ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
МАТЕМАТИКИ

Том 19 Выпуск 2,3

2012

В. В. Бочкарев, Э. Ю. Лернер (Казань, К(П)ФУ). **Ципфовский и нециклофовский законы для однородной марковской цепи.**

В последнее время в приложениях возник интерес к механизмам появления степенных (ципфовских) законов и пределам их применимости. Теоретический результат ниже можно рассматривать как попытку ответить на возникающие вопросы в случае марковской зависимости между соседними символами в «слове». Нас интересуют частоты слов.

Мы рассмотрим произвольную однородную марковскую цепь с дискретным временем и с конечным множеством состояний $E_0, E_1, \dots, E_n$, причем состояние E_0 — поглощающее («пробел»), $E_1, E_2, \dots, E_n$ — невозвратные. Пусть P_0 — матрица переходных вероятностей этой цепи, P — ее субстохастическая подматрица, соответствующая состояниям $E_1, E_2, \dots, E_n$. Обозначим G_0 ориентированный псевдограф с множеством вершин $\{0, 1, \dots, n\}$, ребра (i, j) которого определяются неравенствами $p_{ij} > 0$. Пусть G — подграф графа G_0 с множеством вершин $\{1, 2, \dots, n\}$, включающий в себя все ребра исходного графа G_0 между этими вершинами (полный подграф, порожденный вершинами $\{1, 2, \dots, n\}$). Если H — некоторый полный подграф графа G_0, то обозначим P_H соответствующую подматрицу матрицы P_0: $P_H = (p_{ij})_{i,j \in V(H)}$. Кроме того, положим $P_H(\beta) = (p_{ij}^{\beta})_{i,j \in V(H)}$. Обозначим G' (ациклический) орграф, получающийся из графа G_0 отождествлением вершин и ребер, принадлежащих компонентам сильной связности.

Начальное распределение вероятностей на множестве состояний обозначим $a = (a_0, a_1, \dots, a_n)$. Без ограничения общности будем считать, что у нас отсутствуют состояния, вероятность попадания в которые нулевая в любой момент времени. Мы рассмотрим дискретное вероятностное пространство всевозможных траекторий (слов) $w = (E_{i_1}, E_{i_2}, \dots, E_{i_m})$, где $a_{i_1} > 0$, $E_{i_m} = E_0$, $E_{i_{m-1}} \neq E_0$. Вероятность слова $\mathbf{P}(w)$ определим по формуле $\mathbf{P}(w) = a_{i_1} p_{i_1 i_2} \dots p_{i_{m-1} i_m}$.

Пусть $p(t)$ — вероятность t-го слова в отсортированном по ее убыванию списке всех слов. Нас интересует асимптотика функции $p(t)$ при $t \to \infty$. Мы будем использовать стандартную O-символику. Обозначим C множество всех простых циклов графа G, «вероятность» каждого из них $\mathbf{P}'(c)$ определяется очевидной формулой.

Теорема. *Возможны три альтернативы.*

1. *Если граф G ацикличен, то $p(t)$ финитна.*

2. *Если в графе G имеется вершина, через которую проходит два различных простых цикла, то $p(t) = \Omega(t^{-1/\beta})$, где β — вещественное число, при котором наибольшее по модулю собственное значение матрицы $P_G(\beta)$ равно единице. Отметим, что такое β существует, единственно и принадлежит интервалу $(0, 1)$. Кроме того, $p(t) = o(t^{-1/\beta})$ для любого $\beta > \beta$. При этом $p(t) = \Theta(t^{-1/\beta})$ тогда и только тогда, когда в любом простом пути в графе G содержится не более одной вершины — компоненты сильной связности H графа G, для которой матрица $P_H(\beta)$ имеет собственное число, равное единице.*

3. *Если в графе G есть циклы, причем через каждую вершину графа G проходит не более одного простого цикла, то $p(t) = \Omega(\alpha^t)$ и $p(t) = o\,(t^{-\lambda})$, где λ — любое положительное число, а $\alpha \in (0,1)$ — некоторая константа, зависящая от матрицы P. Последняя альтернатива включает в себя особый случай экспоненциального убывания функции $p(t)$: $p(t) = O\,(e^{-\gamma t})$ для некоторого $\gamma > 0$ тогда и только тогда, когда любой путь в графе G проходит через вершины не более одного цикла. В этом случае $p(t) = \Theta(e^{-\gamma t})$, где γ задается формулой $1/\gamma = -\sum_{c \in C} k(c)/\ln \mathbf{P}'(c)$, здесь $k(c)$ — количество различных слов с неповторяющимися состояниями, проходящими через (некоторые) вершины цикла c.*

Работа выполнена при поддержке РФФИ, проект №12-06-00404а.