



### Grzegorz Dzikowski

Uniwersytet Ekonomiczny w Katowicach  
Wydział Informatyki i Komunikacji  
Katedra Inżynierii Wiedzy  
grzegorz.dzikowski@ue.katowice.pl

### Przemysław Juszcuk

Uniwersytet Śląski  
Wydział Informatyki i Nauki o Materiałach  
Instytut Informatyki  
przemyslaw.juszcuk@us.edu.pl

## ANALIZA KONTEKSTU ZACHOWAŃ e-KLIENTÓW W ZALEŻNOŚCI OD DYNAMIKI ZMIAN W NAWIGACJI INTERNETOWEJ WZGLĘDEM PRZEPROWADZANYCH AKCJI MARKETINGOWYCH

**Streszczenie:** Handel internetowy pozwala na automatyzację wielu procesów marketingowych oraz na pozyskanie cennych danych o zachowaniu klientów i ich nawigacji na stronach internetowych. Przy użyciu technik eksploracji danych można uzyskać pełną analizę zachowań klienta oraz przeprowadzić segmentację populacji. Sam proces segmentacji populacji nie pozwala jednak na określenie celu klienta, gdyż proces nawigacji jest zmienny w czasie i zależy od zewnętrznych czynników. Określenie celu i zrozumienie potrzeby klienta wymusza wprowadzenie analizy kontekstu zachowań e-klienta. Artykuł przedstawia analizę zachowań e-klientów, segmentację populacji oraz analizę kontekstu zachowań względem przeprowadzanych akcji marketingowych.

**Słowa kluczowe:** Web Usage Mining, analiza kontekstu, analiza zachowań.

### Wprowadzenie

Handel internetowy (e-commerce) jest jedną z najszybciej rozwijających się gałęzi rynku zarówno w Polsce, jak i na świecie. E-commerce odgrywa kluczową rolę w wielu modelach biznesowych opracowanych w ostatnim czasie i opartych na danych internetowych. Przy użyciu technik eksploracji danych można uzyskać pełną analizę zachowań klienta oraz przeprowadzić segmentację populacji. Czynności te, powiązane z odpowiednimi akcjami marketingowymi, pozwalają na widoczne zwiększenie sprzedaży on-line. Sam proces segmentacji populacji nie pozwala na określenie celu klienta, gdyż proces nawigacji jest zmienny w czasie i zależy od czynników zewnętrznych. Określenie celu i zro-

rozumienie potrzeby klienta wymusza wprowadzenie analizy kontekstu zachowań e-klienta. Poprawna analiza kontekstu zachowań klientów i implementacja pozyskanej wiedzy w inteligentnych agentach pozwoli na uczłowieczenie automatycznej sprzedaży i zwiększy zaufanie klientów, co przeniesie się na zwiększenie sprzedaży. Zrozumienie celów klientów i ich potrzeb pozwala na dostosowanie się do panujących trendów i wymogów rynku. Przeprowadzona analiza jest oparta na danych – logach serwerowych z okresu jednego roku kalendarzowego.

## 1. Analiza zachowań e-klientów

Pierwszym etapem analizy kontekstu zachowań jest pozyskanie danych obrazujących nawigację e-klientów oraz przeprowadzenie analizy zachowań. Pozyskanie powyższych danych jest kwestią etycznie niejednoznaczną, ponieważ są zbierane i analizowane dane osobowe użytkowników. Przedsiębiorstwa mają pełne prawo do analizy danych zebranych po stronie ich serwera. W przedstawianym projekcie autorzy mieli do dyspozycji rzeczywiste dane w postaci logów serwerowych (85Gb). W celu określenia kontekstu zachowań autorzy analizowali dane z całego roku kalendarzowego, zawierając zarówno okresy intensywnie nacechowane akcjami marketingowymi (np. okres przedświąteczny), jak i przedziały czasu nienaznaczone akcjami marketingowymi. Segmentację populacji przeprowadzono po podzieleniu danych na sesje użytkowników (sesja wygasa po 30 min. bezczynności klienta [Laur, 2003]). W literaturze naukowej wykazano, że wraz z postępem w dziedzinie technik eksploracji danych, możliwym jest wyodrębnienie typowego zachowania podczas określonego czasu [Laur, 2003]. Dane mogą być zgrupowane według zawartości (tekstu i grafiki), struktury (organizacji treści, takich jak hiperłącza), wykorzystania (IP, dane, czas dostępu) i profilu użytkownika (danych rejestracyjnych, profilach osobistych) [Srivastava, 2000]. Dane wyodrębnione z logów serwerowych muszą zostać przetworzone zanim można zastosować techniki eksploracji danych o zachowaniach klientów. Zastosowanie algorytmów uczenia maszynowego dokonuje detekcji wzorców. Do wykrywania wzorców stosowano różne metody. W literaturze najczęściej pojawiają się następujące podejścia do wykrywania wzorców:

