

Додаток Б

Національний університет «Києво-Могилянська академія»

Факультет _____

Кафедра _____

Освітній ступінь _____

Напрямок підготовки _____
(шифр і назва)

Спеціальність _____
(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри _____

“ _____ ” _____ 20__ року

ЗАВДАННЯ

ДЛЯ КВАЛІФІКАЦІЙНОЇ / МАГІСТЕРСЬКОЇ РОБОТИ СТУДЕНТУ

_____ (прізвище, ім'я, по батькові)

1. Тема роботи _____

керівник роботи _____
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затвержені наказом вищого навчального закладу від «__» _____ 20__ року №__

2. Строк подання студентом роботи _____

3. План роботи _____

Міністерство освіти і науки України
Національний університет «Кієво-Могилянська академія»
Факультет інформатики
Кафедра математики

Магістерська робота

освітній ступінь – магістр

на тему: **«МОДЕЛЮВАННЯ СОЦІАЛЬНИХ ПРОЦЕСІВ ЯК
МАРКОВСЬКИХ ПРОЦЕСІВ ПРИЙНЯТТЯ РІШЕНЬ»**

Виконала: студентка 1-го року
навчання,
освітньо-наукової програми
«Прикладна математика», 113

Любима Михайліна Вікторівна

Керівник Чорней Р.К.
кандидат фіз.-мат. наук, доцент

Рецензент _____
(прізвище та ініціали)

Магістерська робота захищена
з оцінкою

Секретар ЕК

«____» _____
20____ р.

Київ – 2023

ВСТУП

Соціологія та математика – не найбільш поширені галузі для співпраці. На відміну від інших дисциплін, що активно впроваджують математичні моделі, а подекуди виступають рушієм у математичному знанні, моделювання соціальних процесів ще донедавна представляло труднощі для дослідників. Усі побудовані моделі були або недостатньо складними, щоб відобразити реальність, або занадто складними, і перевірка за допомогою опитувань або спостережень не могла повністю описати формальну модель.

Одним із містків, що може поєднати і успішно поєднує математику та соціологію, є марковський процес прийняття рішень. Фактично, усі задачі, де органи влади або управління мають розробити оптимальну політику, одразу наштовхують на використання саме цього підходу.

Мета даної роботи – узагальнення існуючого досвіду моделювання соціальних процесів загалом та реалізація моделі поведінки підписників на інтернет-платформах.

Відповідно до цього можна виділити таке наукове завдання:

1. З'ясувати можливі підходи до формалізації соціальних процесів та їх особливостей.
2. Установити особливості стохастичних моделей
3. Виявити, наскільки доцільним є застосування марковських моделей в окремих задачах. Розробити власну інтерпретацію соціального процесу. Оскільки тема для опрацювання надзвичайно актуальна та цікава, результати цієї роботи можна використати для подальших напрацювань та систематизацій «сирих» даних.

Перші два розділи присвячені аналізу теоретичних підходів 20-21 ст. Наведено приклади застосувань моделей з застосуванням теорії динамічних систем, теоретичні засади для перевірки, чи емпіричні дані доцільно представити у вигляді марковського процесу.

Третій розділ – це формулювання марковського процесу прийняття рішень, огляд застосувань та відомих практик, а також прикладу, на мою думку, досить вдалого моделювання процесу.

Четвертий і останній розділ – короткий огляд результатів розробки, узагальнень, та вивчення обраних даних, демонстрація категорій елементів побудованого марковського процесу прийняття рішень.

Окремим завданням було визначити, наскільки можливо реалізувати таку модель за допомогою рішень SAS Base. Код написаний цією мовою, і перевірено використовується для приведення даних до прийнятної моделлю вигляду, проте подальша реалізація та дослідження рекомендуються за допомогою інтерпретованих мов або SAS Viya.

ЗМІСТ

ВСТУП
РОЗДІЛ 1
РОЗДІЛ 2
 2.1
 2.2
РОЗДІЛ 3
РОЗДІЛ 4
ВИСНОВКИ

РОЗДІЛ 1: соціальні процеси як математична структура

Для початку визначимо, що мається на увазі під соціальним процесом. Відповідно до визначення [1] «Соціальний процес – це шаблон спільної діяльності, що виникає регулярно, і завдання соціологів – визначити і описати такі процеси». Він характеризується взаємодією об'єктів, їх станом, або статусом, стадією. Це дуже широке поняття, так, до соціальних процесів відносять міграцію населення, зміну місця роботи, розвиток освіти, зміну ставлення мас до певних явищ, соціальні конфлікти (війни, тероризм, сутички, конфлікт особистих інтересів).

