

СРАВНИТЕЛЬНОЕ МОДЕЛИРОВАНИЕ МАРКОВСКИХ ЦЕПЕЙ В СРЕДЕ ПРОГРАММИРОВАНИЯ R И MICROSOFT EXCEL

Аникин В. И., Козырин Д. И.

Кафедра «Информационный и электронный сервис», Поволжский государственный университет сервиса,
Кафедра «Информационный и электронный сервис», Поволжский государственный университет сервиса
Тольятти, Россия

E-mail: anikin_vi@mail.ru, kozyrin-di@yandex.ru

В Microsoft Excel и R построена табличная модель для преобразования матрицы переходов поглощающей марковской цепи к канонической форме с произвольной размерностью для сравнения сред программирования и выявления их недостатков и преимуществ.

ВВЕДЕНИЕ

Перед началом создания модели всегда встает вопрос: “В какой программной среде создавать модель?” В данной статье будет рассмотрены две программные среды для моделирования марковских цепей: R и Microsoft Excel. Цепь Маркова являются мощным и широко используемым инструментом для анализа множества стохастических (вероятностных) систем во времени [1]. Программа Excel все еще является лидером на рынке программ обработки электронных таблиц, именно она определяет тенденции развития в этой области. Excel позволяет быстро анализировать большие массивы данных. В Excel имеется более 400 математических, статистических, финансовых и других функций, позволяющих строить большие иерархические структуры и обширные модели. Но в настоящее время язык R также является одним из ведущих статистических инструментов в мире. Он активно применяется в молекулярной биологии, науках об окружающей среде, генетике, биоинформатике.

R является доступным и хорошо сбалансированным представлением теории случайных процессов с акцентом на реальных применениях теории вероятностей в естественных и социальных науках. Использование имитации с помощью популярного статистического программного обеспечения R подтверждает теоретические выводы при помощи практических демонстраций [2]. Язык R имеет широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов, на данный момент их насчитывается порядка 12500.

В статье будет рассмотрен один пример разработки в среде Excel и 2 примера в R, так как подробная разработка в Excel были рассмотрены нами в прошлых материалах [3] и [4].

Цель статьи: Разработка моделей марковских цепей в среде R и Excel, сравнение по удобству написания кода и визуализации модели, нахождение общих достоинств и недостатков этих сред программирования.

Для оценки сред программирования была использована готовая модель на Excel из нашей опубликованной статьи [4] (рис. 1).

Рис. 1 – Модель марковской цепи в Microsoft Excel

А именно будут рассмотрены формулы таблиц I-Q, N, NR.

I. СРЕДА РАЗРАБОТКИ MICROSOFT EXCEL

В VBA Excel сначала нужно выделить редактируемый диапазон ячеек и только затем вносить в них формулы, поэтому формула становится длиннее.

I-Q: формула в табличной среде: =R[-4]C-R[-8]C
Формула в VBA:ActiveSheet.Range(Cells(x-p+x-p+13, 5+x), Cells(x+x+x-p-p+12, x+x-p+4)).Select

Selection.FormulaR1C1 = -R[-"& x-p+1 & "]C-R[-"& x-p+x-p+2 & "]C"

N: формула в табличной среде: МОБР(R[-4]C:R[-2]C[2])

Формула в VBA:ActiveSheet.Range(Cells(x-p+x-p+x-p+14, 5+x), Cells(x+x+x+x-p-p-p+13, x+x-p+4)).Select

Selection.FormulaArray = -MINVERSE(R[-"& x-p+1 & "]C:R[-2]C["& x-p-1 & "]")

NR: формула в табличной среде: =МУМНОЖ(RC[-10]:R[2]C[-8];RC[-3]:R[2]C[-2])

Формула в VBA:ActiveSheet.Range(Cells(x+x+x-x-p-p-p+16, x-p+8+p), Cells(x+x+x+x-x-p-p-p-p+15, x-p+p+p+7)).Select

Selection.FormulaArray = -MMULT(RC[-"& x-p+p+5 & "]R["& x-p-1 & "]C[-"& p+6 & "],RC[-"& p+1 & "]R["& x-p-1 & "]C[-"& 2 & "]")

Минусом этой среды является добавление и удаление диапазонов настолько увеличивает время написания кода, что при разработке табличной модели приходится выбирать требуемые размерности заранее и планировать наперед размерность каждой отдельной таблицы, это тема была рассмотрена в прошлой статье по построению итерационной модели марковских цепей в Microsoft Excel с произвольной размерностью матрицы переходов [3].

Преимущества среды разработки: всё, с чем работает пользователь, содержится в едином пространстве таблицы: пометки и комментарии, графики и диаграммы, входные, промежуточные и выходные данные, документация. Поэтому написание алгоритмов в табличном процессоре не требует специальных знаний у пользователя в области программирования.

