Дополнительный материал «Ранжирование страниц»

И. Хованская, Б. Бычков, И. Тельпуховский

23 марта 2015 г.

В этом дополнительном материале мы расскажем об одном из применений линейной алгебры в лице теоремы Фробениса–Перрона. Речь пойдёт о знакомых каждому интернет-страницах. Мы будем свободно пользоваться простейшими понятиями из теории вероятности, например тем, что сумма элементарных исходов одного события равна 1.

Итак, предположим у нас есть простейшая модель интернета, в которой есть n страниц $A_1, \ldots, A_n$ и с каждой страницы стоит ссылка на каждую из n страниц (ссылки на саму себя разрешены). Для каждой ссылки известна вероятность, с которой осуществляется переход, эта вероятность может быть и нулевой.

Составим матрицу R, в k-ом столбце которой, стоят вероятности перейти с k-ой страницы на все остальные. Соответственно, в k-ой строке стоят вероятности перейти на k-ую страницу со всех остальных.

Заметим, что у такой матрицы сумма чисел в каждом столбце равна 1, при этом в каждой строке не равна 1.

Рангом $\mathrm{rk}(A)$ страницы мы будем называть следующую сумму:

$$\mathrm{rk}(A_k) = \sum_{i=1}^{n} x_i p_i,$$

где x_i — вероятность перехода с i-ой страницы на k-ую страницу A_k, а p_i — это ранг i-ой страницы. Из определения легко видеть, что понятие ранга не совсем бессмысленно. Например, если на страницу ведёт только одна ссылка со страницы с большИм рангом, то это может быть «лучше» (ранг этой страницы будет больше), чем если на неё ведёт много ссылок с неизвестных страниц с маленькими рангами.

Определение ранга можно переписать на «матричном» языке следующим образом:

$$R \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Другими словами — вектор рангов страниц является собственным вектором матрицы R с собственным значением 1.

Теперь заметим, что все элементы матрицы R неотрицательны. тогда по следствию из теоремы Фробениуса–Перрона (которое мы не будем здесь подробно обсуждать) у матрицы R есть собственный вектор с собственным значением 1. Тут мы должны заметить, что теорема Фробениуса–Перрона справедлива для матриц с положительными элементами, тогда как мы разрешили в матрице R и нулевые элементы. Оказывается, что для нашего случая это не существенное условие. Такой собственный вектор очень хорошо находится последовательными приближениями.

Мы подготовили программу, расположенную по адресу http://artmikheev.github.io/linalg/, в которой реализован метод последовательных приближений для получения собственного вектора. Возьмём матрицу R и начнём переходить со страницы на страницу с теми вероятностями, которые записаны в матрице: например, пусть

$$R = \begin{pmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{pmatrix}$$

Это значит, что есть две страницы, с первой можно с вероятностью 0.3 перейти на себя и с вероятностью 0.7 перейти на вторую. Со второй с вероятностью 0.6 — на первую и с вероятностью 0.4 — на вторую. Пусть мы находимся на первой странице и сделаем много переходов. Запишем сколько раз мы попали на первую страницу и сколько на вторую. Вектор, составленный из этих чисел делённых на количество переходов и есть приближение к собственному вектору с собственным значением 1.

Для полноты эксперимента мы предлагаем вам самим проверить, что получающийся вектор очень хорошо приближает собственный вектор с собственным значением 1.