



дуг переходов из одного состояния в другое, а также хранение автора и времени изменения позволит увеличить эффективность разработки проекта.

Создание ветвей (branches) проекта позволит создавать различные версии проекта и на этапе слияния в главный проект определять актуальность и правильность каждой ветви. Однако при слиянии возможно возникновение конфликтов, которые будут решаться разработчиками в режиме online.

Таким образом подсистема взаимодействия с пользователем будет оповещать разработчика о конфликтах слияния. Возможность поэтапного хранения состояния графа и просмотра предыдущих версий может быть удобна в процессе обучения студентов для пошагового объяснения алгоритма создания проекта. Все эти меры позволят обеспечить совместную работу над часто меняющейся информацией, идентифицировать автора изменений, хранить полную историю изменения данных, возможность отката к любому моменту изменений данных, реализовать альтернативные, экспериментальные варианты проекта, оповестить других пользователей о произошедших изменениях, а также обеспечить высокое быстродействие при выполнении операция сохранения, получения, слияния версий.

Литература

1 Коварцев А.Н. Методы и средства визуального параллельного программирования. Автоматизация программирования.[Текст]/ А.Н. Коварцев, В.В. Жидченко : учебное пособие. – Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2011. – 168 с.

2 Системы контроля версий [Электронный ресурс]. – URL: <https://hgbook.bacher09.org/html/how-did-we-get-here.html> (дата обращения: 10.02.2017).

К.С. Пивкин

СОЗДАНИЕ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ С ПОМОЩЬЮ ЯЗЫКА R НА ПРИМЕРЕ РОЗНИЧНОГО ПРЕДПРИЯТИЯ

(Удмуртский государственный университет)

Во многих сферах жизнедеятельности предсказание определенной целевой величины стало основой эффективных решений. В частности, в сферах бизнеса данными величинами являются основные показатели процессов, протекающих в них: например, в банковском деле – это вероятность заемщика не выплатить по своим обязательствам, в онлайн-бизнесе – вероятность отклика пользователя, в розничной торговле – величина покупательского спроса на тот или иной товар и т.п. Подобная востребованность в оценке будущих значений ключевых показателей породило массу новых решений и платформ, с помощью которых реализуются современные прогнозные системы [2]. Целью данной работы является описание современной системы прогнозирования, разработанной



с помощью языка программирования R, в приложении к торгово-розничному делу.

Язык R – мультипарадигмальный язык программирования для статистической обработки данных и визуализации. Язык R активно используется для исследований в академической среде и в последнее время получает все большую распространенность для решения бизнес-задач. Особенностью языка является удобство написания высокоуровневых решений на его основе. При этом освоение языка довольно быстро осуществляется и непродвинутым пользователем. Мощными инструментами для разработки приложений также обладают специальные функциональные расширения для языка R – пакеты, количество которых растет с каждым годом активного использования языка [3].

Система прогнозирования – это прежде всего система обработки и анализа данных, на выходе которой исследователь имеет результат в виде модели прогнозирования и предсказания целевой переменной. Подобный класс задач решает такая область методов как Data Mining [1]. В данном контексте разработанная система прогнозирования имеет следующие этапы:

1. Постановка задачи прогнозирования;
2. Сбор данных;
3. Подготовка (предобработка) данных;
4. Выбор подходящего типа модели прогнозирования (линейная регрессия, деревья решений, нейронные сети или т.п.);
5. Подбор метапараметров модели и алгоритма обучения;
6. Обучение модели на данных;
7. Анализ качества модели;
8. Вывод результата (предсказания) в систему поддержки принятия решений на предприятии.

Дополнительное введение последнего пункта связано с тем, что система прогнозирования является частью общей системы управления предприятием.

