby przejścia klienta – na podstawie przedstawionych dokumentów.

Podstawą klasyfikacji jest procedura przedstawiona w podsekcji 2.4. W jej wyniku, podlegający klasyfikacji element (14) zostaje przypisany do konkretnej klasy (skupienia). Zniżka jaka powinna być wówczas zaproponowana została już wyznaczona w poprzedniej podsekcji. Należy podkreślić, iż umożliwia to stosowanie tej części algorytmu w czasie rzeczywistym w trakcie pertraktacji z klientem.

Tym samym zostaje ostatecznie skompletowana formuła procedury będącej przedmiotem badań prezentowanych w niniejszym artykule.

#### 4. WERYFIKACJA NUMERYCZNA

Tematykę poniższej sekcji stanowi weryfikacja poprawności działania prezentowanej w niniejszej pracy procedury wspomagania strategii marketingowej operatora telefonii komórkowej w odniesieniu do klienta korporacyjnego. Została ona przeprowadzona na podstawie rzeczywistych danych uzyskanych od jednego z operatorów telefonii komórkowej w Polsce. W warunkach użyczenia zawarta została klauzula nieujawniania nazwy owego operatora, a także przedstawiania otrzymywanych wyników w formie gwarantującej uniemożliwienie skojarzenia poszczególnych elementów bazy danych z konkretnymi abonentami.

Statystyczne estymatory jądrowe zostały wykorzystane do zagadnień eliminacji elementów odosobnionych, analizy skupień i klasyfikacji. Zastosowanie jednej metodyki do wszystkich wymienionych zadań znacząco ułatwiło jej użycie, interpretacje uzyskanych wyników częściowych, a także umożliwiło wielokrotne wykorzystanie w programie poszczególnych bloków funkcjonalnych o uniwersalnym znaczeniu. Wieloetapowość opracowanej metody umożliwiła dodatkowo naturalny podział programu na niezależne moduły, znacznie ułatwiające zarówno sam proces pisania kodu, jak i weryfikację poszczególnych części.

Użyte zostało jądro normalne, z uwagi na jego dużą efektywność, a jednocześnie różniczkowalność i przyjmowanie dodatnich wartości w całej dziedzinie.

##### 4.1. WSTĘPNA ANALIZA DANYCH

Przystępując do badań dysponowano pełnym zbiorem 2264 klientów rynku biznesowego, na koncie których liczba aktywnych kart SIM była większa lub równa 20. Każdy klient charakteryzowany był przez średni miesięczny przychód na kartę SIM, jego staż w sieci, liczbę kart SIM oraz dodatkowo wielkości pomocnicze określające jego średni abonament, przynależność do jednostek prestiżowych, wielkość przedsiębiorstwa, a także wyróżniające firmy transportowe i typu ochrona-monitoring. Najpierw dokonano standaryzacji pierwszych trzech wielkości.

Po wstępnej analizie struktury danych dokonano eliminacji części elementów bazy danych, w wyniku której zostały usunięte firmy liczące mniej niż 30 kart SIM, co w efekcie zmniejszyło ich liczbę do 1825. Ograniczenie to wynikało z faktu, iż firmy o liczbie kart SIM mieszczącej się w granicach 20-29 nie posiadały typowych cech klienta korporacyjnego. Dzięki temu zwiększyła się jednorodność bazy danych i w konsekwencji osiągnięto lepszą reprezentatywność charakterystyki klientów korporacyjnych.

##### 4.2. ELIMINACJA ELEMENTÓW ODOSOBNIONYCH

W wyniku zastosowania procedury eliminacji elementów odosobnionych, z badanego zbioru usuniętych zostało 186 klientów. Wśród zidentyfikowanych elementów

odosobnionych nie znalazły się jednostki o znaczeniu kluczowym ani wszystkie firmy transportowe lub typu ochrona-monitoring. Zbiór 186 wyeliminowanych abonentów zawierał przedstawicieli wszystkich typów działalności.

#### 4.3. ANALIZA SKUPIEŃ

W kolejnym etapie, na powstałej po wyeliminowaniu elementów odosobnionych próbie, zastosowano algorytm analizy skupień. Wyniki uzyskane dla typowej intensywności modyfikacji parametru wygładzania (ustanawianej przez wartość  $c = 0,5$ , por. [11 – podrozdział 3.1.6, 12]), wskazywały na zbyt dużą ilość skupień o niedużej liczności, ułożonych w obszarach małego zagęszczenia elementów próby, a także zbyt liczne główne skupienie zawierające ponad połowę elementów. Zgodnie z własnościami stosowanego algorytmu [12], wartość ta została powiększona do  $c = 1$ . Uzyskano dzięki temu żądany efekt: liczba „peryferyjnych” skupień istotnie zmniejszyła się, a główne skupienie uległo rozbiciu. Otrzymana liczba skupień była zadowalająca, co spowodowało iż ewentualna zmiana wartości parametru wygładzania okazała się zbędna.

