УДК 004.89 DOI: 10.31673/2412-4338.2022.048489

Шаптала Р.В., Кисельов Г.Д.
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ

ОГЛЯД МЕТОДІВ ЗЛИТТЯ ВЕКТОРНИХ ПРЕДСТАВЛЕНЬ

Анотація. *У даній статті представлено вичерпний огляд найсучасніших методів злиття векторних представлень при обробці природної мови та у машинному навчанні. Злиття векторних представлень має на меті об'єднання кількох векторних представлень слів або документів в єдине представлення, яке може кодувати різні аспекти вхідних даних. Це зазвичай використовується у застосунках мультимодального машинного навчання, де вхідні дані надходять із різних джерел у різних форматах або в ситуаціях, коли векторні представлення вже доступні, і їх потрібно об'єднати в моделі. Стаття охоплює різні методи злиття, включаючи конкатенацію, усереднення, зважене усереднення та підходи на основі нейронних мереж.*

У статті надано детальний аналіз переваг і недоліків кожного методу, а також сценарії, в яких вони є найбільш ефективними. векторні представлення класифіковано за шаром архітектури моделі, на якому відбувається злиття, а також за типом злиття. Більше того, описуються рекомендації вибору методу злиття векторних представлень залежно від обмежень задачі. Також, у статті обговорюються метрики оцінювання, які зазвичай використовуються для оцінки якості результуючих векторних представлень, такі як подібність і точність класифікації.

Загалом, дана стаття є цінним ресурсом для дослідників і практиків у сфері обробки природної мови та машинного навчання, які бажають поглибити своє розуміння методів злиття векторних представлень та їх застосування. Відомості та рекомендації, представлені в цьому документі, можуть допомогти у виборі відповідних методів злиття та покращити ефективність вирішення різноманітних завдань обробки природної мови та машинного навчання. Слідкуючи за останніми розробками в методах злиття векторних представлень, дослідники та практики можуть продовжувати розширювати межі обробки природної мови та машинного навчання.

Ключові слова: *машинне навчання, обробка природної мови, математичне моделювання, нейронні мережі, векторні представлення слів, векторні представлення документів, метрики злиття векторних представлень.*

Shaptala R.V., Kyselov G.D.

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv

REVIEW OF VECTOR EMBEDDINGS FUSION METHODS

Abstract. *This paper presents a comprehensive review of the state-of-the-art techniques for embeddings fusion in natural language processing and machine learning. Embeddings fusion refers to the task of combining multiple word or document embeddings into a single representation that can capture the different aspects of the input data. This is typically used in multimodal machine learning applications where inputs come from different sources with different formats or in situations when embeddings are already available and need to be combined in the model. The paper covers various*

© Шаптала Р.В., Кисельов Г.Д. *2022*

fusion methods, including concatenation, averaging, weighted averaging, and neural network-based approaches.

A detailed analysis of the benefits and limitations of each method, as well as the scenarios in which they are most effective is provided. In the paper vector embeddings fusion methods are categorized by model architecture type as well as by fusion type. Moreover, recommendations to choose the optimal type of fusion method given task limitations are described. In addition, the paper discusses the evaluation metrics commonly used to assess the quality of fused embeddings, such as similarity and classification accuracy.

Overall, this review paper provides a valuable resource for researchers and practitioners in the field of natural language processing and machine learning who wish to deepen their understanding of embeddings fusion methods and their applications. The insights and recommendations presented in this paper can help guide the selection of appropriate fusion methods and improve the performance of various natural language processing and machine learning tasks. By staying up-to-date with the latest developments in embeddings fusion, researchers and practitioners can continue to push the boundaries of natural language processing and machine learning.

Keywords: *machine learning, natural language processing, mathematical modelling, neural networks, word embeddings, document embeddings, embeddings fusion metrics.*

1. Постановка проблеми

У сучасних підходах до обробки природної мови та машинному навчанні, використання векторних представлень є одним з ключових інструментів репрезентації даних. Векторні представлення дозволяють перетворити слова або речення в числові вектори, що можуть бути використані для розв'язання низки важливих завдань, таких як класифікація текстів, машинний переклад чи аналіз тональності текстів [1].

Проте, у багатьох випадках при розв’язанні задач машинного навчання, ознаки, що передаються на вхід моделям повинні кодувати інформацію наявну у кількох джерелах, або розподілену між ними. Тому виникає потреба об’єднувати різні векторні представлення у єдине векторне представлення без втрати сигналу для навчання. Цей процес називається злиттям, і є архітектурним рішенням при побудові та використанні моделей машинного навчання при обробці природних мов.

