УДК 512.8:681.55

О. В. Герман[1], А. А. Дунаев[2]
[1]Белорусский государственный технологический университет
[2]Белорусский государственный университет информатики и радиоэлектроники

ЗАДАЧА ИЗОМОРФИЗМА ГРАФОВ В СИСТЕМЕ НЕЧЕТКОГО РАСПОЗНАВАНИЯ

В работе представлен оригинальный подход к решению задачи определения изоморфизма подграфа графу. Известны квазиполиномиальные алгоритмы установления изоморфизма, однако их нет в случае проверки наличия изоморфизма между подграфом и частью графа. Оригинальность предлагаемого в статье подхода базируется на использовании в качестве инвариантной характеристики графа его максимальных независимых множеств, в частности, определяемых в ходе выполнения процедуры раскраски. Задача заключается в укладке вершин подграфа в построенную систему независимых множеств с предустановленных условий. Очевидно, положительный ответ о возможности укладки предоставляет необходимое, но недостаточное условие изоморфизма. Теоретически можно отыскивать все возможные варианты укладки, однако возможны ситуации, когда число таких вариантов огромное. Поэтому в статье определяется эвристический подход, который строит различные системы независимых множеств, как для исходного графа, так и для его дополнения и транзитивных замыканий. В подавляющем большинстве случаев достаточно использовать дополнительный граф и транзитивное замыкание, получаемое на основе произведения матрицы смежности на нее саму. Практическая апробация подхода свидетельствует о его достаточно высокой эффективности (определение истинного изоморфизма в 80−90% случаев). Подход допускает применение в случае задачи нечеткого изоморфизма. Это значительно расширяет рамки практического использования. Также указывается общий подход к распознаванию образов, искаженных при геометрических преобразованиях типа смещения и поворота.

Ключевые слова: граф, изоморфизм, распознавание образов, нечеткий граф, максимальное независимое множество.

O. V. German[1], A. A. Dunaev[2]
[1]Belarusian State Technological University
[2]Belarusian State University of Informatics and Radioelectronics

GRAPH ISOMORPHISM PROBLEM IN THE FUZZY RECOGNITION SYSTEM

An original approach to graph isomorphism problem is defined. There are known quasi-polynomial computational time algorithms to establish isomorphism (if exists) between two graphs, however these algorithms are not applicable in the case of isomorphism between a graph and some part of the other graph in general. A specificity of the suggested approach is based on the usage of such invariant graph feature as its maximum independent sets of vertices (clusters) defined as a result of a coloring procedure. The problem comes to allocating the nodes of a subgraph into the given independent sets with regards to predefined restrictions. Obviously, the positive answer to that question represents a necessary but not sufficient condition for isomorphism. Theoretically, one could find all possible variants of the vertex distributions into the clusters, however in some cases the total amount of combinations becomes too great. Hence, the article suggests a heuristic-based approach realized through generating a restricted set of independent sets of vertices for original graph, its complementary graph and a transitional closure(s). In the majority of cases it is sufficient to use a complementary graph and its transitional closure found as a result of multiplication of the contiguity graph matrix by itself. A practical approbation testifies to relatively high efficiency of such an approach (the true isomorphism was revealed in 80−90% of the total amount of graphs). The approach allows to treat fuzzy graph isomorphism problem as well. This significantly expands the boundaries of practical utilization of the proposed approach. Also, the paper considers some issues related to pattern recognition with the distortions caused by geometrical transformations, e.g. by shifting or rotation.

Key words: graph, isomorphism, pattern recognition, fuzzy graph, maximum-size independent set.

