

Polsko-Japońska Akademia Techniki Komputerowych



Paulina Adamska

Autoreferat pracy doktorskiej

Reputation Systems on the Web

Promotor: Dr hab. Adam Wierzbicki

Warszawa, 2015

Spis treści

Wstęp	2
Proponowane rozwiązania	3
Model do analizy wpływu systemu reputacyjnego na strategie użytkowników	3
Odporność systemu reputacyjnego na strategie adwersarzy	4
Użyteczność systemu reputacyjnego	4
Podsumowanie	5
Bibliografia.....	5

Wstęp

Ewolucja sieci Internet umożliwiła zwykłym użytkownikom publiczne udostępnianie opracowywanych przez siebie treści, jak również bezpośrednią i anonimową interakcję z innymi członkami wirtualnej społeczności. Zjawisko to zaowocowało pojawieniem się pojęcia "Web 2.0". Oprócz rozlicznych korzyści wynikających z tej zmiany, pojawiły się także wątpliwości związane z jakością prezentowanych treści oraz świadczonych usług. Anonimowość sprzyjała występowaniu tzw. "pokusy nadużycia" [1], wiążącej się z faktem braku bezpośredniego narażenia na ryzyko kary.

Aby zapobiegać wspomnianemu zjawisku, wiele platform wprowadziło *systemy reputacyjne*, jako mechanizmy gromadzące informacje dotyczące wcześniejszych aktywności użytkowników, i prezentujące te dane innym członkom społeczności, umożliwiając tym samym porównywanie twórców treści, czy też dostawców usług. Systemy tego rodzaju stosowane są zarówno przez platformy typu e-commerce (np. platformy aukcyjne, jak Allegro¹ czy eBay²), jak również przez systemy zamieszczające recenzje usług (np. Yelp³, czy Epinions⁴), oceny wiarygodności treści WWW (np. MyWOT⁵, czy Reconcile⁶) lub odpowiedzi na pytania (np. portale z rodziny StackExchange⁷, czy Quora⁸).

Badania naukowe, dotyczące systemów reputacyjnych koncentrują się na różnych aspektach takich systemów:

1. Odporności na działania adwersarzy
2. Motywującym (lub demotywującym) wpływie reputacji na aktywność użytkowników
3. Użyteczności prezentowanych informacji
4. Korelacji reputacji z określonymi cechami użytkowników platform (przykładowo poziomem ich wiedzy)

¹ <http://allegro.pl/>

² <http://www.ebay.com/>

³ <http://www.yelp.com>

⁴ <http://www.epinions.com/>

⁵ <https://www.mywot.com/>

⁶ <http://reconcile.pl/about/>

⁷ <http://stackexchange.com/>

⁸ <https://www.quora.com/>

Badania będące częścią rozprawy koncentrowały się na trzech pierwszych obszarach. Najistotniejszymi ich rezultatami są:

1. Nowy model służący do analizy funkcjonowania określonych systemów reputacyjnych na platformie wspomagającej ocenę wiarygodności treści
2. Propozycja zwiększenia odporności systemu reputacyjnego, projektowanego dla platformy wspomagającej ocenę wiarygodności treści, na działalność adwersarzy
3. Propozycja poprawienia użyteczności informacji, prezentowanych przez system reputacyjny pozwalający na wielokryterialną ocenę pojedynczej aktywności użytkownika.

Opisywane mechanizmy opracowywane były z myślą o ściśle określonych platformach, natomiast obszar ich potencjalnego zastosowania nie musi ograniczać się jedynie do opisanego w rozprawie kontekstu.

Proponowane rozwiązania

Model do analizy wpływu systemu reputacyjnego na strategię użytkowników

Prezentowany w rozprawie model Credibility Game może posłużyć do badania zależności pomiędzy efektywnością systemów reputacyjnych, a aktywnością działających strategicznie członków społeczności. Do przeprowadzonych badań wykorzystana została zarówno analiza teoretyczna, jak i symulacje. Badania dotyczyły w szczególności systemów reputacyjnych implementowanych na platformach wspomagających ocenę wiarygodności stron WWW. Proponowany model pozwala uchwycić sytuacje, kiedy użytkownicy oceniają strony jedynie na podstawie pierwszego wrażenia (bazując głównie na sposobie prezentacji informacji), jak również sytuacji, w której informacja jest oceniana po wcześniejszej empirycznej weryfikacji. Wspomniane pierwsze wrażenie modelowane jest za pomocą sygnału, który jest funkcją dwóch właściwości informacji: jakości prezentacji, i prawdziwości samej treści. Funkcja ta może zostać użyta do modelowania poziomu wiedzy eksperckiej oceniającego użytkownika. Sam algorytm reputacyjny jest prosty, i przypomina wiele mechanizmów w praktyce wykorzystywanych w podobnym kontekście.