Актуальність проблеми полягає в тому, що злиття векторних представлень може покращити точність та ефективність вирішення різних задач обробки природної мови та машинного навчання, додати нові сигнали до існуючих моделей, що здатно суттєво покращити якість систем на їх основі. Класифікація та упорядкування методів злиття векторних представлень дозволяє фахівцям обирати оптимальні методи для їх практичних застосувань, а також створювати нові методи, ітеративно прибираючи недоліки існуючих.

2. Аналіз останніх досліджень і публікацій

З розвитком багатомодального навчання та нейронних мереж багато уваги почали приділяти створенню та аналізу методів злиття векторних представлень. У роботі [2] автори аналізують дані методи у площині практичних застосувань та наводять перелік наборів даних та завдань, які вирішуються ними. Стаття [3] присвячена розробці методів злиття на основі нейронних мереж для побудови представлення користувача у рекомендаційних системах. Автори [4] використовують конкатенацію векторних представлень зі зменшеною розмірністю як метод злиття векторних представлень для системи рекомендації товарів. Робота [5] присвячена категоризації методів злиття векторних представлень у моделях, що застосовуються у біомедичній сфері. Більш глибоко методи злиття на основі нейронних мереж описані у статті [6], у підсумку якої наведені недоліки саме даного типу методів. Для задачі аналізу відео автори [7] будують глибоку архітектуру нейронної мережі, та зазначають, що кращий метод злиття на внутрішніх метриках не завжди відповідає оптимальному у контексті вирішуваної задачі і може призвести до перенавчання. Попри велику кількість нових методів, особливо на основі нейронних мереж, більшість досліджень не систематизує попередні

дослідження та не надає рекомендацій при виборі таких методів, що пропонується у даній статті.

3. Мета і задачі дослідження

Метою даної роботи є комплексний огляд методів злиття векторних представлень. Робота досліджує теоретичну та практичну значимість проблеми та її важливість у різних галузях, таких як обробка природних мов, комп'ютерний зір та машинне навчання. Відповідно до мети сформовані наступні задачі: проаналізувати різні методи злиття векторних представлень, включаючи конкатенацію, усереднення, зважене усереднення та підходи на основі нейронних мереж; оцінити переваги та недоліки кожного методу; узагальнити та упорядкувати методи за категоріями та надати рекомендації з вибору методу злиття векторних представлень у залежності від обмежень вирішуваного завдання; визначити поточні виклики та напрямки майбутніх досліджень у методах злиття векторних представлень.

4. Результати дослідження

Класифікувати методи злиття векторних представлень можна за низкою критеріїв. Для початку наведемо класифікацію за шаром архітектури моделі [2, 5], на якому відбувається злиття:

1. Раннє злиття - векторні представлення об'єднуються перед подальшою обробкою даних;
2. Проміжне злиття - у цьому методі кожне векторне представлення спочатку перетворюється в інший простір, щоб зменшити розмірність вхідних векторів, після чого об'єднані векторні представлення проходять через додаткові шари нейронної мережі, які об'єднують їх в одне злите представлення;
3. Пізнє злиття - векторні представлення об'єднуються на останніх етапах обробки даних, наприклад, перед оцінкою моделі.

За типом злиття можна виділити наступні класи методів:

1. Конкатенація - вектори представлень об'єднуються в один вектор шляхом з'єднання їх у послідовність;
2. Усереднення - вектори представлень об'єднуються шляхом обчислення середнього значення компонентів;
3. Зважене усереднення - вектори представлень об'єднуються зваженим усередненням, де кожен вектор має свій ваговий коефіцієнт;
4. На основі нейронних мереж - вектори представлень об'єднуються за допомогою нейронних мереж.

Кожен з наведених вище класів методів злиття векторних представлень має ряд переваг та недоліків. Конкатенація має наступні переваги: простота (метод досить простий, що дозволяє легко об'єднати два чи більше векторних представлення в одне), незалежність від типу даних (конкатенація дозволяє поєднувати різні типи векторних представлень та створювати нові комбінації, що збільшує різноманітність даних) та гнучкість (можна легко додавати нові векторні представлення до злиття без змін попередніх значень). До недоліків можна віднести розмірність (конкатенація може призводити до збільшення розмірності векторів, що може зробити обробку даних більш складною), однакова вага ознак (різні векторні представлення можуть мати різну важливість для завдання, і конкатенація може не враховувати ці різниці в значеннях), масштаб (векторні представлення можуть мати різні масштаби значень, що може впливати на їхню взаємодію після конкатенації).