Ostatecznie uzyskano podział rozważanej teraz 1639-elementowej próby na 26 skupień o następujących licznosciach: 488, 413, 247, 128, 54, 41, 34, 34, 33, 28, 26, 21, 20, 14, 13, 12, 10, dwa skupienia 4-elementowe, trzy skupienia 3-elementowe, dwa skupienia 2-elementowe i dwa skupienia 1-elementowe. Warto zwrócić uwagę na wyraźnie zarysowane cztery grupy: pierwsza z nich to dwa duże skupienia o licznosciach 488 i 413, następnie dwa średnie 247- i 128-elementowe, po czym małe – dziewięć liczących od 20 do 54 oraz wreszcie 13 skupień zawierających mniej niż 20 elementów.

Następnie przystąpiono do eliminacji skupień o licznosciach mniejszych niż 20, ale z wyłączeniem tych, które zawierają klientów kluczowych (skupienia 14-, 13- i 10-elementowe) i składających się co najmniej w połowie z klientów prestiżowych (skupienie 12-elementowe).

Pozostawione skupienia zawierały firmy transportowe oraz typu ochrona-monitoring, co spowodowało, że warunek zapewnienia reprezentacji tego typu jednostek został spełniony bez dodatkowej ingerencji.

Ostatecznie do dalszej analizy pozostało 17 skupień.

#### 4.4. BUDOWANIE TABLIC OCEN REZYGNACJI

Kolejnym etapem było określenie tablic ocen rezygnacji dla każdego z wyznaczonych powyżej skupień. Podstawą zawartych tam ocen była ekspercka wiedza, oparta na dostępnych charakterystykach jednostek wchodzących w skład poszczególnych skupień. Przykładowo przedstawiona zostanie poniżej tablica (tab. 1) przypisana do skupienia grupującego 54 klientów.

Tabela 1

Tablica ocen rezygnacji dla skupienia 54-elementowego

$(w, \alpha, \beta)$		$d_m$					
		0%	3%	6%	12%	25%	50%
$d_a$	0%	(4,0,2)	(4,0,1)	(4,0,1)	(4,2,1)	(4,2,1)	(4,2,0)
	5%	(4,0,1)	(4,0,1)	(4,1,1)	(3,1,2)	(3,1,2)	(3,2,1)
	10%	(4,0,1)	(4,1,1)	(4,2,1)	(3,2,1)	(3,2,1)	(3,2,0)
	15%	(3,1,2)	(3,2,2)	(3,2,1)	(3,2,1)	(3,2,0)	(3,2,0)
	50%	(3,1,2)	(3,2,1)	(3,2,0)	(2,1,1)	(2,2,0)	(2,2,0)
	99%	(2,1,2)	(2,1,1)	(2,2,1)	(2,2,0)	(2,2,0)	(1,1,0)

Dokonana analiza wchodzących w jego skład elementów wskazała, że skupienie to składa się wyłącznie z klientów dużych, a nawet bardzo dużych. Ponad połowę elementów stanowią tu firmy transportowe. Średnia liczba kart SIM dla tego skupienia wynosi prawie 300, przeciętna wysokość abonamentu kształtuje się na średnim poziomie 54 PLN, aczkolwiek wskaźnik ARPU jest relatywnie duży (około 70 PLN).

Zdaniem eksperta, ryzyko rezygnacji z usług zależy tu głównie od wysokości udzielonego rabatu za minutę połączenia. Dotyczyć on powinien nie tylko połączeń głosowych krajowych, ale w równej mierze połączeń roamingowych. Warto podkreślić jest to, iż w ramach nawet wyższej opłaty abonamentowej, klient – w tym przypadku przeważnie przedsiębiorstwo transportowe – oczekiwać będzie stałych, możliwie najniższych stawek za usługi związane z intensywnością rozmów, a ogólniej transmisji danych, które generują wysokie koszty jednostkom prowadzącym tego typu działalność.