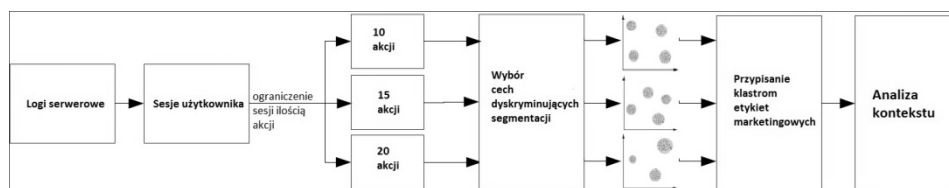
- Segmentacja:
  - Algorytm Leader – segmentacja sesji użytkowników [Yan, 1996],
  - Algorytm partitioning clustering [Cadez, 2000], algorytm Expectation Maximization (EM) oparty na miksturach łańcuchów Markova – segmentacja sesji użytkowników. Każdy łańcuch reprezentuje zachowanie poszczególnej podgrupy,

- Metody model-based clustering [Paliouras, 2000]: metody probabilistyczne, sieci neuronowe, Self-Organizing-Map, metody konceptualne,
- Klasyfikacja:
  - Drzewa decyzyjne, sieci neuronowe, klasyfikatory Bayesa, przykładowo Jording [1999] wykorzystuje algorytm CDL4 do stworzenia listy reguł decyzyjnych wykazującej zainteresowanie użytkownika poszczególnymi tematami,
- Reguły asocjacyjne:
  - Detekcja asocjacji pomiędzy stronami internetowymi opartych na ich występowaniu w sesjach użytkowników [Cooley, 1999],
  - Sieci Baysa w celu detekcji taksonomicznych relacji pomiędzy tematami poszczególnych stron [Schwarzkopf, 2001]
- Techniki sekwencyjne:
  - Techniki deterministyczne oparte na nawigacyjnym zachowaniu użytkownika [Paliouras, 2000],
  - Metody stochastyczne, używające sekwencji stron internetowych, które zostały otwarte w celu przewidywania kolejnych kroków nawigacji [Borges, Levene 1999].

Naszym celem jest autonomiczna detekcja kontekstu zachowań e-klientów. Koniecznym jest więc przeprowadzenie analizy nawigacji klientów oraz ocena ich zachowań, trendów i przyzwyczajzeń. W ten sposób możliwa będzie do przeprowadzenia ocena oczekiwań i motywacji e-klientów.

## 2. Analiza kontekstu zachowań e-klientów

Proces analizy kontekstu zachowań e-klientów został podzielony na etapy (rys. 1). Najpierw należy zidentyfikować klientów internetowych, podzielić nawigacje na sesje użytkowników (jedna sesja zawiera logi serwerowe tego samego klienta, sesja wygasa po 30 min. bezczynności internauty) i następnie dokonać wyboru cech opisujących nawigację po stronach internetowych. Kolejnym krokiem jest detekcja ogólnych zachowań klientów w celu uzyskania zestawu poprawnych zachowań. Następnie otrzymanym klastrom są przypisywane etykiety marketingowe. Ostatnim krokiem jest zdefiniowanie zbioru kontekstów opartych na zbiorze zachowań. Liczba i forma zachowań klientów zależy od cech nawigacyjnych, które udało nam się uzyskać z surowych plików serwerowych.



