Методы и алгоритмы распознавания отношений между именованными сущностями в условиях обучения без примеров для извлечения информации из текстов

Классификация отношений между именованными сущностями является важной задачей в области обработки естественных языков (NLP) и широко применяется в системах информационного поиска, автоматического анализа текстов и построения онтологий. Однако большинство современных моделей требуют обширных размеченных данных, что затрудняет их использование в ситуациях, когда новые классы отношений не представлены в обучающей выборке. В данной работе рассматривается метод классификации отношений в условиях zero-shot learning, при котором модель должна предсказывать связи между сущностями, которые не содержатся в тренировочном наборе данных.

Предложенный подход основан на использовании трансформерной модели Relation Transformer, адаптированной для обработки мультимодальных данных. Векторные представления включают не только эмбеддинги текста, но и текстовое описание классов отношений, что обеспечивает моделирование признакового пространства с учетом семантической структуры данных. Оценка метода проводилась на известных наборах данных WebNLG, NYT и NEREL, последний из которых представляет особый интерес, так как содержит document-level аннотации и ориентирован на русский язык.

В ходе экспериментов продемонстрировано, что предложенная модель показывает высокие показатели точности на полном обучении и успешно обобщает знания в zero-shot сценарии. Хотя метрики классификации при отсутствии обучающих примеров снижаются, анализ эмбеддингов с помощью t-SNE показывает, что модель продолжает формировать осмысленные кластеры отношений, сохраняя структурированную организацию признакового пространства.

Практическая ценность работы заключается в разработке метода идентификации отношений между сущностями без необходимости аннотирования всех возможных классов, что делает предложенный подход полезным при обработке недостаточно размеченных текстов, автоматизированном анализе информации и работе с языками, имеющими ограниченные ресурсы.