

Katarzyna Rostek, Artur Gąsiorkiewicz

Eksploracja danych internetowych jako metoda zwiększenia przewagi konkurencyjnej systemów handlu elektronicznego B2C

Ekonomiczne Problemy Usług nr 68, 90-96

2011

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach
dozwolonego użytku.

KATARZYNA ROSTEK, ARTUR GAŚIORKIEWICZ

Politechnika Warszawska

EKSPLORACJA DANYCH INTERNETOWYCH JAKO METODA ZWIĘKSZENIA PRZEWAGI KONKURENCYJNEJ SYSTEMÓW HANDLU ELEKTRONICZNEGO B2C

Wprowadzenie

Krajowy rynek detalicznego handlu elektronicznego rozwija się w ostatnich latach bardzo dynamicznie, niejako na przekór opiniom o spowolnieniu gospodarczym. Rozwój przedsiębiorczości w tym obszarze, możliwy także dzięki asymilacji funduszy z programów pomocowych, powoduje stałe zwiększanie się i wzrost atrakcyjności oferty handlowej w Internecie. Z kolei rozwój technologii internetowych i ich upowszechnienie się, także w grupach podlegających dotychczas wykluczeniu cyfrowemu, nieustająco kreuje popyt. Oba te procesy powodują, że roczny przyrost wielkości rynku *e-commerce*'owego w segmencie B2C kształtuje się obecnie na poziomie ponad 30%. Przewidywana jest też kontynuacja tego trendu w 2011 roku¹.

Jednym z podstawowych czynników warunkujących efektywność biznesową przedsięwzięcia jest dla przedsiębiorców znalezienie obszaru, w którym możliwe jest zwiększanie przewagi konkurencyjnej. Na rynku *e-commerce*'u, w dobie rozwoju konkurencji bezpośredniej, jak i pośredniej (dostęp do alternatywnych rynków i substytutów) coraz trudniej jest konkurować ceną, a w konsekwencji poziomem marży. Uzyskiwanie przewagi dotyczy więc coraz częściej innych aspektów działalności – jak choćby zwiększania widoczności (*web-visibility*) serwisu WWW, kształtowania użyteczności i dostępności stosowanych rozwiązań, utrzymywania relacji z klientami, oferowania korzyści dodatkowych lub optymalizacji ścieżki konwersji. Wspólnym mianownikiem dla tych działań, szczególnie w przypadku

¹ http://www.forumszerokopasmowe.pl/aktualnosci.php?news_id=1735

systemów *e-commerce*'owych adresowanych do masowego odbiorcy, jest analiza zachowań użytkowników, a podstawowym narzędziem jest w tym zakresie eksploatacja danych internetowych.

1. Konwersja i optymalizacja ścieżki konwersji

Konwersja jest jednym z podstawowych pojęć e-biznesu. Oznacza wymierne ekonomicznie działanie użytkownika, wpisujące się w realizację celów biznesowych przedsięwzięcia, wynikające z podjęcia przez przedsiębiorstwo określonych działań w obszarze marketingu. W przypadku sklepów internetowych najbardziej oczywistym, choć nie jedynym, przykładem konwersji jest sfinalizowanie transakcji zakupu.

Dla celów analitycznych istotna jest średnia wartość konwersji, którą definiuje się zazwyczaj jako średni dochód wynikający z pojedynczej konwersji w określonym czasie. Określenie tego parametru pozwala świadomie zarządzać marketingową działalnością e-sklepu i badać jej efektywność w różnych wymiarach.

Z biznesowego punktu widzenia istnieją dwa główne sposoby optymalizacji sprzedażowych serwisów WWW. Pierwszy z nich polega na podejmowaniu działań służących do zwiększenia ruchu przychodzącego do serwisu WWW – zarówno pierwotnego, jak i wtórnego². Przy założeniu stałego współczynnika i wartości konwersji skutkuje to zwiększeniem przychodów serwisu proporcjonalnym do zmiany wielkości ruchu przychodzącego. Drugi sposób obejmuje działania, których celem jest zwiększanie współczynnika konwersji oraz zwiększanie wartości pojedynczej konwersji. Większość działań w tym zakresie sprowadza się do szeroko pojętej optymalizacji ścieżki konwersji.

W praktyce oba wymienione podejścia do optymalizacji stosowane są równolegle, a podstawowej wiedzy niezbędnej do ich realizacji dostarcza eksploatacja danych internetowych.

