Секция 6. ТЕОРИЯ И МЕТОДИКА ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

УДК 519.85

Алгоритм Дейкстры поиска кратчайшего пути в графе с использованием строковых матриц смежности

Ю.А. Алябышева
АлтГУ, Барнаул

В предыдущей нашей работе [1] было предложено при описании графов использовать матрицы смежности A, элементы которых $a_{i,j}$ представляют собой не числа (обычно целые числа 0, 1 для не взвешенных графов, или веса ребер – вещественные числа для взвешенных графов), а записи, которые можно рассматривать как текст, несущий информацию о первом индексе, о весе ребра и о втором индексе. Если вес ребра равен 0, что соответствует случаю отсутствия ребра, то соответствующий матричный элемент заменяется на элемент Λ – пустое слово, для которого справедливо $\Lambda a = a\Lambda = a$. Матричная алгебра, ориентированная на использование в алгоритмах на графах, в силу нечислового характера элементов матриц смежности оказывается достаточно специфичной.

В заключении работы [1] было указано, что материалы, представленные на предыдущей конференции и посвященные описанию графов на языке матриц смежности со строковыми элементами, могут оказаться полезными для организации учебного процесса по дискретной математике, информатике для отработки известных алгоритмов на графах. В настоящей работе мы это постараемся продемонстрировать и проиллюстрировать на примере задачи о поиске минимального пути из одной заданной вершины x в другую, также наперед заданную, вершину y. Эта задача решается хорошо известным алгоритмом Дейкстры [2–4]. Мы предложим реализацию указанного алгоритма с использованием матриц смежности со строковыми элементами. Классическое описание алгоритма здесь не приводим в силу того, что его можно легко найти в специализированной литературе и многочисленных источниках в сети Интернет (например, в цитировавшихся [2–4]).

Для цельности восприятия материала напомним некоторые основные содержательные моменты предыдущей публикации, необходимые для понимания решения заявленной задачи. Пусть для определенности имеется ориентированный (это требование не принципиально для дальнейшего обобщения), взвешенный, простой (без петель и кратных ребер, будем использовать термин «ребро» вместо часто употребляющегося «дуга») граф с количеством вершин N. Тогда, структура графа в компьютерных программах может быть задана квадратной матрицей смежности A, элементы которой $a_{i,j}$ фиксируют вершину i, вес – $w_{i,j}$, ребро j. Удобно отобразить сказанное записью: $iw_{i,j}j$. Если ребра нет, на месте $a_{i,j}$ ставим Λ. Вес пути, который ищется, очевидно равен сумме весов, входящих в него ребер: $W=\sum w_{i,j}$. Любой путь может быть записан с использованием введенных обозначений следующим образом:

$$(xw_{x,j}j)*(jw_{j,k}\,k)*(kw_{k,l}\,l)* \dots *(p\,w_{p,y}y) \qquad (1)$$

Путь минимальной длины не должен содержать повторно посещенные вершины, то есть в (1) не должно быть кратно встречающихся индексов. Обращаем внимание на последовательную цепочку индексов, описывающих переходы от одного узла к другому. Нетрудно заметить, что подобного рода выражения, полностью аналогичные (1) возникают, когда матрицы перемножаются. В частности, если числовая матрица смежности последовательно умножается сама на себя, порождаются матрицы достижимости вершин графа. Если количество сомножителей в (1) равно числу r, то для получения таких элементов матрицу смежности надо возводить в r-ю степень.

Квадрат матрицы смежности при использовании обычных числовых значений (0, 1) для элементов невзвешенных матриц дает количество способов попасть из одной вершины в другую (это хорошо известный результат), перемещаясь из вершины в вершину по ребрам в количестве два. Куб матрицы смежности дает тот же результат, но двигаться можно уже по ребрам в количестве три и т.д.

При использовании чисел в качестве элементов матрицы смежности теряется информация о путях перемещения по ребрам, а остается только информация о числе путей. Чтобы сохранить информацию о путях перемещения и было предложено использование матриц со строковыми элементами. При этом пришлось переопределить произведение матриц, заменив произведение элементов матриц на их конкатенацию. Тем самым можно легко адаптировать рассматриваемую матричную алгебру для реализации алгоритма Дейкстры. Для применимости предлагаемой реализации потребуется уметь извлекать из записи (1) весовые значения ребер и находить их суммы.

Опишем алгоритм.

Имея матрицу смежности A, нужно строить или вычислять матрицы достижимости A^2, A^3, … A^{N-1}, последовательно умножая предыдущий результат на A:

$$A^{r+1}=A^r * A \qquad (2)$$

Процесс построения пути обязательно завершится, если путь существует. Путь может обнаружиться даже на первом шаге алгоритма, но он может оказаться не минимальным. Уточнение его может продолжаться до рассмотрения матрицы достижимости A^{N-1}. Последний случай отвечает перебору всех вершин графа для построения кратчайшего пути из x в y. Если не применять никаких правил отбора элементов при вычислении матриц достижимости, будет происходить большой рост числа слагаемых в представлении каждого элемента матриц достижимости. Новые пути конструируются в рамках предлагаемого алгоритма. В конечном итоге построены будут все пути, начинающиеся из любой вершины графа и приходящие в любую другую вершину. Фактически так реализуется переборная задача, требующая больших вычислительных затрат. Однако многие из сконструированных путей для целей рассматриваемой задачи будут лишними.