Відповідно, моделювання соціальних процесів можна здійснити за допомогою різних підходів. Водночас, для дослідників додатковою перешкодою є особливості організації та класифікації таких. Часто дані та ефективні методи та поняття важко «перекласти» математичною мовою та надати формальну модель. Тому, не дивлячись на те, що перші спроби моделювання були зроблені ще у 1950-х роках, упродовж другої половини 20 ст. зацікавленість у темі була радше на нейтральному рівні. Окрім того, як часто буває у задачах, що виникають з емпіричних завдань, побудовані моделі фокусувались на конкретних ситуаціях. У роботі [1] указано, що формальне моделювання соціальних процесів не дало вагомого внеску у соціальну науку. На відміну від інших галузей, що почали активно використовувати математичні моделі (такі як біологія, хімія, менеджмент), соціальні процеси затребують більш чіткого планування.

У [2] сформульовані дві причини, чому у роботах 20 ст. та, на думку автора, немає достатньої представленості задачі моделювання соціальних процесів:

1. Розглянута модель достатньо просто пов'язується з наявними емпіричними даними, проте вона недостатньо складна для опису реальності, і тому незастосовна для інших подібних випадків. Зокрема, таке може бути у моделі часових рядів за допомогою лінійної або поліноміальної регресії або однофакторні моделі ARIMA.
2. Розглянута модель настільки складна, що оцінка її параметрів наявними емпіричними можливостями досить складна. Наприклад, для нелінійних математичних моделей, що не мають аналітичного розв'язку. У такому випадку треба додатково користатися чисельним моделюванням.

Соціальні структури відрізняються своєю складністю та багат шаровістю (системи, підсистеми), що, у свою чергу, не знаходило необхідних процедур серед математичного апарату.

Однак, існуючі підходи класифіковані у такі групи:

- Динамічні системи
- Мікроаналітичні симуляційні моделі
- Багаторівневі підходи (мережеві моделі, динамічна теорія ігор, клітинні автомати)
- Методи штучного інтелекту

Описані вище підходи показують гарні результати та достатньо повні, що описати, наприклад, такі задачі:

1. Процес формування (соціальної) думки.

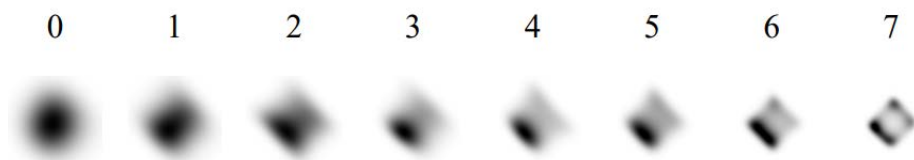
Ця модель розглядає популяцію людей, що можуть мати думку щодо певної проблеми: так чи ні. Залежно від того, наскільки індивідуальна думка залежить від загальної, після певного часу частка «так» бімодально розподілена та врешті досягає 10 або 90 відсотків.

Ця модель ґрунтується на ймовірності переходу окремо взятої людини від «так» до «ні» і навпаки. Після цього задача зводиться до вирішення системи лінійних диференціальних рівнянь.

Цей приклад можна розширити до класу задач, де взаємодіють декілька груп або можна обирати більше ніж один варіант.

2. Зміна ставлення.

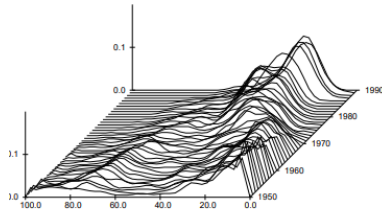
Розглянемо групу людей, де можна кількісно оцінити ставлення до певної теми. Це ставлення розподілено за певним законом. Автори ДЖЕРЕЛО вказують, що тоді нормально розподілені групи розділяються на декілька кластерів. Наприклад, як показано на рисунку, початково некорельовані нормально розподілені ставлення у групі за сім послідовних кроків «збираються» у чотири кластери.



