УДК 519.6

О.А. Яворський, Н.М. Куссуль

ДОСЛІДЖЕННЯ ПРЕДСТАВЛЕННЯ БАГАТОЧАСТКОВИХ ГРАФІВ ЗА ДОПОМОГОЮ ТОПОЛОГІЧНОГО АНАЛІЗУ ДАНИХ

Яворський Олександр Андрійович

Навчально-науковий Фізико-технічний інститут Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»,

orcid: 0009-0001-5175-3825

yaotianjiu@gmail.com

Куссуль Наталія Миколаївна

Навчально-науковий Фізико-технічний інститут Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»,

orcid: 0000-0002-9704-9702

nataliia.kussul@gmail.com

Розглянуто проблему представлення багаточасткових графів для задач машинного навчання (МН) на графах за допомогою методів топологічного МН, зокрема шляхом обчислення персистентних гомологій (ПГ) хмар точок. Розглянуто також векторні представлення графів, отриманих за допомогою білінійних моделей та моделей трансляції, серед яких є модель тензорної декомпозиції TuckeR і моделі зсуву MurE та PairRE. Взято до уваги як повністю експресивні моделі, так і моделі з недоведеним рівнем експресивності. Як приклад багаточасткового графу обрано граф, що має 271 тип вершин та два типи ребер. Обчислення ПГ проведено для кожної моделі. Отримані представлення розбито на два окремих класи. Перший складається лише з векторних представлень вершин, а другий має представлення як вершин, так і одного з типів ребер. Для обох класів обраховано ПГ з максимальним виміром 2, що покриває 1-, 2- та 3-вимірні дірки. Для представлення ПГ обрано персистентні діаграми. Після цього точки отриманих діаграм використано для статистичного аналізу за допомогою обчислення значень коефіцієнтів ексцесу, асиметрії, відхилення та середнього. Дані статистичні характеристики обраховано як для самих моделей, так і для модулів їхніх різниць. Основна мета роботи полягає в тому, аби показати, що різні моделі представлень мають різні характеристики з точки зору ПГ, що вказує на те, що самі моделі не є топологічно еквівалентними, а тому їх вибір принципово впливає на якість та точність вивчення представлень багаточасткових графів. Даний результат досягається шляхом порівняння вищезазначених статистичних параметрів, а також гістограм середніх значень отриманих векторів.

Ключові слова: граф, багаточастковий граф, машинне навчання, топологічний аналіз даних, персистентні гомології.

Вступ

Машинне навчання (МН) як галузь штучного інтелекту в останні роки набуває все більшої популярності. Це обумовлено як зростаючими обчислювальними можливостями та їх доступністю широкому загалу, так і ефективністю даних методів у вирішенні багатьох практичних задач, серед яких — автоматизація, створення рекомендаційних систем, розпізнавання зображень тощо [1].

Машинне навчання можна умовно поділити на дві категорії: статистичне та нейронне, де останнє означає використання штучних нейронних мереж для апроксимацій певної функції, яка має дозволити дати відповідь на задане питання (наприклад, чи належить певний об'єкт до заданої категорії); у такому випадку шукана функція приймає n дискретних значень, що відповідають кількості заданих категорій або міток (labels). Складність такої функції в цілому може бути довільною [2].

Як вхідні значення штучні нейронні мережі зазвичай приймають певні векторні, матричні або тензорні значення, тоді як вихідним значенням часто є оцінка ймовірності або аналогічний вектор чи матриця. Очевидно, що більшість об'єктів реального світу (зображення, текст тощо) не мають векторної чи матричної природи, що створює проблему їхнього представлення. Так, кольорові зображення представляються як триплети матриць RGB, а текст перетворюється на множину векторів, де кожному вектору відповідає певне слово чи словосполучення. Зрозуміло, що такі представлення суттєво впливають на подальшу роботу алгоритмів МН, що робить пошук більш ефективних методів представлень важливою задачею.

Останнім часом, крім питань роботи з чисельними, текстовими та графічними даними, постає питання щодо використання графів для МН. Зокрема, в останні роки набувають популярності так звані графові нейронні мережі [3], які широко використовуються в різних предметних областях. Відомо, що у найпростішому випадку графи можуть бути представлені за допомогою матриці суміжності та/або інцидентності. Втім, такий підхід має суттєві обмеження у випадку роботи з багаточастковими (multipartite) графами [4]. Оскільки такі графи мають особливу практичну цінність, виступаючи моделями мереж взаємодії між різними сутностями, актуальність проблем представлення таких графів є дуже високою.

У даній роботі розглянуто деякі методи представлень багаточасткових графів та проаналізовано їхні результати за допомогою підходів топологічного аналізу даних, зокрема шляхом обчислення ПГ та їхньої оцінки методами статистичного аналізу.