Rys. 1. Proces analizy kontekstu

Źródło: Opracowanie własne.

Kontekst jest przypisany albo do wyszczególnionego zachowania (klastra), albo jest przypisywany zmianie klastra (klastrów) obrazującego zachowanie klienta podczas nawigacji. W celu detekcji kontekstu opartego na zmianie klastra zachowań, koniecznym stało się wprowadzenie elementu czasu. Kontekst zachowania klientów jest więc oparty na sposobie nawigacji w czasie sesji użytkownika. Nawigacja klientów różni się jednak znacząco w zależności od przeprowadzanych akcji marketingowych. W tym celu okres jednego roku został podzielony na podokresy w zależności od intensywności akcji marketingowych, a detekcja kontekstu (cech nawigacji, zbioru zachowań, zbioru kontekstów) była przeprowadzana osobno dla każdego podokresu w jednym roku kalendarzowym. Istotnym podkreślenia jest fakt, iż autorzy otrzymali wraz z logami serwerowymi pełną strategię marketingową przedsiębiorstwa, zawierającą akcje mailowe, wypuszczanie na rynek nowych katalogów, okresy przecen, akcje marketingowe spersonalizowane.

Pierwszym krokiem w naszej architekturze jest wybór cech nawigacji z danych w postaci logów serwerowych. Wyodrębniono wszystkie możliwe cechy nawigacji w oparciu o dostępne dane z logów serwerowych (tab. 1) i następnie, po analizie statystycznej (korelacja, PCA), został wyselekcjonowany zbiór cech, na podstawie którego były realizowane klastry zachowań. Ze względu na ilości danych (ponad 12 mln sesji użytkowników) proces ich obróbki był bardzo czasochłonny. Dane opisują nawigację klientów po sklepie internetowym naszego partnera. Podzielono stronę na 7 różnych kategorii; każda kategoria została podzielona na trzy poziomy hierarchii: sekcja, podsekcja i pod-podsekcja. W otrzymanej hierarchii strona reprezentująca produkt końcowy odpowiada ostatniemu liściu hierarchii – pod-podsekcji kategorii.

Otrzymano 36 cech opisujących nawigację klienta. Cechy te następnie zostały wykorzystane na etapie segmentacji w celu uzyskania zbiorów zachowań. W tab. 1 przedstawiono wybrane cechy nawigacji.

**Tabela 1.** Wybrane cechy nawigacji

ID użytkownika	ID Sesji	Dzień/Miesiąc/Rok
Godzina początku sesji	Godzina końca sesji	Zakończona zakupem
Całkowita wartość	Znany klient	Źródło nawigacji
Całkowity czas	Kod redukcyjny	Ilość tych samych produktów
Ilość stron produktów	Ilość podsekcji	Ilość sekcji
Ilość produktów kupionych	Ilość referencji kupionych	Ilość wszystkich stron
Czas kategoria 1-7	Ilość stron kategoria 1-7	Ilość zmian kategorii
Ilość zmian sekcji	Ilość zmian podsekcji	Ilość zmian podpodsekcji

Źródło: Opracowanie własne.