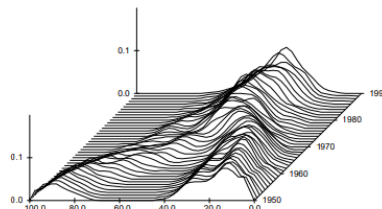
Рух у просторі суджень цієї моделі описується диференціальним рівнянням Лангевіна. Помітимо, що на відміну від попереднього прикладу, де можна було вирахувати загальний розподіл груп, тут можна вирахувати лише розподіл індивідів в межах однієї групи. Тому, на думку авторів вихідного дослідження, неможливо вирахувати ймовірність розпаду на два, три або чотири кластери.

3. Гендерна десегрегація.

Цей приклад описує процес десегрегації у німецьких гімназіях після Другої Світової війни. На рисунку показано розподіл відсотку вчительок у ста п'ятдесяти школах у землі Рейнланд-Пфальц для емпіричних даних та для моделі. Ця модель багаторівнева та включає рівні вчителі/ки – школи – землі. З урахуванням таких припущень: вчителі, що лишають свої посади, заміщуються чоловіками та жінками з однаковою ймовірністю, чоловіки лишаються на посаді вдвічі довше за жінки, і що жінки починають працювати у певній школі з відомим (модельованим) розподілом.



empirical data



simulation results

Як бачимо з графіків, дана симуляція досить добре описує справжні процеси, а можливо, з підбором параметрів, може бути покращена. Оскільки дана модель є певним чином ретроспективною, автори ДЖЕРЕЛЮ не надають роз'яснень, наскільки добре вона працюватиме у майбутньому. Також варто зауважити, що припущення моделі має бути підтверджено політикою шкіл та дотриманням регуляторних заходів, що неможливо перевірити.

4. Самоорганізація.

Розглянуті вище приклади працюють у досить простій конфігурації, адже у відповідних моделях є виключно залежність індивідуальної думки від загальної однорідної популяції незмінного розміру. Завдання можна ускладнити, якщо розглядати неоднорідні моделі або локальні залежності. У прикладах 1 та 2 нова структура (перевага «так» або «ні», кластеризація) виникає як наслідок нелінійних стохастичних процесів, і пояснити їх на макрорівні неможливо.

На момент написання роботи [2] автор вказує, що «For mathematical reasons, there is no possibility of finding simple estimation procedures for most nonlinear stochastic multilevel models analogous to multiple regression or linear structural relationship models.» (англ.) – «З математичних причин немає можливості знайти прості процедури оцінки для більшості нелінійних стохастичних багаторівневих моделей, аналогічних до моделей множинної регресії або лінійних структурних зв'язків»

Висновки до розділу: таким чином, не дивлячись на те, що проблематика соціальних процесів досить актуальна та потребує уваги від дослідників, на кінець 20 ст. найуспішнішими моделями були багаторівневі моделі, що використовують як імовірнісні підходи, так і здобутки теорії динамічних систем. Як і в будь-якій проблематиці моделювання, важливо максимально точно описати систему, що моделюється. І, у силу особливостей соціальних процесів, ця задача є нетривіальною для формальної математичної моделі. Саме тому, хоч перші успішні роботи з теми були написані ще у середині 20 ст., станом на кінець 20 ст. ми маємо найуспішніші приклади для обмеженої кількості задач та для відносно простих конфігурацій. Це підкреслює актуальність та виклик поставленої задачі.

РОЗДІЛ 2: стохастичне моделювання соціальних процесів

2.1 У попередньому розділі були розглянуті та підсумовані здобутки моделінгу загалом для 20 ст. Однак, згадки про випадкові процеси як інструмент є досить рідким явищем. За класифікацією [2] саме марковські процеси та марковські ланцюги зокрема слугували для прогнозування ефектів та аналізу соціальної та фінансової політики.

Виділяється дві основні проблеми, що виникають у спробах моделювання з використанням стохастичних моделей, а отже, і причини їхньої непопулярності (до певного часу):

1. Зазвичай використовуються детальні емпіричні дані щодо початково стану мікроєдиниць (наприклад, індивіди). Вони зазвичай є якісними. Витрати на утримку та збір таких даних надзвичайно високі, тому не дивно, що донедавна обчислювальна здатність та сервери не дозволяли змоделювати притомні моделі.

2. Відсутність універсального ПЗ, яке би могло просто змоделювати процес. Усі програми, що базуються на марковських процесах, були зроблені під певну задачу та писалися з нуля.