Недостаток среды разработки: пользователь пишет код программы, используя низкоуровневые объекты — ячейки, привязанные к сетке. Табличным процессор использует только ячейки и формулы, средств абстракции более высокого уровня нет.

II. СРЕДА РАЗРАБОТКИ R

Программирование в R имеет более привычный вид и математические формулы пишутся короче, но для наглядности был использован графический интерфейс для вывода результатов (рис. 2).

Пример 1. Листинг:

```
k <- matrix(c(0,0.5,0,0.5,0,0.5,0,0,0.5,0,0,0.5,
0,0,0,1,0,0,0,0,1), nrow=5, ncol=5, byrow=TRUE)
q <- matrix(c(0,0.5,0,0.5,0,0.5,0,0.5,0), nrow = 3,
ncol = 3, byrow = TRUE)
i <- matrix(c(1,0,0,0,1,0,0,0,1), nrow = 3, ncol =
3, byrow = TRUE)
r <- matrix(c(0.5,0,0,0,0,0.5), nrow = 3, ncol = 2,
byrow = TRUE)
h <- i-q # I-Q
n <- solve(h) # N
nr <- n
```

I-Q:				N:				NR:			
	i				1				1		
	o				1.5				1	0.5	0.25
	p				2	1.0	2		2	0.50	0.50
					3	0.5	1		3	0.25	0.75

Рис. 2 – Графический вывод таблиц в R

Написание кода модели в среде R происходит быстро и интуитивно понятно все результаты выводятся в консоль, но для визуальной проверки выходных данных нужно использовать графический интерфейс, который использовался для вывода таблиц (рис. 2.)

Пример 2. Найти стационарное распределение погоды Марковской цепи (рис. 3.).

$$P = \begin{matrix} & \begin{matrix} \text{Дождь} & \text{Снег} & \text{Ясно} \end{matrix} \\ \begin{matrix} \text{Дождь} \\ \text{Снег} \\ \text{Ясно} \end{matrix} & \begin{pmatrix} 1/5 & 3/5 & 1/5 \\ 1/10 & 4/5 & 1/10 \\ 1/10 & 3/5 & 3/10 \end{pmatrix} \end{matrix}$$

Рис. 3 – Матрица переходов

Листинг:

```
stationary <- function(mat)
x = eigen(t(mat))$vectors[,1]
as.double(x/sum(x))
mat <- matrix(c(1/5,3/5,1/5,1/10,
4/5,1/10,1/10,3/5,3/10),nrow=3, byrow=T)
> mat
0.2 0.6 0.2
0.1 0.8 0.1
0.1 0.6 0.3
lambda <- stationary(mat)
lambda %% mat
0.1111111 0.75 0.1388889
```

Преимущества среды разработки: большая библиотека статистических методов, имеется большая база дополнительных пакетов для скачивания, в языке R можно взаимодействовать с другими языками программирования, работать можно с различными форматами данных.

Недостаток среды разработки: сложность в использовании для начинающих пользователей, мало документации на русском языке.

III. ВЫВОДЫ

Microsoft Excel и R имеют очень обширный инструментарий для разработки моделей, каждая имеет свои недостатки и преимущества. Если у пользователя небольшой опыт в программировании, то ему хорошо подойдет Microsoft Excel, он дает возможность строить большие модели и визуализировать результаты моделирования, что позволяет даже сторонним людям объяснить принцип работы модели шаг за шагом. Если же нужно работать с данными, часто изменяющими размерность, или использовать для работы средства абстракции более высокого уровня, то лучше подойдет R. Но самым разумным представляется использование этих великолепных программных сред в тандеме, обмениваясь данными между ними посредством csv-файлов или разделяемой базы данных.

IV. СПИСОК ЛИТЕРАТУРЫ

1. Ching, W.-K. Markov Chains: Models, Algorithms and Applications / W.-K. Ching, M. K. Ng // New York: Springer Science + Business Media. – 2006. – 205 p.
2. Dobrow Robert P. Introduction to Stochastic Processes with R / Robert P. Dobrow // Hoboken: Wiley, – 2016. – 504 p.
3. Аникин В. И., Аникина О. В., Козырин Д. И. Итерационные модели марковских цепей в Microsoft Excel / В. И. Аникин. О. В. Аникина. Д. И. Козырин. // Информационные технологии и системы 2017 (ИТС 2017). – Минск: БГУИР, – 2017. – с. 172-173.
4. Аникин В. И., Козырин Д. И. Преобразование матрицы переходов поглощающей марковской цепи к канонической форме в Microsoft Excel / В. И. Аникин. О. В. Аникина. Д. И. Козырин. // Синергетика природных, технических и социально-экономических систем: – Изд-во ПВГУС, – 2018. – 74-79 с.