

Информационно-аналитическая система на основе гибридных моделей прогнозирования показателей социальной сферы РФ

О. В. Китова¹, В. М. Савинова², Л. П. Дьяконова³, А. А. Бесхмельницкий⁴

ФГБОУ ВО «РЭУ им. Г. В. Плеханова»

¹olga.kitova@mail.ru, ²Lesnayapol@yandex.ru, ³ldyak@mail.ru, ⁴phenfeur@gmail.com

Аннотация. В статье рассматривается информационно-аналитическая система, в основе которой лежит гибридная модель прогнозирования. Данная модель включает в себя эконометрические и интеллектуальные методы построения прогнозов.

Ключевые слова: гибридные модели; краткосрочное прогнозирование; искусственные нейронные сети; ANFIS; эконометрические модели; временные ряды

I. ВВЕДЕНИЕ

Современное развитие рыночных отношений, технологий и интеллектуального капитала обусловило необходимость перехода к новому этапу развития экономики – цифровой экономике. Была принята программа «Цифровая экономика Российской Федерации», которая включает в себя основные пути развития государства как в сфере бизнеса, так и социальной сфере, системе государственного управления и общества в целом. В связи с принятием данной программы, а также закона «О стратегическом планировании в РФ» 172-ФЗ от 28.06.2014 появляется необходимость создания новых экономических моделей, которые позволяли бы оперативно принимать решения в условиях неопределенности как на уровне отдельной компании, так и на уровне государства и его регионов.

Одной из наиболее приоритетных задач государства является мониторинг социально-экономических показателей на федеральном и региональном уровнях. Для ее решения создаются ситуационные центры, которые включают в себя оборудование и аналитическое программное обеспечение, необходимое для сбора, анализа и визуализации данных по исследуемым показателям. Создание федеральных и региональных ситуационных центров регламентируется указом Президента Российской Федерации от 25 июля 2013 г. N 648 "О формировании системы распределенных ситуационных центров, работающих по единому регламенту взаимодействия". В условиях цифровой экономики существует необходимость развития моделей и инструментария для сбора, анализа, прогнозирования дан-

ных, которые позволяли бы рассчитывать значения исследуемых показателей с приемлемой точностью, качеством и затратами времени. Это позволит принимать решения и корректировать стратегию как на уровне государства в целом, так и отдельных его регионов (субъектов).

В рамках данной работы исследуются гибридные модели прогнозирования социальной сферы РФ, а также их программная реализация. Существует множество работ российских и зарубежных ученых, посвященных построению гибридных моделей прогнозирования социально-экономических показателей. Например, работы Аверкина А.Н., Ярушева С.А. Демидова Л.А., Пылькина А.Н., Скворцова С.В. и Скворцовой Т.С. [1–4]. В задачи разрабатываемой системы входит автоматизированный подбор моделей для прогнозирования временных рядов на основе предварительной оценки качества исследуемых данных.

II. МЕТОДИКА

В рамках данной работы описывается информационно-аналитическая система, развиваемая на базе гибридных моделей, разработанных коллективом авторов кафедры информатики РЭУ им. Г.В. Плеханова [5]. В качестве исходных данных выступают данные официального сайта Федеральной службы государственной статистики [6] для показателей социально-экономического развития РФ за период с 2000 года по 2016. Создаваемая система должна осуществлять следующие функции:

- загрузку данных (значений временных рядов) в систему;
- процесс очистки, трансформации данных;
- оценку качества полученных данных (расчет ключевых показателей качества) и поиск зависимостей (оценка корреляции);
- процесс подбора моделей прогнозирования;
- разбиение исходных данных на тестовую и обучающую выборки;
- настройку методов обучения;
- процесс обучения выбранной модели (вывод результатов расчета, значений ошибок и т. д.);

Данное исследование выполнено в рамках выполнения научно-исследовательской работы базовой части государственного задания в сфере научной деятельности Министерства образования и науки РФ на тему «Интеллектуальный анализ текстовых данных большого объема в финансах, бизнесе и образовании на основе адаптивных семантических моделей», номер проекта 2.9577.2017/8.9.

- расчет прогнозных значений;
- верификацию моделей (расчет показателей точности и качества);
- визуализацию результатов (диаграммы, графики, olap-кубы и т.д.);
- построение отчетов (отчеты по всем построенным моделям);
- выгрузку данных в удобный формат.

Загрузка данных в модель может осуществляться как из локальных файлов (в формате excel, txt, csv), так и путем выгрузки из различных источников посредством применения API. Данные представлены в системе в форме таблицы.

Оценка качества данных производится путем расчета коэффициентов описательной статистики, поиска выбросов и аномальных значений, выявления пропущенных значений, расчета коэффициентов корреляций. В результате система выдает рекомендации по тому, каким образом лучше изменить данные.

Очистка данных позволяет проводить редактирование выбросов, сглаживание ряда, заполнение пропусков.