² Więcej w: A. Gąsioriewicz: *Czynniki kształtujące ruch (web-traffic) i ich wpływ na poziom konwersji w serwisach WWW systemów handlu elektronicznego w segmencie B2C*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 597, Ekonomiczne Problemy Usług nr 57, WNUS, Szczecin 2010, s. 637.

2. Metody eksploracji danych internetowych wykorzystywane w e-marketingu

Do najczęściej stosowanych metod analitycznych *web data miningu*, które mają zastosowanie w e-marketingu, należy zaliczyć: analizę statystyczną, analizę asocjacji, klasteryzację, klasyfikację oraz wzorce sekwencji³.

Analiza statystyczna służy do przesiewania dużych zbiorów danych w poszukiwaniu zależności i zjawisk występujących w nich najczęściej. Wykorzystuje szereg metod i technik, do których należy zaliczyć analizy: wariancji, korelacji, regresji, czynnikową, dyskryminacyjną, szeregi czasowe oraz kanoniczną. Jest metodą wykorzystywaną do zapoznania się z analizowanym zbiorem danych oraz do wyznaczenia opisu wartości średnich całej populacji. Najczęściej jest to etap wstępny do kolejnych etapów analizy eksploracyjnej, chociaż może być również przydatny do ogólnego wnioskowania, np. identyfikacji obszarów serwisu sprzedażowego, których odwiedzenie prowadzi najczęściej do opuszczenia stron serwisu bez dokonania zakupu (a zatem wymagających dokładniejszej analizy pod kątem funkcjonalności i treści).

Analiza asocjacji służy do wykrywania związków występujących pomiędzy danymi. Są to zależności polegające na współwystępowaniu pewnych elementów w analizowanych danych⁴. Przykładem może być zjawisko polegające na tym, że pewne towary są często kupowane jednocześnie, a klienci serwisu A są często klientami także serwisu D i K. Reguły asocjacyjne są wykorzystywane do definiowania znaczenia i wzajemnych powiązań obiektów z tego samego zbioru danych, a przyjmują postać wyrażenia:

$$p = > q,$$

gdzie:

p – zbiór wartości warunkujących reguły,

q – zbiór wartości warunkowanych reguły.

Reguła asocjacyjna może przyjmować postać następującego opisu: *Klienci, którzy kupili komputer, nabyli również oprogramowanie MS Office 2010*. W regule asocjacyjnej znaczące jest również to, dla jak licznej grupy klientów ta reguła jest prawdziwa. Należałoby ją zatem wzbogacić o wielkość wsparcia⁵ oraz wiarygodność⁶ reguły. Jej nowy zapis mógłby zatem wyglądać następująco: *15% klientów kupuje jednocześnie komputer i oprogramowanie, a w tej grupie aż 88% klientów*

³ Ch. Zhang, Z. Liu, J. Yin: *The Application Research on Web Log Mining in E-Marketing*, 2nd International Conference on e-Business and Information System Security (EBISS), Wuhan 2010.

⁴ P. Cichosz: *Systemy uczące się*, WNT, Warszawa 2000.

⁵ Wsparcie reguły asocjacyjnej – liczba współwystąpień kompleksu p i q w stosunku do liczby wszystkich przykładów zbioru.

⁶ Wiarygodność reguły asocjacyjnej – liczba współwystąpień kompleksu p i q w stosunku do liczby wszystkich wystąpień kompleksu p.

kupuje oprogramowanie MS Office 2010, gdzie 15% jest wartością wsparcia, a 88% oznacza poziom wiarygodności reguły.

Reguły asocjacyjne znajdują zastosowanie w projektowaniu katalogu produktów i usług, segmentacji klientów oraz opracowywaniu pakietów promocyjnych. Często służą też do optymalizacji sprzedaży poprzez oferowanie dóbr komplementarnych (*cross-selling*⁷).

Klasteryzacja (grupowanie) jest jedną z metod analizy danych, której celem głównym jest podział rozpatrywanego zbioru obiektów na grupy (klastry), w taki sposób, aby każda z grup była możliwie jednorodna (tzn. zawierała elementy podobne do siebie), a jednocześnie poszczególne klastry były jak najbardziej zróżnicowane. Idea klasteryzacji opiera się na analizie wariancji⁸. Dąży się do takiego podziału zbioru danych, aby wariancja wewnątrzgrupowa w każdym z klastrow była możliwie mała, a jednocześnie wariancja międzygrupowa była możliwie duża. W e-marketingu klasteryzacja jest wykorzystywana między innymi do grupowania klientów i ich preferencji, a następnie do wykorzystywania tej wiedzy podczas tworzenia programów marketingowych.