W rozpatrywanym przypadku oceny rezygnacji z usług uwzględniają – poprzez średnie wartości dla małych zniżek – niedużą możliwość utraty tego typu klienta, głównie wynikłą stąd, iż zmiana operatora dla jednostek prowadzących działalność transportową jest wyjątkowo kłopotliwa, ze względu na dużą mobilność użytkowników poszczególnych numerów.

#### 4.5. WYZNACZANIE FUNKCJI CELU DLA POSZCZEGÓLNYCH SKUPIEŃ

Kolejnym etapem prezentowanej procedury było wyznaczenie dla każdego elementu tablic ocen rezygnacji wartości funkcji celu (13), a następnie wskazanie tych, dla których osiąga ona wartości najmniejsze. Dzięki szczególnej postaci liczb L-R danej wzorami (10)-(12), w przypadku gdy występujące tam funkcje przynależności mają postać liniową, ściślej opisaną następującą formułą

$$L(x) = \begin{cases} 0 & \text{gdy } x \in (-\infty, -1) \\ 1+x & \text{gdy } x \in [-1, 0] \end{cases} \quad (15)$$

$$R(x) = \begin{cases} 1 - x & \text{gdy } x \in [0, 1] \\ 0 & \text{gdy } x \in (1, \infty) \end{cases} \quad (16)$$

to funkcję celu można wyrazić dogodnym do obliczeń wzorem analitycznym. Należy zaznaczyć iż postać liniowa (15)-(16) została również wskazana przez eksperta.

Niech zatem, dla dowolnie ustalonego skupienia,  $i_1, i_2, \dots, i_I$  oznaczają indeksy tych elementów zbioru danych, które zostały do niego zaliczone, natomiast  $I$  oznacza jego licznosc. Uwzględniając zależności (10)-(12) oraz (15)-(16) funkcja celu przyjęła następującą postać analityczną:

$$\begin{aligned} \mathcal{P}(d_a, d_m; x_i, y_{i,1}, w, \alpha, \beta) &= \frac{\alpha(3w - \alpha) + \beta(\beta + 3w)}{3(\alpha + \beta)} \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (y_{i,1} + b) + \\ &+ \left( 6 - \frac{\alpha(3w - \alpha) + \beta(\beta + 3w)}{3(\alpha + \beta)} \right) \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (d_a \cdot y_{i,1} + d_m \cdot b), \end{aligned} \quad (17)$$

gdy  $\alpha, \beta > 0$ , a także dla wartości indeksów  $\alpha = 0$  albo  $\beta = 0$  odpowiednio

$$\begin{aligned} \mathcal{P}(d_a, d_m; x_i, y_{i,1}, w, 0, \beta) &= \frac{\beta(\beta + 3w)}{3\beta} \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (y_{i,1} + b) + \\ &+ \left( 6 - \frac{\beta(\beta + 3w)}{3\beta} \right) \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (d_a \cdot y_{i,1} + d_m \cdot b) \quad \text{dla } \beta > 0, \end{aligned} \quad (18)$$

$$\begin{aligned} \mathcal{P}(d_a, d_m; x_i, y_{i,1}, w, \alpha, 0) &= \frac{\alpha(3w - \alpha)}{3\alpha} \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (y_{i,1} + b) + \\ &+ \left( 6 - \frac{3w - \alpha}{\alpha} \right) \sum_{i \in \{i_1, i_2, \dots, i_I\}} x_{i,3} (d_a \cdot y_{i,1} + d_m \cdot b) \quad \text{dla } \alpha > 0. \end{aligned} \quad (19)$$

Zgodnie z powyższym wobec każdego z 17 skupień obliczone zostały wartości wszystkich 36 pól związanej z nimi tablicy ocen rezygnacji. Następnie dla każdego skupienia wyznaczana została wartość minimalna oraz dwie wartości jej najbliższe. Odpowiadające im pary rabatów i upustów wskazały na wartość najkorzystniejszą z punktu widzenia strategii operatora, a także na dwie niewiele „gorsze”, których celem było uelastycznienie procesu negocjacji. W przypadku tych dwóch ocen podany także został relatywny wzrost wskaźnika jakości wobec wartości najmniejszej. Wyraźnie zauważalna była zgodność przedstawionych tam wartości z dokonaną wcześniej analizą ekspercką, dotyczącą poszczególnych skupień.

Na tym etapie została zakończona faza dotycząca wstępnej analizy i przetwarzania danych, w rezultacie której uzyskano wyniki umożliwiające zakwalifikowanie rozważanego klienta do odpowiedniego skupienia oraz wskazanie najkorzystniejszych w jego przypadku wariantów upustów i rabatów.