Введение. Очевидный практический интерес представляет задача распознавания нечетких, фрагментарных и искаженных объектов (например, лиц на размытых фотографиях, или идентификация по фрагментам изображений). Для решения этой задачи применяют различные подходы, например, статистические методы на основе использования теоремы Байеса, нейросетевые модели, технику вейвлет-преобразований, логико-алгебраические методы и т. д. Особое место занимает подход на основе решения задачи об изоморфизме (под)графа графу. Его достоинство усматривается в том, что свойство изоморфизма «инвариантно» относительно растяжений, сжатий, поворотов, смещений графов, а также сохраняет математически неизменную формулировку при распознавании изоморфизма подграфа графу. Кроме того, «не далеко» уходит от основной формулировки задача о нечетком изоморфизме. Основная трудность состоит в том, что задача изоморфизма подграфа графу в общем случае весьма сложна с вычислительной точки зрения, хотя ее NP-полнота не доказана. Согласно результату Ладнера, если P ≠ NP, то задача об изоморфизме, по-видимому, не является ни P-полной, ни NP-полной. Вместе с тем известны два метода, которые, как полагают, являются наилучшими с точки зрения вычислительной сложности. Это метод Люка [1] и метод Бабайя [2, 3]. Последний позиционируется как метод с квазиполиномиальной вычислительной сложностью. При этом заметим, что оба указанных метода предназначены для решения задачи об изоморфизме двух графов, а не об изоморфизме подграфа некоторой части другого графа. Тем более, они не применимы для решения задачи о нечетком изоморфизме. Таким образом, задача, указанная в заголовке этой статьи, все еще достаточно актуальна и требует существенных усилий.

В настоящей статье предлагается подход к решению задачи об изоморфизме (под)графа графу на основе использования множества конструктивно определяемых инвариантных характеристик графа. Сам подход эвристический, т. е. обеспечивает лишь необходимую (но недостаточную в общем случае) часть условий, гарантирующих изоморфизм. Вместе с тем наши условия легко определимы и формализуемы системой логических уравнений. Главное в том, что они допускают сравнительно простое обобщение на «нечеткий вариант задачи об изоморфизме». Сообщается о результатах предварительной апробации, из которых можно полагать общую эффективность предложенного подхода.

Основная часть. Пусть граф-оригинал есть G. Любое изображение, например, буква или подпись может быть заменена графом, узлы которого соответствуют характерным элементам изображения. Так, в случае буквы таковыми являются конечные точки, точки пересечения с линиями сетки, на которую накладывается изображение. В частности, для лица характерными точками являются концевые точки бровей, ноздри, мочки ушей, ямка подбородка и др. Пусть H – подграф, соответствующий части образа. Спрашивается, изоморфен ли H некоторому подграфу или всему графу G.

Одним из способов задания графа является его матрица смежности M. Это симметричная 0,1-матрица. Можно определить на этой матрице различные максимальные независимые множества (вершин). Каждое такое множество (будем называть их также кластерами) можно раскрасить в один цвет, так что любые две вершины графа либо попадут в одно независимое множество, либо нет (определенно нет, если вершины связаны ребром). Таким образом возникает следующая идея: для данного разбиения вершин графа G на независимые кластеры попытаться «уложить» вершины графа H «по аналогии». Для этого нужно обеспечить выполнение следующих правил:

$R1$) каждая вершина графа H должна войти в состав одного из кластеров графа G;

$R2$) вершины графа H, соединенные ребром, должны входить в разные кластеры;

$R3$) в формируемый кластер для H может попасть не более числа вершин, равных размеру этого кластера, определенного из G.

Дальнейшее изложение проще провести на примере. Пусть матрица M_G смежности графа G имеет следующий вид:

	1	2	3	4
1	0	1	1	1
2	1	0	0	1
3	1	0	0	1
4	1	1	1	0

Пусть матрица M_H смежности графа H имеет такой вид:

	1	2	3	4
1	0	1	0	1
2	1	0	1	1
3	0	1	0	1
4	1	1	1	0

Определим, являются ли эти графы изоморфными. Вершины M_G можно разбить на следующие кластеры: {1}, {2, 3}, {4}. Соответственно, присвоим этим кластерам цвета: «к», «ж», «с». Теперь сформулируем математическую задачу о возможности «уложить» вершины второго графа по аналогии с учетом правил $R1$−$R3$. Введем двоичные переменные: $x_{i,j} = 1$, если вершина i попадает в кластер j; $x_{i,j} = 0$ в противном случае.