У свою чергу, усереднення зручне через наступні ознаки: зменшення розмірності (усереднення допомагає зменшити розмірність векторів, що полегшує їх обробку та зберігання), інтерпретація (усереднення дозволяє отримувати більш зрозумілі результати, оскільки кінцевий вектор містить середнє значення властивостей, що представлені в початкових векторах) та зменшення впливу шуму (усереднення може допомогти зменшити вплив шуму на результат, оскільки він буде зменшуватися з кожним доданим вектором). Серед

недоліків усереднення можна виділити втрату інформації (усереднення може привести до втрати інформації, оскільки кінцевий вектор містить лише середнє значення властивостей, що представлені в початкових векторах), різні масштаби (усереднення може результувати у недостатньо вагомому внеску від векторів з меншими значеннями, що може вплинути на кінцевий результат) та не ефективність для нелінійних відношень (усереднення не ефективне для виявлення нелінійних залежностей між властивостями, що представлені в початкових векторах).

Зважене усереднення усуває недолік однакової ваги векторних представлень попередніх методів, адже у ньому додається механізм контролю ваги кожного вектора відповідно до його важливості в контексті завдання. Це може підвищити точність результатів, так як важливі векторні представлення матимуть більший вплив на кінцевий результат. Недоліки даного методу все ще включають втрату інформації та різні масштаби векторів, але до них додається потреба у налаштуванні нових вагів та складність обчислень якщо векторних представлень, що приймають участь у злитті багато.

Методи злиття векторних представлень на основі нейронних мереж значно більш складні та потужні, ніж інші. Так до їх переваг можна віднести автоматичне виокремлення ознак (нейронні мережі можуть автоматично виокремлювати важливі ознаки з кожного векторного представлення та комбінувати їх для отримання кращих результатів), гнучкість (метод злиття векторних представлень за допомогою нейронних мереж може бути налаштований для вирішення різних завдань і використовувати різні архітектури мереж для досягнення кращих результатів) та здатність до роботи з неструктурованими даними (нейронні мережі можуть бути застосовані для обробки складних даних, таких як зображення або текст, та використовувати векторні представлення для збільшення точності результуючого рішення). У свою чергу дані методи мають набір обмежень, а саме: потреба в навчанні (навчання нейронних мереж може вимагати багато часу та ресурсів для отримання оптимальних результатів), складність розуміння (нейронні мережі важко інтерпретувати, а їхні результати важко пояснити) та необхідність великої кількості даних (для навчання нейронних мереж може знадобитися велика кількість даних, які можуть бути не доступні, або вимагають значні ресурси для отримання).

Вибір методу злиття векторних представлень залежить від конкретної задачі та обмежень, які вона має. Відповідно до даних обмежень, а також враховуючи переваги та недоліки кожного методу можна сформулювати наступні рекомендації вибору методу злиття векторних представлень:

1. Якщо наявні обсяг даних або потужність обчислювальних ресурсів обмежені, метод усереднення може бути оптимальним варіантом. Він зазвичай працює швидше за інші методи, не збільшує розмірність вихідного вектору та не вимагає навчання чи налаштування;
2. Якщо вхідні дані багатомодальні, але кожна модальність має різну важливість для вирішення задачі, метод зваженого усереднення дозволить це врахувати. Він забезпечує збереження та контроль сигналу кожного з векторів вхідних представлень;
3. Якщо у моделі закладена можливість обробки багатовимірних векторів, слід обрати метод конкатенації. Він дозволяє зберегти всі властивості вхідних векторів і збільшити розмірність результату без втрати інформації для наступних шарів моделі;
4. Якщо вхідні дані складаються з неструктурованих даних, методи на основі нейронних мереж можуть забезпечити найбільшу точність результуючого рішення через здатність до адаптації. Вони можуть автоматично вибирати найкращий спосіб злиття векторів та дозволяють отримати високу якість вихідного представлення.

Оцінка якості злиття векторних представлень зазвичай залежить від конкретної задачі, для якої векторні представлення використовуються. Вибір метрик для оцінки якості злиття векторних представлень повинен враховувати конкретну задачу та обмеження використання векторних представлень. Наприклад, для задачі класифікації текстів можна використовувати метрики, що вимірюють точність класифікації з використанням різних методів злиття [8].

Проте, існують загальні метрики якості, які можуть бути використані для порівняння різних методів злиття. Однією з таких метрик є косинусна відстань між векторними представленнями [9]. Ця метрика вимірює кут між векторами у просторі векторів, що представляють слова або об'єкти. Чим менше кут між векторами, тим більш схожі слова або об'єкти. Іншою метрикою якості є внутрішня якість злиття, яка оцінюється шляхом вимірювання відстані між векторами одного слова або об'єкта, які були представлені різними методами злиття. Ця метрика дозволяє оцінити, наскільки добре метод злиття зберігає важливі риси векторного представлення [10].

5. Висновки і перспективи подальших досліджень.