Окрім того, існувала і задача представлення соціальних процесів як марковських структур з неперервним часом, адже з емпіричних міркувань дослідник не може спостерігати ймовірнісну матрицю переходу у кожен момент часу. З цього природньо випливають дві задачі ДЖЕРЕЛО

а) перевірити, чи могли емпіричні спостереження виникнути як частина марківської структури

б) ідентифікувати, чи спостереження узгоджуються з більше ніж однією марківською структурою.

2.2.

Розглянемо випадковий процес, що задається матрицею переходу P , має скінченну множину станів S , $|S| < \infty$. Також що виконується:

$$\frac{dP(t)}{dt} = QP(t), \quad P(0) = I \quad (1)$$

$Q, P(t)$ – матриці розмірності $r \times r$.

Твердження 1. Якщо Q має вигляд:

$$q_{ij} \geq 0, i \neq j, q_{ii} \leq 0, \quad \sum_{j=1}^r q_{ij} = 0, 1 \leq i, j \leq r$$

Наприклад, якщо

$$Q = \begin{pmatrix} -1,805 & 1,718 & 0,087 \\ 0,044 & -1,784 & 1,740 \\ 2,262 & 0,017 & -2,279 \end{pmatrix}$$

То матриця $P(t)$, $t > 0$, яка є розв'язком РІВНЯННЯ, є матрицею переходу ланцюга Маркова з неперервним часом. Тобто, елемент $p_{ij}(t)$ матриці P – це ймовірність, що перебуваючи у стані s_i у час 0 опиниться у стані s_j у час t . Матрицю Q називають матрицею інтенсивності. Позначимо клас таких матриць \underline{Q} .

Матриця Q надає інформацію щодо структури групи. Елементи цієї матриці можна тлумачити так:

1. $q_{ij}/-q_{ii}$ - ймовірність, що індивід у стані s_i перейде у стан s_j , за умови, що перехід відбудеться.

2. $1/-q_{ii}$ – очікувана тривалість перебування у стані s_i .

Розв'язком (1) є матриця наступного вигляду:

$$P(t) = e^{Qt}, \quad t > 0$$

Де експонента певної квадратної матриці A визначається таким чином:

$$e^A = \sum_{k=0}^{\infty} A^k/k!$$

При цьому очевидно, що для практичних задач досить неефективна така аналітична презентація. Зокрема, ми не можемо точно визначити обернену функцію – логарифм матриці, а отже, ряд не збігається для матриць переходу, для яких виконується марківська умова.

Більше того, рівняння $\hat{P} = e^Q$ може мати декілька розв'язків $Q \in \underline{Q}$.

Це дуже важливо в соціологічних дослідженнях, оскільки ДЖЕРЕЛЮ «the conventional strategy in using Markov models for theory construction emphasizes decomposing the q_{ij} elements of Q among theoretically postulated effect parameters» (англ.) – «традиційна стратегія використання марківських моделей для побудови теорії наголошує на розкладанні елементів q_{ij} матриці Q на теоретично визначені параметри ефекту».

Отже, у першу чергу треба забезпечити відновлення матриці Q .

2.3. Проблема вбудованості

Нас цікавить, які матриці переходу сумісні з марковським процесом.

Коротко сформулюємо основні твердження [1]

Твердження 2. Матрицю

$$\hat{P} = \begin{pmatrix} \hat{p}_{11} & \hat{p}_{12} \\ \hat{p}_{21} & \hat{p}_{22} \end{pmatrix}, \hat{p}_{ij} \geq 0, 1 \leq i, j \leq 2, \sum_j \hat{p}_{ij} = 1$$

Можна подати у вигляді $e^Q, Q \in \underline{Q}$ тоді і лише тоді, коли $\hat{p}_{11} + \hat{p}_{22} > 1$.

Це дуже потужне твердження, але, на жаль, воно оперує лише матрицями розмірності 2.