## 4.6. KLASYFIKACJA KLIENTA

W celu ilustracji uzyskiwanych w ostatnim etapie wyników, poniżej rozważono rzeczywistego klienta, którego poddano procedurze klasyfikacji, zgodnie z metodyką przedstawioną w podsekcji 2.4. Klient ten został zaliczony do przykładowo rozważanego w podsekcji 4.4, 54-elementowego skupienia. Było to przedsiębiorstwo transportowe, działające na rynku krajowym oraz w Europie Wschodniej. Klient ten oczekiwał, w ramach renegotjacji warunków umowy, obniżenia kosztów miesięcznych. Duża wartość jego ARPU jest wynikiem przede wszystkim rozmów prowadzonych za granicą. Abonent został scharakteryzowany przez następujący wektor wielkości:

$$\tilde{x} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix} = \begin{bmatrix} 35 \\ 80 \\ 250 \end{bmatrix} \quad (20)$$

Jak wspomniano, po zastosowaniu procedury klasyfikacji klienta tego przypisano do 54-elementowego skupienia. W przypadku tego skupienia wskazane zostały następujące zniżki:

- wariant najkorzystniejszy – upust na abonament  $d_a = 15\%$ , rabat za minutę połączenia  $d_m = 25\%$ ;
- I wariant alternatywny – upust na abonament  $d_a = 10\%$ , rabat za minutę połączenia  $d_m = 12\%$ , wartość wskaźnika jakości większa o 0,75%;
- II wariant alternatywny – upust na abonament  $d_a = 15\%$ , rabat za minutę połączenia  $d_m = 6\%$ , wartość wskaźnika jakości większa o 2,46%.

Klient ten był zaklasyfikowany do skupienia, do którego należała również duża część innych jednostek transportowych, dzięki czemu oceny eksperta, na podstawie których wyznaczane były wartości funkcji celu, uwzględniły preferencje właściwe ich specyfice. Zaproponowane zniżki dla abonamentu na poziomie 10 i 15% oraz dla stawki za minutę 6%, 12% i 25% mogą być dla niego zadowalające przy założeniu, że korekcie podlegają także kwoty generowane przez klienta w roamingu.

## 5. KOŃCOWE KOMENTARZE I PODSUMOWANIE

W ostatniej dekadzie ubiegłego wieku rozpoczął się gwałtowny rozwój rynku telefonii komórkowej w Polsce. Obecnie nie widać przesłanek do oczekiwania spowolnienia tych procesów, natomiast można spodziewać się wzrostu konkurencji, poprzez pojawianie się nowych operatorów i zaostrzanie rywalizacji cenowej. Bardzo istotne w związku z tym staje się poszukiwanie nowoczesnych rozwiązań o charakterze systemowym, które pozwalają na tworzenie narzędzi skutecznych w procesie wspomagania decyzji dotyczącej określenia optymalnej strategii postępowania wobec konkretnego klienta, w celu maksymalizacji zysku operatora przy jednoczesnym możliwie dużym zadowoleniu abonentów.

Przedmiotem prezentowanego artykułu były opracowanie i weryfikacja procedury realizującej powyższy cel wobec klienta korporacyjnego, możliwej do zastosowania

w czasie rzeczywistym podczas negocjacji. Metodę oparto na współczesnych metodach technik informacyjnych w zakresie analizy i eksploracji danych. W szczególności zastosowano koncepcję statystycznych estymatorów jądrowych oraz elementy logiki rozmytej. Estymatory jądrowe zostały wykorzystane do zagadnień eliminacji elementów odosobnionych, analizy skupień i klasyfikacji. Zastosowanie do wszystkich tych zadań jednolitej metodyki znacząco ułatwiło jej użycie, interpretacje uzyskiwanych wyników częściowych, a także umożliwiło wielokrotne wykorzystanie w programie poszczególnych bloków programowych<sup>2</sup>.

Powstała metoda pozwala operatorowi sieci komórkowej aktywnie reagować na potrzeby zgłaszane przez klienta w procesie prowadzenia z nim negocjacji. Opracowany algorytm, przede wszystkim dzięki wzbogaceniu wiedzą eksperta, nie stanowi jedynie matematycznego aparatu służącego do kalkulacji powszechnie analizowanych w takich sytuacjach wskaźników lub weryfikowania wielkości abonamentów dla dostępnych planów taryfikacyjnych, na podstawie czego budowane są oferty utrzymaniowe i przejściowe, jak to ma powszechnie miejsce w dotychczasowej praktyce branżowej. Koncepcja przedstawionego rozwiązania nie opiera się także na popularnej obecnie analizie *churnu*, czyli wskaźnika określającego poziom rezygnacji z usług. Podstawą zaproponowanej w niniejszej pracy metody są informacje zawarte w bazie danych klientów, a także wiedza eksperta o rynku, których połączenie z nowoczesnymi technikami informacyjnymi, stworzyło możliwość opracowania metody o charakterze systemowym pozwalającym na efektywne utrzymywanie dotychczasowych i pozyskiwanie nowych klientów z zapewnieniem maksymalnego zysku operatora.

