ПРОБЛЕМА ИЗОМОРФИЗМА ГРАФОВ В СИСТЕМЕ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

А. А. Дунаев, О. В. Герман
Кафедра ИТАС, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: alexd.by@gmail.com, ovgerman@tut.by

В работе представлен оригинальный подход к решению задачи определения изоморфизма графов, используемый в системе распознавания изображений. Оригинальность предлагаемого в статье подхода базируется на хешировании структуры графа, используя в качестве инвариантной характеристики графа кратчайшие расстояний между всеми вершинами.

Описание задачи

В теории графов изоморфизмом графов называется биекция множества вершин графа G на множество вершин графа H, сохраняющая отношение смежности. Другими словами, для любых вершин u и v графа G их образы и смежны в H тогда и только тогда, когда u и v смежны в G. Отношение изоморфизма графов является эквивалентностью, т.е. оно симметрично, транзитивно и рефлексивно. Следовательно, множество всех графов разбивается на классы так, что графы из одного класса изоморфны, а графы из разных классов не изоморфны. Из определения следует, что изоморфные графы могут различаться лишь обозначениями вершин и ребер, так как у них должно быть равное число вершин и ребер, соответствующие друг другу вершины обязаны иметь одинаковые степени и полустепени, и, разумеется, совершенно все равно, какую геометрическую реализацию графа выбирать для его изображения.

I. Области применения задачи изоморфизма графов

На практике необходимость проверки изоморфизма графов возникает при решении задач химической информатики, математической химии, автоматизации проектирования электронных схем, оптимизация программ, но наибольший интерес представляют задачи распознавания образов. Распознавание образов — это отнесение исходных данных к определенному классу с помощью выделения существенных признаков, характеризующих эти данные, из общей массы несущественных данных. На основании информации о классе, может быть создан эталонный граф, с которым в последствии будет сравниваться исходный граф, используя метод решения задачи изоморфизма графов.

II. Сложность решения задачи изоморфизма графов

Как говорилось выше, два графа G(V, E) и H(V*, E*) изоморфны, если существует взаимнооднозначное соответствие между множествами вершин V и V*, такое, что любые две вершины в графе G смежны тогда и только тогда, когда смежны соответствующие им вершины в графе H. Проблема определения, являются ли два графа изоморфными, неожиданно оказывается трудной. Для произвольных графов все известные алгоритмы, гарантирующие правильный ответ, экспоненциальные. Пусть |V|=n. Проставим в графе G нумерацию вершин [1,2,...,n] и составим для него матрицу смежности C. Затем вершины графа H пронумеруем в соответствии с некоторой перестановкой списка [1,2,...,n] и составим его матрицу смежности C*. Если C=C*, то графы изоморфны, если нет, то снова нумеруем вершины графа H в соответствии с другой перестановкой списка [1,2,...,n], снова составляем матрицу смежности, и процесс сравнения матриц повторяется. Понятно, что если графы не изоморфны, то придется проверить все n! перестановок из n элементов, т.е. сложность этого алгоритма составляет T(n)=O(n!). Как следствие, исследователи часто отказываются от поиска эффективного алгоритма изоморфизма и взамен этого строят простые процедуры, от которых ожидается хорошая работа в большинстве случаев. Кроме того, разрабатываются алгоритмы полиномиальной сложности, пригодные для специальных классов графов, появляющихся в некоторых приложениях.

Инвариантом графа G называется параметр, имеющий одно и то же значение для всех графов, изоморфных графу G. Среди самых очевидных инвариантов отметим следующие:

- Число вершин |V|=n;
- число ребер |E|=m;
- число компонент связности;
- «последовательность степеней», т.е. список степеней всех вершин в убывающем порядке значений.