Następnym krokiem po wyborze cech jest wykonanie segmentacji w celu otrzymania zbioru zachowań. Segmentacja jest przeprowadzana dla różnych przedziałów roku różniących się intensywnością akcji marketingowych. Kontekst może wynikać ze zmiany zachowania klienta podczas nawigacji, dlatego przeprowadzono ograniczenia na ilość akcji w sesji i osobno przeprowadzono analizę zachowań dla każdego ograniczenia. Ograniczenia zostały dobrane zgodnie z częstotliwością krótkich sesji w bazie i wynoszą 10, 15, 20 akcji (rys. 2a). Progi 10 i 20 akcji zostały wybrane ze względu na fakt, iż segmentacja sesji krótszych niż 10 akcji nie była znacząca, a tylko 10% wszystkich dostępnych sesji jest dłuższe niż 20 akcji. Segmentacja (metoda Warda, kryterium Least Squares, parametr Cubic Criterion Cut-off < 3) została przeprowadzona dla każdego progu i dla różnych okresów w roku kalendarzowym. Klastry, które zostały zewalutowane ze statystycznego punktu widzenia, były następnie analizowane przez ekspertów marketingowych naszych partnerów handlowych w celu przypisania etykiet marketingowych najodpowiedniejszym z marketingowego punktu widzenia klastrów. W ten sposób otrzymano klastry dla każdego progu z etykietami marketingowymi. Współrzędne klastrów (wartości cech nawigacji) i ich liczba jest różna w zależności od progu (10, 15, 20 akcji).

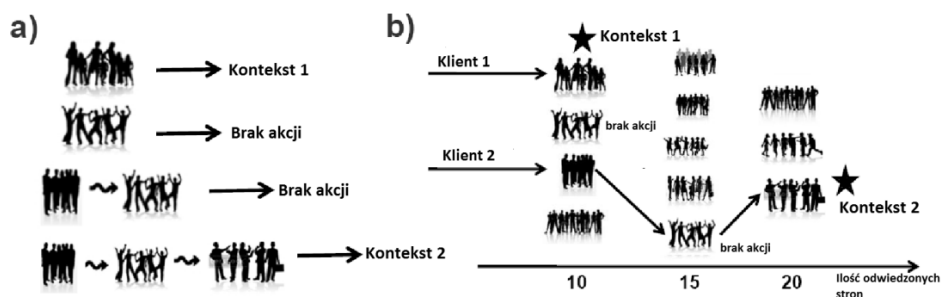
Jak już wspomniano, brano pod uwagę różne okresy roku. Można zauważyć, że główne zmiany w ogólnej nawigacji klientów (kontekstu) podczas całego roku zależą od obecności lub nieobecności działań marketingowych. Z tego powodu poprzedni proces został powtórzony dla różnych okresów. W następnym rozdziale przedstawiono porównanie dla dwóch okresów – okresu bez akcji marketingowych i okresu silnie nacechowanego działaniami handlowymi.

Ostatnim krokiem w naszym systemie jest przydzielenie kontekstu biorąc pod uwagę etykiety marketingowe otrzymanych klastrów segmentacji nawigacji e-klientów (rys. 2a). Kontekst może dotyczyć jednego klastra (jeśli klient jest w klastrze, dla którego kontekst został przydzielony) lub kontekst może odnosić

się do zmiany zachowania klienta podczas nawigacji, co przejawia się zmianą klastra w kolejnym progu. Przydzielenie kontekstu zostało oparte na scenariuszach sprzedaży, które są interesujące z punktu widzenia marketingowego. Ogólna postać zbioru kontekstów jest następująca:

$$(C_{p10} \in c_i \wedge C_{p15} \in c_j \wedge C_{p20} \in c_k) \rightarrow \text{Kon} \{Id, Zc\}$$

gdzie,  $C_{p10}$ ,  $C_{p15}$ ,  $C_{p20}$  – profil klienta po 10, 15, 20 akcjach na stronie;  $c_i$ ,  $c_j$ ,  $c_k$  – zbiór istniejących klastrów dla poszczególnych progów; Kon – etykieta kontekstu; Id – Id użytkownika; Zc – zbiór cech nawigacji.



**Rys. 2.** a) Przydzielenie kontekstu etykietom zachowań nawigacji, kontekst jest przypisywany albo bezpośrednio do klastra zachowania, albo jest przypisywany do zmiany zachowań nawigacji – do zmiany klastrów zachowań; b) Detekcja kontekstu autonomicznie po analizie sesji klienta po 10, 15 i 20 akcjach

Źródło: Opracowanie własne.