Твердження. Необхідні умови вбудованості:

1. Якщо $\hat{p}_{ij}(t_1) = 0$, то $\hat{p}_{ij}^{(n)}(t_1) = 0 \forall n \in \mathbb{N}$. І навпаки, якщо $\hat{p}_{ij}(t_1) \neq 0$, то

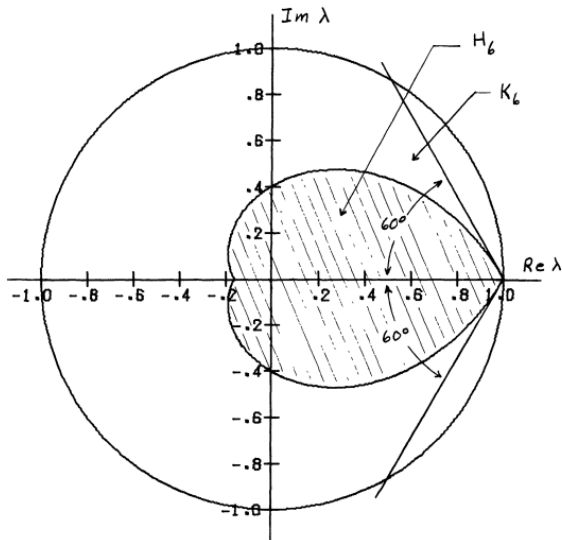
$\hat{p}_{ij}^{(n)}(t_1) \neq 0 \forall n \in \mathbb{N}$.

2. $\det \hat{P} > 0$

3. $\forall \lambda_i$ - власні числа матриці $\hat{P} |\lambda_i| \neq 1$, окрім, можливо, $\lambda_i = 1$. Будь-яке від'ємне власне число матриці \hat{P} має парну кратність.

4. Власні числа матриці \hat{P} вміщуються в область H_r , що має кордон у вигляді кривої $x(v) + iy(v)$:

$$\begin{aligned} x(v) &= \exp\left(-v + v \cos \frac{2\pi}{r}\right) \cos\left(v \sin \frac{2\pi}{r}\right) \\ y(v) &= \exp\left(-v + v \cos \frac{2\pi}{r}\right) \sin\left(v \sin \frac{2\pi}{r}\right) \end{aligned}$$



Приклад: регіон для

розмірності 6×6

2.4. Процес перевірки, чи емпірично визначена матриця \hat{P} відображає матрицю переходу марківської моделі з неперервним часом, тобто чи $\log \hat{P} \in \underline{Q}$

2.4.1. Для різних власних чисел матриці \hat{P} :

1. Перевірити, чи виконуються необхідні умови 2 і 3.
2. Перевірити умову 4.
3. Якщо власні числа матриці дійсні та додатні, то обчислити $\log \hat{P}$ за однією з формул:

А) За допомогою ряду (якщо він збігається):

$$\frac{1}{t} \log \hat{P}(t) = \frac{1}{t} \sum_{k=0}^{\infty} \frac{(-1)^{k-1} (\hat{P}(t) - I)^k}{k}$$

Б) формула Сильвестра для різних власних чисел:

Якщо $A - r \times r$ матриця з різними власними числами $\lambda_1, \dots, \lambda_r$. Функція f визначена на підмножині комплексних чисел \tilde{C} , так, що $\forall i = 1, \dots, r \lambda_i \in \tilde{C}$ та функція неперервна у цих точках. Тоді:

$$f(A) = \sum_{i=1}^r f(\lambda_i) \prod_{j \neq i} \frac{(A - \lambda_j I)}{(\lambda_i - \lambda_j)}$$

Тоді, поклавши $f(A) = \log A$ та $A = \hat{P}$,

$$\log(\hat{P}) = \sum_{i=1}^r \log \lambda_i \prod_{j \neq i} \frac{(\hat{P} - \lambda_j I)}{(\lambda_i - \lambda_j)}$$

В) трансформація діагоналізації:

Якщо представити квадратну матрицю $A r \times r$ у вигляді

$$A = HDH^{-1}$$

Де H – необоротна матриця, що складається зі стовпців - власних векторів матриці A та що відповідають власним числам у матриці D , а

$$D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r \end{pmatrix}$$

Тоді якщо функція f однозначна:

$$f(A) = Hf(D)H^{-1}$$

Де:

$$f(D) = \begin{pmatrix} f(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f(\lambda_r) \end{pmatrix}$$

Якщо f – багатозначна, то розбиваємо результат відповідно до однозначної гілки значення функції та отримуємо:

$$f_\alpha(A) = Hf_\alpha(D)H^{-1}$$

Де:

$$f_\alpha(D) = \begin{pmatrix} f_{\alpha_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{\alpha_r}(\lambda_r) \end{pmatrix}$$

Де f_{α_i} - будь-яка з однозначних гілок функції f . Отримаємо різні значення функції $f(A)$ у залежності від набору $(f_{\alpha_1}, f_{\alpha_2}, \dots, f_{\alpha_r})$.

