



АНАЛИЗ СРЕДСТВ ЯЗЫКА ПРОГРАММИРОВАНИЯ R

LANGUAGE FEATURES ANALYSIS OF THE R PROGRAMMING LANGUAGE

В. С. Мамедов

Донской государственной технической университет, Ростов-на-Дону, Российская Федерация

Relecor3@gmail.com

Проведен анализ средств языка программирования R, отмечены достоинства среды R. Представлена реализация нейронных сетей на языке R, описан алгоритм, реализующий параллелизацию вычислений. Проведено сравнение языка программирования R с другими языками программирования.

Ключевые слова: Язык R, параллелизация, нейронные сети.

Введение. Язык программирования R. R — язык, позволяющий пользователю использовать операторы циклов, чтобы последовательно анализировать несколько наборов данных. Также язык R предоставляет возможность объединить в отдельную программу различные статистические функции для проведения более сложного анализа.

R — язык со многими функциями для выполнения статистического анализа и графического отображения результатов, которые визуализируются в собственном окне и могут быть сохранены в различных форматах. Результаты статистического анализа могут быть отображены на экране, а промежуточные результаты (*P-values*, коэффициент регрессии и т. п.) могут быть сохранены в файле и использоваться для последующего анализа.

Rgui — это стандартный графический интерфейс, встроенный в R по умолчанию. Эта оболочка имеет вид командной строки в окне, называемом консолью. Командная строка работает по принципу «вопрос — ответ». Однако для записи сложного алгоритма команд в *Rgui* существует дополнительное скриптовое окно, где пишется программа (скрипт). Третьим элементом данной оболочки является графический модуль, который появляется при необходимости отображения графиков. На приведенном ниже рисунке продемонстрирована полная версия *Rgui*: консоль (слева), скриптовое окно и графический модуль (справа).

Rstudio — интегрированная среда разработки (*IDE*). В отличие от *Rgui*, у данной оболочки есть заранее разделенные области и дополнительные модули (например, история команд, рабочая область). По мнению некоторых пользователей, *Rstudio* имеет более удобный интерфейс, упрощающий работу с R. Ряд особенностей, таких как цветовая подсветка и автоматическое завершение кода, удобная навигация по скрипту и другие, делают *Rstudio* привлекательной не только для новичков, но и для опытных программистов.

V. S. Mamedov

Don State Technical University, Rostov-on-Don, Russian Federation

Relecor3@gmail.com

The article provides language features analysis of the R programming language, outlines the advantages of the R environment. The paper presents the implementation of neural networks in the R language, describes the algorithm executing the parallelization of computations. The article includes the comparison of the R programming language with other programming languages.