Korzystając z opisanego modelu, zweryfikowane zostały następujące hipotezy:

1. Poziom wiedzy eksperckiej użytkowników oceniających treści jest kluczowym czynnikiem, wpływającym wybór strategii twórców informacji

Zgodnie z intuicją model pokazuje, że w społeczności laików nieuczciwi twórcy treści faktycznie zamieszczają nieprawdziwe treści. Z kolei zwiększenie liczby ekspertów motywuje do zmiany zachowania.

2. Kiedy sygnał dotyczący jakości informacji nie jest precyzyjny, a system reputacyjny posiada wystarczającą ilość informacji (nie występuje zjawisko określane mianem *whitewashingu*), użytkownicy korzystają także z reputacji podczas podejmowania decyzji o akceptacji, lub odrzuceniu informacji.

3. Reputacja bazująca na rzetelnej ewaluacji informacji (w odróżnieniu od powierzchownej oceny opartej o pierwsze wrażenie) jest korzystna zarówno dla producentów treści, jak i dla konsumentów.

Odporność systemu reputacyjnego na strategię adwersarzy

W literaturze przeanalizowanych zostało wiele strategii adwersarzy oraz metod ich zwalczania [3, 4, 5, 6], natomiast proponowany mechanizm został zaprojektowany na potrzeby platformy wspomagającej ocenę wiarygodności treści, i koncentruje się na jednej, charakterystycznej dla tego rodzaju systemów, technice. System ma za zadanie wykrywać użytkowników, których celem jest jak najszybsze i jak najłatwiejsze uzyskanie wysokiej reputacji. W przypadku wspomnianej platformy najprostszą techniką jest kopiowanie istniejących ocen treści, i zgłaszanie ich jako własnych opinii [2]. Dzięki tego rodzaju zabiegom użytkownik może później wystawiać fałszywe oceny wybranym treściom, a oceny te będą traktowane jak opinie eksperta.

Przedstawiona w pracy technika wykrywania adwersarzy bazuje na koncepcji sugerowania użytkownikom treści do ewaluacji. Część z nich prezentowana jest z zafałszowaną oceną, dzięki czemu platforma może wykryć, czy użytkownik rzetelnie zweryfikował treść, czy jedynie skopiował widoczną (fałszywą) opinię innych użytkowników. Na podstawie zebranych informacji system oblicza prawdopodobieństwo niepożądanego zachowania w przyszłości. Jeśli przekracza ono pewien określony próg, użytkownik uznawany jest za adwersarza, i platforma może odpowiednio zredukować wpływ jego opinii na zagregowaną ocenę treści.

Użyteczność systemu reputacyjnego

W celu poprawy jakości prezentowanych informacji, w części systemów reputacyjnych zdecydowano się wprowadzić możliwość wielokryterialnej oceny aktywności użytkowników. Przykładowo, platforma aukcyjna eBay, w 2007 roku wprowadziła system Detailed Seller Ratings (DSR). System ten miał służyć jako źródło dodatkowych danych dotyczących transakcji. Definiuje on cztery kryteria oceny sprzedających: zgodność przedmiotu z opisem, czas i koszt dostawy, oraz komunikacja z klientem. Kiedy kupujący określa ogólny poziom satysfakcji z przeprowadzonej transakcji, może także ocenić sprzedającego w każdej kategorii DSR, używając 5-stopniowej skali. Średnia takich ocen w każdej kategorii pojawia się na stronie profilu użytkownika, obok standardowej reputacji przedstawianej w postaci liczby całkowitej. Zaprezentowane w takiej formie informacje nie pozwalają na automatyczne przygotowanie prostego rankingu sprzedawców. Nie uwzględniają też indywidualnych preferencji użytkownika, dotyczących istotności poszczególnych kryteriów. W rozprawie zaproponowany został operator pozwalający na agregację tych informacji z uwzględnieniem personalizacji wag określonych kryteriów.

Dodatkowym problemem, związanym z zastosowaniem DSR jest fakt, że statystyki pochodzące z tego systemu są monitorowane nie tylko przez kupujących, ale także przez samą platformę, dla której stanowią podstawę do przyznania sprzedawcy statusu PowerSeller. Status ten daje sprzedawcy pewne przywileje, zatem jest on zainteresowany, aby jego statystyki nie spadły poniżej określonego progu, który z kolei wyznaczany jest na podstawie wszystkich ocen. Ponieważ, jak wykazują analizy, dla pozytywnie ocenianych transakcji ocena poniżej 4 jest niezwykle rzadka, próg wymagany do utrzymania statusu PowerSeller jest wysoki. Z tego względu niedoświadczony użytkownik wystawiając (teoretycznie pozytywną) ocenę 4 może spowodować utratę przywilejów przez uczciwego sprzedającego. Proponowany w rozprawie mechanizm przechowuje informację o tym, jak

użytkownicy wykorzystują skalę ocen, i w razie potrzeby koryguje oceny przed przystąpieniem do wyliczania zagregowanej wartości.