Оскільки логарифм – багатозначна функція:

$$\log_n z = \log|z| + i(\theta + 2\pi n), n \in \mathbb{Z}, z = a + bi \in \mathbb{C}, \theta = \arg(z)$$

Кожне значення n породжує значення функції логарифма (гілка логарифма). Тоді, застосовуючи формулу діагоналізації:

$$\log_n \hat{P} = H \log_n D H^{-1}$$

Де:

$$\log_n D = \begin{pmatrix} \log_{n_1} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \log_{n_r} \lambda_r \end{pmatrix}$$

Для того, щоб перевірити, чи $\log \hat{P} \in \underline{Q}$ треба визначити, чи принаймні одна гілка логарифму задовольняє цю умову.

4. Якщо \hat{P} має комплексні власні числа, вони мають бути спряженими парами. Для кожної такої спряженої пари $(\lambda, \bar{\lambda})$, в залежності від гілки логарифму n , має виконуватись умова Рунберга:

$$\pi \left(\frac{1}{2} + \frac{1}{r} \right) \leq \arg(\log_n \lambda) \leq \pi \left(\frac{3}{2} - \frac{1}{r} \right)$$

Якщо вона виконується, то обчислити $\log \hat{P}$ за допомогою розкладу в ряд або за формулою Сильвестра.

Перевірити, чи $\log \hat{P} \in \underline{Q}$. Повторити обчислення для усіх гілок логарифма, що задовольняють умову Рунберга.

2.4.2. Складніші конфігурації.

Описаний у попередньому пункті алгоритм хоч і добре описаний, та має ряд недоліків:

1. Потребує обчислювальних здатностей та ефективного програмування.

2. На практиці матриця \hat{P} буде лише оцінкою реальних значень, адже матиме місце, по-перше, похибка вибірки, по-друге, похибка вимірювань. Тоді, якщо для матриці \hat{P} умови не виконані, модифікувати матрицю \hat{P} так, щоб для модифікованої матриці $\log \tilde{P} \in \underline{Q}$. Існують класичні методи вибору матриці \tilde{P} ДЖЕРЕЛО.

3. Навіть якщо власні числа \hat{P} різні, але близькі до рівних (із заданою точністю), ми з обчислювальної точки зору буде правильніше виправити її на матрицю з повторюваними власним числами. Тоді дещо змінюється алгоритм (зокрема, формула Сильвестра та обчислення функції від матриці J , що фігурує у діагональному розкладі замість D).

4. Дані можуть бути неточні, цензуровані, приховані. Це породжує велику варіацію та не дає однозначного алгоритму для виправлення матриці \hat{P} .

Отже, запропонований метод потребує кропіткої реалізації та передбачає лише часткову автоматизацію.

Інший підхід, що запропонований, зокрема у ДЖЕРЕЛО – використати теорію інформації. Метод полягає в оцінці параметрів та порівнянні з отриманою матрицею.

Так, автори пропонують змоделювати рух клієнтів від компанії до компанії як марківський ланцюг. Враховуючи інформацію про частку ринку протягом певного часу, можна оцінити \hat{P} за допомогою описаної вище процедури. Для цього або потрібно принаймні r спостережень та багато припущень. Якщо ж маємо менше r переходів, для оцінки ймовірностей можна використати ентропію. Шляхом максимізації функції ентропії як функції від ймовірностей (формулу можна коригувати в залежності від задачі, але в класичному вигляді – ентропія за Шенноном):

$$\begin{cases} H_i = - \sum_{j=1}^r p_{ij} \log p_{ij} \rightarrow \max \\ \sum_j p_{ij} = 1 \end{cases}$$

А також враховуючи додаткові обмеження, що передбачені марківською моделлю.

РОЗДІЛ 3: марковські процеси прийняття рішень входять у гру

Не дивлячись на відносну непопулярність марковських процесів як моделей соціальних процесів у попередньому столітті, знаходимо цікаві та амбіційні застосування у нещодавніх роботах ДЖЕРЕЛА.

Така зміна, на мою думку, пояснюється, по-перше, розвитком технічного устаткування, по-друге, широким застосуванням марковських ланцюгів в економічній теорії, по-третє, успіхами алгоритмів навчання з підкріпленням, що породжує природню ідею застосувати подібний підхід до інших задач.

