

Symulacja migracji ludności na terenie Stanów Zjednoczonych

Projekt na metody stochastyczne

Filip Ręka
Daniel Kuc
Piotr Rzeźnik

A
G
H

19 maja 2023

Wstęp

Migracja jest powszechna w dzisiejszym świecie. Ludzie migrują ze względu na pracę, szkołę, studia, rodzinę, a także przez wiele innych czynników. Migrację można podzielić ze względu na kilka aspektów, ale w tym przypadku najważniejszy jest podział na migrację wewnętrzną i zewnętrzną. Ta praca będzie opierała się na migracji wewnętrznej, czyli w obrębie granic jednego państwa. Stany Zjednoczone zostały wybrane ze względu na bardzo dużą mobilność ich ludności, czyli skłonność do zmiany miejsca zamieszkania [1].

Przegląd literatury

Temat migracji pojawia się w literaturze od wielu lat i jest przedmiotem badań w wielu obszarach naukowych. Michael Greenwood w swoim artykule rozważał trendy, kierunki i powody wewnętrznej migracji w Stanach Zjednoczonych w latach 1965-1983 [2]. Bardzo często badany jest wpływ zmiany klimatu i zagrożeń z niego płynących na migrację ludności [3]. Brany jest pod uwagę wzrost poziomu wody w morzach, powodzie, a także susze [4, 5, 6]. Często migracja jest zestawiana z rozprzestrzenianiem się chorób. Były przeprowadzone badania na temat wpływu migracji na rozprzestrzenianie się np. malarii lub koronawirusa [7, 8]. Do modelowania samej migracji wykorzystywane od dawna był wykorzystywany model grawitacyjny, który przewiduje stopień migracji pomiędzy dwoma miejscami i opiera się na prawie powszechnego ciężenia. Model ten jest nadal popularny w ostatnich latach [9]. Rozszerzeniem modelu grawitacyjnego jest model radiacyjny. Także bazuje on na zjawisku fizycznym, tym razem na rozchodzeniu się fal w próżni [10]. Może on zostać zgeneralizowany także na inne cechy dwóch miejsc np. ich poziom zurbanizowania [11]. W badaniach nad migracją ludności w sieci miejsc używano już łańcuchów Markowa do symulowania przepływu ludności między miastami [12]. Do przewidywania kierunków migracji ludności były użyte także metody uczenia maszynowego. W artykule Caleb Robinson i Bistra Dilkina zostały porównane ze sobą modele grawitacyjne, radiacyjne oraz modele uczenia maszynowego XGBoost i sieć neuronowa (MLP) [13].

Dane użyte przy realizacji projektu

Cechy, które zdecydowaliśmy się użyć to:

1. Procent ludzi w danym hrabstwie z wykształceniem wyższym
2. Zmiana indeksu HPI (*House price index*), mówiący o zmianie cen mieszkań
3. Średni przychód w hrabstwie
4. Bezrobocie w hrabstwie
5. Stan zdrowia
6. Stan zdrowia psychicznego
7. Ilość przestępczości

8. Wartość dotycząca problemów ze znalezieniem zamieszkania
9. Indeks nierówności społecznych
10. Wartość opisująca dostępność miejsc do ćwiczeń

Dane pochodzą z instytucji rządowych, przez co mamy pewność że są wiarygodne. Opisują stan rzeczy na rok 2016.

Aby móc wylosować cechy konkretnej osoby, należy najpierw znaleźć odpowiednią dystrybucję do każdej z cech. Nie możemy *a-priori* założyć że wszystkie dane pochodzą na przykład z rozkładu normalnego, gdzie średnia to wartość cechy danego hrabstwa.

Na stan obecny, udało nam się odszukać odpowiednie dystrybucje, które reprezentują dane na poziomie całych stanów zjednoczonych, co zostało pokazane na grafice poniżej.



Kolejnym krokiem w projekcie, jest dokonanie parametryzacji dystrybucji dla każdego hrabstwa, tak aby wartość oczekiwana zgadzała się z wartością, która mamy w danych, a kształt krzywej był jak najbardziej podobny do tej, która została przedstawiona dla wszystkich hrabstw.