Klasyfikacja polega na ustaleniu przynależności obiektów lub wzorców do określonej kategorii – klasy⁹. Na podstawie zbioru obiektów budowany jest model, który służy do klasyfikowania nowych obiektów w zbiorze lub zrozumienia istniejącego podziału obiektów na predefiniowane klasy. Do klasyfikacji obiektów wykorzystywane są takie metody, jak: naiwny klasyfikator bayesowski, perceptron, algorytm k-najbliższych sąsiadów, drzewa decyzyjne czy sieci bayesowskie. Klasyfikacja klientów ułatwia przygotowanie dedykowanych kampanii marketingowych, jak również wskazanie najważniejszej grupy klientów dla przygotowywanej akcji promocyjnej.

Różnica pomiędzy klasyfikacją a klasteryzacją polega na tym, że w procesie klasyfikacji obiekty przydzielane są do predefiniowanych kategorii według przyjętego kryterium klasyfikacyjnego, a w procesie klasteryzacji obiekty są dzielone na możliwie homogeniczne i różniące się między sobą grupy, dla których następnie znajduje się opis charakterystyki.

Wzorce sekwencji są odkrywane podczas analizy zdarzeń, które wystąpiły w określonym czasie, w celu znalezienia zależności pomiędzy występowaniem tych zdarzeń w czasie. Przykładem wzorca sekwencji jest śledzenie nawigacji użytkownika po stronach internetowych i znajdowanie zależności typu: użytkownik, który

⁷ *Cross-selling* – strategia sprzedaży polegająca na oferowaniu dodatkowych produktów do produktu podstawowego, zwiększająca łączny dochód ze sprzedaży.

⁸ Analiza wariancji – metoda statystyczna służąca do badania obserwacji, które zależą od jednego lub wielu działających równocześnie czynników. Metoda ta wyjaśnia, z jakim prawdopodobieństwem wyodrębnione czynniki mogą być powodem różnic pomiędzy obserwowanymi średnimi wartościami w grupie.

⁹ P. Cichosz: *Systemy uczące się...*, *op. cit.*

odczytał stronę A i B, przejdzie następnie do strony D i F. Odnajdywanie wzorców sekwencji prognozujących zachowania klientów pomaga dostarczać ukierunkowane na nich usługi marketingowe oraz lepiej spełniać ich wymagania użytkowe.

3. Zastosowania eksploracji danych internetowych w e-marketingu

Najważniejsze zastosowania eksploracji danych internetowych obejmują:

- optymalizację działań promocyjnych,
- personalizację przekazu,
- optymalizację użyteczności i dostępności serwisu WWW,
- analizę biznesową¹⁰.

Optymalizacja działań promocyjnych w wyniku eksploracji danych internetowych opiera się w głównej mierze na poszukiwaniu zależności pomiędzy parametrami działań promocyjnych (nośniki reklamowe, grupy docelowe, kreacje reklamowe, strony docelowe/*landing pages*) a uzyskanym poziomem konwersji. Zidentyfikowane zależności pozwalają na przeniesienie ciężaru realizowanych kampanii promocyjnych do obszarów, które charakteryzują się wyższą efektywnością biznesową. W praktyce oznacza to zazwyczaj optymalizację kampanii reklamy internetowej.

Personalizacja przekazu, oparta na eksploracji danych internetowych, uwzględnia preferencje, zachowania oraz prywatne zainteresowania klienta. Podstawowym zadaniem personalizacji przekazu jest służenie wsparciem klientowi w wyborze produktów i usług firmy najlepiej dopasowanych do jego potrzeb. Badania dowiodły, że przygotowanie aplikacji internetowej wykorzystującej model personalizacji przekazu może zwiększyć liczbę zakupów sieciowych od 2% do 8%¹¹. Według autorów badań wśród innych korzyści zastosowania personalizacji przekazu należy wymienić:

- znaczne obniżenie kosztów promocji,
- zwiększenie zainteresowania produktem,
- personalizację oferty marketingowej,
- zwiększenie satysfakcji i lojalności klientów,
- skuteczne promowanie sprzedaży typu *cross-selling* i *up-selling*¹².