Na rys. 2b) przedstawiono działanie systemu w czasie rzeczywistym. System na bieżąco analizuje nawigację klientów, porównując współrzędne cech ze zbiorem zachowań utworzonym po 10, 15 i 20 akcjach klienta na stronie internetowej. W momencie detekcji zbieżności nawigacji ze zbiorem kontekstów, generowana jest informacja zawierająca Id użytkownika oraz zbiór cech nawigacji analizowanego w danej chwili e-klienta.

### 3. Porównanie kontekstu okresów z i bez akcji marketingowych

Jak już wspomniano, analizę przeprowadzono w różnych okresach roku i potwierdzono przypuszczenia, iż kontekst zachowań klientów zależy od intensywności akcji marketingowych. Z tego powodu nasze ostateczne rozwiązanie zawiera różne segmentacje (zbiory zachowań i zbiory kontekstów) dla każdego

progu i dla różnych okresów w roku kalendarzowym. Przed dokonaniem porównania, należy podkreślić jeden istotny punkt – nie istnieją obiektywne kryteria pozwalające na ocenę wyników segmentacji (uczenia maszynowego bez nadzoru), dlatego nasze wyniki zostały zewaluowane ręcznie przez ekspertów handlowych naszych partnerów przemysłowych. Poza oceną otrzymanych klastrów, przypisano im etykiety marketingowe oraz określono na ich podstawie kontekst zachowań e-klientów. W celu zobrazowania różnic kontekstu w zależności od intensywności akcji marketingowych przedstawiono porównanie dwóch okresów miesięcznych. Pierwszy miesiąc, który odpowiada okresowi bez żadnych działań handlowych odpowiada miesiącu kwiecień (989 793 sesji użytkowników). Miesiąc reprezentujący okres akcji marketingowych przypadł na lipiec (1 228 971 sesji).

Przedstawione poniżej porównanie różnic pomiędzy okresami odnosi się do progu ograniczającego sesję do 15 akcji, jednak te same różnice występują dla wszystkich progów ograniczeń. W okresach o podobnym natężeniu akcji marketingowych odnajdujemy podobne klastry i odpowiadające im etykiety marketingowe dla każdego progu (10, 15, 20). W innych okresach parametry klastrów i ich etykiety znacznie się różnią. Wynika stąd, iż pomimo wyboru tak dużej ilości charakterystycznych dla e-commerce cech nawigacji (36 cech), trudno jest ustalić jeden zestaw klastrów i zbioru kontekstów poprawnie opisujących cały rok kalendarzowy. Jest to spowodowane różnicą w scenariuszach sprzedaży w odrębnych okresach czasowych. Liczba klastrów reprezentujących nawigację e-klientów dla miesiąca kwiecień (brak akcji marketingowych) wynosi: dla progu 10 akcji – 6 klastrów, dla progu 15 akcji – 8 klastrów i progu 20 akcji – 8 klastrów. Odpowiednio dla lipca (okres z działaniami marketingowymi) ilość klastrów to: próg 10 – 7, próg 15 – 8, próg 20 – 8 klastrów.

Porównanie wyników z opisanych miesięcy przedstawiono w tab. 2. Wynika z niej, że wartość cech dla wszystkich klastrów dla miesiąca lipca w porównaniu z miesiącem kwiecień jest:

- większa dla cech: ilość stron kategoria przecena, ilość produktów kupionych, całkowita wartość, zakończona zakupem, ilość zmian kategorii, ilość stron kategoria sprzedaż bezpośrednia, ilość podsekcji, źródło nawigacji equal mail;
- mniejsza dla cech: ilość stron kategoria sklep, ilość zmian sekcji, ilość zmian pod-podsekcji, całkowity czas.

Otrzymane rezultaty pokazują niezgodność w nawigacji klientów w różnych okresach roku. Na przykład w kwietniu klienci spędzają więcej czasu, aby osiągnąć 15 akcji (stron internetowych). Może być to kierowane szeroką gamą produktów i brakiem kierowania klientów na konkretne produkty. W lipcu klienci kupują więcej produktów i jest większy odsetek sesji zakończonych za-

kupem, jednak łączna kwota całkowitego zakupu zwiększa się nieznacznie. Wy tłumaczeniem może być zakup najtańszego produktu, przy zachowaniu stałej wydatków. Klienci także częściej dokonują zakupu za pomocą szybkiego zamówienia lub listy mailowej, są to bowiem akcje marketingowe. Główną różnicą w zachowaniu nawigacji dla tych dwóch okresów jest to, że klienci w lipcu częściej zmieniają produkty w kategorii, a rzadziej porównują produkty końcowe tej samej rodziny.

