

Tomasz Zdziebko

Identyfikacja preferencji użytkowników serwisów handlu elektronicznego

Ekonomiczne Problemy Usług nr 88, 136-143

2012

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.

TOMASZ ZDZIEBKO

Uniwersytet Szczeciński

IDENTYFIKACJA PREFERENCJI UŻYTKOWNIKÓW SERWISÓW HANDLU ELEKTRONICZNEGO

Wprowadzenie

Rosnące wymagania klientów, spadek lojalności i nasilenie konkurencji zmuszają firmy do podejmowania coraz dalej idących działań mających na celu zdobycie i utrzymanie klientów. Potrzeba ta jest szczególnie istotna w obszarze handlu elektronicznego, gdzie konkurencja znajduje się o jedno kliknięcie myszki dalej. Wyższe z reguły koszty pozyskania klientów niż ich utrzymania przyczyniają się do podejmowania działań mających na celu przywiązanie klientów do firmy. Koncepcja zarządzania relacjami z klientami (ang. Customer Relationship Management – w skrócie CRM) stanowi odpowiedź na negatywne zmiany w środowisku funkcjonowania firm¹. Filozofia CRM polega na stosowaniu indywidualnych strategii w stosunku do każdego klienta w celu nie tylko zdobycia, ale przede wszystkim utrzymania długoterminowego z nim kontaktu. Stąd istotą CRM jest budowa spójnego, bogatego w szczegóły wizerunku klienta oraz organizacja pracy pozwalająca na możliwie indywidualne traktowanie każdego klienta. Dlatego, jak zauważa Philip Kotler, zrozumienie wartości postrzeganych przez klienta i jemu dostarczanych może stanowić trwałe źródło przewagi konkurencyjnej².

¹ M. Szymura-Tyc, *Zarządzanie przez wartość dla klienta – budowa wartości firmy*, Materiały konferencyjne Drugiej Międzynarodowej Konferencji Zarządzania i Technologii Informatycznych „Przedsiębiorstwo Przyszłości”, Warszawa 2001 (dostęp: 2011-11-22). Dostępny w Internecie: http://www.zti.com.pl/institut/pp/referaty/ref4_full.html.

² P. Kotler, *Marketing*, Rebis, Poznań 2005, s. 70.

Piotr Drygas twierdzi, że klienci internetowi są bardzo wymagający i wykazują klarownie określone preferencje³. Istotę wyzwania, przed jakim stoi e-handel, trafnie ujmuje Joanna Harasymów, twierdząc, że podstawą zdobycia lojalnego klienta w Internecie jest przede wszystkim poznanie jego preferencji i dostosowanie się do nich⁴. Dostosowanie to określane jest mianem adaptacji. Jednym z jej rodzajów jest personalizacja, która w najszerszym ujęciu definiowana jest jako każda akcja, która dostosowuje informacje lub usługi oferowane przez stronę internetową do potrzeb konkretnego użytkownika, wykorzystując wiedzę na temat zachowania użytkowników i ich preferencji w połączeniu z zawartością i strukturą strony⁵. Przykładem praktycznego zastosowania personalizacji są systemy rekomendacyjne, których zadaniem jest dostarczanie użytkownikom rekomendacji produktów, informacji i usług jak najlepiej dopasowanych do ich osobistych preferencji i potrzeb. Przykłady praktycznego zastosowania systemów rekomendacyjnych obejmują rekomendację filmów, książek i muzyki, czy też wszystkich produktów dostępnych w sklepach internetowych. Dostarczanie skutecznych rekomendacji wymaga jak najlepszego poznania indywidualnych potrzeb i preferencji e-klientów. Do pozyskiwania preferencji w obszarze serwisów internetowych wykorzystywane są techniki sprzężenia zwrotnego (ang. *relevance feedback*), które mogą realizować ten proces w sposób jawny (ang. *explicit*) lub niejawny (ang. *implicit*).

1. Techniki pozyskiwania preferencji użytkowników serwisów internetowych

Jawne pozyskiwanie preferencji polega na bezpośrednim pytaniu użytkowników o nie, gdyż – jak zauważa Jakob Nielsen – to właśnie użytkownicy z reguły wiedzą najwięcej o własnych preferencjach, potrzebach i celach⁶. Pozyskiwanie preferencji realizowane jest najczęściej w trakcie pierwszego kontaktu użytkownika z systemem, np. poprzez wypełnianie formularzy rejestracyjnych, jak również w trakcie współpracy z systemem. W handlu elektronicznym klienci mogą dokonać oceny produktu w skali punktowej, bądź też zamieścić swoją opinię. Niestety w wielu przypadkach pytanie użytkowników o ich preferencje jest niemożliwe lub niepożądane. Problem ten dotyczy przede wszystkim serwisów internetowych, któ-

³ P. Drygas, *Zarządzanie marketingowe przedsiębiorstwem handlu internetowego*. Electronic Commerce: Teoria i Zastosowania, Gdańsk 2002.