С точки зрения инструментов разработанного решения, архитектура системы прогнозирования выглядит следующим образом:

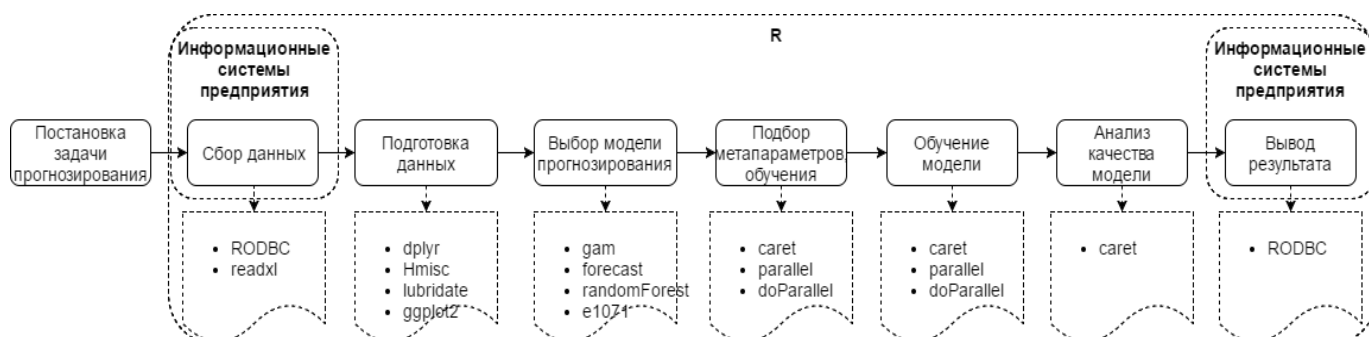


Рисунок 1 – Архитектура системы прогнозирования на языке R
(с указанием доп. пакетов)



Как видно, возможности языка R охватывают все этапы работы системы прогнозирования. При этом, при подготовке решения не обошлось без использования пакетных расширений, широко распространённых среди разработчиков в среде R. Условно их можно разделить на 4 группы:

- Пакеты, которые осуществляют функции интеграции для сбора данных: *RODBC* (для интеграции с SQL-базами данных), *readxl* (для сбора данных с внешних Excel-файлов).
- Пакеты, которые осуществляют функции обработки и подготовки данных к моделированию: *dplyr* (облегчает фильтрацию, агрегацию и иные манипуляции с данными), *lubridate* (облегчает работу с временными данными), *Hmisc* (обработка пропущенных значений), *ggplot2* (позволяет визуализировать необходимые связи в данных).
- Пакеты, которые отвечают за создание и оценку качества моделирования: *gam* (обобщенные аддитивные модели), *forecast* (модели временных рядов), *randomForest* (реализация метода «случайный лес»), *e1071* (реализация нескольких методов, в частности машины опорных векторов), *caret* (отвечает за организацию моделирования и оценку качества результата);
- Пакеты, которые отвечают за повышение эффективности реализуемого программного комплекса: *parallel*, *doParallel* (позволяет ускорить процедуры за счет параллелизации процессов).

Использование пакетов для языка R позволило сделать решение всеобъемлющим и гибким с точки зрения настройки возможных результатов.

Особенностью данного решения является использование не одной модели прогнозирования, а композиции (ансамбля) нескольких методов. При этом, сам выбор методов подтверждается выводами статистического и визуального анализа, который совершается средствами R. Также проводится процедура разделения выборки на обучающую (*train*) и тестовую (*test*), процедура перекрестной проверки (или кросс-валидации) для подбора метапараметров моделей и само их обучение. Все это позволяет совершить в эффективной высокоуровневой форме пакет *caret*.

Выбранные методы объединяются в ансамбль следующим образом:



Рисунок 2 – Конструкция ансамбля методов



Инструментально данный подход также реализуется с помощью R: для ускорения процедуры создания нескольких моделей используются пакеты для параллелизации вычислений (*parallel*, *doParallel*), а для создания ансамбля – тот же пакет *caret* и его расширение *caretEnsemble*, в котором реализованы функции предназначенные для создания и оценки требуемых композиций.

Следует отметить, что реализация системы прогнозирования на языке R позволяет заменить уже существующую на розничном предприятии. В качестве оценки результата двух систем используется среднеквадратичная ошибка:

$$MSE = \frac{1}{n} \times \sum_i^n (y_i - \hat{y}_i)^2,$$

где y_i – фактические значения целевой переменной (покупательского спроса), \hat{y}_i – оценка прогнозной величины, n – количество элементов в выборке. Для справедливости результата MSE оценивалась строго на результатах тестовой (проверочной) выборки. По этому показателю качество системы прогнозирования на предприятии улучшилось более чем на 20% по сравнению с существующей системой.