Cechy najbardziej dyskryminujące okres bez akcji marketingowych w kolejności malejącej to: źródło nawigacji, ilość stron kategoria sklep, ilość zmian pod-podsekcja, ilość stron kategoria różne, ilość sekcji, ilość stron produktów, ilość podsekcji, ilość zmian kategorii, zakończona zakupem. Cechy najbardziej dyskryminujące okres z akcjami marketingowymi to: źródło nawigacji, ilość stron kategoria przeceny, ilość stron kategoria sklep, ilość stron kategoria różne, ilość podsekcji, godzina początku sesji, ilość stron kategoria koszyk, ilość stron produktów oraz ilość zmian kategorii. Tab. 2 przedstawia współrzędne klastrów dla najbardziej dyskryminujących cech dla progę 15 akcji dla miesięcy kwiecień i lipiec (wartości binarne przyjmują wartość 0 – nie, 1 – tak; pozostałe wartości zostały znormalizowane do przedziału 1-10).

**Tabela 2.** Cechy dyskryminujące klastery dla ograniczenia sesji 15 akcjami dla dwóch miesięcy

Cecha	Okres bez akcji marketingowych – Kwiecień
Częstotliwość klastra	C1=1.13, C2=2.40, C3=1.54, C4=1.26, C5=1.48, C6=3.34, <b>C7=8.12</b> , C8=1.03
Ilość stron kategoria sklep	C1=6.50, C2=9.16, C3=6.20, C4=6.71, C5=4.39, C6=1, <b>C7=9.60</b> , C8=3.73
Ilość stron kategoria przecena	C1=1.01, C2=1.01, C3=1.01, C4=1.01, C5=1, C6=1, <b>C7=1.01</b> , C8=1.01
Źródło nawigacji = Mail	C1=0, C2=0, C3=0, C4=1, C5=0.28, C6=1, <b>C7=0</b> , C8=0.08
Ilość stron kategoria różne	C1=4.81, C2=2.51, C3=6.24, C4=4.60, C5=10, C6=2.19, <b>C7=1.77</b> , C8=6.76
Ilość stron kategoria koszyk	C1=2.65, C2=1.07, C3=1.13, C4=1.47, C5=1.07, C6=1, <b>C7=1.07</b> , C8=2.33
Ilość stron produktów	C1=1.83, C2=2.99, C3=10, C4=2.40, C5=2.59, C6=4.91, <b>C7=5.65</b> , C8=4.55
Ilość sekcji	C1=3.33, C2=6.46, C3=3.83, C4=4.57, C5=1, C6=7.54, <b>C7=10</b> , C8=2.82
Ilość podsekcji	C1=2.60, C2=1.75, C3=1, C4=3.6, C5=2.78, C6=5.67, <b>C7=5.53</b> , C8=1.66
Ilość zmian kategorii	C1=4.61, C2=2.31, C3=6.91, C4=5.57, C5=4.82, C6=1.37, <b>C7=1</b> , C8=5.15
Zakończona zakupem = nie	C1=1, C2=1, C3=1, C4=1, C5=1, C6=1, <b>C7=0.99</b> , C8=0
Całkowita wartość	C1=1, C2=1, C3=1, C4=1, C5=1, C6=1, <b>C7=1</b> , C8=10
	Okres z akcjami marketingowymi – Lipiec
Częstotliwość klastra	<b>C1=5.56</b> , C2=2.09, C3=5.55, C4=8.17, C5=1, C6=1.65, C7=10, C8=2.28
Ilość stron kategoria sklep	<b>C1=1.53</b> , C2=1, C3=9.06, C4=1.42, C5=2.88, C6=2.75, C7=9.03, C8=4.56
Ilość stron kategoria przecena	<b>C1=9.23</b> , C2=10, C3=1.89, C4=9.95, C5=2.39, C6=2.64, C7=2.16, C8=2.72
Źródło nawigacji = Mail	<b>C1=1</b> , C2=0, C3=0.83, C4=0, C5=0.10, C6=0.01, C7=0, C8=0.05
Ilość stron kategoria różne	<b>C1=2.09</b> , C2=1, C3=2.17, C4=1.29, C5=6.25, C6=2.57, C7=1.67, C8=7.32
Ilość stron kategoria koszyk	<b>C1=1.18</b> , C2=1, C3=1.03, C4=1.07, C5=5.84, C6=10, C7=1.14, C8=1.39
Ilość stron produktów	<b>C1=3.33</b> , C2=1, C3=6.62, C4=5.44, C5=1.96, C6=1.96, C7=5.92, C8=4.44
Ilość sekcji	<b>C1=1.98</b> , C2=1, C3=6.60, C4=1.91, C5=3.73, C6=3.73, C7=7.17, C8=6.53
Ilość podsekcji	<b>C1=6.52</b> , C2=10, C3=4.43, C4=6.49, C5=1.13, C6=1.17, C7=2.81, C8=1.79
Ilość zmian kategorii	<b>C1=3.29</b> , C2=1.43, C3=3.29, C4=2.44, C5=10, C6=7.38, C7=3.27, C8=7.93
Zakończona zakupem = nie	<b>C1=1</b> , C2=0.99, C3=1, C4=0.99, C5=0.08, C6=0.99, C7=1, C8=0.99
Całkowita wartość	<b>C1=1</b> , C2=1, C3=1, C4=1, C5=7.78, C6=1.11, C7=1, C8=1