Podsumowanie

Odpowiednio zaprojektowany system reputacyjny jest istotnym mechanizmem, wspomagającym właścicieli platform online w kształtowaniu i promowaniu określonych wzorców zachowań wśród użytkowników. Stworzenie takiego, poprawnie działającego systemu nie jest zadaniem trywialnym chociażby ze względu na różnorodność celów i strategii członków społeczności online. Jednocześnie wprowadzanie reguł, które nie zostały uprzednio kompleksowo przetestowane w kontrolowanych warunkach, może być krokiem ryzykownym, ze względu na fakt częstego przywiązywania ogromnej wagi do wypracowanej reputacji i wiążącego się z nią statusu w społeczności (szczególnie wśród najbardziej aktywnych użytkowników). Z tego względu istotne jest opracowywanie narzędzi, które (choć w ograniczonym stopniu) pozwolą wstępnie określać, jaki wpływ na funkcjonowanie platformy online mogą mieć zmiany, zarówno te wprowadzane przez jej właścicieli, jak i te dotyczące samej struktury społeczności (np. pojawienia się nowej strategii adwersarza). W tym właśnie celu może być wykorzystany proponowany w pracy model.

Oprócz samego modelu, w rozprawie zaprezentowany został konkretny mechanizm, pozwalający na ograniczanie wpływu wybranej strategii adwersarza na poprawne działanie platformy wspomagającej ocenę wiarygodności informacji w sieci. Proponowane rozwiązanie zostało przetestowane za pomocą symulacji. Rezultaty tej wstępnej weryfikacji wydają się obiecujące, a w przyszłości mają szansę zostać zweryfikowane także eksperymentalnie, ponieważ mechanizm został zaimplementowany na publicznie dostępnej platformie do oceny wiarygodności, która powstała w ramach projektu Reconcile⁹.

Aby udostępnić możliwie jak najbardziej szczegółowe informacje dotyczące aktywności użytkowników, systemy reputacyjne często pozwalają oceniać udostępniane treści/świadczone usługi uwzględniając wiele kryteriów. Taki sposób oceny implementowany jest zarówno przez platformy aukcyjne, jak również wspomnianą wcześniej platformę do oceny wiarygodności treści na stronach WWW. Dla tego rodzaju systemów wyzwaniem jest prezentowanie informacji o dostawcach usług/autorach treści w przejrzysty, umożliwiający automatyczne wygenerowanie rankingu sposób. Dodatkowym atutem może być także uwzględnianie indywidualnych preferencji konsumentów, dotyczących istotności poszczególnych kryteriów oceny. Tego rodzaju zagadnieniom poświęcona jest trzecia część badań, opisanych w rozprawie. Przeprowadzone analizy pokazują, że dzięki zdefiniowaniu odpowiednich operatorów możliwe jest sprostanie powyższym wyzwaniom, a tym samym zdjęcie z użytkowników-konsumentów konieczność analizy i porównywania dwóch oddzielnych metryk - ogólnej reputacji (na platformie eBay określanej mianem Feedback Score) i szczegółowych statystyk (na platformie eBay określanej mianem DSR).

Bibliografia

[1] Dembe, A. E. and Boden, L. I. (2000). Moral hazard: a question of morality? *NEW SOLUTIONS*, 10(3):257–279.

⁹ <http://reconcile.pjwstk.edu.pl/>

- [2] Liu, X., Nielek, R., Adamska, P., Wierzbicki, A., and Aberer, K. (2015). Towards a highly effective and robust web credibility evaluation system. *accepted to Decision Support Systems*.
- [3] Vu, L., Zhang, J., and Aberer, K. (2014). Using identity premium for honesty enforcement and whitewashing prevention. *Computational Intelligence*, 30(4):771–797.
- [4] Harmon, A. (2004). Amazon glitch unmask war of reviewers. *The New York Times*.
- [5] Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, 104(8):2421–55.
- [6] Douceur, J. R. (2002). The sybil attack. In *Revised Papers from the First International Workshop on Peer-to-Peer Systems, IPTPS '01*, pages 251–260, London, UK, UK. Springer-Verlag.