Poprawność opracowanej tu metody została zweryfikowana z użyciem danych rzeczywistych. Analiza uzyskanych rezultatów potwierdziła ich poprawność zarówno w zakresie interpretacji poszczególnych skupień, jak i wyników otrzymanych na przykładach nowych klientów. Wyraźnie można było zauważyć, iż w skład poszczególnych skupień zaliczone zostały jednostki podobne z punktu widzenia operatora, a niekoniecznie rodzaju reprezentowanej przez nie branży. Przykładowo, firmy typu ochrona-monitoring ulokowały się pojedynczo w wielu różnych skupieniach, a w przeciwieństwie, jednostki transportowe praktycznie skupiły się w nielicznych skupieniach, zwłaszcza jednym, w którym stanowiły połowę jego liczości. Jedno ze skupień zawierało swoiste „zbiorowisko” różnorodnych jednostek, natomiast inny obejmował firmy „zaniedbane”, dla których nie wynegocjowano dotychczas korzystnych warunków umów. Najważniejszą wielkością stanowiącą o skupianiu się jednostek była wartość średniego przychodu na kartę SIM, a w dalszej kolejności liczba kart SIM, i wreszcie staż w sieci. Także rozkład liczości skupień był zbliżony z oczekiwaniami. Dwa duże zawierały 27% i 23%

<sup>2</sup> Warto w tym miejscu wspomnieć, iż w literaturze czasem rekomenduje się dla potrzeb procedury klasyfikacji (podsekcja 2.4) metody wyznaczania wartości parametru wygładzania estymatora jądrowego oparte na kryterium minimalizacji błędu klasyfikacji (por. [10] – podrozdział 2.4), odmienne od ogólnie stosowanych, wykorzystujących kryterium scałkowanego błędu średniokwadratowego (por. [11] – podrozdział 3.1.5). Tych pierwszych nie da się jednak użyć do algorytmu analizy skupień (podsekcja 2.3). W prezentowanych badaniach konsekwentnie stosowane były zatem metody wykorzystujące kryterium scałkowanego błędu średniokwadratowego. W przeciwnym bowiem przypadku, użycie różnych kryteriów (a zatem różnych wartości parametru wygładzania, czyli w konsekwencji także różnych estymatorów jądrowych  $\hat{f}$ ) do wyznaczania wzorców poszczególnych klas za pomocą algorytmu analizy skupień, po czym do dokonywanej na podstawie tych wzorców klasyfikacji, mogłoby w praktyce skutkować trudnymi do przewidzenia rezultatami.

i kolejne dwa o średniej wielkości 14% i 7% elementów początkowej bazy – były to elementy o typowych charakterystykach. Reszta skupień zawierała mniej niż 3%, najczęściej nietypowych, specyficznych firm. Także kończące proces weryfikacji przykłady wyników uzyskiwanych dla nowych klientów dawały wskazania właściwych – z punktu widzenia ich specyfiki i uwarunkowań operatora – propozycji rabatów i upustów.