Зокрема, маємо ряд застосувань марковських процесів прийняття рішень ДЖЕРЕЛО

- збирання врожаю, сільське господарство, фінанси та інвестиції, закупівлі, інвентаризація та виробництво, стимулювання збуту та спорт
- внутрішня безпека:
 - o перевірка контрольно-пропускних пунктів безпеки в аеропорту, метою якої є максимально збільшити загальну очікувану безпеку при визначенні оптимальних завдань перевірки пасажирів;
 - o вибір маршруту – мінімізація ефекту визначених параметрів, наприклад, вартості перевезень;
 - o задача оптимального захисту від атак з перешкодами в когнітивних радіомережах, у яких зловмисники прагнуть перешкодити каналу зв'язку користувача шляхом створення інтерференції.
 - o розподіл ресурсів правоохоронних органів у ланцюгах незаконного постачання наркотиків – максимізувати очікувану винагороду, отриману від арешту злочинців.

З такого погляду, підхід марковських процесів прийняття рішень досить просто уявити як засіб менеджменту або розробки політики компанії, системи рекомендацій тощо.

Як відомо, марковський процес прийняття рішень формулюється в термінах станів, дій, переходів станів, функцій винагород та політики дій. Припускається, що існує певний усталений зв'язок між поточним станом, дією, що виконується, та потенційними станами.

Формалізуючи, марковський процес прийняття рішень – це набір $(S, A, P_a(\cdot, \cdot), R(\cdot, \cdot), \gamma)$

- S – множина станів (скінченна)
- A – множина дій (скінченна)
- $P_a(s, s') = P(s_{t+1} = s' | s_t = s, a_t = a)$
- $R(s, s')$ нагорода від переходу $s \rightarrow s'$
- γ – коефіцієнт знецінювання.

Як приклад формалізації соціального процесу можна навести аналіз поведінки користувачів онлайн-навчальної системи для програмування ДЖЕРЕЛО.

Дії користувачів у системі розбиті на 2 великі категорії: створення контенту (коментування коду, поширення, оцінка, тощо – те, що видно у спільноті) та споживання контенту (перегляд коду, запуск коду, тощо – те, чого спільнота не бачить). Кожна з них поділена на 5 категорій. Активність створення контенту покладається як множина A – дій, а споживання – S , тобто станів. Елементи цих

множин – п’ятимірні вектори. Кожен елемент складається з інтенсивності відповідної складової. Основна задача – визначити, яку найкраще дію обрати, коли спільнота у заданому стані споживання контенту з урахуванням витрат.

Приклад у ТАБЛИЦІ

comment	reply	share	grade-comm	grade code	Observations
0	0	0	0	0	1621
0	0	1	0	0	109
0	0	2	0	0	54
1	0	0	0	0	23
0	0	0	0	1	12

action

Таблиця 4: Найпоширеніші дії; 0 означає низький рівень активності, 1 середній рівень і 2 високий рівень. Дія – це сума всіх дій усіх користувачів.

Спостереження представляють загальну кількість спостережень цієї дії в наборі даних.

Після цього функція переходу ймовірностей визначається емпірично як частота виникнення наборів $s \times a \rightarrow s'$, $s, s' \in S, a \in A$. Якщо у даних такого переходу нема, встановлюється ймовірність наступних станів за допомогою пріоритету Діріхле, тобто ймовірність даного стану дорівнює частці всіх спостережень стану, що відповідають цьому стану. Він дорівнюватиме нулю, якщо цей стан не спостерігався.

РОЗДІЛ 1: результати дослідження

4.1. Формулювання задачі

Частково надихнувшись попереднім прикладом, я досліджую, чи придатний до моделювання процес набуття підписників на платформі YouTube. По-перше, його можна оцінити за допомогою кількісних показників у вільному доступі.

По-друге, дизайн моделі подібний до прикладу. По-третє, «класичним рецептом» привертання уваги до відео є активність аудиторії – реакції (вподобайки) та коментарі. Їх безпосередній вплив і досліджується.

Розглядаємо показники найпопулярнішої тисячі каналів на платформі станом на червень, вересень та грудень 2022 року.

Множина станів – динаміка кількості підписників на каналі. За кількістю канали були розбиті на категорій – від найменшої до найбільшої. Таким чином, стан – одновимірний вектор.