⁴ J. Harasymów, *Kształtowanie lojalności klientów sklepów internetowych* (dostęp: 2011-02-24). Dostępny w Internecie: http://www.biznesfirma.pl/Kszaltowanie_lojalnosci_klientow_sklepow_internetowych,14770.html.

⁵ M. Eirinaki, M. Vazirgiannis, *Web mining for web personalization*. ACM Transactions on Internet Technology (TOIT). ACM. February 2003, Vol. 3, Issue 1.

⁶ J. Nielsen, *Personalization is over-rated*. Alertbox (dostęp: 2010-06-15). Dostępny w Internecie: <http://www.useit.com/alertbox/981004.html>.

rych użytkownicy chcą z reguły od razu przystąpić do realizacji swoich celów. Jak wykazują badania, pytanie użytkowników o preferencje zakłóca realizację zadań, a konsekwencji nawet ich złości i denerwuje⁷. Obserwacja zachowania się użytkowników prowadzi do wniosku, iż są oni niechętni podejmowaniu dodatkowych działań, jeśli nie uważają, iż mogą osiągnąć z nich korzyści⁸. Niestety prowadzi to bardzo często do całkowitej rezygnacji z wyrażania swoich opinii o oglądanych towarach⁹. Jeśli jednak użytkownicy wykażą chęci do dokonywania ocen, to oceniają znacznie mniejszą liczbę produktów niż w rzeczywistości oglądają¹⁰.

Ze względu na wspomniane wady podejścia jawnego w praktyce stosuje się podejście niejawne polegające na obserwacji zachowania użytkowników, na podstawie którego wyciągane są wnioski o preferencjach i zainteresowaniach użytkowników. Chociaż techniki te są z reguły mniej dokładne, są one pozbawione najistotniejszych wad technik jawnych, ponieważ odkrywanie preferencji odbywa się w sposób niewidoczny dla użytkownika, nierozpraszający go i niewymagający od niego żadnych dodatkowych działań. Do wnioskowania o preferencjach użytkowników wykorzystywane są różnorodne metody obserwacji, pozwalające na monitorowanie zachowania należącego do różnego rodzaju kategorii. Najbardziej podstawowe zastosowanie ma założenie, że każdy zasób (strona) otwarty przez użytkownika świadczy o jego zainteresowaniach. Wykorzystane do tego celu mogą zostać logi serwera WWW. Jednakże możliwość śledzenia zachowania użytkowników w serwisach internetowych dostarcza pełniejszego obrazu ich zachowania. Technologie i języki tworzące serwisy WWW takie jak: HTML, DOM, JavaScript pozwalają na śledzenie każdego ruchu myszą. Powszechnie przyjęty wzór klasyfikacji rodzajów zachowania użytkowników serwisów WWW, wykorzystywanych przez techniki niejawnego sprzężenia zwrotnego, można znaleźć w pracy Douglasa. W. Orada i Jinmooka Kima¹¹. Klasyfikacja ta w wersji rozszerzonej została przedstawiona przez Diane Kelly¹². Zawiera ona pięć kategorii rodzajów zachowania: Ba-

⁷ M. Kellar et al., *Effect of task on time spent reading as an implicit measure of interest*. American Society for Information Science and Technology 2004, Volume 41, Issue 1, s. 168–175.

⁸ J. Grundin, *Groupware and Social Dynamics: Eight Challenges for Developers*. Communications of the ACM, January 1994. Vol. 37, Issue 1, s. 35–92.

⁹ C. Avery, R. Zeckhauser, *Recommender systems for evaluating Computer Messages*. Communications of the ACM. March 1997, Vol. 40, Issue 3, s. 40–88.

¹⁰ B. Sawar et al., *Using filtering agents to improve prediction quality in the GroupLens Research Collaborative Filtering System*. Conference on Computer Supported Cooperative Work. ACM, New York 1998.

¹¹ D.W. Orada, J. Kim, *Modeling Information Content Using Observable Behavior*. Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology, USA 2001, s. 38–45.

