

ВЭБОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЗАИМОДЕЙСТВИЯ В ГРУППЕ НА ОСНОВЕ ЦЕПЕЙ МАРКОВА

Фильченко Д.В., ассистент; Быков Д.В., студент

Вэбометрическое моделирование является современным методом анализа, прогнозирования и оптимизации web-сайтов на основе вэбометрической статистики. С точки зрения web-маркетинга, важной характеристикой функционирования web-сайтов является их взаимодействие с другими участниками заданного тематического сегмента web. Одним из показателей такого взаимодействия часто считают количество входящих (inlinks) и исходящих ссылок (outlinks) для каждого web-сайта внутри тематической группы.

Основываясь на работах [1, 2], Web предлагается моделировать с помощью цепи Маркова: $M = (S, T)$, где S – множество Web сайтов (состояний) группе ($|S| = m$), T – множество гиперссылок между Web сайтами (переходов), каждой из которых соответствует некоторая вероятность. Тогда интегрированным показателем взаимодействия участников группы можно считать вектор $\mathbf{p} = (p_1, p_2, \dots, p_m)$, координаты которого также называют рангом Web сайта [1]:

$$\mathbf{p} = \mathbf{A}'\mathbf{p}, \mathbf{A} = \{a_{ij}\}_{i,j=1}^m : a_{ij} = \begin{cases} 1/O_j, & (i, j) \in T, \\ 0, & (i, j) \notin T, \end{cases} \quad (1)$$

где O_j – количество outlinks j -го участника группы.

Применение модели (1) на практике, когда число m достаточно велико, сопряжено со следующими трудностями. Во-первых, идентификация матрицы \mathbf{A} требует автоматизации. Во-вторых, система (1) для единственности решения требует регуляризации: согласно теореме эргодичности, матрица \mathbf{A} быть стохастической, а цепь M – неприводимой и апериодичной. В-третьих, стоит вопрос выбора эффективного метода решения системы (1). В данной работе предложены способы решения этих проблем и проведена апробация описанного подхода на реальных вэбометрических данных.

1. B. Liu, et al., *Web Data Mining* (N.Y.: Springer: 2006).
2. G. Xu, et al., *Web Mining and Social Networking: Techniques and Applications* (N.Y.: Springer: 2011).