Keywords: the R language, parallelization, neural networks

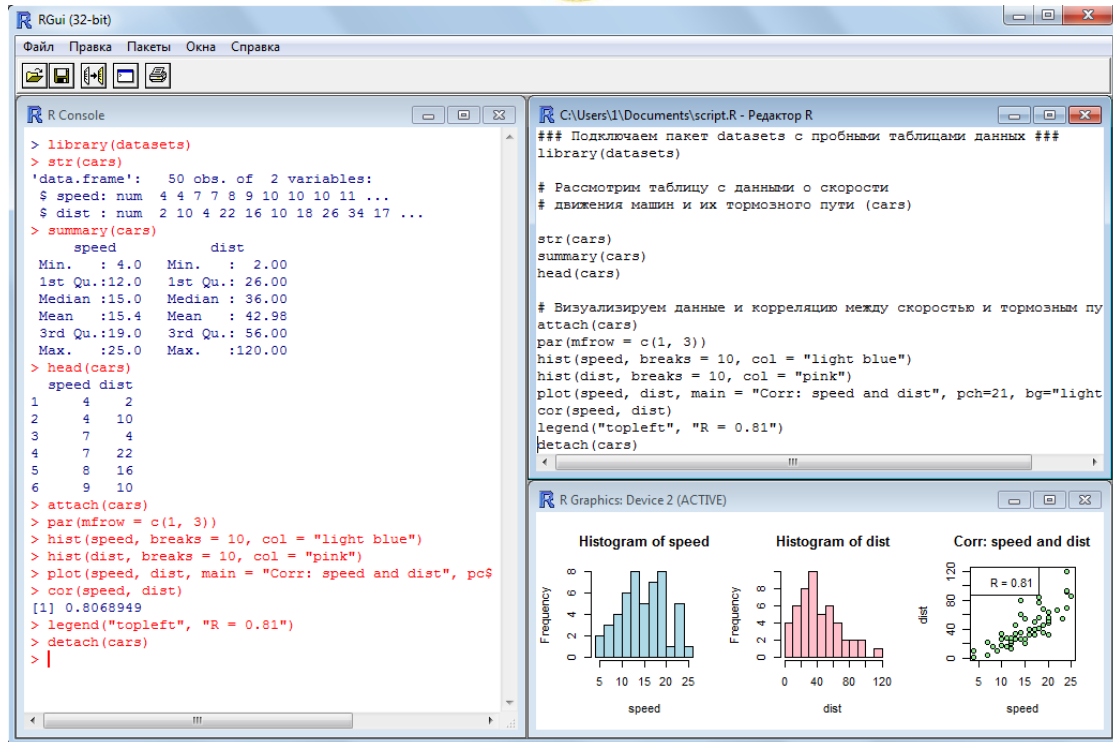


Рис. 1. Среда разработки Rgui

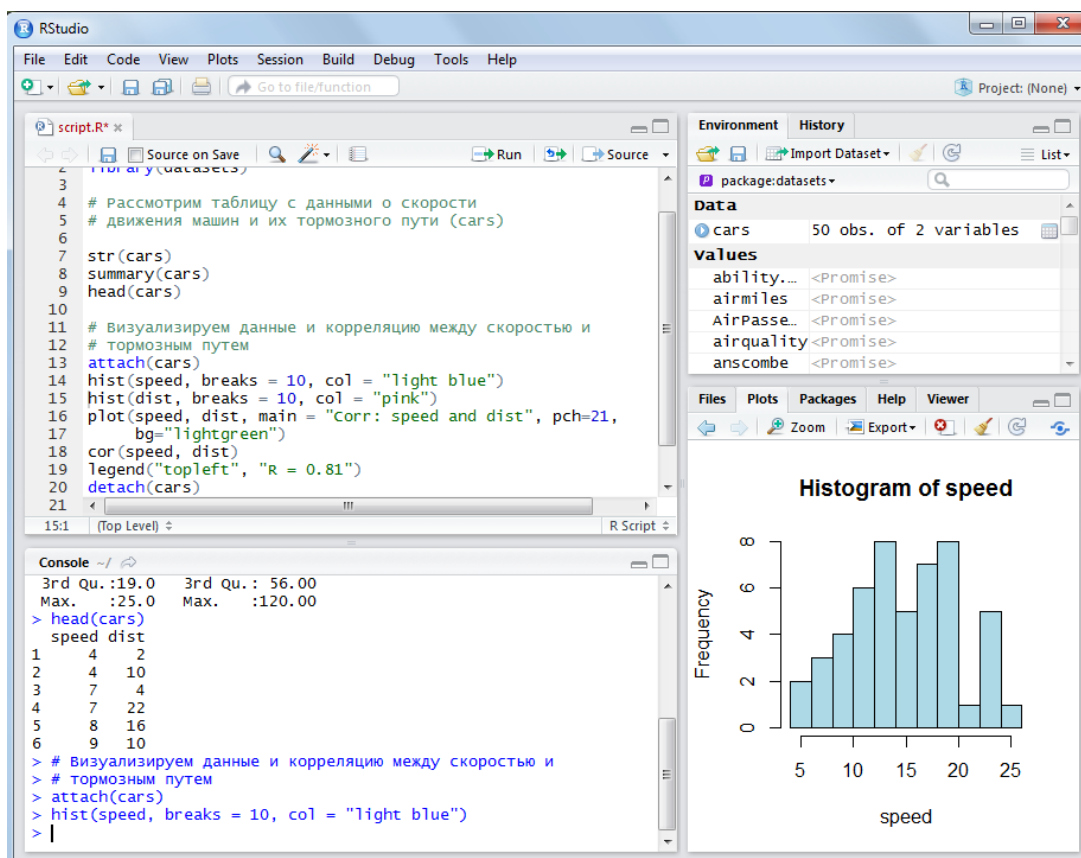
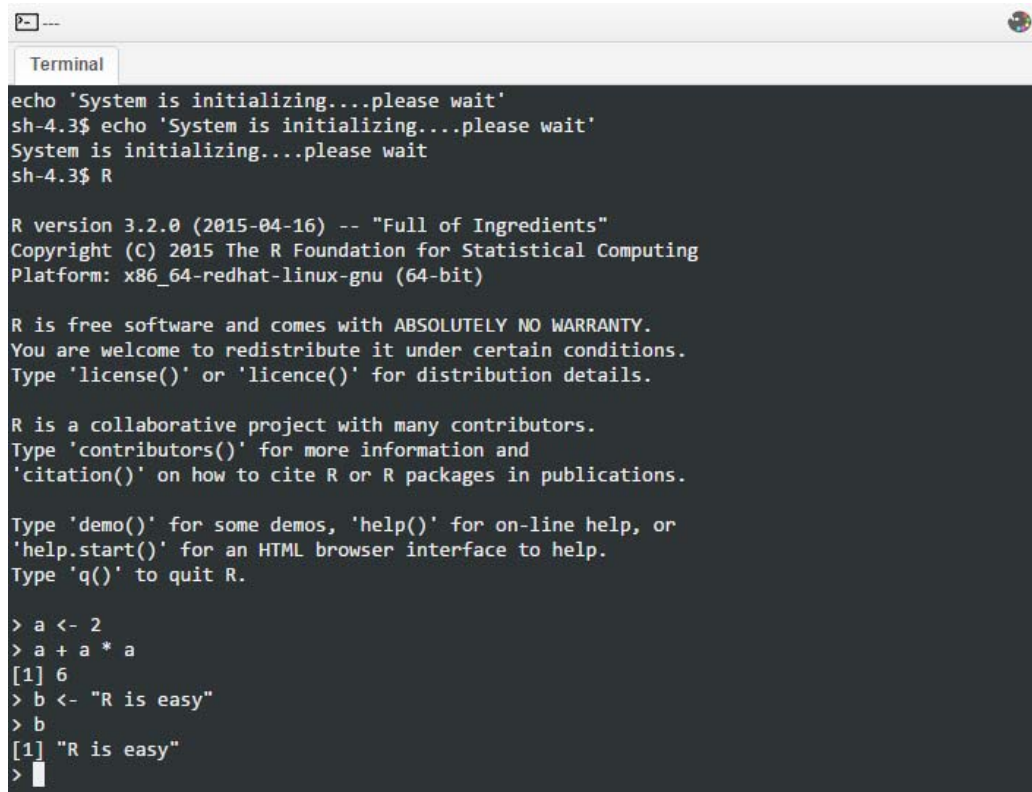