Realizacja działań opartych na personalizacji przekazu wymaga szczególnej ostrożności nie tylko ze względu na pewne ograniczenia wynikające z przepisów

¹⁰ Por. M. Łapczyński: *Web Usage Mining*, <http://www.statsoft.pl/czytelnia.html>

¹¹ Ch. Zhang, Z. Liu, J. Yin: *The Application Research...*, *op. cit.*

¹² *Up-selling* – strategia sprzedaży polegająca na oferowaniu klientowi tego samego produktu, ale w wersji poprawionej, wzbogaconej, lepszej jakościowo i oczywiście droższej.

dotyczących ochrony danych osobowych, ale także z uwagi na wysoką wrażliwość klientów na kojarzenie danych biznesowych z ich profilami (także anonimowymi).

Badanie aktywności klientów na poszczególnych stronach serwisu, analiza scenariuszy i punktów porzucenia oraz wykrywanie zależności asocjacyjnych dotyczących tego, jak często klient korzysta z serwisu, ile czasu jest aktywny w serwisie, w jaki sposób nawiguje po stronach serwisu i do ilu podstron jest skłonny zagłębiać się w treści serwisu (współczynnik „lepkości” serwisu), stanowią podstawę prac nad optymalizacją użyteczności i dostępności serwisu WWW. Podstawowym ich celem jest implementacja zmian w zakresie funkcjonalności serwisu WWW, która skutkować ma zwiększeniem współczynnika konwersji (np. ulepszenie systemu wyszukiwania, dodanie funkcjonalności ułatwiających wybór właściwego produktu, dodanie możliwości zakupu bez rejestracji konta użytkownika, dopasowanie architektury informacji do profilu użytkownika). Realizowane są także zmiany pozafunkcyjne, np. eliminacja technologii ograniczających dostępność serwisu WWW na rzecz rozwiązań poszerzających grono odbiorców serwisu.

Analiza biznesowa, stanowiąca fragment większego obszaru określanego mianem *Business Intelligence*¹³, przeprowadzana jest na podstawie wszelkich dostępnych danych o klientach, o dokonywanych przez nich zakupach oraz o aktywności użytkowników w serwisie. Pozyskana w ten sposób wiedza jest wykorzystywana w e-marketingu do realizacji metod optymalizacji sprzedaży (*cross-selling* i *up-selling*) oraz utrzymania (przeciwdziałania utracie) klientów.

Podsumowanie

Przedstawione w artykule metody eksploracji danych internetowych i ich zastosowania nabierają w ostatnim czasie szczególnie istotnego znaczenia na rynku *e-commerce*'u na skutek wzrostu stopnia konkurencji i wyczerpywania się możliwości konkurowania poziomem marży. Zarządzający systemami *e-commerce*'owymi w segmencie B2C powinni położyć szczególny nacisk na wykorzystanie technologii *web data mining* do optymalizacji biznesu, szczególnie w przypadku przedsięwzięć internetowych kierowanych do masowego odbiorcy. Efektywność podjętych działań stanowić może ponadto interesujący obszar badań naukowych.

¹³ *Business Intelligence* – proces analizy i raportowania danych z wykorzystaniem nowoczesnych metod i narzędzi wielowymiarowej analizy danych.

Literatura

1. Cichosz P.: *Systemy uczące się*, WNT, Warszawa 2000.
2. Gąsiorkiewicz A.: Czynniki kształtujące ruch (web traffic) i ich wpływ na poziom konwersji w serwisach WWW systemów handlu elektronicznego w segmencie B2C, *Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 597, Ekonomiczne Problemy Usług nr 57, WNUS, Szczecin 2010, s. 637–645.*
3. Zhang Ch., Liu Z., Yin J.: *The Application Research on Web Log Mining in E-Marketing*, 2nd International Conference on e-Business and Information System Security (EBISS), Wuhan 2010.

**WEB DATA MINING AS A FACTOR FOR BUSINESS ADVANTAGE
ON B2C E-COMMERCE MARKET****Summary**

The paper presents main web data mining methods which can be used for increasing of business advantage in B2C e-commerce sector. Authors pay some attention to practical implementation of these methods, including: promotion optimisation, personalized communication, web-usability and web-accessibility enhancement, business analysis.

Translated by Artur Gąsiorkiewicz