УДК-519.7

Метод кластеринга, индуцированный монотонными системами

Мгеладзе А.П.
Грузинский Технический Университет, г. Тбилиси

Резюме:

В статье представлены результаты разработки метода кластеринга, источником которого служит теория монотонных систем. В работе широко используются понятия и правила общей алгебры, подчеркивающие алгебры, подчеркивающие алгебраический подход при разработке указанного метода. Приводится наглядный пример – результат программной реализации алгоритма, соответствующего использованному методу.

Ключевые слова: *метод, кластеринг, монотонные системы, граф близости, отображение, решетка, классы факторрешетки, конгруэнция.*

Введение

Актуальность математических методов структуризации социально – экономической информации не снимается с повестки дня. Одним из методов, в котором косвенно предусмотрены высокий, прикладного характера наукаемкий спрос потребителей информационного общества, является „теория монотонных систем“. Сопоставляя данный метод с другими методами упомянутого направления, можно обратить внимание на тот факт, что в теории монотонных систем, элемент системы осмысливается не как отдельно взятый, изолированный, а как часть целостной системы во взаимоотношении с другими элементами. В связи с этим приобретает важное значение определения места, веса каждого элемента в „коллективе“ других элементов системы (подсистемы), что можно выразить количественным показателем. Таким образом, в теории монотонных систем появилась скалярная функция, называемая монотонной (весовой) функцией. Подобная трактовка элементов в теории монотонных систем дает возможность для полного, всестороннего описания структуры обрабатываемых данных.

Теорию монотонных систем, на которой основан метод кластеризации, конструктивно можно разделить на три, главные части: 1. свойство монотонности; 2. особые подсистемы монотонных систем называемие ядрами и являющиеся краеугольным камнем теории; 3. экстремальные свойства ядер, которые характеризующие центральную теорему монотонных систем. Интересно, что число ядер не задается априори, а вычисляется в процессе поиска решении экстремальной задачи. В алгоритме кластеринга, на основе монотонных систем и алгебраического подхода, каждое ядро служит определенным источником для очередного уровня кластеринга иерархической группировки. Кстати, иерархический кластеринг имеет свое преимущество, в частности, в результате разбиения, дается возможность общего визуального представления о стратификационной структуре совокупности обрабатываемых данных (дендрограмма), в которой показано исчерпывающее описание взаимосвязи разбиения на разных слоях, оценка степени принадлежности каждого элемента к фиксированному классу. Наблюдая над иерархической структурой, можно раскрыть некоторые отношения, а в результате и неизвестные свойства исследуемых данных.

Обратим внимание на часто встречающиеся операции и символы общей алгебры, т.к. вызывает интерес какие идеи и конструкции абстрактной (общей) алгебры и на каких именно фазах алгоритмов используются в математической теории распознавания. В работе [1] приводятся выводы общей теории распознающих алгоритмов. Оказалось, что множество распознающих алгоритмов является алгеброй, причем операции этой алгебры обладают набором свойств, позволяющих детально изучить множество распознающих алгоритмов.

Алгебраические методы позволяют эффективно решить задачу выбора экстремального алгоритма.

Мы показали пример использования в кластер-анализе алгебраического подхода (подходящим и наглядным примером безусловно является настоящий алгоритм). Но можно поступить и обратно - пользоваться идеями и понятиями классификации в общей алгебре. Напомним некоторые определения (см. [2]). Отношение р называется эквивалентностью, если оно удовлетворяет условиям рефлексивности, симметричности и транзитивности. Если р эквивалентность на множестве А, $a \in A$, то подмножество

$$\rho(a)=\{x \,/ x \in A,\, a\rho x\} \tag{1}$$

называется классом эквивалентности. Множество различных классов эквивалентности является разбиением множества *A* и наоборот. Множество классов эквивалетности ρ на множестве *A*, или что то же самое, множество подмножеств, образующих соотвествующие разбиения, называется фактор- множеством множества *A* и обозначается через *A*/*p*.Приведем теперь конкретные примеры конкретных универсальных алгебр. Пусть *S* полугруппа. Бинарное отношение на *S* называется стабильным слева (справа), если для любых $a,b,c \in S$, из условия $a\rho b$ следует $ca\rho cb(ac\rho bc)$. Отношение стабильное слева и справа, называется стабильным отношением. Стабильную эквивалентность определим как конгруэнцию. Эквивалентность на полугруппе будет стабильной тогда и только тогда, когда для любых $a_1, a_2, b_1, b_2 \in S$ из условий $a_1\rho b_1$ и $a_2\rho b_2$ следует $a_1\, a_2\;\; \rho b_1\, b_{\,2}$ Докажем указанную теорему.

