

**РАЗРАБОТКА МЕТОДИКИ ОЦЕНКИ КАЧЕСТВА  
АБСТРАКТИВНОЙ СУММАРИЗАЦИИ НАУЧНЫХ СТАТЕЙ SMOSA  
(SUMMARIZATION METRIC OF SCIENTIFIC ARTICLE)**

**Горобец И.С. (СПБГУТ), Громов Р.Д. (СПБГУТ)**

**Научный руководитель – доктор технических наук, декан Зикратов И.А.  
(СПБГУТ)**

**Введение.** Абстрактная суммаризация - один из мощнейших инструментов автоматизированной обработки текстовой информации, представляющий собой генерацию краткого описания исходного текста. Однако разработка и совершенствование моделей абстрактной суммаризации требует наличия надежных и объективных метрик для оценки их качества. Существующие методы оценки такие как ROUGE [1], TER [2], chrF++ [3] не применимы в большинстве реальных задач, так как завязаны на сравнении с эталоном. С другой стороны безреференсные метрики, основанные на нейросетевых подходах, такие как MaTese [4], MS-COMET [5] и REUSE [6], порождают проблему интерпретируемости (объяснимости) результатов. В рамках данной статьи предлагается более интерпретируемая и универсальная метрика, способная оценивать качество абстрактной суммаризации.

**Основная часть.** Для решения проблемы интерпретируемости предлагается разбить единый оценочный механизм на отдельные компоненты, каждый из которых отражает определенный аспект качества текста. В контексте оценки качества сформированного краткого описания, можно выделить два основных параметра: читабельность текста и соответствие фактов.

Для оценки читабельности был собран и размечен датасет содержащий аннотации научных статей и оценок их читабельности, складывающихся из ответов эксперта на вопросы об этом тексте. На эмбедингах аннотаций была обучена модель SGDR regressor [7], предсказывающая оценку, выставленную экспертом.

Для оценки соответствия фактов между исходным текстом и сгенерированным кратким описанием разработан алгоритм, основанный на использовании векторных представлений предложений. В ходе исследования было протестировано несколько критериев отбора наиболее значимых предложений из исходного текста и выбран лучший. Также был разработан алгоритм, сопоставляющий отобранные предложения, предложениям сгенерированного краткого описания.

В основу предлагаемой метрики оценки качества абстрактной суммаризации легли два ключевых компонента: модель оценки читабельности текста и алгоритм оценки фактической согласованности. Метрика представляет собой значение от 0 до 1, полученное путем усреднения нормализованных компонентов, отражающее качество абстрактной суммаризации.

**Выводы.** Для анализа полученной метрики, она была применена к выборке из 1000 примеров датасета “neuralwork/archiver”. Среднее значение составило 0,905. Данный результат свидетельствует о валидности предложенной метрики: полученное значение достаточно высоко, что указывает на хорошее соответствие краткого описания исходному тексту, однако не достигает максимума, что отражает наличие недостатков в ручных аннотациях.

Также в отличие от других способов оценки качества абстрактной суммаризации предложенная метрика позволяет явно указывать на недостатки в сгенерированном кратком описании: отсутствие каких-либо значимых фактов и качество написания текста. Данная особенность является преимуществом в случаях внедрения в реальные системы - добавление агента, который будет исправлять недостатки, выявленные метрикой, может заметно повысить качество абстрактной суммаризации.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1. Lin C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries // Text Summarization Branches Out. Barcelona, Spain: Association for Computational Linguistics, 2004. P. 74–81.
2. Snover M. et al. A Study of Translation Edit Rate with Targeted Human Annotation // Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers. Cambridge, Massachusetts, USA: Association for Machine Translation in the Americas, 2006. P. 223–231.
3. Popović M. chrF++: words helping character n-grams // Proceedings of the Second Conference on Machine Translation. Copenhagen, Denmark: Association for Computational Linguistics, 2017. P. 612–618.
4. Perrella S. et al. MaTESe: Machine Translation Evaluation as a Sequence Tagging Problem // Proceedings of the Seventh Conference on Machine Translation (WMT) / ed. Koehn P. et al. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, 2022. P. 569–577.
5. Kocmi T., Matsushita H., Federmann C. MS-COMET: More and Better Human Judgements Improve Metric Performance.
6. Ananya Mukherjee, Shrivastava M. REUSE: REference-free UnSupervised quality Estimation Metric. Unpublished, 2022.
7. SGDRegressor [Electronic resource] // scikit-learn. URL: [https://scikit-learn/stable/modules/generated/sklearn.linear\\_model.SGDRegressor.html](https://scikit-learn/stable/modules/generated/sklearn.linear_model.SGDRegressor.html) (accessed: 07.12.2024).