

АЛГОРИТМ ФОРМИРОВАНИЯ РЕКОМЕНДАЦИЙ ХУДОЖЕСТВЕННЫХ ПРОИЗВЕДЕНИЙ НА ОСНОВЕ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТОВ С ПРИМЕНЕНИЕМ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

Анджаев Б. М.¹

Научный руководитель – канд. техн. наук, доцент Харитонов А.Ю.¹

¹Университет ИТМО

mail81025@mail.ru

Введение

Рекомендательные системы художественных произведений в большинстве используют метаданные и поведенческие характеристики пользователей, не учитывая содержательные особенности текста. Но художественные произведения обладают сложной нарративной структурой, большой длиной и выраженной эмоциональной динамикой, что затрудняет их формальный анализ. Развитие трансформерных архитектур и моделей семантических эмбедингов[1,2] открывает возможность построения алгоритмов рекомендаций, учитывающих не только формальные характеристики произведения, но и его смысловую и эмоциональную близость к предпочтениям пользователя.

Основная часть

В работе предложен алгоритм формирования рекомендаций художественных произведений, основанный на иерархической сегментации текста и формировании семантических представлений сегментов с использованием нейросетевых моделей эмбедингов [2]. Художественное произведение разбивается на логические фрагменты (главы, сцены, эпизоды), для которых с помощью моделей эмбедингов формируются векторные представления. Суммаризация и retrieval-отбор ключевых фрагментов, позволяют учитывать ограничения контекстного окна и сохранять связность нарратива [3]. Чтобы повысить устойчивость анализа длинных текстов, используется комбинация локальной суммаризации сегментов и retrieval-ориентированного отбора наиболее значимых фрагментов. Итоговое представление произведения формируется на основе агрегированных пересказов и их семантических векторов. Такой подход снижает влияние ограничений контекстного окна моделей и повышает связность итогового описания. Сравнение произведений в векторном пространстве производится по метрикам семантической близости для формирования рекомендаций на основе смыслового и эмоционального сходства.

Выводы

Проведен анализ и сравнение существующих моделей интерпретации и рекомендации. Разработан алгоритм рекомендаций художественных произведений на основе семантического анализа текста с применением нейросетевых моделей.

Литература

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). Attention Is All You Need. In: Guyon, I., Luxburg, U., Bengio, S. et al. (eds) Advances in Neural Information Processing Systems 30 (2017). Curran Associates, Inc. <https://arxiv.org/abs/1706.03762>.
2. Reimers, N., Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics. <https://arxiv.org/abs/1908.10084>.

3. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: Advances in Neural Information Processing Systems 33 (2020). Curran Associates, Inc. <https://arxiv.org/abs/2005.11401>.

Автор _____ Анджаев Б.М.

Научный руководитель _____ Харитонов А.Ю.