Графи та їх представлення

У загальному випадку можна говорити, що граф G складається з множини вершин $V(G)$ та ребер $E(G)$ [5]. Тоді k-частковий граф G_k визначається як граф, де $c(v)_i \neq c(v)_j$ з $C_V(G)$ та/або $c(e)_i \neq c(e)_j$ з $C_E(G)$, де c та C позначають клас або тип окремої вершини (ребра) та множину таких класів для всіх вершин (ребер) відповідно. Інакше кажучи, вершини (ребра) такого графу мають внутрішні нееквівалентні розбиття, тобто елементи цих множин належать до різних класів. Наприклад, 2-частковий граф університетської авдиторії має два типи вершин: хлопці та дівчата.

У роботі розглянуто декілька методів представлення таких графів за допомогою методів тензорної декомпозиції (білінійні моделі) та переносу (трансляції). Дані методи зазвичай використовуються для роботи з так званими графами знань (knowledge graphs), у яких кожен елемент може бути представлений у вигляді трійки, де h, t позначають елементи (або сутності, об'єкти), а r_i — i-й зв'язок, тобто (h, r_i, t), ребро між ними. Для білінійних моделей характерне представ-

лення відношень як лінійних трансформацій, що діють на вектори сутностей. У трансляційних же методах, за певною аналогією з лінгвістичними моделями, використовуються вектори зсуву (offset vectors), аби позначити відношення, що «зсуває» один об'єкт до іншого.

Як представника тензорного підходу розглянемо алгоритм TuckeR [6]. У даному алгоритмі декомпозицію Такера [7] використано для вирішення завдання про передбачення значень ребер графа. Таке завдання є одним із типових завдань МН на графах. У його межах, маючи набір значень вершин та ребер, необхідно передбачити замасковане (невідоме) значення ребра.

Декомпозицію Такера можна визначити наступним чином. Маючи вхідний тензор 3-го рангу X з $R^{I_x J_x K}$, розкладаємо його на (тензорний) добуток базового тензора C та факторних матриць M_1, M_2, M_3. Факторні матриці можна уявити як напрямок за колонками, рядками та глибиною тензора для M_1, M_2, M_3 відповідно. Базовий тензор менший за початковий тензор, а сама декомпозиція не є унікальною. Оскільки метою алгоритму TuckeR є передбачення значень ребер, вихідним значенням є ймовірність того, що певна трійка $(h,\ r,\ t)$ є правдивою:

$$X = C \times_1 M_1 \times_2 M_2 \times_3 M_3.$$

Для отримання цієї ймовірності вводиться поняття скорингової функції phi. У машинному навчанні скорингові функції використовують для порівняння якості вихідних значень моделей. Мета скорингової функції в даній задачі — дати оцінку правдоподібності отриманої трійки:

$$\varphi(h, r, t) = W \times_1 h_S \times_2 w_r \times_3 t_o.$$

Тут $h,\ t$ позначають рядки з матриці E, що зберігає представлення окремих елементів графу, аналогічно w_r з R відповідає матриці відношень, а W позначає базовий тензор. Важливо, що даний алгоритм повністю експресивний, тобто (теоретично) має змогу вивчати (а тому і передбачати) будь-які типи відношень між елементами, наприклад, «один до одного», «один до багатьох» тощо.

Наступним розглянемо алгоритм MurE [8]. Даний підхід належить до трансляційних методів репрезентації. Як і у випадку TuckeR, ставимо на меті віднайти значення маскованих ребер. Для цього використовуємо скорингову функцію:

$$\varphi(h, r, t) = -d(h^r,\ t^r)^2 + b_t + b_h = -d(P * \eta,\ \tau + \rho)^2 + b_h + b_t.$$

У даному рівнянні $d(h_r, t) = d_E(h_r, t_r)^2$, тобто це евклідова відстань між векторами представлень об'єктів відповідно до відношення, що їх пов'язує; b_h та b_t — вектори зміщення (biases); P та ρ позначають діагональну матрицю відношень та вектор зсуву відповідно. Не дивлячись на простоту, даний алгоритм показує доволі високу точність у порівнянні з такими класичними алгоритмами, як TransE [9], DistMult [10] чи ComplEx [11]. Щодо експресивності моделі, наразі теоретично не доведено, чи є дана модель повністю експресивною. Варто зазначити, що досягнення повної експресивності залежить не тільки від самої моделі, а й від певних зовнішніх умов. Показано, що модель може досягти такого якісного рівня лише у випадку, якщо простір вкладення (embedding dimension) дорівнює $N/32$, де N — кількість об'єктів у датасеті [12]. Зважаючи на те, що графи знань можуть зберігати сотні тисяч та мільйони різних об'єктів, досягнути відповідної «просторовості» обчислювально неможливо або щонайменш неефективно з точки зору витрат ресурсів та часу.