Ситуацию спасают правила отбора. Опишем их. При r=1 в матрице A – первом сомножителе в (2) нужно оставить только те элементы, которые имеют первый индекс равный x, что отвечает стартовой вершине, то есть остается только одна строка в матрице смежности. Второй сомножитель A берется полным. Вычислять в A^2 нужно только строку x. В итоге в матрице достижимости A^2 получим отличными от Λ все вершины, в которые можно добраться по одному ребру (если в конкатенации участвовал элемент Λ), или по двум ребрам. Если элементы матрицы A^2 содержат по два и более слагаемых, то нужно оставить только то из них, которое содержит меньшую сумму весов. Это слагаемое отвечает более короткому пути. Слагаемые, в которых дважды повторяются уже посещенные вершины можно сразу отбрасывать, исключив из рассмотрения.

На следующем этапе полученная строка матрицы достижимости A^2 снова умножается на полную матрицу смежности A. Поскольку нас интересуют только пути, начинающиеся в x, в A^2 опять оставляется только строка с номером x. В результате в матрице A^3 происходит следующее: а) появляются элементы, описывающие попадание из x в новые вершины, которые ранее были недоступны (элементы, стоящие в строке x равные Λ в матрице достижимости A^2 превращаются в A^3 в элементы, аналогичные выражению (1) с тремя сомножителями); б) изменяются элементы матрицы A^2, ранее ставшие отличными от Λ, добавляется новый способ попадания в уже посещенную вершину, за счет путешествуя из x по большему числу ребер. В новой матрице достижимости в элементах, отвечающих случаю б) производим отбор тех слагаемых, которые описывают более короткий путь из вершины *x*. Обращаем внимание на то, что в матрицах достижимости A^r будут присутствовать элементы аналогичные (1), но содержащие разное число сомножителей.

Процесс построения новых матриц достижимости более высокого порядка нужно продолжать до тех пор, пока степень матрицы достижимости не станет равной N-1. Элемент, стоящий в *x*-ой строке и в *y*-ом столбце, будет решать поставленную задачу. Этот элемент также будет содержать и всю информацию о промежуточных вершинах, которые были посещены при реализации пути минимальной длины. Если этот элемент останется равным Λ, то для ориентированных графов это означает, что вершина *y* из вершины *x* недостижима.

Замечание 1. Если ищется минимальный путь в неориентированном графе, то можно одновременно отслеживать как прямой ход (движение от *x* к *y*), так и обратный ход (движение от *y* к *x*). Для этого нужно задействовать дополнительную строку с номером *y* в матрицах достижимости.

Замечание 2. Можно сократить количество операций по вычислению сумм весовых множителей в выражениях типа (1), если хранить в одномерном массиве результаты суммирования весовых множителей путей, проанализированных на предыдущих этапах реализации алгоритма.

Замечание 3. Получение элементов в матрицах достижимости и отбор нужных слагаемых можно реализовывать в рамках однопроходных алгоритмов, сокращая вычисления.

Нам представляется, что описанная реализация алгоритма Дейкстры достаточно наглядна, а матрицы смежности со строковыми элементами могут быть использованы в процессе реализации других алгоритмов из теории графов.

Библиографический список

1. Алябышева Ю.А., Веряев А.А. Использование строковых матриц смежности в алгоритмах на графах // «МАК-2017» : сборник трудов Всероссийской конференции по математике. Барнаул, 29 июня – 1 июля 2016 г. – Барнаул : Изд-во Алт-ун-та, 2017. – С. 216–220.

2. Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К. Алгоритмы: построение и анализ. – Москва : Изд-во «И.Д. Вильямс», 2013. – 1328 с.

3. Клейнберг Д., Тардос Е. Алгоритмы: разработка и применение. – Санкт-Петербург, 2016. – 800 с.

4. Алгоритм Дейкстры // Заголовок с экрана [Электронный ресурс] // Режим доступа: https://neerc.ifmo.ru/wiki/index.php?title =Алгоритм_Дейкстры.

УДК 004.85

Прикладное направление тематического моделирования в учебном процессе

Н.С. Бабкина, Л.Л. Смолякова
АлтГУ, г. Барнаул

Ключевые слова: методы анализа текстовых данных, тематическое моделирование, обработка учебно-методических материалов, вычислительный эксперимент

Целью исследования является изучение возможностей современных методов анализа текстовых материалов их систематизация, сжатие для возможности быстрого восприятия и передачи, как во время проведения учебных занятий, так и при дистанционной форме организации учебного процесса. Переход от анализа конкретного текста к анализу коллекций текстов существенно расширяет возможности изучения и практического применения знаний о значении и употреблении слов естественного языка.

В последнее время популярным направлением извлечения информации из информационных потоков является использование различных статистических методов для обработки текста. Одним из таковых методов является метод тематического моделирования, позволяющий построить модель коллекции текстовых документов, определяющую тематическую направленность каждого из них [1].

Тематическое моделирование определяется как способ построения модели коллекции текстовых документов, которая позволяет перейти от набора документов к темам, определяющих содержание каждого документа. Тематическое моделирование способствует семантическому анализу коллекции текстовых документов [2].

Тематические модели — это модели со скрытыми переменными, для выявления которых лучше всего подходит нечеткая кластеризация, при которой любое слово или документ с некоторой вероятностью относится к нескольким темам [2].

Наиболее популярные в настоящий момент методы тематического моделирования можно разделить на две основных группы — алгебраические и вероятностные (генеративные) [1; 3; 4]. К алгебраическим моделям относятся стандартная векторная модель текста VSM (Vector