Теперь можно записать систему логических условий для второй матрицы M_H. Очевидно, что каждая вершина второго графа H может соответствовать только тем вершинам графа G, которые имеют степень не ниже степени этой вершины в H. Таким образом, например, вершина 4 графа H может соответствовать любой из вершин 1, 4 графа G. С учетом этого вершины 1 и 3 графа H соответствуют кластеру «ж» графа G. Для остальных составляем систему:

$$\begin{aligned} x_{2,\text{к}} + x_{4,\text{к}} &= 1, \\ x_{2,\text{с}} + x_{4,\text{с}} &= 1, \\ x_{2,\text{к}} + x_{2,\text{с}} &= 1, \\ x_{4,\text{к}} + x_{4,\text{с}} &= 1. \end{aligned} \qquad (1)$$

Эта система дает несколько допустимых решений. Каждое решение легко проверить на предмет, определяет ли оно изоморфизм обоих графов. В самом деле достаточно построить матрицы смежности с учетом полученного в решении соответствия вершин графов G и H. Если полученные матрицы полностью совпадают, то изоморфизм установлен. Так, в нашем случае возьмем, например, решение: $x_{2,\text{к}} = 1$, $x_{4,\text{с}} = 1$, $x_{1,\text{ж,}} = 1$, $x_{3,\text{ж,}} = 1$. Легко убедиться, что оно действительно задает изоморфизм. И, напротив, если ни одно из допустимых решений не устанавливает требуемого для изоморфизма соответствия, то графы не изоморфны. Количество выполняющих интерпретаций для составляемых по указанным правилам систем может быть огромным. Поэтому используем эвристический подход для сужения пространства поиска. Суть его состоит в том, что составляем системы, подобные (1), также для дополнительных графов и транзитивных замыканий. Если получаемые при этом системы выполнимы, то делаем предположительный вывод о существовании изоморфизма между графами (условия выполнимости систем необходимы для изоморфизма, но, очевидно, недостаточны). Для проверки практической эффективности проводится практическая обкатка метода для случайно генерируемых графов, результаты которой сообщаются в заключительной части.

Обратимся теперь к нечетким графам. Базируемся на алгоритме отыскания максимального нечеткого независимого множества, описанного в [4]. Дадим определения.

Ребро графа будем считать нечетким, если определенно неизвестно, входит оно в граф или нет. Ребро считаем четким, если определенно известно, что оно входит в граф. Граф с нечеткими ребрами кодируем матрицей смежности A с элементами $a_{i,j}$, такими, что

$$a_{i,j} = 1, \text{ если ребро } (i, j) \text{ четкое;}$$

$$a_{i,j} = 0,5, \text{ если ребро } (i, j) \text{ нечеткое.}$$

Определение.

А. (Нечетким) независимым множеством нечеткого графа назовем любое множество его вершин, никакие две из которых не связаны четким ребром (таким образом, допускается связь нечеткими ребрами).

Б. Ядром нечеткого независимого множества Ψ считаем подмножество его вершин $\Psi_C \subseteq \Psi$, никакие две из которых не соединены нечетким ребром.

Определение. Пусть для нечеткого графа определены два независимых множества Ψ_1, Ψ_2. Говорим, что Ψ_1 максимально предпочтительнее (*м*-предпочтительнее) Ψ_2, если размер его ядра больше размера ядра Ψ_2.

Определение. Нечеткое независимое множество Ψ называется *м*-максимальным, если оно имеет ядро максимального размера, а при условии равенства размеров ядер содержит наибольшее общее число вершин.