Процесс подбора модели прогнозирования реализуется посредством модуля NFIS (adaptive neuro-fuzzy inference system), созданного авторами на языке C#. Модель ANFIS, или Адаптивная нейро-нечеткая система вывода, представляет собой гибрид на основе систем нечеткого вывода и искусственных нейронных сетей. В модуле NFIS на основе оценки качества данных (показатели качества входных данных), количества данных, периодичности (годы, кварталы, месяцы), наличия корреляции с другими временными рядами подбирается наиболее релевантная модель прогнозирования для имеющихся рядов. В то же время у эксперта остается возможность выбрать модель вручную, не опираясь на рекомендации модуля.

Тестовая и обучающая выборки настраиваются пользователем (пропорция разделения, отбор примеров случайно или последовательно и т.д.)

После выбора модели осуществляется выбор метода обучения и настройка его параметров.

Далее производится непосредственное обучение модели, результаты которого выводятся с помощью выбранных пользователем визуализаторов (таблица, график, диаграмма). Пример построения графика приведен на рис. 1.

Расчет прогноза подразумевает выбор пользователем периода прогнозирования и расчет прогнозных значений.

После всех приведенных этапов производится верификация, визуализация и создание отчетов о построенных моделях.



Рис. 1. Визуализатор: график сравнения фактических данных и результата расчета модели

В настоящий момент в систему включены модели линейной множественной регрессии, нейронные сети (персептроны и рекуррентные сети) и генетические алгоритмы. В дальнейшем планируется расширить этот список моделей с целью обеспечения качественными прогнозами всей совокупности прогнозируемых показателей.

Разрабатываемая система позволяет выполнять расчеты моделей как для одного показателя, так и для системы взаимосвязанных показателей. Созданные модели хранятся в базе данных моделей и могут быть вызваны пользователем в необходимое время.

Система позволяет выполнять детализацию и агрегирование данных, что дает возможность осуществлять расчет прогнозов показателей в режиме лет, кварталов, месяцев и проводить их сравнение.

III. РЕЗУЛЬТАТЫ

На рис. 2 представлены возможности системы, а также взаимодействие ее компонентов с внешними агентами (диаграмма «use-case»).

Разрабатываемая система будет иметь возможность интеграции с Ситуационным центром РЭУ. Разработанная архитектура будет предоставлять возможность работы с любого устройства, поддерживающего современные браузеры. Серверная часть будет иметь возможность запуска под любой современной операционной системой, поддерживающей приложения DotNet (.Core). В качестве СУБД решения может использоваться PostgreSQL или MS SQL Server (в зависимости от возможностей и предпочтений). В архитектуру разрабатываемого продукта изначально заложена возможность горизонтального и вертикального масштабирования и поддержка распределения нагрузки. Поддерживаемая нагрузка без потери производительности на одном экземпляре сервера (нода кластера) ожидается не менее 1000 одновременно работающих пользователей. Для сложных операций вычисления предусматривается очередь выполнения и выделение ресурсов (квот) для пользователей с расчетом предполагаемого затрачиваемого времени исходя из ресурсов, располагаемых системой.

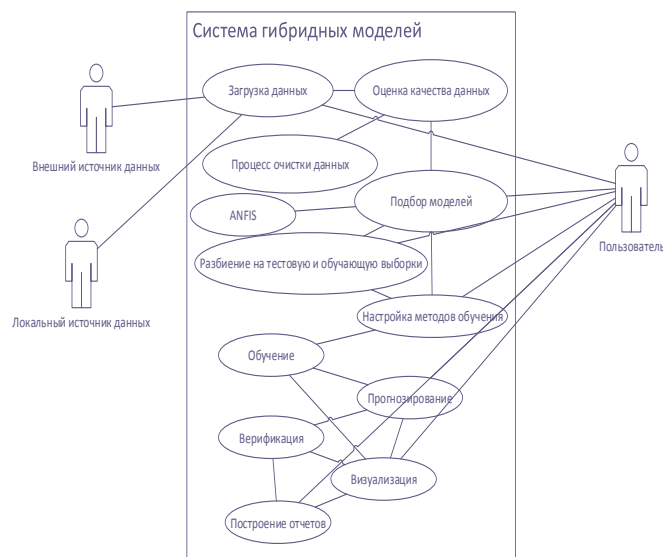


Рис. 2. Диаграмма прецедентов в нотации UML разрабатываемой системы

Компоненты системы представлены на рис. 3.

В настоящий момент уже разработано 3 компонента системы:

- модуль загрузки данных;
- модуль моделей множественной линейной регрессии;
- модуль искусственных нейронных сетей (персептрон).

В ближайшее время планируется завершение разработки модуля регрессионных деревьев решений (CART).

