

# Wstęp do Sztucznej Inteligencji

Projekt 2: *System rekomendacji lokali gastronomicznych.*

Łozińska Magdalena, Łuczaj Tomasz, Łuka Bartosz,  
Sulik Michał, Tomulewicz Szymon

Prowadzący: dr Anna Maria Radzikowska

23.03.2021

## 1 Co to są systemy rekomendacyjne?

Systemy rekomendacyjne to programy starające się przewidzieć które z dostępnych przedmiotów najbardziej odpowiadałyby danemu użytkownikowi. Powstały po to, by ułatwić ludziom odnalezienie się w ogromnej ilości informacji, a także ulepszyć doświadczenia klienta na danej platformie.

Podstawą działania takich systemów jest zbieranie informacji, które może być wykonywane na dwa sposoby: jawne i domniemane. Jawne zdobywanie danych polega na zapytaniu wprost, na przykład wykorzystując formę ankiety, o preferencje danej osoby lub ocenę jakichś przedmiotów. Niejawne zbieranie danych może mieć formę zapisywania aktywności użytkownika, przedmiotów, które kupuje, jak wysoko je ocenia czy ile czasu spędza na oglądaniu danej strony.

Dwie główne techniki są wykorzystywane przez systemy rekomendacji. Pierwsza z nich, oparta na użytkowniku, zapisuje i interpretuje dane na temat człowieka, jego upodobania, historię zakupów czy podobieństwo do innych klientów. Druga, oparta na zawartości, skupia się na informacjach o przedmiotach – gatunek książki, muzyki, rodzaj restauracji, autor obrazu. Na ich podstawie stara się kojarzyć ze sobą przedmioty o wspólnych cechach, by następnie zaproponować je klientowi, o którym wie, że kiedyś spodobał mu się podobny przedmiot.

Niektóre techniki wymagają zatem czasu od założenia przez użytkownika konta, aby móc sprawnie i skutecznie proponować produkty, podczas gdy na przykład ankieta pytająca o wymienienie ulubionych książek, umożliwia całkiem sprawne przefiltrowanie dostępnych przedmiotów już na początku historii danego klienta.

W tym projekcie wykorzystana jest technika oparta na zawartości. Dane zbierane są w sposób jawny i zapisywane jako atrybuty, na podstawie których odbywa się wyszukiwanie.

## 2 Jaki mamy problem?

Postawionym problemem jest znalezienie  $k$  najbardziej podobnych lokali gastronomicznych do jednego wprowadzonego przez użytkownika lokalu w określonym przez niego mieście. Restauracje są wybierane z danej bazy lokali, gdzie każda z nich posiada dodatkowe atrybuty (np. pub, pizzeria, kawiarnia, bar, restauracja włoska, restauracja hinduska, kuchnia regionalna, kuchnia polska itp.) oraz miasto, w którym się znajduje.

Dana jest baza lokali gastronomicznych ( $R = \{r_1, \dots, r_n\}, n \geq 50$ ) wraz ze zbiorem ich atrybutów ( $A = \{a_1, \dots, a_m\}, 10 \leq m \leq 15$ ) oraz podaną lokalizacją lokalu (miasto). Każdy z lokali może mieć kilka atrybutów. Każda restauracja reprezentowana jest jako wektor zero-jedynkowy o długości  $m$  (liczbie dostępnych atrybutów), gdzie 1 na  $i$ -tym ( $i \in \{1, \dots, m\}$ ) miejscu oznacza, że restauracja posiada  $i$ -ty atrybut a 0, że go nie posiada.

Na wejściu użytkownik podaje lokal, do którego chciałby znaleźć podobne, liczbę  $k$  oznaczającą ile lokali ma zostać zarekomendowanych oraz miasto, do którego mają być zawężone rekomendacje.

Na wyjściu użytkownik otrzymuje listę lokali gastronomicznych podobnych do tego podanego na wejściu.

## 3 Jak go rozwiązujemy?

W celu rozwiązania opisanego powyżej problemu znajdowania najbardziej podobnych lokali gastronomicznych, zastosowano technikę filtrowania na podstawie zawartości (ang. *content-based filtering*) wspomaganą metodą najbliższych sąsiadów.