Наостанок розглянемо ще одну трансляційну модель, яка має один з найвищих рівнів експресивності, — PairRE [13]. Підхід даного алгоритму полягає в наступному: маючи два пов'язаних елемента, h та t, будуємо їхнє відображення шляхом обчислення добутку Шура (добутку Адамара) відповідно до відношення, що їх пов'язує, отримуючи вектори в евклідовому просторі. Відповідні вектори позначимо як r^H та r^T. Далі необхідно вирахувати відстань між отриманими векторами. Ця відстань фактично буде відігравати роль індикатора ймовірності того, що обрана трійка є істиною. Для цього ми хочемо, аби $h \circ r^H \approx t \circ r^T$. Тоді ця скорингова функція виглядатиме наступним чином:

$$\varphi_r(h,t) = \| h \circ r^H - t \circ r^T \|.$$

У даному рівнянні $\|h\|^2$ та $\|t\|^2 = 1$. Візуалізацію даного алгоритму наведено на рис. 1 [13].

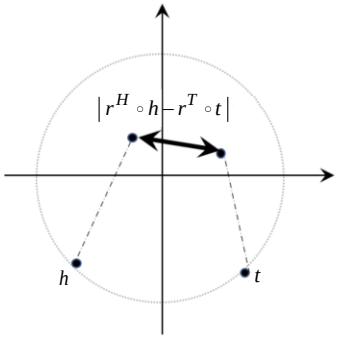

Рис. 1

Наразі ми не розглядаємо питання розрізнення неізоморфних графів, яке постає під час побудови представлень графів. Утім, варто зазначити, що як методи декомпозиції, так і методи трансляції посилюють дану проблему, адже представлення відбувається шляхом умовної сюр'єкції з множини векторних просторів ребер та вершин до одного векторного простору представлень.

Персистентні гомології

У даній роботі зроблено спробу проаналізувати різницю між описаними вище методами репрезентації багаточасткових графів за допомогою топологічних властивостей цих репрезентацій. Для цього зручно використовувати такий інструмент топологічного аналізу даних, як ПГ. Неформально ПГ дозволяють виділити геометричні константи, які формуються у так званих хмарах точок (point clouds). Насправді такі хмари зазвичай є множиною векторів, яка отримується шляхом репрезентації елементів заданої множини об'єктів (графів у нашому випадку). Для виділення цих геометричних властивостей використовується поняття симпліційного комплексу. Так, k-симплексом є опукла оболонка з $k+1$ вершинами. Наприклад, 0-, 1- та 2-симплексами є точка, ребро та трикутник відповідно [14].

Тоді алгоритм обчислення ПГ містить почергову побудову k-симплексів шляхом поєднання точок з хмари точок. Ясно, що в процесі такого об'єднання виникають певні геометричні структури, найбільш цікавою серед яких є дірка, адже вона характеризує топологічні властивості простору через гомологічні групи. Чим довше така дірка зберігається під час побудови k-симплексів, тим більше вона є персистентною. Таким чином, ПГ дозволяє обчислити топологічні властивості простору, маючи лише невелику вибірку з його представників.

Зазвичай результат обчислення ПГ для заданої хмари точок представляється у вигляді так званих персистентних діаграм, як показано на рис. 2. На зображенні цифрою 1 позначена персистентність усіх знайдених 0-вимірних дірок, а цифрою 2 — 1-вимірних. Чим далі точки знаходяться від діагоналі, тим важливішими вони є, адже зберігаються найдовше. Аналіз даних за допомогою ПГ широко використовується в МН для виконання цілої низки завдань, у першу чергу — для передобробки даних у проблемах фізики, хімії та біології [15–18]. Однією з особливостей даного підходу є можливість працювати з широким класом даних, враховуючи випадки, коли дані подаються у змішаному форматі, наприклад у вигляді картинок та тексту (мультимодальність), що важливо для роботи з графовими структурами, адже вони часто використовуються для зберігання різного типу даних.

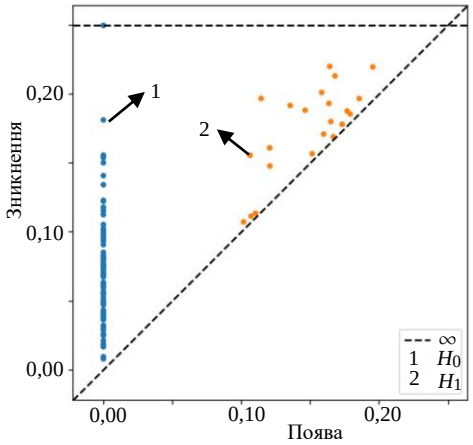

Рис. 2

Опис експерименту з аналізу репрезентацій за допомогою персистентних гомологій

Програмну реалізацію було здійснено в Python 3.10.2. Для реалізації алгоритмів було обрано бібліотеку pykeen 1.10.1, а ПГ обраховано у бібліотеці giotto-tda 0.6. Обчислення здійснювались у хмарі Google Colab із GPU A100 та 16 GB RAM. Експеримент проводився для датасету Countries [19], який складається з 1,158 трійки, маючи 271 унікальну сутність (країну) та 2 типи відношень («сусід» та «знаходиться в»). Тренування відбувалося протягом 100 епох, розмір простору вкладення (embedding dimension) був обраний за 128. Для навчання і тестування обирались 80 % та 20 % загального датасету відповідно.

