
ИССЛЕДОВАНИЕ РАЗЛИЧНЫХ МЕТОДОВ ПАРСИНГА МУЛЬТИМЕДИЙНЫХ ДАННЫХ И АНАЛИЗ ИХ КЛЮЧЕВЫХ РАЗЛИЧИЙ ДЛЯ ЗАДАЧИ ПОДБОРА РЕКОМЕНДАЦИЙ СОГЛАСНО КАТЕГОРИЗАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Оганесьянц Константин Павлович,

студент, магистрант

1 курс, факультет "Отдел магистратуры", кафедра «Медиа-технологии»

Донской государственной технической университет

Россия, г. Ростов-на-Дону

Кудинов Никита Георгиевич,

руководитель Центра R&D, Технопарк Магика,

генеральный директор, ООО «ЗРЕНИЕ 2.0» (резидент Фонда «Сколково»)

аспирант по направлению «Медиакоммуникация и журналистика»

Донской государственной технической университет

Россия, г. Ростов-на-Дону

Аннотация

В данной работе проводится сравнительный анализ и исследование различных методов парсинга медиаконтента. Основная задача данного исследования – выявление наиболее эффективного инструмента, для создания персонализированных рекомендаций. Оценка производится на основе точности, скорости и обучаемости каждого метода. Работа направлена на предоставление детального анализа существующих подходов к парсингу мультимедийных данных, без углубления в разработку новых методов.

Ключевые слова: мультимедийные данные, виртуальная реальность, дополненная реальность нейронные сети, таргетированные рекомендации, парсинг данных.

INVESTIGATING DIFFERENT MULTIMEDIA PARSING METHODS AND ANALYZING THEIR KEY DIFFERENCES FOR THE TASK OF RECOMMENDATION MATCHING ACCORDING TO ARTIFICIAL INTELLIGENCE CATEGORIZATION

Konstantin P. Oganesyants,

student, master's student

1st year, Faculty "Master's Degree Division", Department "Media Technologies"

Don State Technical University

Russia, Rostov-on-Don

Nikita G. Kudinov,

Head of R&D Center, Technopark Magika,

General Director, LLC "Sight 2.0" (Skolkovo Foundation resident)

Postgraduate student in the field of Media Communication and Journalism

Don State Technical University

Russia, Rostov-on-Don

ABSTRACT

This study analyzes the effectiveness of multimedia parsing techniques in the context of Augmented Reality (AR). The main objective of the study is to compare different parsing methods using neural networks to achieve targeted recommendations in an AR application. The evaluation is done on the basis of accuracy, speed and learnability of each method. The work aims to provide a detailed analysis of existing approaches to parsing multimedia data in the virtual reality domain, without delving into the development of new methods.

Keywords: multimedia data, virtual reality, augmented reality neural networks, targeted recommendations, data parsing.

В современном мире все большее значение приобретают технологии, связанные с дополненной реальностью (AR). В связи с этим вопросы обработки и последующего анализа мультимедийных данных становятся особенно актуальными, ведь в AR-среде особенно важны визуальные элементы, большую часть которых составляют изображения и видеоматериалы. Данное исследование направлено на изучение существующих методов парсинга. Так же, помимо сферы AR быстрое развитие на сегодняшний день получила область искусственного интеллекта. Поэтому для данной статьи были выбраны алгоритмы парсинга, которые максимально используют все преимущества нейронных сетей.[2] В отличие от традиционных методов, такой подход предоставляет автоматизированный инструментарий распознавания предоставляемой информации, что способствует более эффективному предоставлению информации конечным пользователям.[3] В рамках данного исследования подробно рассматриваются различные методы парсинга мультимедийных данных, с оценкой их эффективности с точки зрения точности, скорости и обучаемости.

В рамках приложения дополненной реальности (AR) предполагается, что пользователь взаимодействует с окружающим миром через камеру устройства. Суть процесса заключается в том, что при наведении камеры на интересующий объект искусственный интеллект распознает содержимое кадра и генерирует набор описательных характеристик. Этот набор представляет собой ключевую информацию о содержании сцены, запечатленной камерой. Задача, рассматриваемая в данной работе, заключается в обработке полученного списка признаков (далее - тегов) и последующем создании персонализированных рекомендаций, тематически связанных со сканируемым объектом интереса. Основной целью является выявление наиболее эффективных подходов, способных обеспечить формирование персонализированных рекомендаций на основе контента, зафиксированного камерой виртуальной реальности.[4] Достижение этой цели будет обеспечиваться в первую очередь тем, насколько подходящий контент будет предложен пользователю после распознавания интересующего его объекта. Эта задача

решается путем создания эффективного инструмента для парсинга данных. В следующем тексте будет рассмотрено несколько различных подходов к процессу парсинга.

Для задачи выделения скрытых тем в коллекции текстовых документов Topic Modeling с использованием Latent Dirichlet Allocation (LDA): Latent Dirichlet Allocation (LDA). Одним из основных этапов решения задачи, рассматриваемой данной статьей является выявление ключевых тем в созданном наборе тегов. Данная модель существенно упрощает данную задачу, так как предполагает, что каждый документ (в данном случае, текстовый тег) является смесью нескольких тем, а каждая тема – совокупность распределения слов. Из основных плюсов данного подхода можно выделить: интерпретируемость – LDA предоставляет интерпретируемые темы, что облегчает понимание содержания текстов, способность выявления скрытых тематических паттернов, а так же широкое применение в различных областях. Основными минусами данной модели принято считать высокую зависимость от качества картинки – ее точность снижается в соответствии с повышением уровня «зашумленности» изображения, а так же немаловажным является факт высокой зависимости от подаваемых на вход параметров.[1] Наиболее подходящим сценарием использования данной модели является задача каталогизации большого объема фотографий в социальных сетях или изображений в базе данных. LDA может помочь автоматически выявить темы, характеризующие содержание изображений, такие как "природа", "города", "люди", что упростит процесс организации и поиска по данным.