Эвристики для решения задач изоморфизма обычно состоят в попытках показать, что два рассматриваемые графа не изоморфны. Для этого составляется список различных инвариантов в порядке, определяемом сложностью вычисления инварианта. Затем последовательно сравниваются значения параметров графов. Как только обнаруживаются два различных значения одного и того же параметра, приходят к заключению, что данные графы не изоморфны. Множество инвариантов, которое позволило бы этой процедуре установить изоморфизм графов за полиномиальное время, наз. кодом графа. К сожалению, на сегодня шний день такое множество не найдено. По существу, эвристический алгоритм рассматриваемого типа сводится к сравнению неполных кодов двух графов. Конечно, рассмотрение большого числа инвариантов увеличивает вероятность правильного заключения об изоморфизме при совпадении всех значений параметров, но в общем случае ничего не гарантирует.

III. Описание алгоритма считывания изображения в графы

Основные этапы алгоритма считывания изображения в графы:

– Изображение считывается как набор пикселей;
– считанные цветные пиксели по степени яркости преобразуются в двухцветные: черные и белые;
– пустые, белые края изображения обрезаются;
– изображение делится сеткой размера N*N на ячейки;
– создается граф из N вершин, который будет соответствовать изображению;
– вершины n и k соденяются ребром с весом равным проценту (от 0 до 1) черных пикселей в ячейке в строке n и столбце k;
– выбирается несколько уровней (значений в диапозоне от 0 до 1): 0.0 , 0.25, 0.5, 0.75;
– для каждого уровня строится граф на основе полученого ранее по следующему правилу: если вес ребра меньше значения уровня, то ребро удаляется, иначе вес ребра меняется на 1.

IV. Описание алгоритма решения задачи изоморфизма графов

Рассматриваемый алгоритм основан на сравнения результатов хеширования графов. Основные этапы алгоритма решения задачи изоморфизма графов:

– По графу составляется матрица смежностей, состоящая из 0 и 1;
– все 1, обозначающие наличие ребра между вершинами, заменяются на среднее арифметическое степеней связаных вершин;
– строится матрица кратчайших расстояний между любыми двумя вершинами, получается симметричная матрица на главной диагонали которой 0;
– по каждой строке матрицы вычисляется среднее арифметическое и составляется последовательность из средних кратчайших расстояний;
– получаеная последовательность дробных чисел упорядочивается по возрастанию.

Полученая последовательность дробных чисел является «хеш-кодом» структуры рассматриваемого графа. Теперь для того, чтобы определить являются ли графы изоморфными достаточно сравнить их «хеш-коды» по алгоритму, описаному выше.

V. Описание системы распозавания изображений

При инициализации системы загружаются изображения-образцы, с которыми в последствии будут сравниваться изображения. Каждое изображение считывается в набор графов с соответствующим уровнем. Каждый граф хешируется и полученый «хеш-код» с уровнем графа сохраняется в базу. В результате получается база с «хед-кодами» для каждогого уровня по каждому изображению-образцу В последствии любое изображение можно хешировать через преобразование в графы и сравнить с «хеш-кодами» хранящимися в базе.

VI. Преимущество данного подхода

К преимуществам описанной системы распознавания изображений можно отнести следующее:

– Имеющиеся в изображении шумы или перпады цвета не учитываются;
– в базе хранятся только «хеш-коды» графов, которые по объему намного меньше чем исходное изображение;
– сравнение «хеш-кодов» состоящих из последовательности десятичных чисел происходит мгновенно.

1. Luks E.M. Isomorphism of graphs of bounded valence can be found in polynomial time //Journal of Computational Sciences. № 25(vol.1), 1982, p.p.42-65.
2. Babai L. Moderately exponential bound for graph isomorphism. // Proceedings of the International FCT-Conference on foundations of computer theory. – London, 1981, p.p. 34-50.
3. Babai L. Random graph isomorphism // L. Babai, Erdos P., Selkow M. SIAM Journal of Computing, № 9 (vol. 3), 1980, p.p.628-635.
4. Герман Ю.О. Поиск максимального независимого множества в нечетком графе. // Ю.О. Герман, О.В. Герман. – М. Прикладная информатика. № 2(56), т. 10. – 2015. – с.132-139.