¹² D. Kelly, *Implicit Feedback: Using Behavior to Infer Relevance*. New Directions in Cognitive Information Retrieval. The Information Retrieval Series, 2005, Vol. 19, Section IV, s. 169–186.

danie (*Examine*), Zachowywanie (*Retain*), Tworzenie odwołania (*Reference*), Adnotowanie (*Annotate*), Tworzenie (*Create*). Stosunkowo najczęściej obserwowanymi rodzajami zachowania są te należące do klasy Badanie, gdyż z reguły użytkownicy wykonują działania mające na celu lepsze poznanie obiektu – zbadanie go. Celem działań należących do kategorii Zachowywanie jest utrwalenie obiektu w celu jego późniejszego użycia, np. poprzez: zapis, wydruk lub dodanie do zakładek. Zachowania z kategorii Tworzenie odwołania zmierzają do utworzenia odwołania – powiązania do obiektu, np. hiperłącza. Adnotowanie polega na wzbogacaniu obiektu o własne komentarze lub też jego organizowanie w hierarchii. Stosunkowo rzadko obserwowanymi rodzajami zachowania są te, które polegają na stworzeniu nowego obiektu. Ich wystąpienie świadczy zdecydowanie pozytywnie o zainteresowaniu użytkowników. W zależności od możliwości technicznych i samej specyfikacji zachowania ich obserwacja może być prowadzona w obrębie klasy obiektów, obiektu lub segmentu. Na bazie danych uzyskanych w trakcie obserwacji zachowania użytkowników obliczane są tzw. niejawne wskaźniki zachowania (ang. *implicit interest indicators*). W literaturze można spotkać szereg propozycji różnorodnych wskaźników. Niestety występuje spory niedostatek badań w zakresie ich praktycznego zastosowania w procesie predykcji preferencji e-użytkowników. Co więcej, wyniki badań są miejscami rozbieżne. W szczególności spory niedobór badań występuje w obszarze serwisów handlu elektronicznego.

2. Procedura badawcza

W celu wypełnienia opisanej powyżej luki badawczej autor opracował i przeprowadził wstępne badanie poznawcze oraz stworzył niezbędne narzędzia. Na potrzeby badania zostało opracowane autorskie rozszerzenie (BPE) dla przeglądarki FireFox. Wybór tej technologii został podyktowany wysoką popularnością przeglądarki oraz dobrze udokumentowaną procedurą tworzenia rozszerzeń. Aby przystąpić do badania konieczne jest zainstalowanie rozszerzenia w przeglądarce FireFox. Należy zauważyć, że każdy użytkownik ma możliwość wyłączenia rozszerzenia lub jego deinstalacji w dowolnym momencie.

W trakcie oglądania stron z produktami rozszerzenie monitorowało aktywność uczestników badania w obrębie pięciu serwisów. Na potrzeby badania wybrano pięć sklepów internetowych: agito.pl, komputronik.pl, electro.pl, morele.net, merlin.pl. Jako kryterium wyboru przyjęto popularność serwisu oraz możliwość implementacji monitorowania zachowania użytkowników. Należy zauważyć, że rozszerzenie BPE monitoruje aktywność użytkowników tylko w obrębie tych sklepów. Na podstawie zgromadzonych danych rozszerzenie wyliczało kilkadziesiąt wskaźników opisujących zachowanie użytkowników, jak też parametry odwiedzonych stron (tabela 1). Dziewięć z przedstawionych wskaźników stanowi autorskie propozycje.

