



KAŻDA NASZA AKTYWNOŚĆ POZOSTAWIA CYFROWY ŚLAD



tekst: Agnieszka Niewdana



Dr inż. Roman Simiński
Instytut Informatyki
Wydział Nauk Ścisłych i Technicznych
Uniwersytet Śląski
roman.siminski@us.edu.pl

Komputery są w naszym życiu wszechobecne. Wykonują dla nas potrzebne działania, rejestrując duże ilości informacji, które łatwo jest obecnie pozyskiwać, przetwarzać i magazynować. Informacje rejestrują zarówno nasze urządzenia komputerowe, zapisując je na lokalnych nośnikach danych (dyski, pamięci typu *flash*), jak również w tzw. chmurach obliczeniowych. Korzystanie z systemów komputerowych po zalogowaniu się do odpowiedniego konta i usługi pozwala nam wykonywać działania spersonalizowane. Coraz częściej korzystamy z możliwości zapisywania naszych danych „gdzieś w internecie”.

Spersonalizowanie zasobów oraz usług bazujących na chmurach sprawia jednak, że informacje o nas, naszych zasobach oraz aktywnościach, są skrzętnie zapisywane przez oprogramowanie zarządzające dostępem do chmur obliczeniowych i wszelkich innych usług internetowych. Każda nasza aktywność pozostawia cyfrowy ślad – przykładowo, po korzystaniu z portali społecznościowych, sklepów internetowych czy usług finansowych. Warto pamiętać, że taki ślad zostawiamy również, dokonując fizycznych zakupów w sklepach, korzystając z kart płatniczych i kart lojalnościowych. Właściwie nie pozostawiamy go wyłącznie wtedy, gdy realizujemy płatności gotówkowo.

Niezależnie od tego, czy ślad cyfrowy pozwala na naszą identyfikację, czy też pozostajemy anonimowi, zawiera on informacje o naszej aktywności. O tym co, kiedy i za jaką kwotę zakupiliśmy, jaką transakcję finansową przeprowadziliśmy, co wyszukiwaliśmy w internecie, gdzie i jak przemieszczaliśmy się, jakie zrobiliśmy zdjęcia (oraz kiedy i gdzie). Ślady naszej aktywności zarejestrowane w systemach komputerowych wiążą się z szeregiem istotnych problemów związanych między innymi z poufnością informacji o nas. Nieuprawnione ich wykorzystanie np. w celach przestępczych to rzecz możliwa oraz groźna. Informacje zarejestrowane przez systemy informatyczne mogą jednak służyć wielu korzystnym dla nas działaniom.

Informacje zapisane przez system bankowy w pewnym określonym czasie są nie tylko wprost zapisem pojedynczych

zdarzeń, ale również, w sposób niejawny, stanowią odzwierciedlenie procesów zachodzących w zakresie naszych rejestrowanych na koncie środków finansowych. Szczegółowa analiza większej liczby zdarzeń z pewnego okresu pozwolić może na wydedukowanie wielu często zaskakujących informacji o rzeczywistych zdarzeniach z naszego życia. Przykładowo, ustanie comiesięcznych stałych wpływów z pewnego konta może świadczyć o potencjalnej utracie pracy, ale jednocześnie rozpoczęcie przepływu stałych, comiesięcznych przychodów o wyższym poziomie świadczyć może po prostu o tym, że zmieniliśmy pracę na lepszą.

Analiza anonimowych zakupów w sklepie samoobsługowym może nieść wiele informacji istotnych dla właściciela sklepu. Zawartość koszyków może pozwolić np. na wytypowanie grup towarów kupowanych najczęściej razem. Ich identyfikacja może być wykorzystana do takiej kompozycji i ekspozycji towarów, aby te kupowane najczęściej były rozlokowane w bezpośredniej bliskości.

Przykłady te pokazują proste zastosowania bardzo istotnej obecnie dziedziny sztucznej inteligencji, jaką jest inteligentna analiza danych nazywana również nauką o danych (ang. *data science*). Koncepcja wydobywania wie-

i prowadzącego działalność naukową w USA. Algorytmy uczenia maszynowego bazują na przykładach, z których w sposób automatyczny mają się czegoś nauczyć. *Nauczyć* oznacza utworzyć pewien opis zawierający wcześniej nieznaną wiedzę na temat prawidłowości, relacji, tendencji występujących w danych uczących. Koncepcja maszynowego uczenia miała w sposób inny niż algorytmiczny nauczyć maszynę sposobu rozwiązania problemu. Wyniki działania algorytmu mogą być różne, bywają to najczęściej drzewa decyzyjne (ang. *decision trees*) lub reguły decyzyjne (ang. *decision rules*). W obu przypadkach wyniki działania algorytmu uczenia maszynowego pozwalają na próbę sklasyfikowania nowych przypadków.

Uczenie maszynowe w istocie pozwala nam odkryć wiedzę na temat rozwiązywanego problemu. Rozwijając i uogólniając tę koncepcję, dochodzimy do koncepcji eksploracji danych (ang. *data mining*), czasem określanego też jako *drążenie danych*. Celem eksploracji danych jest wykrycie nieznannej wcześniej, użytecznej i nietrywialnej wiedzy zapisanej niejawnie w danych. Ponieważ pozyskanie danych do eksploracji może wymagać dodatkowych czynności, np. oczyszczenia i przygotowania, a wynik eksploracji – oceny i weryfikacji, szerszy

Pierwsza i druga dekada XXI wieku to czas wielkiego postępu w zakresie technologii informatycznych. Komputery stały się szybsze, a ich zdolność do przechowywania i przetwarzania różnorodnych informacji – ogromna. Codziennie robimy zdjęcia, prowadzimy rozmowy za pomocą różnych komunikatorów, przesyłamy dokumenty, płacimy kartą za zakupy. Nie zastanawiamy się nad tym, gdzie te dane są zbierane i jak mogą zostać wykorzystane.

dziny z danych nie jest nowa, wywodzi się z koncepcji maszynowego uczenia (ang. *machine learning*). Pierwsze znane i z sukcesem zastosowane metody rozwijane były w drugiej połowie ubiegłego stulecia. Najbardziej znane algorytmy to ID3, C4.5, C5.0 Quinlana, AQ Ryszarda Michalskiego, Polaka mieszkającego

proces, obejmujący eksplorację jako jeden z elementów – nazywany jest *odkryciem wiedzy w danych* (ang. *knowledge discovery in data*). Obecnie możliwości pozyskiwania zbiorów danych mogących zawierać ukrytą, a istotną wiedzę są relatywnie łatwe i powszechne, a zbiory te mogą być objętościowo duże.