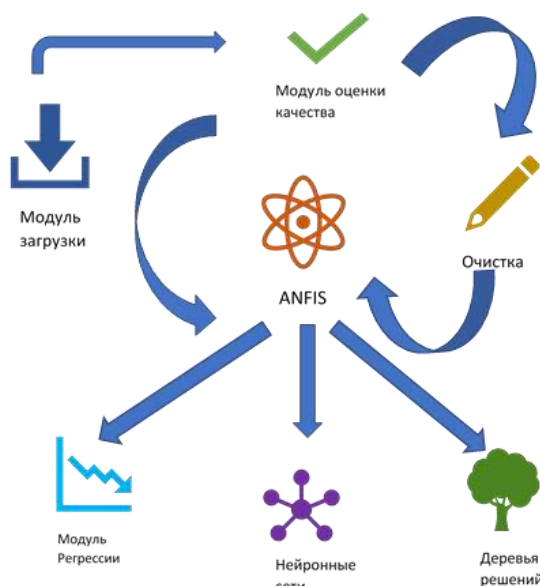


Рис. 3. Компоненты гибридной системы прогнозирования

IV. ОБСУЖДЕНИЕ

В РЭУ им. Г.В. Плеханова успешно работает Ситуационный центр социально-экономического развития регионов Российской Федерации. Основными целями центра являются:

- развитие фундаментальных исследований в области социально-экономического развития РФ и ее регионов;
- предоставление актуальной статистической информации по экономическому развитию как регионов, так и страны в целом;
- выполнение экспертно-аналитических и консалтинговых работ по актуальным направлениям социально-экономического развития России и субъектов российской федерации.

В настоящее время базовой информационно-аналитической системой ситуационного центра является система Contour BI. Данная система позволяет осуществлять оперативный анализ данных (OLAP), имеет инструменты графической визуализации, а также позволяет строить отчетность. Система Ситуационного центра взаимодействует с базами данных Федеральной службы государственной статистики, наполняя внутреннее хранилище данных актуальными значениями исследуемых показателей. Однако в системе Ситуационного центра отсутствует аналитический модуль, который позволял бы осуществлять прогнозирование исследуемых показателей с использованием как классических эконометрических методов, так и методов искусственного интеллекта.

Разрабатываемая информационно-аналитическая система на основе гибридных моделей будет иметь возможность интеграции с системой Ситуационного центра РЭУ посредством API, что позволит получать из системы Contour BI необходимые данные, и затем посредством применения системы гибридных моделей осуществлять прогнозирование необходимых показателей. Интеграция с системой Ситуационного центра позволит экспертам, специалистам и учащимся РЭУ им. Г.В. Плеханова не только получать из достоверных источников необходимую информацию и строить OLAP-кубы, но и проводить прогнозирование в режиме «Что-Если» с возможностью доступа к рассчитанным прогнозам и построенным моделям.

V. ВЫВОДЫ

В настоящее время на кафедре информатики РЭУ им. Г.В. Плеханова разрабатывается информационно-аналитическая система на основе гибридных моделей прогнозирования, которая позволит загружать, обрабатывать данные, рассчитывать показатели их качества, а также осуществлять построение прогнозов временных рядов.

Благодаря модулю NFIS разработанная система позволит не только строить качественные и точные прогнозы, но и позволит подбирать наиболее релевантную модель в зависимости от параметров временного ряда. В дальнейшем для оценки влияния новостного и политического фона на исследуемые показатели планируется внедрение технологии Web Mining.

СПИСОК ЛИТЕРАТУРЫ

- [1] Аверкин А.Н., Прокопчина С.В. Мягкие вычисления и измерения // Интеллектуальные системы (МГУ). 1997. Т. 2. Вып. 1–4. С. 93–114.
- [2] Sergey A. Yarushev, Alexey N. Averkin, Time Series Analysis Based on Modular Architectures of Neural Networks, *Procedia Computer Science*, Volume 123, 2018, Pages 562-567
- [3] Averkin A. N., Yarushev S. Hybrid approach for time series forecasting based on ANFIS and Fuzzy Cognitive Maps //Soft Computing and Measurements (SCM), 2017 XX IEEE International Conference on. IEEE, 2017. Pp. 379-381.
- [4] Демидова Л.А., Пылькин А.Н., Сковцов С.В., Сковцова Т.С. Гибридные модели прогнозирования коротких временных рядов. М: Горячая линия – Телеом, 2015. 206 с.
- [5] Danko T.P., Kitova O.V., Kolmakov I.B., Dakonova L.P., Grishina O.A., Sekerin V.D. Hybrid intelligent system of forecasting of the socio-economic development of the country/ *ijaber*, Vol. 14, No.9. 2016. P. 5755-5766.
- [6] Официальный сайт Федеральной службы государственной статистики [Электронный ресурс]. Федеральная служба государственной статистики, 1999–2016. URL: www.gks.ru (дата обращения: 20.04.2016).