Рис. 2. Среда разработки Rstudio

R в терминале *Linux/UNIX*. Данный вариант предпочтителен для анализа большого объема данных через сервер, суперкластер или суперкомпьютер. Большинство из них работают на операционных системах класса *Linux/UNIX*, доступ к которым осуществляется через терминал команд.



```
Terminal
echo 'System is initializing...please wait'
sh-4.3$ echo 'System is initializing...please wait'
System is initializing...please wait
sh-4.3$ R

R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-redhat-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> a <- 2
> a + a * a
[1] 6
> b <- "R is easy"
> b
[1] "R is easy"
>
```

Рис. 3. Среда разработки в терминале *Linux*

Достоинства среды *R*:

- является бесплатной и кроссплатформенной;
 - имеет богатый арсенал статистических методов;
 - обладает качественной векторной графикой;
 - насчитывает более 7000 проверенных пакетов;
 - является гибкой в использовании:
- позволяет создавать и редактировать скрипты и пакеты,
 - взаимодействует с другими языками, например, с *C*, *Java* и *Python*,
 - может работать с форматами данных для *SAS*, *SPSS* и *STATA*;
 - объединяет активное сообщество пользователей и разработчиков;
 - поддерживает регулярные обновления и тех. Поддержку;
 - обрабатывает большое количество данных (до 1 Гб);
 - дает возможность использовать ее как программу (на основе готовых пакетов);
 - позволяет создавать таблицы произвольной структуры;
 - параллельные вычисления (пакеты *foreach* и *doParallel*).