Kroki przy realizacji projektu

Wybór cech wpływających na migrację w danym hrabstwie

Mając dane dotyczące ilości osób które migrują z danego hrabstwa do innych hrabstw, jesteśmy w stanie dowiedzieć się, które cechy z jaką siłą wpływają na migrację ludności. Oczywiście jest to, że w każdym hrabstwie będą to inne cechy, dlatego prezentujemy takie podejście do rozwiązania tego problemu.

Każde hrabstwo jest przedstawione przy pomocy wektora cech o stałym rozmiarze N . Mając dane dotyczące ilości osób, które z rozważanego hrabstwa wyemigrowały, jesteśmy w stanie policzyć siłę "przyciągania" w następujący sposób:

- h_A - macierz cech hrabstwa z którego emigrację rozważamy
- h_1, h_2, \dots, h_k - macierze cech hrabstw, do których mieszkańcy hrabstwa A emigrowali

Dla każdego hrabstwa od 1 do k posiadamy dane dotyczące ile osób z hrabstwa A wyemigrowało do każdego z pozostałych. Wpiszmy te dane do wektora d . Aby obliczyć prawdopodobieństwa stosunku migracji używamy funkcji softmax, która zamieni nasz wektor w wektor prawdopodobieństw.

$$t = \text{softmax}(d) \quad (1)$$

Następnie jesteśmy w stanie policzyć “wirtualne” hrabstwo, w następujący sposób:

$$h_{vA} = \sum_{i=1}^k t_i h_i \quad (2)$$

“Wirtualne” hrabstwo jest średnim hrabstwem, do którego przeciętna osoba z danego hrabstwa będzie chciała emigrować.

Jeżeli każde hrabstwo jest opisywane przez wektor o stałej długości, jesteśmy w stanie nanieść wszystkie hrabstwa na N -wymiarową przestrzeń. Osoba zamieszkująca dane hrabstwo jest opisywana przy pomocy takiej samej liczby cech co wektor dla hrabstwa. Osoby, które są bardziej skłonne do migracji, po naniesieniu ich wektora cech na przestrzeń, będą znajdowały się dalej od rodzinnego hrabstwa, odległość ta determinuje skłonność do migracji. Podejście to nie bierze pod uwagę czynników, które mogą wpłynąć na miejsce docelowe. Przykładowo osoba znajdująca się w hrabstwie o niskim PKB, będzie bardziej skłonna do emigracji do tych hrabstw, które również mają niskie PKB, co jest raczej nieintuicyjne, ponieważ będzie znajdowała się w przestrzeni bliżej nich.

Aby zaradzić temu problemowi wykorzystamy sieć neuronową, która zostanie nauczona aby zamieniać wartości danego hrabstwa na wartości hrabstwa “wirtualnego”. W ten sposób, wiedząc z jakiego hrabstwa pochodzi, będziemy mogli użyć modelu, a następnie nanieść na przestrzeń zamienione wartości danej osoby.

Aby lepiej zachować lokalne warunki, nie możemy nauczyć jednej sieci na całe Stany Zjednoczone (która może pominąć lokalne potrzeby osób emigrujących), oraz nie możemy nauczyć jednej sieci dla danego hrabstwa (ponieważ mamy tylko jedną próbkę do nauki). Zamiast tego, pogrupujemy wektory hrabstw przy pomocy algorytmu K-means i wytrenujemy jedną sieć neuronową na każdy klastr. W ten sposób zachowamy balans pomiędzy lokalnością sieci oraz ilością danych do nauki.

Przebieg symulacji

1. Wylosowanie ustalonej liczby przedstawicieli z każdego hrabstwa
2. Każdą wylosowaną osobę przepuszczamy przez sieć neuronową, która zależy do odpowiedniego klastra
3. Dla każdego mieszkańca obliczmy jego “odległość” od hrabstwa rodzinnego i na podstawie tej odległości decydujemy czy będzie on migrował czy nie
4. Sprawdzamy ile osób wyemigrowało z danego hrabstwa i na tej podstawie losujemy bez zwracania właśnie taką liczbę emigrantów
5. Dla każdej wylosowanej osoby obliczamy odległość w przestrzeni do każdego innego hrabstwa i na podstawie tej odległości losujemy jedno hrabstwo do którego reprezentant wyemigruje

6. Ponownie losujemy wybrane zmienne cech (takie jak np. zarobki) inne pozostawimy bez zmian, a jeszcze inne ustawiamy na takie, jakie występują w obecnym miejscu zamieszkania
7. Przez ustaloną liczbę kroków symulacji wykonujemy polecenia zaczynając od numeru 2.