Полученная система прогнозирования спроса функционально интегрируется с глобальной информационной системой предприятия. В онлайн-режиме и статических отчетах результат прогноза используется для заказа товара у поставщика, планирования показателей продаж (количественных и стоимостных), проработки маркетинговых стратегий и многих других стратегических мероприятий.

На данный момент, для итогового внедрения системы прогнозирования требуются следующие действия по созданию документальной базы:

- создание спецификации модели с описанием используемых данных и их источников;
- создание технологической карты описанной системы прогнозирования;
- дополнительная структуризация самого скрипта для упрощения восприятия.

После документального описания система будет внедрена в промышленную эксплуатацию на одном из предприятий розничной торговли, а также с учетом универсальности системы – может быть внедрена и на других.

Проделанная работа позволяет сделать выводы, что использование языка R не накладывает какие-либо ограничения на работу системы принятия решений на розничном предприятии. Более того, из-за высокоуровневой и богатой системы пакетных расширений данный язык программирования становится более эффективным инструментом для решения экономико-математических задач, чем существующие. Ввиду использования языка программирования в качестве основы решения, создаются модульные структуры (согласно архитектуры на рис. 1), которые просто интерпретировать и отлаживать в будущем. Отмечено, что интеграция с информационной системой предприятия легко настраивается при соблюдении формальных правил по написанию запросов для настройки «вход-выход». В ходе разработки также было обнаружено, что в случае воз-



никновения проблемы обработки «больших данных» для языка R предоставлен программный интерфейс для популярного фреймворка Apache Spark. Это позволит с минимальными трудозатратами скорректировать архитектуру существующего решения, в случае появления нового массива данных, который будет влиять на качество решения поставленной задачи.

Литература

1. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям (+CD) [Текст] / Н.Б. Паклин, В.И. Орешков. – СПб. : Питер, 2013. – 704 с.: ил.
2. Чучуева, Ирина. Сравнение программных продуктов для анализа данных: R, MATLAB, SciPy, MS Excel, SAS, SPSS, Stata [Электронный ресурс] / И. Чучуева. – Электрон. текстовые дан. – Москва: [б.и.], 2012. – Режим доступа: <http://www.mbureau.ru/blog/sravnienie-programmnyh-produktov-dlya-analiza-dannyh-r-matlab-scipy-ms-excel-sas-spss-stata>, свободный.
3. De Vries, Andrie. On the growth of CRAN packages. [Электронный ресурс] / A. de Vries. – Электрон. текстовые дан. – USA, 2016. – Режим доступа: <https://www.r-bloggers.com/on-the-growth-of-cran-packages/>, свободный.

В.С. Плахов, А.В. Благов

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ РАСЧЁТА И ПРОГНОЗИРОВАНИЯ ФИНАНСОВОГО СОСТОЯНИЯ УНИВЕРСИТЕТА

(Самарский университет)

Введение

Университет представляет собой сложную с финансовой точки зрения систему, бюджет которой складывается из государственных субсидий выделяемых Министерством образования и науки Российской Федерации согласно существующим правилам [1,2], а также платных услуг (образовательных, научных, прочих). В состав университета, как правило, входят институты либо факультеты, имеющие определенную финансовую и управленческую самостоятельность, в которых ведется подготовка обучающихся по различным специальностям. При этом существует большое количество данных (стоимость той или иной группы специальностей, повышающие коэффициенты, количество победителей олимпиад и т.д.) влияющих на бюджет института (факультета), да и вуза в целом.

Актуальной является задача разработки информационной системы, учитывающей все параметры, определяемые установленными правилами, производящей расчёт финансового состояния, а также выносящей прогнозы и сценарии принятия решения.

1. Исследование основных параметров влияющих на финансовое состояние вуза

Для определения параметров разрабатываемой информационной системы рассмотрим подробнее предметную область.