Tabela 1

Wskaźniki zachowania i parametry stron rejestrowane przez rozszerzenie BPE

Parametr	Objaśnienie
mark	ocena zainteresowania produktem
familiar	wcześniejsza znajomość produktu
Parametry określające cechy strony produktu	
document_length	liczba znaków tekstu zawartego na stronie
desc_length	liczba znaków opisu produktu zawartego na stronie
review_length	liczba znaków opinii o produkcie zawartego na stronie produktu
recommend_length	liczba znaków tekstów dotyczących rekomendowanych produktów
image_number	liczba zdjęć produktu zawartych na stronie
page_height	wysokość zawartości strony w pikselach
Parametry określające czasy interakcji	
page_time	czas otwarcia strony
tab_activ_time	czas aktywności karty
user_activ_time	czas aktywności użytkownika
prod_desc_time	czas przebywania kursora w obrębie opisu produktu
prod_recommend_time	czas przebywania kursora w obrębie rekomendacji
prod_review_time	czas przebywania kursora w obrębie opinii o produkcie
prod_image_time	czas przebywania kursora w obrębie zdjęć produktu
prod_other_time	czas przebywania kursora w obrębie pozostałego obszaru
Parametry określające zachowanie użytkowników	
mouse_distance	dystans kursora, o jaki został przesunięty kursor myszy
vertical_scroll	dystans przewijania strony w pionie
horizontal_scroll	dystans przewijania strony w poziomie
mouse_clicks	liczba kliknięć myszy
lb_mouse_clicks	liczba kliknięć lewego klawisza myszy
rb_mouse_clicks	liczba kliknięć prawego klawisza myszy
mb_mouse_clicks	liczba kliknięć środkowego klawisza myszy
copycut_action	liczba zdarzeń kopiowania/wycinania
select_action	liczba zdarzeń zaznaczania
select_text_size	liczba znaków zaznaczonego tekstu
keydown_single	liczba zdarzeń wielokrotnego wciśnięcia klawisza
keydown_repeatable	liczba zdarzeń pojedynczego wciśnięcia klawiszy
find_action	liczba akcji wyszukiwania
print_action	liczba akcji drukowania
bookmark_action	liczba akcji tworzenia zakładki
save_action	liczba akcji zapisu
resize_action	liczba akcji modyfikowania rozmiaru wyświetlanego dokumentu
search_referral	produkt odszukany poprzez wyszukiwarkę
Relatywne parametry określające zachowanie użytkowników	
rel_page_time	relatywny czas otwarcia
rel_user_activ_time	relatywny czas aktywności użytkownika
rel_tab_active_time	relatywny czas aktywności karty
rel_prod_desc_time	relatywny czas przebywania kursora w obrębie opisu produktu
rel_prod_recommend_time	relatywny czas przebywania kursora w obrębie rekomendacji
rel_prod_review_time	relatywny czas przebywania kursora w obrębie opinii o produkcie
rel_prod_image_time	relatywny czas przebywania kursora w obrębie zdjęć produktu
rel_mouse_distance	relatywny dystans kursora
rel_vertical_scroll	relatywny dystans przewijania strony w pionie
rel_horizontal_scroll	relatywny dystans przewijania strony w poziomie

Źródło: opracowanie własne.

W momencie opuszczania strony zawierającej informacje o produkcie respondenci byli pytani jawnie o ocenę zainteresowania produktem znajdującym się na tej stronie. Ocena ta wyrażana była w skali pięciostopniowej 1–5, gdzie wartości 1 przypisano oznaczenie „nie interesujący” a wartości 5 – „bardzo interesujący”. Uczestnik badania mógł również określić to, czy znał wcześniej oglądany produkt.

W trakcie korzystania z rozszerzenia BPE wszystkie dane gromadzone są na dysku lokalnym użytkownika. W momencie zakończenia badania użytkownik może skorzystać z opcji przesłania zgromadzonych danych na serwer badawczy. Procedura badawcza oraz rozszerzenie BPE zostały opracowane w taki sposób, aby zapewnić jak najbardziej naturalne warunki środowiska pracy użytkowników e-sklepów.

3. Wyniki badania wstępnego

W toku prowadzonych badań mających na celu opracowanie modelu klasyfikacyjnego preferencji e-klientów w kwietniu 2011 roku przeprowadzono badania wstępne. W badaniu uczestniczyło 26 studentów Wydziału Informatyki ZUT, którzy dokonali oceny 448 towarów. Studenci uczestniczyli w badaniu w trakcie zajęć z zakresu e-biznesu. Przed rozpoczęciem badania zostali poproszeni o poszukiwanie interesujących ich produktów w pięciu sklepach biorących udział w badaniu. Respondenci nie zostali poinformowani o zakresie monitorowanych aktywności przez rozszerzenie BPE. Miało to na celu ograniczenie świadomego wpływu na wyniki badania. Dane zgromadzone w badaniu zostały przekształcone w opisany poniżej sposób. Zmienna objaśniana marka – ocena zainteresowania została przekształcona ze skali pięciostopniowej na skalę dwustopniową. Wartości 1, 2 i 3 zostały przekształcone na wartość 0 – wyrażającą brak zainteresowania i stosunkowo niskie zainteresowanie. Natomiast wartości 4–5 zostały przekształcone na wartość 1 – oznaczającą wysoki poziom zainteresowania. Do budowy modelu klasyfikacyjnego preferencji wykorzystane zostały drzewa klasyfikacyjne. Jako metody walidacji użyto metody k-krotnej walidacji krzyżowej ($k = 10$). Uzyskany model drzewa cechuje 76,8-proc. dokładność klasyfikacji, 75,3-proc. czułość oraz 78,3-proc. specyficzność. Oznacza to, że drzewo przewiduje stosunkowo dobrze przypadki rzeczywistego zainteresowania lub jego braku. Drzewo składa się z siedmiu poziomów, dziesięciu węzłów, szesnastu liści oraz maksymalnie trzech gałęzi. Na podstawie przeglądu postaci drzewa można wyszczególnić najistotniejsze zmienne: `recommend_length`, `mouse_distance`, `desc_length`, `verical_scroll`, `page_height`, `rel_tab_activ_time` i `review_length`. Analizując uzyskane drzewo, należy zauważyć, że: długość rekomendacji, wysokość strony i długość opisu cechuje dodatnie powiązanie z zainteresowaniem użytkowników.