Отримані після навчання алгоритмів вектори було розбито на два окремих датасети. Перший містив ПГ для хмари точок кожної країни, до якої входили 0-, 1- та 2-вимірні дірки. Другий датасет містив ПГ для дірок аналогічного виміру, але хмари точок були сформовані кожною країною та відношенням «сусід». Значення ПГ були записані у матричному форматі. Таким чином, ми перейшли від векторів вкладень країн та відношень до матриць ПГ, що обчислені за векторами вкладень.

У результаті ми отримали два датасети. Перший складався зі значень ПГ, обрахованих для вкладення окремої країни, а другий — зі значень ПГ, обрахованих для хмари точок, що містили окрему країну та відношення «сусід». У результаті було отримано 542 (271*2) матриці ПГ. Розмір матриці визначався як $N \times D$*2, де N — довжина вхідного представлення (у нашому випадку це 128), а D — кількість вимірів, у яких відбувався обрахунок гомологій (у нашому випадку це 3); мультиплікатор 2 означає, що кожен вимір записується як точка у дво-

вимірній системі координат, як можна побачити на рис. 2. Для спрощення роботи було вирішено використати статистичні методи, як описано на рис. 3.

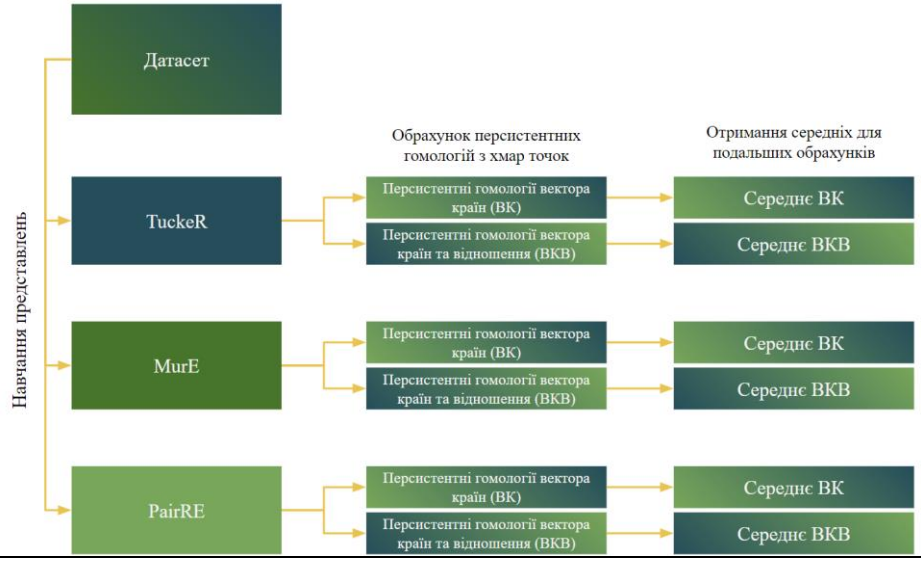

Рис. 3

На першому кроці було обчислено середнє значення для всіх ПГ, що дозволяло працювати лише з 542 окремими значеннями. Далі для цих значень було розраховано коефіцієнти ексцесу (γ_2), асиметрії (γ_1) та відхилення (σ). Для візуальної оцінки результатів було побудовано гістограми, які відображають частоту, з якою певне середнє зустрічається в датасеті.

$$\gamma_1 = \mu_3 / \sigma^3,$$

$$\gamma_2 = (\mu_4 / \sigma^4) - 3,$$

$$\sigma = \left(\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \right)^{1/2}.$$

Як зазначено вище, γ_1 та γ_2 позначають коефіцієнти асиметрії та ексцесу відповідно, σ — стандартне відхилення, а μ — середнє.

