

Рецензия на статью

Табаченкова Андрея

«Оценка сложности отдельных объектов в рамках
вложенных моделей данных текстовой разметки»

«ASSESSING THE DIFFICULTY OF SPANS
WITHIN NESTED TEXT MARKUP DATA MODELS»

Статья посвящена разработке методов оценки сложности текстовых фрагментов для задач цифровых гуманитарных наук. Предлагается подход, сочетающий обработку много-пользовательской разметки, векторизацию фрагментов и применение различных методов оценки сложности. В работе систематизированы подходы к оценке сложности объектов, включая метод Гауссиана, метод отступов и метод дисперсии градиентов.

Эксперименты демонстрируют эффективность методов на данных Kaggle NER Corpus и ADE, с выделением метода Гауссиана как наиболее точного по метрике ROC-AUC. Основными достоинствами работы являются системный подход к решению задачи, глубокая проработка адаптации методов к специфике текстовых данных и убедительная экспериментальная проверка.

В качестве ограничения можно отметить тестирование на ограниченном числе наборов данных и влияние переобучения модели на результаты оценки сложности. Рекомендуется расширить валидацию методов на более разнообразных наборах данных, а также углубить анализ интерпретируемости результатов для лучшего разделения объективной сложности фрагментов и артефактов обучения модели.

Рецензент,
Овсиенко Олеся,
Студентка ВМК МГУ