Technika ta wymaga dwóch rodzajów informacji w celu wyznaczenia odpowiedniej rekomendacji - charakterystyki obiektów, które będą rekomendowane oraz profilu użytkownika. W przypadku tego projektu preferencje użytkownika są określane na podstawie jednej restauracji, którą podaje jako dane wejściowe aplikacji. Zadaniem algorytmu jest znalezienie  $k$  restauracji o najbardziej zbliżonych atrybutach, znajdujących się w konkretnym mieście. Problem ten sprowadza się zatem do wyznaczenia lokali o takich wektorach cech, które są najbardziej podobne do wektora restauracji zadanej przez użytkownika.

Można to wyliczyć za pomocą miary podobieństwa kosinusowego (ang. *cosine similarity measure*) [1]. Podobieństwo dwóch restauracji wyznaczamy ze wzoru:

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}, \quad (1)$$

gdzie  $\vec{a}, \vec{b}$  - wektory atrybutów,  $\cdot$  iloczyn skalarny wektorów,  $|\vec{a}|$  - długość wektora w przestrzeni euklidesowej.

Otrzymane podobieństwo jest wartością z przedziału od 0 do 1, a zbieżność restauracji jest tym większa, im większa jest obliczona wartość [2]. Zatem po obliczeniu miary podobieństwa pomiędzy knajpą zadaną przez użytkownika a każdym innym miejscem znajdującym się w bazie pozostaje wybrać  $k$  miejsc o największych wynikach. Są to miejsca zwrócone klientowi przez system rekomendacyjny.

### 3.1 Przykład działania metody

W celu zaprezentowania działania algorytmu zastosujemy go ręcznie na poniższym przykładzie, dla  $k = 3$ , oznaczającego liczbę podobnych restauracji, które nas interesują.

Weźmy 8 restauracji i profil użytkownika, opisane następującymi atrybutami:

- $PL$  - Kuchnia polska
- $IT$  - Kuchnia włoska
- $JP$  - Kuchnia japońska
- $US$  - Kuchnia amerykańska

- $P_i$  - Przyjazne pieskom
- $D_z$  - Przyjazne dzieciom

Dla przejrzystości wartości atrybutów równe 0 zostały zamienione na puste pola, a restauracje posortowane po wartości metryki podobieństwa. Dla dalszego uproszczenia przykładu, wszystkie restauracje znajdują się w tym samym mieście, co wybrana przez użytkownika (inaczej miałyby wartość metryki równą 0).

Tabela 1: Przykładowe wartości atrybutów i metryki podobieństwa.

Nazwa	PL	IT	JP	US	Pi	Dz	Wartość
U Szwejka	1				1		0.8165
Duże Szynce	1					1	0.8165
Super Bistro	1	1		1	1	1	0.7746
Chłopskie Jadło	1			1		1	0.6667
Pizzeria Mamma		1				1	0.4082
Ciao Napoli		1			1		0.4082
Sushi Maestro			1				0.0
Ave! Pizza		1					0.0
<i>Wybrana</i>	1				1	1	1.0

Algorytm wybierze i zwróci *U Szwejka*, *Duże Szynce* i *Super Bistro*, ponieważ te restauracje mają najwyższe wartości metryki podobieństwa w stosunku do wybranej przez użytkownika restauracji (W Tab. 1 oznaczone jako *Wybrana*). Można zauważyć, że miara podobieństwa kosinusowego (Wzór 1.) jest dobrze określona, ponieważ wartość podobieństwa każdej restauracji ze sobą samą wynosi 1.0 (jak widać w polu *Wartość* restauracji wybranej przez użytkownika), czyli zgodnie z definicją, najwyższą możliwą wartość.

## Literatura

- [1] Google Developers. Candidate generation overview. <https://developers.google.com/machine-learning/recommendation/overview/candidate-generation>, 2020. Dostęp: 2021-03-23.
- [2] D. Janach, M. Zanker, A. Felfering, and G. Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, 2011.