Przedstawione wyniki stanowią owoc badań wstępnych, wymagający pogłębienia oraz dalszej ich weryfikacji.

Literatura

1. Avery C., Zeckhauser R., *Recommender systems for evaluating Computer Messages*. Communications of the ACM. March 1997, Vol. 40, Issue 3, s. 40–88.
2. Drygas P., *Zarządzanie marketingowe przedsiębiorstwem handlu internetowego*. Electronic Commerce: Teoria i Zastosowania, Gdańsk 2002.
3. Eiriniaki M., Vazirgiannis M., *Web mining for web personalization*. ACM Transactions on Internet Technology (TOIT). ACM. February 2003, Vol. 3, Issue 1.
4. Grundin J., *Groupware and Social Dynamics: Eight Challenges for Developers*. Communications of the ACM, January 1994. Vol. 37, Issue 1, s. 35–92.
5. Harasymów J., *Kształtowanie lojalności klientów sklepów internetowych* (dostęp: 2011-02-24). Dostępny w Internecie: http://www.biznesfirma.pl/Kształtowanie_lojalnosci_klientow_sklepow_internetowych,14770.html.
6. Kellar M. et al., *Effect of task on time spent reading as an implicit measure of interest*. American Society for Information Science and Technology 2004, Vol. 41, Issue 1, s. 168–175.
7. Kelly D., *Implicit Feedback: Using Behavior to Infer Relevance*. New Directions in Cognitive Information Retrieval. The Information Retrieval Series, 2005, Vol. 19, Section IV, s. 169–186.
8. Kotler P., *Marketing*, Rebis, Poznań 2005, s. 70.
9. Nielsen J., *Personalization is over-rated*. Alertbox (dostęp: 2010-06-15). Dostępny w Internecie: <http://www.useit.com/alertbox/981004.html>.
10. Orad D.W., Kim J., *Modeling Information Content Using Observable Behavior*. Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology, USA 2001, s. 38–45.
11. Sawar B. et al., *Using filtering agents to improve prediction quality in the GroupLens Research Collaborative Filtering System*. Conference on Computer Supported Cooperative Work. ACM, Nowy Jork 1998.
12. Szymura-Tyc M., *Zarządzanie przez wartość dla klienta – budowa wartości firmy*. Materiały konferencyjne Drugiej Międzynarodowej Konferencji Zarządzania i Technologii Informatycznych „Przedsiębiorstwo Przyszłości”. Warszawa 2001 (dostęp: 2011-11-22). Dostępny w Internecie: http://www.zti.com.pl/instytut/pp/referaty/ref4_full.html.

E-CUSTOMERS PREFERENCES' IDENTIFICATION

Summary

Increasing competition and decreasing customers' loyalty requires companies to take upon activities aiming toward building long term relationships with customers. These is especially crucial objective for e-commerce sector. Personalization is one of the techniques used nowadays for building individual customers' experience. This process requires to possess good knowledge of their preferences. One of techniques used to obtain such knowledge is implicit feedback. This article proposes new procedure for obtaining e-commerce customer's behavior data and learning about their preferences from them. Main contribution of this procedure is extension for FireFox browser which allow to run study in user's natural environment. Extension created for this study allows to monitor variety of events, which are used for calculating implicit interest indicators. This article describes results of preliminary study for modeling e-customers preferences with implicit feedback techniques.

Translated by Tomasz Zdziebko