Ilustracja mechanizmu projektu

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}, \mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix}, \mathbf{D} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}, \mathbf{E} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}, \mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} \quad (3)$$

A, B, C, D, E oraz F są przykładowymi hrabstwami, mające odpowiadające im cechy zapisane w wektorach po prawej stronie znaku równa się.

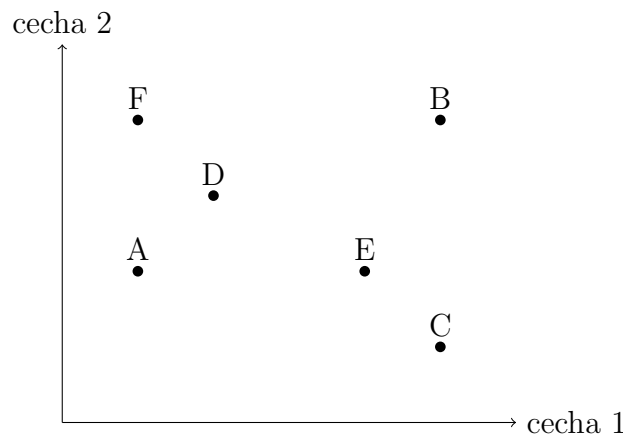
Na potrzeby prezentacji, założmy, że osoby z hrabstwa A migrowały w następujących ilościach: x_1, x_2, x_3 kolejno do hrabstw B, C oraz D.

Policzmy “wirtualne” hrabstwo V_A dla A:

$$\begin{aligned} t &= \text{softmax}([x_1, x_2, x_3]) \\ V_A &= t_1 B + t_2 C + t_3 D \\ V_A &= \begin{bmatrix} v_{a1} \\ v_{a2} \\ \vdots \\ v_{aN} \end{bmatrix} \end{aligned} \quad (4)$$

Analogiczne operacje zostaną wykonane aby policzyć “wirtualne” hrabstwo dla każdego z naszych hrabstw.

Nanieśmy teraz nasze hrabstwa na przestrzeń. Dla uproszczenia wizualizacji pokażemy tylko ją dla dwóch wymiarów.



Mając hrabstwa naniesione na płaszczyznę, możemy zająć się treningiem modeli. Zaczynamy najpierw od podziału hrabstw na klastry. W pokazanym przypadku będziemy mieli dwa klastry, gdzie do pierwszego będą należały hrabstwa A, D i F, a do drugiego

klastra C, E oraz B. Jak zostało to powiedziane wcześniej, klastrowanie odbywa się przy pomocy algorytmu K-means.

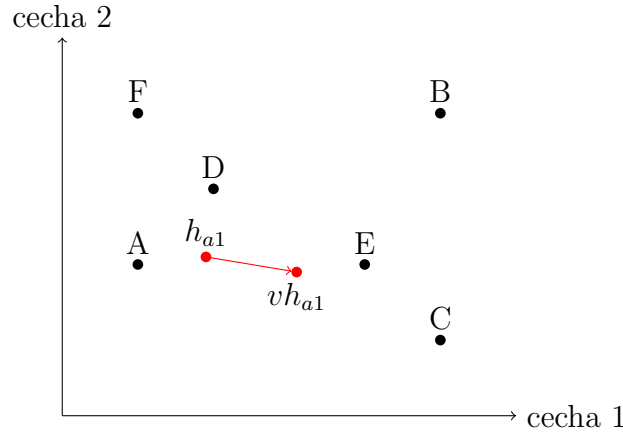
Przy pomocy danych z klastrów tworzymy dwie sieci neuronowe NN_1 i NN_2 , gdzie pierwsza uczy się na danych wejściowych A, D, F i wyjściowych V_A, V_D i V_F , a druga jako wejście dostaje wektory B, E, C a jako wyjście odpowiadające im V_B, V_E i V_C .

Po wytrenowaniu obu sieci mamy już podstawy aby zacząć losować ludzi z danych hrabstw.

Przyjrzyjmy się osobie p_{a1} , która została wylosowana z hrabstwa A. Z hrabstwem A jest skojarzony odpowiedni wielowymiarowy wspólny rozkład prawdopodobieństwa f_A , który jest w odpowiedni sposób sparametryzowany, co zostało opisane w sekcji poświęconej temu tematowi.

$$\begin{aligned} h_{a1} &\sim f_A \\ v h_{a1} &= NN_1(h_{a1}) \end{aligned} \tag{5}$$

Kolejnie osoba jest przekształcana przy pomocy sieci neuronowej oraz jest nanaszana na przestrzeń hrabstw.



