ИТЕРАЦИОННЫЕ МОДЕЛИ МАРКОВСКИХ ЦЕПЕЙ В MICROSOFT EXCEL

Аникин В. И., Аникина О. В., Козырин Д. И.
Кафедра «Информационный и электронный сервис», Поволжский государственный университет сервиса,
Кафедра «Прикладная математика и информатика», Тольяттинский государственный университет,
Кафедра «Информационный и электронный сервис», Поволжский государственный университет сервиса
Тольятти, Россия
E-mail: anikin_vi@mail.ru, blue-waterfall@yandex.ru, kozyrin-di@yandex.ru

В Microsoft Excel построена итерационная табличная модель финитной марковской цепи без программирования на языке VBA. Модель позволяет исследовать временную эволюцию множества марковских цепей с фиксированной или изменяющейся при итерациях матрицей вероятностей переходов в системах с небольшим числом состояний.

Введение

Марковские цепи находят широкое применение для моделирования стохастических процессов во многих областях человеческой деятельности: экономике, социологии, биологии, генетике, лигвистике, физике, информатике и др. [1]. Пять впечатляющих примеров эффективного практического использования марковских цепей приведено, в частности, в работе [2].

Целью работы являлась разработка итерационной табличной модели марковских цепей в Microsoft Excel, основываясь на оригинальной авторской технологии моделирования вычислительных алгоритмов в электронных таблицах [3]. Очень простая, но не итерационная табличная модель знаменитого алгоритма PageRank, базирующегося на марковских цепях, приведена в работе [4].

Авторская табличная модель по своим возможностям значительно превосходит модель из работы [4] и является универсальной в том смысле, что без всякой коррекции она может использоваться для моделирования многообразия дискретных марковских цепей (за исключением скрытых), даже когда количество состояний марковской цепи в ходе итераций динамически изменяется, как, например, в задаче о случайных блужданиях. В таком случае в модели нужно лишь исправить формулу, по которой в каждой итерации вычисляется текущее число состояний системы.

I. Структура и работа итерационной табличной модели марковских цепей

Структура созданной итерационной табличной модели марковских цепей показана на рис.1. Математически модель описывается формулами:

$$\mathbf{H}(\mathrm{i}) = \mathbf{H}(\mathrm{i}\text{-}1)*\mathbf{H}(0)$$
$$\mathbf{p}(\mathrm{i}) = \mathbf{p}(\mathrm{i}\text{-}1)*\mathbf{H}(0)$$

где $\mathbf{H}(0)$, $\mathbf{p}(0)$ – исходные матрица вероятностей переходов и вектор распределения вероятностей марковской цепи, $\mathbf{H}(\mathrm{i})$, $\mathbf{p}(\mathrm{i})$ – то же на i-ой итерации.

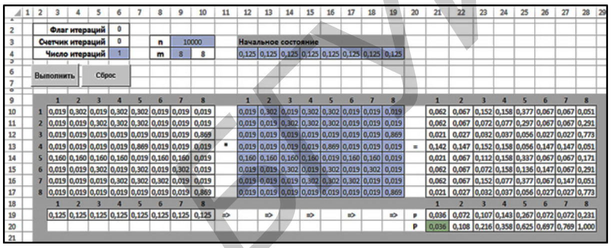

Рис. 1 – Итерационная табличная модель марковской цепи с 8-ю состояниями.

В табличной модели на рис.1 этим матрицам и векторам соответствуют интервалы ячеек: H(0) – интервал R10C12:R17C19, H(i-1) – интервал R10C3:R17C10, H(i) – интервал R10C21:R17C28, p(0) – интервал R4C12:R4C19, p(i-1) – интервал R19C3:R19C10, p(i) – интервал R19C21:R19C28. Интервал ячеек R20C21:R20C28 содержит интегральное распределение вероятностей марковской цепи на i-ой итерации.

Кнопка *Выполнить* запускает цикл итераций, кнопка *Сброс* приводит марковскую цепь в исходное состояние.

Рис. 1 содержит лишь часть созданной нами табличной модели. Дополнительно модель реализует: 1) процесс эволюции марковской цепи (10000 переходов), 2) сбор статистики состояний системы в этом процессе, 3) расчет матрицы вероятностей в этом процессе методом Монте-Карло.

Именованные ячейки модели: Start-R2C6, cnt-R3C6, Nr-R4C6, n-R3C9, m-R4C10.