Аналіз результатів експерименту

З'ясовано, що розподіли середніх ПГ для всіх алгоритмів мають суттєві відмінності, як показано на гістограмах на рис. 4 (значення було помножено на 103 для зручності у разі оригінальних розподілів, різниці подані без змін). Було обчислено коефіцієнти ексцесу та асиметрії, що наведені в табл. 1 та 2. Як можна побачити, розподіли PairRE найбільше відрізняються від TuckeR та MurE; разом з тим PairRE має найбільший коефіцієнт ексцесу, що вказує на те, що даний алгоритм є більш робастним до викидів. Дійсно, зазвичай більшість елементів діаграми ПГ тяжіють до координати (0,0) і лише невелика кількість елементів визначається як персистентна. Разом з тим велика потужність персистентної множини може вказувати на те, що репрезентація видалась більш змішаною, тобто такою, що порівняно велика кількість елементів має координати, які суттєво відмінні від (0,0). Така ситуація може вказувати на те, що геометрична структура є в певному сенсі однорідною, тобто наявні «рівномірно» розташовані персистентні елементи. Іншими словами, позитивним є результат, коли гістограма не схожа на таку, що приналежність елемента до квартиля є майже рівноймовірною, як при MurE.

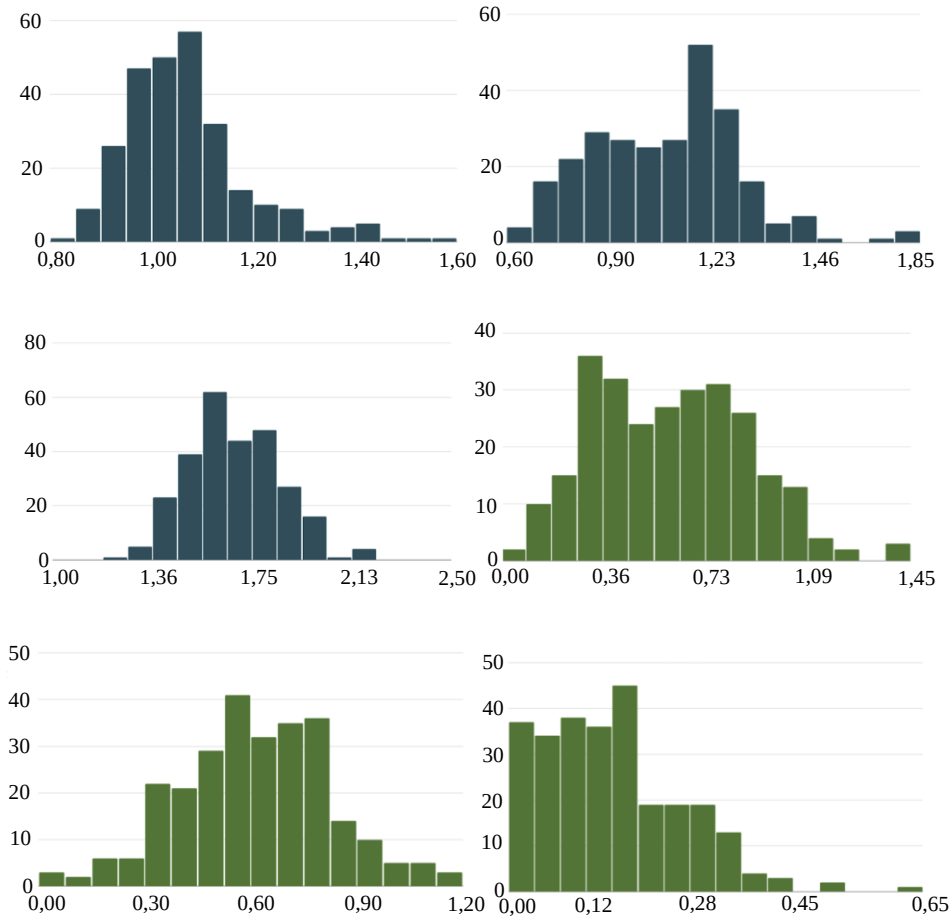

Рис. 4

Із цього випливає, що дисперсія та відхилення мають бути «найвужчими» для PairRE, що і підтверджується експериментально. Цікаво те, що хоча відхилення абсолютних значень різниць MurE та PairRE (MurE – PairRE) є найменшим, як наведено у табл. 2, дисперсія має найбільше значення, що вказує на суттєву різницю між елементами розподілів, а відповідно, і між геометричними властивостями хмар точок для цих моделей. Зазначимо, що у даному випадку результати ПГ містили в собі лише 0-вимірні значення, адже на вхід подавався вектор, а не матриця.

