

Автор:

Швейкин Владислав Витальевич

студент

ФГАОУ ВО «Самарский национальный

исследовательский университет

им. академика С.П. Королева»

г. Самара, Самарская область

АЛГОРИТМ ССЫЛОЧНОГО РАНЖИРОВАНИЯ PAGERANK

Аннотация: в данной статье рассматривается принцип работы алгоритма ссылочного ранжирования PageRank.

Ключевые слова: ссылка, граф, ребро, ранжирование, PageRank.

PageRank – ранжирования ссылок, применяющийся к набору документов, которые связаны друг с другом гиперссылками. Алгоритм каждому из них ставит в соответствие некоторое число, определяющее важность объекта относительно других. В частном случае, алгоритм применяется к любому связанному набору объектов, который можно представить в виде графа. Объектами являются вершины графа, а ссылками – ребра.

Рассмотрим реализацию алгоритма. Пусть u – вершина графа. Тогда F_u – множество вершин, на которые ссылается u , и B_u – множество вершин, которые, в свою очередь ссылаются на u . Пусть $N_u = |F_u|$ – набор ребер из u . Введем коэффициент затухания $c < 1$ (по умолчанию 0,85). Он необходим, чтобы вершина, являющаяся стоком, не получала полный вес от вершины, соединенной с ней дугой. Обозначим, что вершина A ссылается на вершину B , если A – начало дуги (A,B) , а B – конец.

$$R(u) = (1 - c) + c \sum_{v \in B_u} \frac{R(v)}{N_v}$$

Следовательно, для того чтобы вычислить вес вершины A нужно знать значения всех вершин, с которыми эта вершина соединена, веса которых, в свою очередь, будет зависеть от других вершин, которые будут являться началами дуг.

Вес PageRank, передаваемый с вершины В на вершину А уменьшается с каждой ссылкой на какую-либо другую вершину. Это означает, что PageRank, можно представить в виде меры ее голоса; страница может разделить этот голос между одной или многими вершинами, но суммарный всех вершин остается неизменным.

Пример работы алгоритма

На первом шаге алгоритма инициализируем веса каждой из вершин.

Для простоты вычислений, выбираем вес вершин равный единице (рисунок 1).

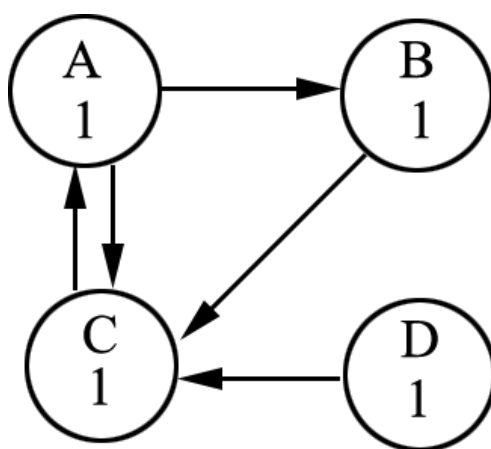


Рис. 1. Инициализация весов вершин графа

На втором шаге, используя коэффициент затухания, разделим сохранившийся вес на количество ссылок. Затем необходимо подсчитать итоговый вес, который должен быть добавлен ко всем вершинам. Таким образом, значение веса, передаваемое вершиной А после затухания равно $1 * 0,85 = 0,85$. Из вершины А исходят две дуги, поэтому, после итерации, прибавим $0,85 / 2 = 0,425$ к весам вершин В и С соответственно. Вершина В является началом одной дуги, поэтому В передаст $0,85$ вершине С в конце второй итерации. Из вершины С исходит одна дуга, следовательно, она передаст вес $0,85$ вершине А. Аналогично вершина D передаст вес $0,85$ вершине С. Добавим результаты вычислений к начальным весам и представим результаты вычислений в виде графа (рисунок 2).

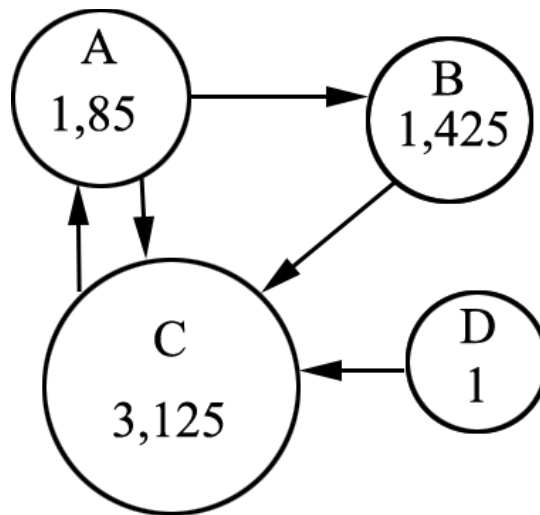


Рис. 2. Второй шаг выполнения алгоритма

На рисунке 2 видно, что наибольший приоритет имеет вершина C. Поскольку все страницы начали с одинакового значения, для повышения точности результата повторим алгоритм.

Вес узла A равен 1,85. Вес, который можно передать другим вершинам, после применения коэффициента затухания составит $1,85 * 0,85 = 1,5725$. Так как из вершины A выходят 2 дуги, по окончании итерации по 0,78625 к весу получат вершины B и C. Вершина B, имеющая одну исходящую дугу передаст $1,425 * 0,85 = 1,21125$ вершине C по окончании итерации. Узел C тоже является началом одной дуги и передаст $3,125 * 0,85 = 2,65625$ вершине A. Вершина D имеет одну дугу, поэтому она передаст 0,85 вершине C.

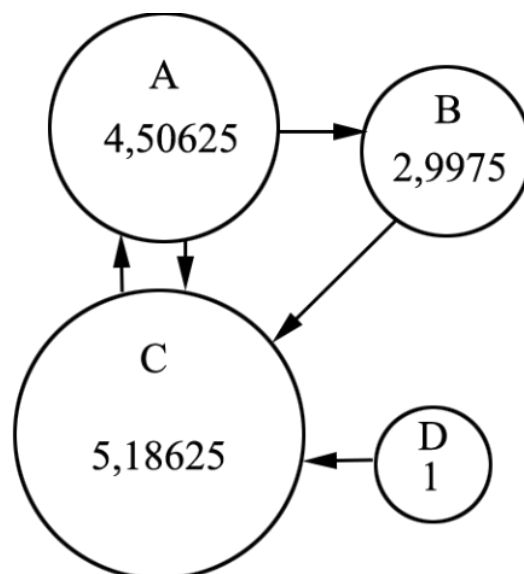


Рис. 3. Результирующий граф

На получившемся графе наглядно показано, что вершина С имеет наибольший вес, вершина А – следующая по величине. На практике необходимо повторить эти действия десятки раз, чтобы гарантировать высокую точность результатов. Результирующий граф изображен на рисунке 3.