Формулы в ячейках модели:
R3C6:=ЕСЛИ(Start=1;cnt+1;0)
R4C10:=ЕСЛИ(Start=1;R4C9;R4C9)
R10C3:R17C10:=ЕСЛИ(И(R9C<=m; RC2<=m);ЕСЛИ(Start=1;RC[18];RC[9]);)
R10C21:R17C28:=МУМНОЖ(СМЕЩ(R10C3; 0;0;m;m);СМЕЩ(R10C12;0;0;m;m))
R19C3:R19C10:=ЕСЛИ(R9C<=m;ЕСЛИ(Start=1; RC[18];R[-15]C[9]);)
R19C21:R19C28:=МУМНОЖ(СМЕЩ(R19C3;0;0;1;m); СМЕЩ(R10C12;0;0;m;m))

```
R20C21:=R[-1]C
R20C22:R20C28:=ЕСЛИ(RC[-1]<1;RC[-1]+R[-
1]C;)
R24C22:R10023C22:=СЛЧИС()
R24C23:R10023C23:=ЕСЛИ(R18C[-2]<=m;
ЕСЛИ(RC[-1]<R20C21;1;ЕСЛИ(RC[-1]<R20C22;
2;ЕСЛИ(RC[-1]<R20C23;3;ЕСЛИ(RC[-1]<R20C24;
4;ЕСЛИ(RC[-1]<R20C25;5;ЕСЛИ(RC[-1]<R20C26;
6;ЕСЛИ(RC[-1]<R20C27;7;8)))))));)
R24C24:R10023C24:=ЕСЛИ(R[1]C23>=RC23+1;
RC[-1]& "&R[1]C[-1];
ЕСЛИ(R[1]C23=RC23;RC[-1]& "&
R[1]C[-1];ЕСЛИ(R[1]C23<=RC23-1;
RC[-1]& "&R[1]C[-1];“-")))
R10027C21:R10027C28:=ЕСЛИ(R10025C<=m;
СЧЁТЕСЛИ(СМЕЩ(R24C23;0;0;n;1);
R[-2]C)/n;)
R10030C21:R10037C28:=ЕСЛИ(И(R10029C<=m;
RC20<=m)СЧЁТЕСЛИ(СМЕЩ(R24C24;0;0;n-
1;1);
RC20& "&R10029C);)
R10040C21:R10047C28:=ЕСЛИ(И(R10039C<=m;
RC20<=m);ЕСЛИ(СУММ(R[-10]C21:
R[-10]C28)>0;R[-10]C/
СУММ(R[-10]C21:R[-10]C28);0);)
R10049C21:R10049C28:=СМЕЩ(R10039C20;
R10039C;R10039C)
```

II. Марковская цепь алгоритма PageRank

Работу созданной табличной модели марковской цепи проиллюстрируем на примере ранжирования страниц простого web-сайта, содержащего 7 страниц: 1-Staff, 2-Student, 3-Alumni, 4-Library, 5-Home, 6-Admin, 7-Dept (рис.2) [4], - алгоритмом PageRank.

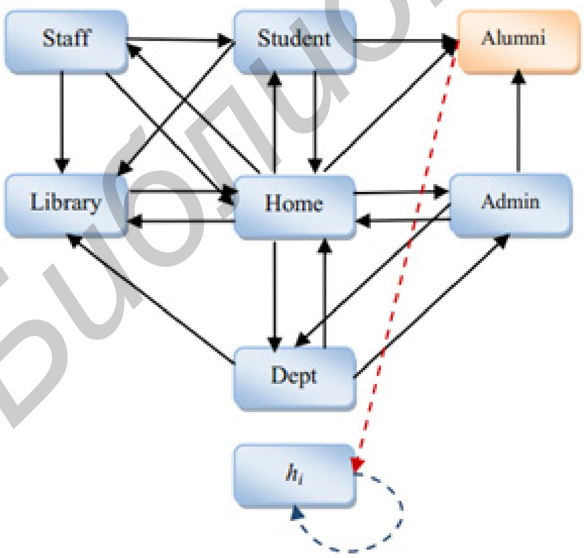

Рис. 2 – Граф переходов состояний марковской цепи с гипотетическим узлом hi.

Матрица вероятностей переходов по страницам этого сайта не является стохастической, т.к. на графе (рис.2) имеется тупиковый узел Alumni. Существует несколько методов устранения этой проблемы. Мы воспользовались методом добавления на граф виртуального поглощающего узла hi. Кроме того, для устранения тупиков, зацикливаний и гарантированного существования стационарного состояния марковского процесса, авторы алгоритма PageRank предложили добавлять к исходной матрице переходов дополнительную демпфирующую матрицу с небольшим весовым коэффициентом (обычно 0.15). Эта матрица обеспечивает гарантированную вероятность перехода марковской цепи из любого состояния в любое другое состояние, т.е. превращает марковскую цепь в неприводимую. Матрица переходов результирующей марковской цепи показана на рис.1 в интервале ячеек R10C12:R17C19. Начальное распределение вероятностей в цепи задано в интервале ячеек R4C12:R4C19.

Модельный эксперимент показал, что после 28 итераций достигается стационарное состояние, равное 0.039, 0.050, 0.068, 0.079, 0.142, 0.054, 0.054, 0.513, т.е. страницы сайта ранжируются в последовательности: Home, Library, Admin, Alumni, Student, Dept, Staff.

III. Заключение

Созданную итерационную табличную модель дискретных марковских цепей можно эффективно использовать в лабораторном практикуме по ряду вузовских дисциплин для изучения студентами закономерностей и алгоритмов моделирования стохастических процессов, а также в научных исследованиях для тестирования и подгонки матрицы вероятностей переходов моделируемой системы.

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. Hilgers, P. V. The Five Greatest Applications of Markov Chains / P. V. Hilgers, A. M. Langville // Charlston SC: Proceedings of the Markov Anniversary Meeting. -– p. 155-168
3. Аникин, В. И. Моделирование и визуализация информационной структуры алгоритмов в Microsoft Excel / В. И. Аникин, О. В. Аникина // Математическое и программное обеспечение систем в промышленной и социальной сфере. –№1(4). – 2014. – с. 21-28
4. Davey, B. A. Google PageRank / B. A. Davey. -– Melbourne: La Trobe University. – 2013. -– 23 p.
5. Proceeding of mini–symposium on biological nomenclature in the 21st centry [Electronic resource] / Ed. J. L. Reveal. – College Park M.D., 1996. – Mode of access: http://www.inform.ind.edu/PBIO/. – Date of access: 14.09.2012.
6. Kumar, P. R. Application of Markov Chain in the PageRank Algorithm [Electronic resource] / P. R. Kumar, A. K. Singh, K. L. Goh - Сайт ResearchGate. -– Mode of access: https://www.researchgate.net/publication/256089617_Application_of_Markov_Chain_in_the_PageRank_Algorithm. – Date of access: 01.09.2017.