Таблиця 1

Параметр/Модель	**TuckeR**	**MurE**	**PairRE**
Коефіцієнт ексцесу	– 0,134	0,329	2,380
Коефіцієнт асиметрії	0,284	0,319	1,289
Відхилення	0,030	0,050	0,015
Середнє	0,002	0,001	0,001

Таблиця 2

Параметр/ Різниця моделей	**TuckeR – MurE**	**TuckeR – PairRE**	**MurE – PairRE**
Коефіцієнт ексцесу	– 0,364	0,076	0,734
Коефіцієнт асиметрії	0,300	– 0,065	0,776
Відхилення	0,075	0,046	0,012
Середнє	0,599	0,613	0,014

Розглянемо ще один датасет, де на вхід подавалася матриця, перша колонка якої містила вектор країни, а друга — вектор відношення «сусід». Як зазначено вище, початковий датасет складався з 271 типу вершини (1 тип — 1 країна) та 2 типів ребер (відношень між вершинами): «сусід» та «знаходиться в». Репрезентація відношення «сусід» відбувається шляхом вивчення того, як ребра з даною міткою поєднуються з вершинами. Усі алгоритми представлення графів, які розглядаються в даній статті, здатні до репрезентації як вершин, так і ребер, тому вихідним результатом, як і у випадку країн, є вектор довжиною 128 елементів. Варто зазначити, що загалом не кожен алгоритм представлень графів здатен вивчати як ребра, так і вершини (наприклад, Node2Vec [20]), а тому даний підхід не можна використовувати в усіх можливих випадках.

Оскільки відношення «сусід» вивчалося для графу в цілому, а не для кожної вершини, його векторна репрезентація є однаковою для всіх країн, адже всі країни мали хоча б одного сусіда. Результати аналогічних розподілів представлено на рис. 5. Можна звернути увагу на те, що розподіли мають структуру, близьку до нормальної, окрім алгоритму TuckeR, що має суттєвий асиметричний правий хвіст та найнижчий коефіцієнт ексцесу. Разом з тим, як можна побачити з табл. 3, відхилення мають суттєво менші значення, як і дисперсія для всіх алгоритмів. Це пов’язано з фактичною «нормалізацією», розмиттям векторів країн шляхом додавання вектора відношень. Для даного типу даних існують також 1-вимірні елементи діаграм ПГ. Утім, більш складні відношення не утворюються. Що ж стосується різниць, то, як можна побачити в табл. 4, відношення моделей, що передбачувано, зберігається, і дисперсія пари «MurE – PairRE» є найбільшою. Однак структурно всі три розподіли відповідають випадку з більшою ентропією, адже мають більш рівномірну структуру значень.

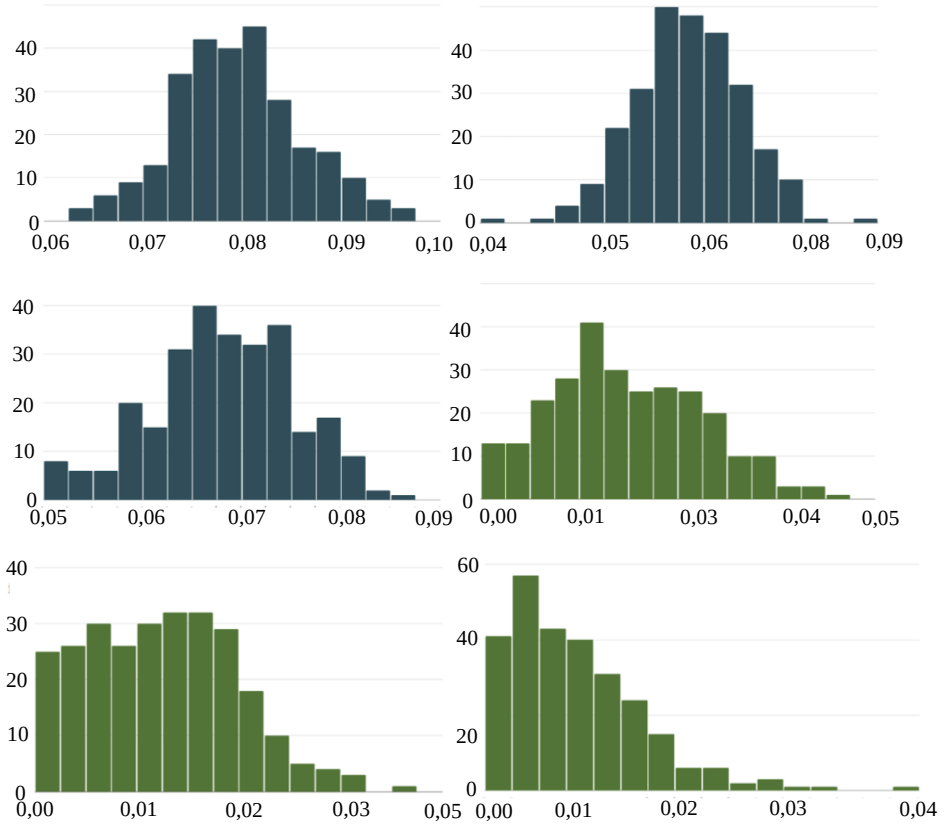

Рис. 5

ISSN 2786-6491

Таблиця 3

Параметр/Модель	TuckeR	MurE	PairRE
Коефіцієнт ексцесу	0,036	0,315	– 0,272
Коефіцієнт асиметрії	0,153	– 0,143	– 0,235
Відхилення	0,00004	0,00005	0,00004
Середнє	0,080	0,061	0,066