С учетом введенного понятия *м*-максимального множества будем далее ассоциировать это множество с кластером. Таким образом, мы получаем задачу разбиения множества вершин графа на минимальное число максимально независимых множеств. Каждое из этих множеств играет роль кластера, а все его вершины окрашиваем в один и тот же цвет. Теперь можно сформулировать идею подхода к задаче нечеткого изоморфизма. Как и в случае с четким изоморфизмом, находим (нечеткие) кластеры, представляющие *м*-максимальные множества вершин. Составляем систему типа (1) для проверки условия, что вершины второго графа можно разбить в подобные кластеры. Если такое разбиение удается найти, то заключение об изоморфизме графов получает эмпирическое подтверждение, но требует аналогичных проверок на дополнительных графах и транзитивных замыканиях.

Заключение. Предложенные методы могут непосредственно применяться в различных интеллектуальных системах. Например, можно использовать описанный подход для идентификации подписи по ее фрагменту. В равной степени это относится и к фотографии. Однако в отношении идентификации лиц дело упирается в способ проецирования оригинала. Метод изоморфизма графов не может быть применен в этом случае непосредственно. С другой стороны, если изображение не подверглось «топологическим искажениям», но его фрагменты выцвели, или изменили окраску, то при пересечении с линиями сетки следует определять нечеткую оценку наличия пересечения. В этом случае описанный в статье подход вполне применим. При топологических искажениях следует «привязываться» не к точкам пересечения с сеткой (сетку вообще неуместно использовать), а выделять характерные точки изображения (уголки глаз, ноздри, уголки губ) и «накладывать» это изображение на оригинал. В результате получим картинку совмещения двух многоугольников, где точность совмещения следует оценить. Однако это – уже другая парадигма.

Разработанная программа позволила получить результаты практического использования метода. В случае достаточно больших графов (более 20 вершин) степень получения точного результата определяется в районе 90%. Для целей практического применения этого может быть во многих случаях достаточно. Для небольших графов точность решения существенно падает, однако для таких графов можно использовать методы, основанные на переборе [4].

Литература

1. Luks E. M. Isomorphism of graphs of bounded valence can be found in polynomial time // Journal of Computational Sciences. 1982. Vol. 1, no. 25. 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. 34–50.

3. Babai L., Erdos P., Selkow M. Random graph isomorphism // SIAM Journal of Computing. 1980. Vol. 3, no. 9. P. 628–635.

4. German Yu., German O. Search for the maximum-size independent set in a fuzzy graph // Applied Informatics. 2015. Vol. 10, no. 2. P. 132–139.

References

1. Luks E. M. Isomorphism of graphs of bounded valence can be found in polynomial time. *Journal of Computational Sciences*, 1982, vol. 1, no. 25, pp. 42–65.

2. Babai L. Moderately exponential bound for graph isomorphism. *Proceedings of the International FCT-Conference on foundations of computer theory*. London, 1981, pp. 34–50.

3. Babai L., Erdos P., Selkow M. Random graph isomorphism. *SIAM Journal of Computing*, 1980. vol. 3, no. 9, pp. 628–635.

4. German Yu., German O. Search for the maximum-size independent set in a fuzzy graph. *Applied Informatics*, 2015, vol. 10, no. 2, pp. 132–139.

Информация об авторах

Герман Олег Витольдович – кандидат технических наук, доцент кафедры информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: ovgerman@tut.by

Дунаев Александр Александрович – аспирант. Белорусский государственный университет информатики и радиоэлектроники (220600, г. Минск, ул. П. Бровки, 6, Республика Беларусь). E-mail: alexander.dunaev@googlemail.com

Information about the authors

German Oleg Vitoldovicz – PhD (Engineering), Assistant Professor, the Department of Information Systems and Technologies. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: ovgerman@tut.by

Dunaev Alexander Alexandrowicz – PhD student. Belarusian State University of Informatics and Radioelectronics (6, P. Brovki str., 220600, Minsk, Republic of Belarus). E-mail: alexander.dunaev@googlemail.com

Поступила 01.03.2016