Параллелизация вычислений в *R*. Учитывая, что вычисления по нескольким независимым моделям хорошо распараллеливаются, используем возможность, предоставляемую языком *R*, и со-

сделаем кластер из нескольких ядер нашего процессора или компьютеров локальной сети, независимо от ОС этих компьютеров.

`library(doParallel)` ; загрузка нужных библиотек

`library(foreach)`

`puskCluster<-function(){` ; определение кол-ва ядер на ПК

`cores<-detectCores()`

`cl<-makePSOCKcluster(cores)` ;

`registerDoParallel(cl)` ; создание кластера

`clusterSetRNGStream(cl)`; добавление в каждый поток генератора случайных чисел

`return(cl)`

`}`

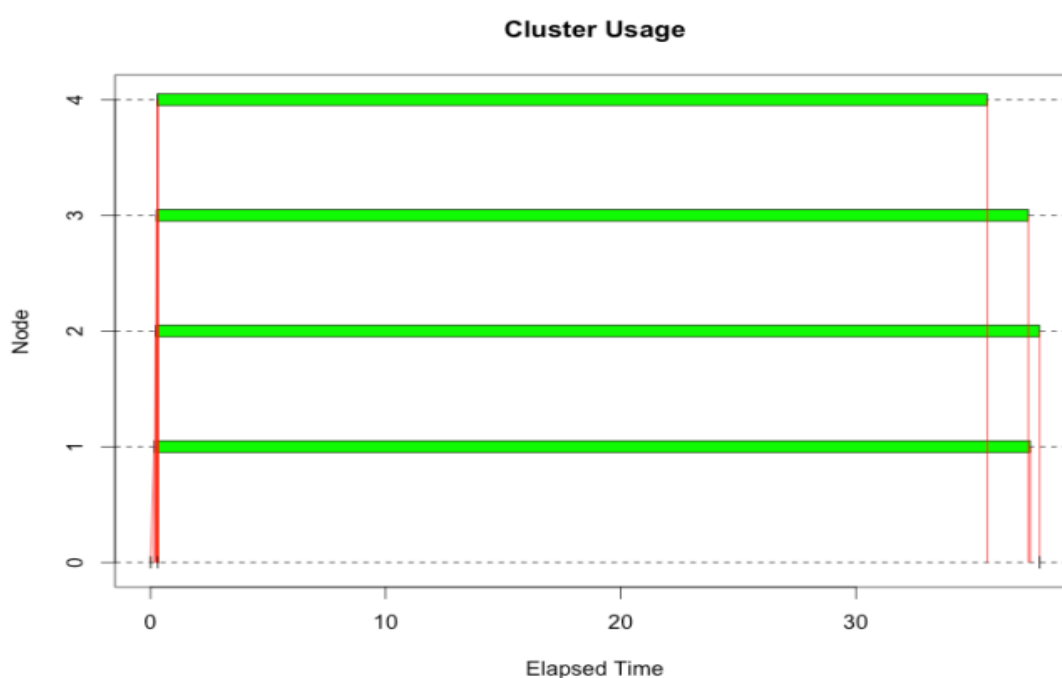


Рис. 4. Реализация нейронных сетей на языке R

Опишем пакеты языка R для реализации нейронных сетей:

«*deepnet*» — простой пакет, реализующий модель *DN SAE* и *DN SRBM*. При реализации пакета длина набора входных данных при обучении без учителя и с учителем одинакова. Отсутствует возможность провести тонкую настройку в два этапа. Предназначен для освоения и опробования моделей на начальном этапе.

«*darh*» — очень развитый и широкий пакет моделирования для *DN SRBM*. Построен на базе оригинальных программ Хинтона на языке *m* для *MatLab*.

«*H2O*» — профессиональный пакет, предназначен для обучения моделей глубоких сетей (и не только) на «больших наборах данных» (>1 Гб), записанных в *csv*-файлах.

Программная реализация (индикатор и эксперт). Опишем программную реализацию индикатора и эксперта, использующих сеть для получения торговых сигналов. Возможна реализация в двух вариантах:

1. Работа по обучению нейросети проводится в *Rstudio* вручную. После получения приемлемых результатов нейросеть сохраняется в соответствующем каталоге. Затем запускается эксперт

и индикатор на графике, загружается обученная сеть. Индикатор на каждом новом баре готовит вектор новых входных данных и передает их эксперту. Эксперт предъявляет данные нейросети, получает сигнал и далее его выполняет. Задача индикатора подготовить и передать эксперту новые входные данные на каждом новом баре и, что более важно, выводить на график сигналы, полученные от нейросети при предсказании. Визуальный контроль, как показывает практика, является наиболее эффективным способом оценки работы нейросети.