Таблиця 4

Параметр/ Різниця моделей	TuckeR – MurE	TuckeR – PairRE	MurE – PairRE
Коефіцієнт ексцесу	– 0,417	– 0,291	2,162
Коефіцієнт асиметрії	0,319	0,360	1,242
Відхилення	0,00009	0,00007	0,00004
Середнє	0,018	0,014	0,005

Висновок

У даній роботі проаналізовано вплив різних методів представлення багаточасткових графів на топологічні характеристики хмар точок, які сформовано векторами вкладень. Було обрано три алгоритми, TuckeR, MurE та PairRE, де перший відноситься до білінійних моделей, а інші — до методів трансляцій. Для аналізу хмар точок обрано підхід на основі ПГ. Експеримент проводився на датасеті Countries, зважаючи на його невеликий розмір і простоту інтерпретації. Для подальшого розгляду впливу елементи з персистентних діаграм було розділено на елементи, отримані виключно для векторів представлень країн, та елементи, побудовані за допомогою поєднання представлення окремої країни та відношення «сусід».

Показано, що для випадку векторів країн найбільше значення ексцесу досягається алгоритмом PairRE, що вказує на кращу робастність алгоритму. Це підтверджується тим, що, як відомо, даний алгоритм має кращі результати для передбачення значень ребер у порівнянні з іншими алгоритмами. Крім цього, дисперсія для даного методу була найнижчою. Разом з тим найбільшу різницю у статистичних параметрах середніх значень векторів було знайдено між алгоритмами PairRE та MurE. Це можна пояснити недостатньою експресивністю MurE у порівнянні з двома іншими підходами.

З'ясовано, що використання загального вектора відношення допомагає отримати ближчі до нормального розподіли для випадку векторів країн, тоді як для різниць методів структура розподілів стала більш складною, адже коефіцієнти ексцесу знизились.

Додатково показано, що поєднання векторів представлень окремих країн та загального відношення дозволяє нормалізувати представлення об'єктів. Варто зазначити, що в деяких випадках це може суттєво погіршити навчання мережі, адже знижує розрізнювальну потужність алгоритму. Втім, у випадках невеликих датасетів, таких як Countries, використання цього відношення для подальших передбачень є доцільним, оскільки дозволяє запобігти перенавчанню (overfitting).

Таким чином, показано, що персистентні діаграми можуть бути використані для аналізу методів представлення графів знань та аналізу топологічних властивостей відповідних хмар точок. Це дозволяє говорити про можливість використання даного підходу для порівняння моделей у випадку, коли їхні результати схожі на певному класі завдань чи датасетів, адже вони надають можливість оцінити, наскільки топологічно стабільними є вкладення графів. Так, вкладення можна назвати більш стабільним, якщо розподіл середніх значень елементів діаграм за ви-

міром має вищий коефіцієнт ексцесу та низьку дисперсію, адже це вказує на те, що кількість персистентних елементів для кожного випадку мінімальна, а більшість знайдених елементів є «шумом».

O. Yavorskyi, N. Kussul

STUDYING THE MULTIPARTITE GRAPH REPRESENTATIONS WITH TOPOLOGICAL DATA ANALYSIS

Oleksandr Yavorskyi

National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»,

yaotianjiu@gmail.com

Nataliia Kussul

National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»,

nataliia.kussul@gmail.com

The problem of representation of multipartite graphs for graph learning problems using topological machine learning methods, in particular, by calculating the persistent homologies of point clouds, is considered. We consider vector representations of graphs obtained using bilinear models and translation models, including the tensor decomposition model TuckeR, and the shift models MurE and PairRE. Both the fully expressive models and models with unproven expressivity levels are taken into account. As an example of a multipartite graph, a graph with two hundred and seventy-one types of vertices and two types of edges was chosen. We calculate persistent homologies for each model. The resulting representations are divided into two separate classes. The first one consists of vector representations of vertices only, while the second one has both vertex and edge representations. For both classes, persistent homologies with a maximum dimension of two are computed, covering one-, two-, and three-dimensional holes. To represent the persistent homologies, we chose persistent diagrams. After that, the points of the obtained diagrams are used for statistical analysis by calculating the values of kurtosis, skewness, deviation, and mean. These statistical characteristics are calculated both for the models themselves and for the modules of their differences. The main goal of this paper is to show that different representation models have different characteristics in terms of persistent homologies, which indicates that the models themselves are not topologically equivalent and therefore their choice fundamentally affects the quality and accuracy of studying representations of multipartite graphs. This result is achieved by comparing the above statistical parameters, as well as histograms of the average values of the obtained vectors.

Keywords: graph, multipartite graph, machine learning, topological data analysis, persistent homology.

