

Метод интеграции структурированных и неструктурированных данных, текстовых массивов для решения задач прогнозирования рынка с учетом оценки перспектив научно-технологического развития

Авторы: к.ф.-м.н. Д.А. Девяткин, к.ф.-м.н. И.В. Соченков, В.В. Жебель

Предложены подходы к совместному использованию структурированных баз производства и международной торговли продовольственной продукцией, материалов профильных СМИ, баз научно-технических документов для прогнозирования рынка с учетом перспектив научно-технологического развития в области производства, хранения и транспортировки продукции. Для анализа сообщений, связанных с состоянием рынка, совместно используются векторные представления текстов (эмбединги), полученные с помощью мультязычных сетей-Трансформеров, и показатели для оценки уровня эмоционального напряжения и тональности этих текстов. Для автоматизированного выявления перспективных технологий переработки, транспортировки, хранения осуществляется анализ массива патентов (ФИПС и USPTO) с помощью методов тематического моделирования и методов кросс-языкового поиска тематически-близких документов. Перечисленные методы используются для оценки значений качественных индикаторов, отражающих состояние рынка, наличие или отсутствие угрозы санкционных ограничений, выявления перспектив внедрения новых технологий, способствующих повышению уровня производства и снижению издержек.

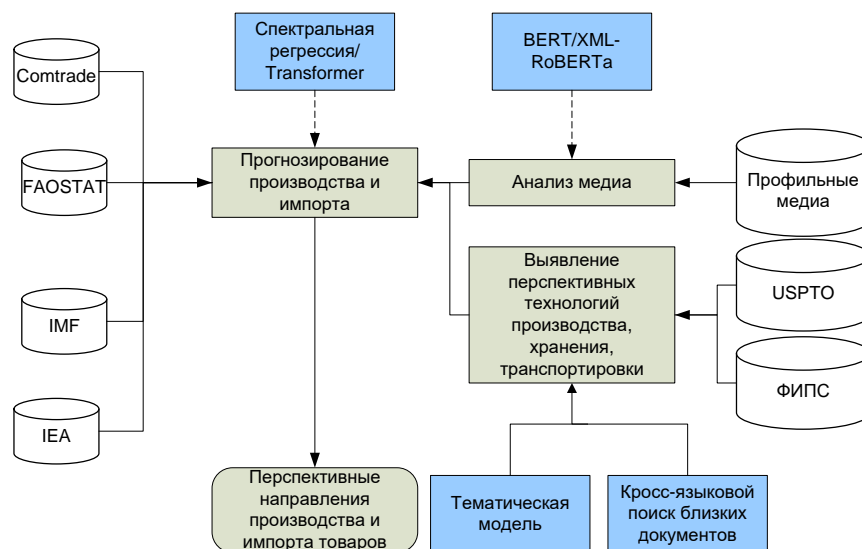


Рис. Схема метода интеграции структурированных и неструктурированных данных, текстовых массивов для решения задач прогнозирования рынка

С помощью модели спектральной регрессии на основе нейронной сети с архитектурой «Трансформер» и с учетом значений сформированных качественных индикаторов осуществляется прогнозирование торговых потоков и цен на продовольственном рынке. Сеть-трансформер позволяет учитывать сложные взаимосвязи между рассматриваемыми индикаторами и прогнозируемыми величинами. В отличие от предложенных ранее подходов на основе рекуррентных сетей предлагаемый метод лишен недостатков, связанных с «размыванием градиента», что позволяет увеличить точность долгосрочных прогнозов.

Публикации:

1. Отмахова Ю. С., Девяткин Д. А. Моделирование агропродовольственных цепей поставок с применением машинного обучения и агент-ориентированных моделей // Информационное общество, 2024. №. 4. С. 21-32.
2. Zhebel, V. V., Devyatkin, D. A., Zubarev, D. V., Sochenkov, I. V. Approaches to Cross-Language Retrieval of Similar Legal Documents Based on Machine Learning // Scientific and Technical Information Processing, 2023. Vol. 50. Iss. 5. P. 494-499.