Źródło: Opracowanie własne.



Dla każdego klastra, dla każdego progu i dla dwóch okresów przypisano etykiety marketingowe w celu określenia kontekstu. Dla przedstawionych dwóch miesięcy stworzono 26 kontekstów zachowań klienta.

Prezentujemy pierwszy przykład, który odnosi się do klastra C1<sub>15</sub> z ograniczeniem 15 akcji na sesję i obejmującego okres z intensywnymi akcjami marketingowymi (tab. 2). Przyporządkowana etykieta marketingowa jest następująca: klaster dotyczy sesji, w której dostęp nie jest bezpośredni, lecz sesja jest wyzwana wiadomością e-mail i ukierunkowuje klienta na specjalną sekcję sklepu. Klient ukierunkowany kampanią marketingową prezentuje wyższy poziom zaangażowania niż w przypadku wejścia na stronę z wyszukiwarki. Po pierwszych 15 akcjach żaden produkt nie został przeniesiony do koszyka. Kontekst utworzony bezpośrednio dla tego klastra wskazuje na potencjalną konieczność dialogu w celu zachęcenia klienta do dokonania zakupu.

Drugi przykład dotyczy zmiany zachowania klienta podczas nawigacji (zmiany klastra). Kontekst dotyczy przejścia z klastra C3<sub>10</sub> (segmentacja zrealizowana po 10 akcjach internauty) poprzez C7<sub>15</sub> (segmentacja po 15 akcjach) do klastra C2<sub>20</sub> (segmentacja po 20 akcjach). Przykład dotyczy okresu bez akcji komercyjnych: Podczas dwóch pierwszych ograniczeń sesji (10 i 15 akcji) klient miał zachowanie osoby znającej swoje potrzeby, która sprawdza precyzyjne produkty ze zbliżonych kategorii, która spędza czas na nawigacji na stronach dotyczących podobnych produktów. Przy ostatnim progu, czyli po 20 akcjach, nawigacja ulega zmianie i wskazuje na ogólną nawigację po sklepie bez głównej idei zakupu. Klient przybył spontanicznie na stronę i ma *a priori* jasne wyobrażenie o pożądanym produkcie (podobne produkty przeglądane). Po około 6 min., klient zaprzestaje wglębienia się w opisy produktów i rozprasza się w sekcji na najwyższym poziomie. Kontekstem w tym przykładzie jest ryzyko opuszczenia strony przez klienta i konieczność wywołania akcji wspomagających konsultacje produktów poprzednio przeglądanych w celu zatrzymania klienta i zakończenia sesji sprzedażą.