*Instytut Badań Społecznych, PAN*

#### LITERATURA

- [1] Barnett V., Lewis T., [1994], *Outliers in Statistical Data*, Wiley, Chichester.
- [2] Burnett K., [2002], *Relacje z kluczowymi klientami. Analiza i zarządzanie*, Oficyna Ekonomiczna, Warszawa.
- [3] Dyche J., [2002], *CRM. Relacje z klientami*, Helion, Warszawa.
- [4] Fodor J., Roubens M., [1994], *Fuzzy Preference Modelling and Multicriteria Decision Support*, Kluwer, Dordrecht.
- [5] Fukunaga K., Hostetler L.D., [1975], The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Transactions on Information Theory*, 21(1975)32-40.
- [6] Hołubowicz W., Płóciennik P., [1994], *Cyfrowe systemy telefonii komórkowej GSM 900, GSM 1800, UMTS*, Holkom, Poznań.
- [7] Hryniewicz O., [2004], *Wykłady ze statystyki dla studentów informatycznych technik zarządzania*, Wydawnictwo WSISiZ, Warszawa.
- [8] Jagoda A., de Villepin M., [1993], *Mobile communications*, Wiley, Chichester.
- [9] Kacprzyk J., [1986], *Zbiory rozmyte w analizie systemowej*, PWN, Warszawa.
- [10] Krzyśko M., Wołyński W., Górecki T., Skorzybut M., [2008], *Systemy uczące się. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości*, WNT, Warszawa.
- [11] Kulczycki P., [2005], *Estymatory jądrowe w analizie systemowej*, WNT, Warszawa.
- [12] Kulczycki P., Charytanowicz M., A Complete Gradient Clustering Algorithm Formed with Kernel Estimators, *Applied Mathematics and Computer Science*, 19(2009), w druku.
- [13] Lake N., [2005], *Planowanie strategiczne w firmie*, Onepress, Gliwice.
- [14] Silverman B.W., [1986], *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- [15] Wand M.P., Jones M.C., [1995], *Kernel Smoothing*, Chapman and Hall, London.
- [16] Wesolowski K., [1999], *Systemy radiokomunikacji ruchomej*, WKiŁ, Warszawa.

Praca wpłynęła do redakcji w marcu 2009 r.

#### METODA WSPOMAGANIA STRATEGII MARKETINGOWEJ OPERATORA TELEFONII KOMÓRKOWEJ

#### Streszczenie

Przedmiotem prezentowanych badań jest zadanie wspomagania decyzji w zakresie wyznaczania strategii postępowania wobec klienta korporacyjnego – abonenta sieci telefonii komórkowej. Badania przeprowadzono na rzeczywistej bazie danych, uzyskanej od jednego z polskich operatorów sieci GSM. Wzmiankowana strategia została określona z zastosowaniem metodyki statystycznych estymatorów jądrowych, wykorzystywanej tu do zagadnień wykrywania elementów nietypowych (odosobnionych), analizy skupień (klasteryzacji) i klasyfikacji. Z kolei elementów logiki rozmytej użyto do oceny możliwych do zaproponowania kombinacji

zniżek, natomiast teoria preferencji rozmytych została zastosowana przy wyznaczaniu wariantu postępowania jak najkorzystniejszego dla operatora w sensie maksymalizacji spodziewanego zysku.

**Słowa kluczowe:** analiza i eksploracja danych, techniki informacyjne, analiza systemowa, strategia marketingowa, rynek telefonii komórkowej, statystyczne estymatory jądrowe, logika rozmyta.

#### A METHOD FOR SUPPORTING THE MARKETING STRATEGY OF A MOBILE PHONE NETWORK PROVIDER

##### Summary

The subject of the research presented in this paper is the task of decision support concerning strategy of behavior towards a corporate client – a mobile phone network subscriber. Investigations were carried out on a real database made available by one of the Polish GSM network operators. This strategy was defined with the statistical kernel estimators methodology, applied to the problem of discovering atypical elements (outliers), clustering and classification. In turn, elements of fuzzy logic were used to assess different possible combinations of discounts to offer, while the theory of fuzzy preferences was applied in selecting the most advantageous – in the sense of maximizing expected profit – variant of behavior for the operator.

**Key words:** data analysis and mining, information technology, systems analysis, marketing strategy, mobile phone market, statistical kernel estimators, fuzzy logic.

### Errata do nr 2/2009, tom 56

W artykule P. Kulczycki, K. Daniel, *Metoda wspomagania strategii marketingowej operatora telefonii komórkowej* powinno być:

1. s. 118 – w trzeciej i czwartej linii od dołu, „podziału zbioru danych – na przykład określonych w postaci próby losowej  $x_1, x_2, \dots, x_m$  – na podgrupy” zamiast: „podziału zbioru danych – na przykład określonych w postaci próby losowej  $x_1, x_2, \dots, x_m$  na podgrupy”,

2. s. 119 – w końcowej części wzoru (5)  $k = 0, 1, \dots$ , zamiast  $k = 1, 1, \dots$ ,

3. s. 120 – w czwartej linii  $m_1, m_2, \dots, m_J$ , zamiast  $m_1, m_2, \dots, m_j$ , a także  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_J$ , zamiast  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_j$ ,

4. na s. 133 – Instytut Badań **Systemowych** PAN, zamiast Instytut Badań Społecznych PAN, dwukrotnie w przypadku afiliacji obojga Autorów.