Na powyższym rysunku przedstawiono wszystkie hrabstwa, wylosowaną osobę, a jej transformację przy pomocy sieci neuronowej przedstawiono przy pomocy strzałki.

Następnie znajdujemy k najbliższych sąsiadów (hrabstw) dla punktu $v h_{a1}$ i obliczamy do nich odległości. Oznaczmy je d_1, d_2, \dots, d_k . Losujemy do którego z hrabstw osoba h_{a1} wyemigruje z prawdopodobieństwami $\text{softmax}([d_1, d_2, \dots, d_k])$.

Kroki są powtarzane przez ustaloną liczbę kroków, lub do momentu, w którym ustalony próg odległości.

Bibliografia

- [1] Wikipedia contributors. *Migracja ludności — Wikipedia, wolna encyklopedia*. [Online;] 2023. URL: https://pl.wikipedia.org/wiki/Migracja_ludno%C5%9Bci.
- [2] Michael J. Greenwood. “Human migration: theory, models, and empirical studies.” In: *Journal of regional science* 25 4 (1985), pp. 521–44.
- [3] Rebecca Parrish et al. “A Critical Analysis of the Drivers of Human Migration Patterns in the Presence of Climate Change: A New Conceptual Model”. In: *International Journal of Environmental Research and Public Health* 17.17 (2020). ISSN: 1660-4601. DOI: 10.3390/ijerph17176036. URL: <https://www.mdpi.com/1660-4601/17/17/6036>.
- [4] Sabine Perch-Nielsen. “Understanding the effect of climate change on human migration”. In: (2004). URL: <https://www.research-collection.ethz.ch/>.
- [5] Pietro De Lellis, Manuel Ruiz Marín, and Maurizio Porfiri. “Modeling Human Migration Under Environmental Change: A Case Study of the Effect of Sea Level Rise in Bangladesh”. In: *Risk Analysis* (2021).
- [6] Sibren Isaacman, V. F.-M., and E. F.-M. “Modeling Human Migration Patterns During Drought Conditions in La Guajira, Colombia”. In: *Conference on Computing and Sustainable Societies (COMPASS '18) Proceedings* (2018).
- [7] N.N.A.B.M.E.K.H. Rejeki. “A Time Series Analysis: Weather Factors, Human Migration and Malaria Cases in Endemic Area of Purworejo, Indonesia, 2005-2014”. In: *Iranian Journal of Public Health* (2018).
- [8] T. C. F. Y. L. Z. Z. H. Zhan C. “Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data.” In: *PLOS ONE* (2020).
- [9] Jacques Poot Omoniyi Alimi Michael P. Cameron David C. Maré. “The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science”. In: *Investigaciones Regionales - Journal of Regional Research IZA DP No. 10329* (2016).
- [10] Maritan Amos Simini Filippo González Marta C. and Barabási Albert-László. “A Universal Model for Mobility and Migration Patterns”. In: *Nature* 484.7392 (2012), pp. 96–100. ISSN: 1476-4687. DOI: 10.1038/nature10856. URL: <https://doi.org/10.1038/nature10856>.
- [11] Legara Erika Fille Alis Christian and Monterola Christopher. “Generalized Radiation Model for Human Migration”. In: *Scientific Reports* 11.1 (2021), p. 22707. ISSN: 2045-2322. DOI: 10.1038/s41598-021-02109-1. URL: <https://doi.org/10.1038/s41598-021-02109-1>.
- [12] J. Pan and A. Nagurney. “Using Markov chains to model human migration in a network equilibrium framework”. In: *Mathematical and Computer Modelling* 19.11 (1994), pp. 31–39. ISSN: 0895-7177. DOI: [https://doi.org/10.1016/0895-7177\(94\)90014-0](https://doi.org/10.1016/0895-7177(94)90014-0). URL: <https://www.sciencedirect.com/science/article/pii/0895717794900140>.
- [13] Caleb Robinson and Bistra Dilkina. *A Machine Learning Approach to Modeling Human Migration*. 2017. arXiv: 1711.05462 [cs.SI].