Максим Анатольевич Сторчевой

Национальный исследовательский университет «Высшая школа экономики» (Санкт-Петербург)

[mstorchevoy@hse.ru](mailto:mstorchevoy@hse.ru)

**Использование моделей ИИ для расчета рейтинга ESG**

Целью исследования является разработка методы машинного обучения для обработки открытой текстовой информации о российских компаниях и составления рейтинга ESG. Для проведения экспериментов были использованы нефинансовые отчеты, новости и веб-сайты ведущих российских компаний за период 2019–2022 годы. Разработаны два подхода к составлению рейтинга ESG на основе ИИ: 1) многокритериальное ранжирование TOPSIS на основе tf-idf для топиков ESG в отчетах компаний, 2) обучение моделей ИИ для предсказания рейтингов по методике внешних рейтеров RAEX, РСПП и НКР.

Ключевые слова: ESG, КСО, этика, рейтинг, ИИ.

Исследование выполнено в рамках Программы фундаментальных исследований НИУ ВШЭ.

Рейтинг ESG позволяет оценить качество управления компанией в области экологии, социальным обязательствам и принципам корпоративного управления. Идея такой оценки для инвесторов была предложена еще в 2004 г. рабочей группой ООН и затем реализована множеством независимых рейтинговых агентств: Bloomberg, S&P Dow Jones Indices, MSCI ESG Rating, CDP, Sustainalytics, Refinitiv, China Securities Index, Sino Securities Index, Risk Insights и др. Большинство из них использует экспертную оценку, но некоторые задействуют технологии ИИ (например, Clarity AI, RiskInsights). В России существует несколько агентств и общественных организаций, которые занимаются составлением рейтингов ESG ручным методом: RAEX, РСПП, АКРА, НКР и др., но технологии искусственного интеллекта пока не использует никто.

Методика расчета рейтинга ESG предполагает разработку структуры различных индикаторов по разделам E (экология), S (социальная политика) и G (корпоративное управление), по которым выставляют баллы. Итоговая оценка по каждой переменной рассчитывается путем усреднения полученных баллов (иногда с использованием весов). Использование экспертов при выставлении баллов по каждому индикатору позволяет с одной стороны добиться высокой степени надежности оценки, но с другой стороны сильно удорожает процесс оценки. Разработка компьютерных алгоритмов, которые позволят рассчитывать ESG-рейтинг для компаний на основе лингвистического анализа, снизит необходимость привлечения экспертов.

В зарубежной литературы уже есть примеры исследований, изучающих ESG на основе NLP [1; 2; 3; 4; 5].

В нашем исследовании мы использовали следующую последовательность шагов для построения рейтинга.

1. Выделение топиков ESG. Для разработки структуры топиков ESG мы вначале воспользовались автоматическим моделированием топиков (например, Top2Vec) – было выделено около 300 топиков, из которых экспертным образом мы оставили только 45 топиков, которые хорошо покрывали все области ESG, которые можно обнаружить в отчетах (см. рис. 1).

Изображение выглядит как текст, снимок экрана, меню, Шрифт

Автоматически созданное описание

Рис. 1. Топики ESG.

2. Матрица признаков. Далее методами машинного обучения была подготовлена матрица признаков, которая определила коэффициенты для каждого слова, показывающие вероятность его попадания в том или ином топике (см. рис. 2). С помощью данной матрицы признаков мы можем с уверенностью приписывать топики ESG различным фрагментам текста из нефинансовых отчетов, новостей или веб-сайтов.

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рис. 2. Матрица признаков.

3. Рейтинг через ранжирование TOPSIS. Затем мы попробовали ранжировать отчеты по степени, в которой они более полно описывают свою деятельность в области ESG, т. е. насколько все топики полно представлены в ее отчете. Для этого мы рассчитываем метрику TF-IDF для отдельного топика, подсчитывая его относительную редкость для всей базы отчетов и частоту вхождения в отчет отдельной компании, и далее применяем многокритериальное ранжирование TOPSIS для упорядочения компаний по данным 45 метрикам. В результате мы сформировали новый рейтинг ESG, который при сравнении с другими рейтингами ESG показывает похожую динамику (см. рис. 3).

Изображение выглядит как текст, диаграмма, линия, График

Автоматически созданное описание

Рис. 3. Сравнение динамики рейтинга RAEX и рейтинга на основе TOPSIS.

4. Предсказание эталонных рейтингов. Мы также попробовали обучить модели ИИ предсказывать рейтинги ESG от RAEX, НКР и РСПП на основе анализа отчетов, веб-сайтов и новостей. По данным текстам для каждой компании также вначале выделяются топики ESG и затем модель учится предсказывать эталонные рейтинги ESG на основе полученных датасетов. В результате у нас получились следующие показатели метрики ROC-AUC по каждому источнику данных и по каждому рейтеру (см. рис. 4).

Изображение выглядит как текст, число, снимок экрана, Шрифт

Автоматически созданное описание

Рис. 4. Точность ROU-AUC для различных типов данных и рейтеров.

Результатом данной методики будет новый рейтинг ESG на основе искусственного интеллекта, который обобщает методы оценки разных российских рейтеров.

Литература

1. Aiba Y., Ito T., Ibe Y. Network Structure in ESG Ratings Suggests New Corporate Strategies: Evolving AI Technology to Quantify Qualitative Data. Security Analysts Journal. 2020. vol. 16. pp. 3–15.

2. Cheng T.–T., Tsai Y.–H., Lai Ch., Hwang S.–Y. Mining Impacts of CSR Disclosure on Firm Performance. PACIS 2023 Proceedings. 2023. pp. 188.

3. Marcelo G.–B., Espinosa-Leal L. Natural Language Processing Methods for Scoring Sustainability Reports –A Study of Nordic Listed Companies. Sustainability. 2022. vol. 14, no. 15. pp. 9165.

4. Takuya K., Nozaki M. A Text Mining Model to Evaluate Firms’ ESG Activities: An Application for Japanese Firms. Asia-Pacific Financial Markets. 2020. vol. 27, no. 4. pp. 621–632.

5. Tremblay M. Ch., Parra C., Castellanos A. Analyzing Corporate Social Responsibility Reports Using Unsupervised and Supervised Text Data Mining. Lecture Notes in Computer Science. 2015. vol. 9073. pp. 439–46.