Множина дій – активність глядачів, категоризовано перегляди, коментарі та вподобайки, отже, дії – двовимірні вектори.

	name	cat	s2	a21	a22	a23	s3
1	#RefugioMental	Movies	1	3	2	2	1
2	0214mex	Music & Dance	1	4	4	3	1
3	1032010814	DIY & Life Hacks	1	1	2	1	1
4	123GO!Hindi		2	3	1	2	1
5	1MILLIONDanceStudio		1	1	1	2	1
6	5-MinuteCrafts		2	2	2	1	3
7	5-MinuteCraftsDIY	Music & Dance	1	0	1	0	1
8	5-MinuteCraftsFAMILY		1	0	1	0	1
9	5-MinuteCraftsLIKE		1	0	0	0	2
10	5-MinuteCraftsPLAY		1	1	1	0	1
11	5-MinuteCraftsRecycle		2	1	1	0	2
12	5-MinuteRecipes		1	1	1	0	1
13	A4	Animation	3	4	4	4	3
14	ABCNews	News & Politics	2	0	2	0	1
15	ABSCBNOnline		3	0	0	0	2
16	AM3NIC	Food & Drinks	1	2	2	3	1

Функція P обрахована як частота унікальних наборів $(s_2, a_{21}, a_{22}, a_{23}, s_3)$:

	tuple	COUNT	PERCENT
1	10001	13	2.5590551181
2	10002	1	0.1968503937
3	10003	1	0.1968503937
4	10101	16	3.1496062992
5	10102	5	0.9842519685
6	10103	1	0.1968503937
7	10111	8	1.5748031496
8	10112	1	0.1968503937
9	10211	3	0.5905511811
10	11011	1	0.1968503937
11	11101	4	0.7874015748
12	11111	9	1.7716535433
13	11112	5	0.9842519685
14	11121	1	0.1968503937
15	11123	1	0.1968503937
16	11211	19	3.7401574803
17	11212	4	0.7874015748

Продовження таблиці

147	34403	1	0.1968503937
148	34441	3	0.5905511811
149	34442	5	0.9842519685
150	34443	18	3.5433070866

Функцію винагород визначимо за допомогою значення рівній $R(s, s') = \sum_{s' \text{ - досяжна з } s} p_{s \rightarrow s'} \cdot s'$, тобто очікуваній динаміці динаміці набутих підписників. Дана модель знаходиться на початковій стадії, і наразі було побудовано процес стандартизації даних та представлення у вигляді, що сходиться з марковським процесом прийняття рішень.

Очікувані покращення:

- Більше даних протягом часу
- Врахування інших дій(перегляди сторінки, взаємодія, час перегляду)
- Більше каналів (оскільки дані були сильно неповні, їх довелося урізати)
- Обрахування оптимальних наборів дій

Також дана модель може дати поштовх для симуляції дій.

Для перевірки пропонується:

- Встановити доцільність функції переходу, зокрема, методами, описаними у 2 розділі
- Змоделювати процес та порівняти з результатами
- Використати вдубовані можливості SAS (експериментальну процедуру ENTROPY) та порівняти з отриманими вручну.

ВИСНОВКИ

Список використаних джерел

1. Singer, Burton & Spilerman, Seymour. (1976). The Representation of Social Processes by Markov Models. *American Journal of Sociology - AMER J SOCIOL.* 82. 10.1086/226269.
2. Troitzsch, Klaus. (1995). *Multilevel Process Modeling in the Social Sciences: Mathematical Analysis and Computer Simulation.*
3. Gintis, H. 2013. Markov models of social dynamics: Theory and applications. *ACM Trans. Intell. Syst. Technol.* 4, 3, Article 53 (June 2013), 19 pages. DOI: <http://dx.doi.org/10.1145/2483669.2483686>
4. Yee-King, Matthew & d'Inverno, Mark. (2016). Stimulating collaborative activity in online social learning environments with Markov decision processes. 10.13140/RG.2.1.3592.0242.
5. Lim GJ, Desai SS. Markov decision process approach for multiple objective hazardous material transportation route selection problem. *Int J Oper Res* 2010;7(4):506–29.
6. N. Orkun Baycik, Thomas C. Sharkey, Chase E. Rainwater, A Markov Decision Process approach for balancing intelligence and interdiction operations in city-level drug trafficking enforcement, *Socio-Economic Planning Sciences*, Volume 69, 2020, <https://doi.org/10.1016/j.seps.2019.03.006>.
7. SAS Institute Inc. 2014. *SAS/ETS® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.