Для более корректной обработки текста существует несколько различных подходов.

Одним из основных является методика преобразования текстового материала в векторное представление. Такой подход существенно упрощает поиск смысловых связей между различными частями текста. Данная методика используется такими крупными компаниями, как Google, Facebook и Microsoft. Одним из наиболее популярных инструментов такого преобразования и последующего анализа полученных данных можно считать Embedding на основе Word2Vec. Он основан на идее, что слова, используемые в схожих контекстах, имеют близкие векторные представления.[3] В контексте обработки текстовых тегов, Embedding на основе Word2Vec может помочь преобразовать слова в векторы, сохраняя семантическую связь между ними. LDA (Latent Dirichlet Allocation) обладает несколькими существенными преимуществами. Во-первых, модель предоставляет интерпретируемые темы, что значительно облегчает понимание содержания текстовых данных. Кроме того, LDA способна выявлять скрытые тематические паттерны, такие как такие как "король - мужчина + женщина = королева". Однако необходимо выделить несколько ключевых ограничений данной модели. Первым и основным недостатком можно считать игнорирование семантики фраз. В модели Word2Vec слова рассматриваются независимо друг от друга, что может привести к игнорированию семантической связи между словами в фразах. Так же, ключевым ограничением данного подхода на основе векторного представления можно считать сильное ограничение контекстного смысла.[5]

Главным инструментом выявления взаимосвязей между объектами можно считать статистический подход. На нем в той или иной степени основывается большинство существующих моделей искусственного интеллекта и обработки данных. Поэтому одним из наиболее эффективных инструментов определения содержания текстовых данных можно считать TF-IDF (Term Frequency-Inverse Document Frequency). Он основывается на вычислении индексов частоты, с которыми слова встречаются в текстовых данных, которые передаются в качестве входных данных. Ключевой особенностью этого подхода является эффективное выделение взаимосвязей между различными частями входных данных. Главным преимуществом такого подхода является простота интерпретации результатов, однако необходимо подчеркнуть, что данная модель не учитывает семантическую связь

между словами и может быть неэффективным в задачах, где важен контекст. Преимущества модели включают в себя простоту интерпретации результатов. TF-IDF чаще всего используется для взвешивания тегов в задачах, таких как организация фотографий по ключевым словам. Это упрощает процесс классификации и поиска в больших объемах данных.[1]

В заключении необходимо выделить оптимальное решение задачи парсинга мультимедийных данных в среде дополненной реальности. По результатам проведенного анализа нескольких популярных методов решения видно, что наибольшую эффективность обеспечит подход, использующий метод Embedding на основе Word2Vec. Этот метод обеспечивает семантическую близость между словами, учитывает контекст и смысл тегов, что особенно важно для эффективного подбора рекомендаций. Этот подход также анализирует семантическое сходство между тегами, что может обеспечить более эффективное выделение «общей» темы, которой является отсканированный объект интереса.

Список литературы:

1. Xie J. et al. Webpage Intelligent Parsing Algorithm Based on Text and Symbol Density // Academic Journal of Computing & Information Science. – 2022. – Т. 5. – №. 4. – С. 18-21.
2. Yuan J. et al. Attention-based neural tag recommendation // Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part II 24. – Springer International Publishing, 2019. – С. 350-365.
3. Ma C. et al. Content Feature Extraction-based Hybrid Recommendation for Mobile Application Services // Computers, Materials & Continua. – 2022. – Т. 71. – №. 3.
4. Dong J. et al. Personalized recommendation system based on social tags in the era of Internet of Things // Journal of Intelligent Systems. – 2022. – Т. 31. – №. 1. – С. 681-689.
5. Luo Y. et al. pRide: Privacy-preserving ride matching over road networks for online ride-hailing service // IEEE Transactions on Information Forensics and Security. – 2018. – Т. 14. – №. 7. – С. 1791-1802

References:

1. Xie J. et al. Webpage Intelligent Parsing Algorithm Based on Text and Symbol Density // Academic Journal of Computing & Information Science. – 2022. – Т. 5. – №. 4. – С. 18-21.
2. Yuan J. et al. Attention-based neural tag recommendation // Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part II 24. – Springer International Publishing, 2019. – С. 350-365.
3. Ma C. et al. Content Feature Extraction-based Hybrid Recommendation for Mobile Application Services // Computers, Materials & Continua. – 2022. – Т. 71. – №. 3.
4. Dong J. et al. Personalized recommendation system based on social tags in the era of Internet of Things // Journal of Intelligent Systems. – 2022. – Т. 31. – №. 1. – С. 681-689
5. Luo Y. et al. pRide: Privacy-preserving ride matching over road networks for online ride-hailing service // IEEE Transactions on Information Forensics and Security. – 2018. – Т. 14. – №. 7. – С. 1791-1802