ПОСИЛАННЯ

1. Bertolini M., Mezzogori D., Neroni M., Zammori F. Machine learning for industrial applications: a comprehensive literature review. *Expert Systems with Applications*. 2021. Vol. 175. P. 1–29. DOI: https://doi.org/10.1016/j.eswa.2021.114820
2. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators. *Neural Networks*. 1989. Vol. 2. P. 359–366. DOI: https://doi.org/10.1016/0893-6080(89)90020-8
3. Scarselli F., Gori M., Tsoi A., Hagenbuchner M., Monfardini G. The graph neural network model. *IEEE Transactions on Neural Networks*. 2009. Vol. 20. P. 61–80. DOI: https://doi.org/10.1109/TNN.2008.2005605
4. Xu J., Chen J., You S., Xiao Z., Yang Y., Lu J. Robustness of deep learning models on graphs: a survey. *AI Open.* 2021. Vol. 2. P. 69–78. DOI: https://doi.org/10.1016/j.aiopen.2021.05.002

5. Bollobás B. Modern graph theory. Graduate texts in mathematics. New York : Springer Link, 1998. Vol. 184. P. 394. DOI: https://doi.org/10.1007/978-1-4612-0619-4
6. Balazevic I., Allen C., Hospedales T. TuckER: tensor factorization for knowledge graph completion. *2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. PRC : Hong Kong, 2019. P. 5185–5194. DOI: https://doi.org/10.18653/v1/D19-1522
7. Tucker L. The extension of factor analysis to three-dimensional matrices. *Contributions to Mathematical Psychology*. 1964. P. 109–127.
8. Balažević I., Allen C., Hospedales T. Multi-relational poincaré graph embeddings. *NIPS'19: 33rd International Conference on Neural Information Processing Systems*. Canada : Vancouver, 2019. P. 4463–4473. DOI: https://doi.org/10.48550/arXiv.1905.09791
9. Bordes A., Usunier N., Garcia-Durán A., Weston J., Yakhnenko O. Translating embeddings for modeling multi-relational data. *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*. USA : New York, 2013. Vol. 2. P. 2787–2795. DOI: https://doi.org/10.5555/2999792.2999923
10. Yang B., Yih W., He X., Gao J., Deng L. Embedding entities and relations for learning and inference in knowledge bases. *CoRR.* 2014. P. 12. DOI: https://doi.org/10.48550/arXiv.1412.6575
11. Trouillon T., Welbl J., Riedel S., Gaussier E., Bouchard G. Complex embeddings for simple link prediction. *33rd International Conference on Machine Learning.* USA : New York, 2016. Vol. 48. P. 2071–2080. URL: https://proceedings.mlr.press/v48/trouillon16.html.
12. Wang Y., Gemulla R., Li H. On multi-relational link prediction with bilinear models. *32nd AAAI Conference on Artificial Intelligence*. USA : New Orleans, 2018. Vol. 32, N 1. P. 8. DOI: https://doi.org/10.1609/aaai.v32i1.11738
13. Chao L., He J., Wang T., Chu W. PairRE: knowledge graph embeddings via paired relation vectors. 2020. P. 10. DOI: https://doi.org/10.48550/arXiv.2011.03798
14. Edelsbrunner H., Harer J. Computational topology: an introduction. *American Mathematical Society*. 2010. P. 241. ISBN: 1470467690.
15. Cang Z., Wei G.-W. Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction. *International Journal for Numerical Methods in Biomedical Engineering*. 2018. Vol. 34. N. 2. P. 1–25. DOI: https://doi.org/10.1002/cnm.2914
16. Xin B., Huang J., Zhang L., Zheng Ch., Zhou Y., Lu J., Wang X. Dynamic topology analysis for spatial patterns of multifocal lesions on MRI. *Medical Image Analysis.* 2022. Vol. 76. P. 1–18. DOI: https://doi.org/10.1016/j.media.2021.102267
17. Gao D., Chen J., Dong Z., Lin H. Connectivity-guaranteed porous synthesis in free form model by persistent homology. *Computers* & *Graphics*. 2022. Vol. 106. P. 33–44. DOI: https://doi.org/10.1016/j.cag.2022.05.018
18. Bouchard G., Singh S., Trouillon T. On approximate reasoning capabilities of low-rank vector spaces. *2015 AAAI Spring Symposium on Knowledge Representation and Reasoning (KRR): Integrating Symbolic and Neural Approaches*. USA : Palo Alto, 2015. 9 p.
19. Countries dataset. URL: https://pykeen.readthedocs.io/en/stable
20. Grover A., Leskovec J. node2vec: scalable feature learning for Networks. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. USA : San Francisco, 2016. P. 855–864. DOI: https://doi.org/10.1145/2939672.2939754

Отримано 24.07.2023