Таким чином, у статті систематизовано методи злиття векторних представлень за типом злиття незалежно від прикладної області чи задачі, виділено їх переваги та недоліки, а також надано рекомендації з вибору методу злиття векторних представлень враховуючи обмеження конкретних задач. За допомогою результатів даного дослідження можна вибрати оптимальний метод злиття для практичних задач або покращити якість існуючих моделей машинного навчання через заміну відповідного методу.

У подальшому дослідження можна розвивати у кількох напрямках. Перший з них – це систематизація найбільшої категорії методів злиття векторних представлень, а саме на основі нейронних мереж. Такий напрям дозволить обирати оптимальні архітектури для вирішення задач, де наявно багато даних та обчислювальних ресурсів. Іншим напрямом дослідження може бути порівняння метрик для оцінки методів злиття векторних представлень та дослідження кореляції між загальними метриками та метриками залежними від конкретних задач. Це дозволить отримати рекомендації по їх використанню залежно від вимог розв’язуваних проблем.

Список літератури:

1. Camacho-Collados J., Pilehvar M. Embeddings in Natural Language Processing. Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts. 2020. P. 10 – 15.

2. Ramachandram D., Taylor G. W. Deep Multimodal Learning: A Survey on Recent Advances and Trends. IEEE Signal Processing Magazine. 2017. №34(6). P. 96 – 108.

3. Blandfort P., Karayil T., Raue F., Hees J., Dengel A.R. Fusion Strategies for Learning User Embeddings with Neural Networks. 2019 International Joint Conference on Neural Networks (IJCNN). 2019. P. 1 – 8.

4. Wróblewska A., Dąbrowski J., Pastuszak M., Michałowski A., Daniluk M., Rychalska B., Wieczorek M., Sysko-Romańczuk S. Designing Multi-Modal Embedding Fusion-Based Recommender. Electronics. 2022. №11(9). P. 1391.

5. Stahlschmidt S. R., Ulfenborg B., Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics. 2022. №23(2). P. 1 – 15.

6. Gao J., Li P., Chen Z., Zhang J. A survey on deep learning for multimodal data fusion. Neural Computation. 2020. №32(5). P. 829 – 864.

7. Shvetsova N., Chen B., Rouditchenko A., Thomas S., Kingsbury B., Feris R. S., Kuehne H. Everything at once-multi-modal fusion transformer for video retrieval. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 20020 – 20029.

8. Qiu S., Chang G. H., Panagia M., Gopal D. M., Au R., Kolachalama V. B. Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2018. №10. P. 737 – 749.

9. McFee B., Lanckriet G., Jebara T. Learning Multi-modal Similarity. Journal of machine learning research. 2011. №12(2). P. 491 – 523

10. Li Q., Xiong Q., Gao M., Ji S., Yu Y., Wu C. Multi-view heterogeneous fusion and embedding for categorical attributes on mixed data. Soft Computing. 2020. №24. P. 10843 – 10863.

References:

1. Camacho-Collados J., Pilehvar M. Embeddings in Natural Language Processing. Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts. 2020. P. 10 – 15.

2. Ramachandram D., Taylor G. W. Deep Multimodal Learning: A Survey on Recent Advances and Trends. IEEE Signal Processing Magazine. 2017. №34(6). P. 96 – 108.

3. Blandfort P., Karayil T., Raue F., Hees J., Dengel A.R. Fusion Strategies for Learning User Embeddings with Neural Networks. 2019 International Joint Conference on Neural Networks (IJCNN). 2019. P. 1 – 8.

4. Wróblewska A., Dąbrowski J., Pastuszak M., Michałowski A., Daniluk M., Rychalska B., Wieczorek M., Sysko-Romańczuk S. Designing Multi-Modal Embedding Fusion-Based Recommender. Electronics. 2022. №11(9). P. 1391.

5. Stahlschmidt S. R., Ulfenborg B., Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics. 2022. №23(2). P. 1 – 15.

6. Gao J., Li P., Chen Z., Zhang J. A survey on deep learning for multimodal data fusion. Neural Computation. 2020. №32(5). P. 829 – 864.

7. Shvetsova N., Chen B., Rouditchenko A., Thomas S., Kingsbury B., Feris R. S., Kuehne H. Everything at once-multi-modal fusion transformer for video retrieval. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 20020 – 20029.

8. Qiu S., Chang G. H., Panagia M., Gopal D. M., Au R., Kolachalama V. B. Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2018. №10. P. 737 – 749.

9. McFee B., Lanckriet G., Jebara T. Learning Multi-modal Similarity. Journal of machine learning research. 2011. №12(2). P. 491 – 523

10. Li Q., Xiong Q., Gao M., Ji S., Yu Y., Wu C. Multi-view heterogeneous fusion and embedding for categorical attributes on mixed data. Soft Computing. 2020. №24. P. 10843 – 10863.