## Podsumowanie

W artykule przedstawiono autonomiczny system detekcji kontekstu zachowań e-klientów. Zrealizowany system automatycznie analizuje sesje użytkowników w czasie rzeczywistym i dostarcza informacji inteligentnym agentom o zachowaniu, gustach i nawigacji e-klientów. W celu detekcji kontekstu są analizowane sesje klientów. Wyszczególniono 36 cech nawigacji. Przedstawiono segmentację populacji wspomaganą scenariuszami sprzedaży. Dla dwóch okre-

sów skrajnie różniących się ilością akcji marketingowych otrzymano 45 klastrów. Na podstawie tych danych otrzymano 26 różnych kontekstów zachowań klientów. Przedstawiona w artykule analiza porównuje konteksty nawigacji w różnych okresach roku kalendarzowego. Pod uwagę zostały brane okresy w zależności od częstotliwości akcji marketingowych. Kolejnym etapem badań jest analiza danych z kolejnych lat nawigacji w celu sprawdzenia zgodności otrzymanego zbioru kontekstów i detekcji zmian nawigacji w kolejnych latach.

## Literatura

- Borges J., Levene M. (1999), *Data mining of user navigation patterns*, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego.
- Cadez I., Heckerman D., Meek C., Smyth P., White S. (2000), *Visualization of navigation patterns on a web site using model based clustering*, Technical Report MSR-TR-00-18, Microsoft Research.
- Cooley R., Mobasher B., Srivastava J. (1999), *Data preparation for mining World Wide Web browsing patterns*, „Journal of Knowledge and Information Systems”, Vol. 1, No. 1, s. 5-32.
- Jording T. (1999), *A temporary user modeling approach for adaptive shopping on the web* [w:] Proceedings of the 2nd Workshop on Adaptive Systems and User Modeling on the WWW, UM99, Banff, Canada.
- Laur P.-A., Teisseire M., Poncelet P. (2003), *Web usage mining: extraction, maintenance and behaviour trends* [w:] IICAI'03: Indian International Conference of Artificial Intelligence, s. 493-506.
- Paliouras G., Papatheodorou C., Karkaletsis V., Spyropoulos C.D. (2000) *Clustering the users of large web sites into communities* [in:] Proceedings of International Conference on Machine Learning (ICML), Stanford, California.
- Schwarzkopf E., (2001), *An adaptive web site for the UM2001 conference* [in:] Proceedings of the UM2001 Workshop on Machine Learning for User Modeling.
- Srivastava J., Cooley R., Deshpande M., Tan P.-N. (2000), *Web usage mining: discovery and applications of usage patterns from web data*, SIGKDD Explorations, University of Minnesota, manuscript.
- Yan T.W., Jacobsen M., Garcia-Molina H., Dayal U. (1996), *From user access patterns to dynamic hypertext linking*, WWW 5/Computer Networks 28.

**E-CUSTOMERS BEHAVIORS CONTEXT ANALYSIS BASED  
ON THE DYNAMICS OF CHANGES IN WEB NAVIGATION  
DUE TO MARKETING ACTION PERFORMED**

**Summary:** E-commerce allows to automate marketing processes and to gain valuable data about customer behavior and their navigation on the website. Using data mining techniques, we can get a complete analysis of customer behavior and to segment the population. However, population segmentation process does not identify the customer, because the navigation process is unpredictable over time and depends on external factors. This article presents an analysis of the behavior of e-customer, segmentation of population and analysis of the context of population behavior towards marketing actions.

**Keywords:** Web Usage Mining, Customer Behavior Analysis, Context Analysis.