2. Эксперт и индикатор запускаются на графике. При первом запуске индикатор передает эксперту подготовленный набор входных и выходных данных. Эксперт запускает обучение, тестирование и выбор лучшей нейросети. После этого работа продолжается по первому варианту. Такой вариант исполнения дает возможность подключать к одному эксперту несколько индикаторов, расположенных на разных символах, и, соответственно, работать на них. Ниже представлена схема взаимодействия индикатора и эксперта.

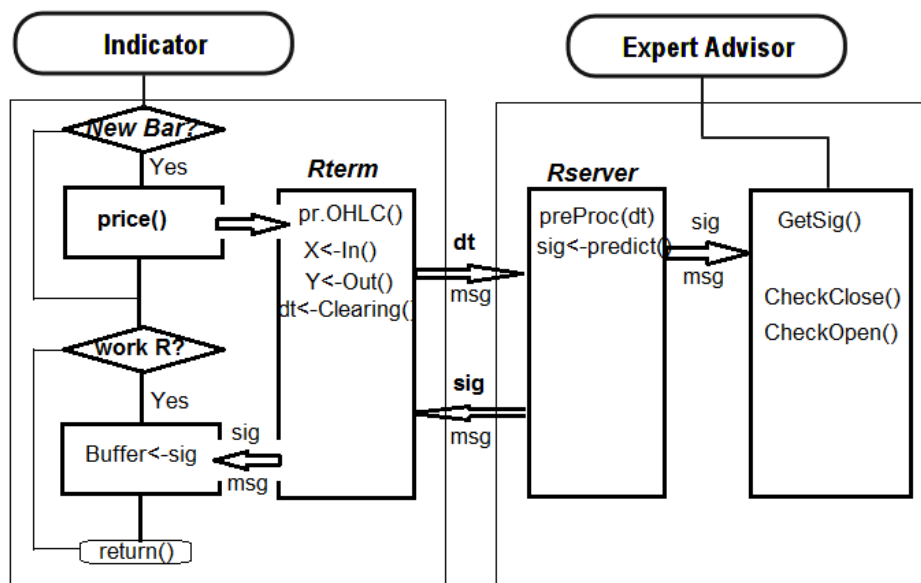


Рис. 5. Схема взаимодействия индикатора и эксперта

Сравнение языка R с другими языками программирования.

Суммирование столбцов

| R | C# |
|--|--|
| <code>p.y<-apply(tbl,2,sum)/sum(tbl)</code> | <pre> double sum = 0; foreach(DataGridViewRow r in data- GridView1.Rows) { sum += Convert.ToDouble(r.Cells[0].Value); } MessageBox.Show(sum.ToString()); </pre> |

ГСЧ

| R | C# |
|---|----|
|---|----|

| | |
|--|---|
| $> rfunc(n, p [1], p [2]...)$ <i>n</i> -число данных <i>func</i> - закон вероятности <i>p</i> -значение параметров закона | <code>Random rnd=new Random();</code> <code>rnd.next(10);</code> |
|--|---|

| Закон | Параметры |
|--------------------------------------|--|
| Гауссовское распределение | <code>rnorm(n, mean=0, sd=1)</code> |
| Экспоненциальное распределение | <code>rexp(n, rate=1)</code> |
| Распределение Пуассона | <code>rpois(n, lambda)</code> |
| Распределение Коши | <code>rcauchy(r, location=0, scale=1)</code> |
| Распределение Стьюдента (<i>t</i>) | <code>rt(n, df)</code> |
| Распределение Фишера (<i>F</i>) | <code>rf(n, df1, df2)</code> |
| Распределение Пирсона (X^2) | <code>rchisq(n, df)</code> |

Библиографический список

1. Третье поколение нейросетей: глубокие нейросети [Электронный ресурс] / MQL5. — Режим доступа: https://www.mql5.com/ru/articles/1103#3_1 (дата обращения : 01.05.16).
2. R-FAQ (часто задаваемые вопросы) [Электронный ресурс] — Режим доступа: https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f (дата обращения : 01.05.16).