Пусть *S* полугруппа (коммутативная), и на *S* задана конгруэнция. Тогда для $a_{\,1} a_2 b_1 b_2 \in S$ имеет место

$$\begin{cases} a_1\rho b_1 \\ a_2\rho b_2 \end{cases} \tag{2}$$

Из свойства рефлексивности следует

$$a_2\rho a_2\,. \tag{3}$$

Рассмотрим систему отношений

$$\begin{cases} a_1\rho b_1 \\ a_2\rho b_2. \end{cases} \tag{4}$$

Так как полугруппа стабильна, поэтому

$$a_1 a_2 \rho b_1 a_2 \tag{5}$$

Условия коммутативности отношения (5) превращается в отношение:

$$a_1 a_2 \rho a_2 b_{1_1}$$

Рассмотрим отношение

$$a_1\rho b_1\,.$$

Это отношение стабильно. Отсюда следует, что оно стабильно справа

$$a_1 b_2 \rho b_1 b_2\,.$$

И, наконец, из свойства транзитивности имеем

$$a_1 a_2 \rho a_1 b_2 \quad \text{и} \quad a_1 b_2 \rho b_1 b_2\,,$$

тогда

$$a_1 a_2 \rho b_1 b_2\,.$$

В результате получим

$$a_1 a_2 \rho b_1 b_2\,.$$

Возьмем обратное условие - пусть справедливо следующее отношение, если $a_1\rho b_1$ и $a_2\rho b_2$, тогда

$$a_1 a_2 \rho b_1 b_2 . \qquad (6)$$

Докажем, что эквивалентность на полугруппе стабильна.

Из свойства рефлексивности имеем

$$b_2 \rho b_2 . \qquad (7)$$

Рассмотрим систему отношений

$$\begin{cases} a_1 \rho b_1 \\ b_2 \rho b_2 . \end{cases}$$

Из условии теоремы следует

$$a_1 b_2 \rho b_1 b_2 \qquad (8)$$

Обьединим (6) и (8), и используя свойства транзитивности, получим доказательство теоремы, если $a_1 a_2 \rho b_1 b_2$ и $b_1 b_2 \rho a_1 b_2$, то $a_1 a_2 \rho a_1 b_2$.

В результате имеем $\begin{cases} a_1 \rho a_1 \\ a_2 \rho b_2 \end{cases}$, но тогда $a_1 a_2 \rho a_1 b_2$ (6), поэтому эквивалентность стабильна слева.

Аналогично доказывается стабильность справа. Если ρ конгруэнция на полугруппе S, то фактор множества обозначается как S/ρ и само это отношение является полугруппой.

Подобное определение и осмысление конгруэнции приводятся и в других универсальнх алгебрах [3], в частности - решетки, что согласуется с общим понятием конгруэнции.

Введем следующие обозначения

$$a \wedge b = \inf\{a, b\}$$
$$a \vee b = \sup\{a, b\}$$

и будем называть операцию $\wedge$ пересечением, а $\vee$-обьединением. В решетках они также являются бинарными операциями и, будучи примененные к паре $a, b \in L$, снова дают элемент из решетки L, Отношение эквивалентности θ на решетке L называется конгруэнцией, если из соотношений:

$$\{a_1 \equiv b_1(\theta)$$
$$\{a_2 \equiv b_2(\theta)$$

следует:

$$a_1 \wedge a_2 \equiv b_1 \wedge b_2(\theta)$$
$$a_1 \vee a_2 \equiv b_1 \vee b_2(\theta)$$

Если $a \in L$, то через $[a]\theta$ обозначим смежный класс, содержащий элемент a:

$$[a]\theta = \{\ x \mathrm{I} x \equiv a(\theta)\ \}.$$

Пусть L решетка, θ конгруэнция на ней. Обозначим через L/θ множество смежных классов разбиения решетки *L*, инициированный конгруэнцией θ:

$$\mathrm{L}/\theta = \{[a]\theta \ \mathrm{I}\ a \in L\}.$$

Если определим операции $\wedge$ и $\vee$ на L/θ тогда, как и в случае полугрупп, L/θ сама будет решеткой. L/θ множество называется факторрешеткой решетки L по конгруэнции θ. Изложение примеров использования понятий классификаций в общей алгебре можно продолжить: изучение внутренней структуры универсальных алгебр путем разбиения универсальных алгебр соответствующего типа, смежные классы по некоторой подалгебре, отображение φ однотипных универсальных алгебр, где φ гоммоморфное наложение, Kerφ-ядро гомоморфизма, а A/Kerφ факторалгебра. Все эти примеры ведут к сближению таких двух направлении науки, как теоретическая - в виде общей алгебры и прикладная в виде раздела искусственного интеллекта-распознавания образов (цель распознавания в общих чертах та же - разработать принципы и методы классификации). В результате появления общеалгебраического языка в математической теории распознавания, в частности, в методе кластеризации на основе монотонных систем, есть надежда, что это будет способствовать взаимообогощению и развитию этих направлений науки.

Что касается включения в метод кластеринга на основе монотонных систем и алгебрайческого подхода элементов теории графов [4], причина этого состоит в идентичности построения графа π близости, с одной стороны, и весовой функции монотонных систем парных связей - с другой.

Основная часть

Приступая к основной части статьи, приведем некоторые определения из теории графов, что поможет нам в определении графа π близости. Пусть V –непустое множество, $V^{(2)}$ - множество всех его двухэлементых подмножеств. Пара (V,E), где E произвольное подмножество множества $V^{(2)}$ называется графом. Множество $V=\{v_1,v_2,...v_n\}$, именуемое как множество вершин, а $E=\{e_{12},e_{13},e_{ij}...e_{nn-1}\}$ - как множество ребер. Если мощность $|E|=m$, то $m\le n(n-1)/2$. Каждое ребро e_{ij} есть определенная пара вершин $v_i,v_j\in V$, что выражается равенством: $e_{ij}=v_iv_j$. Пусть $m=n(n-1)/2$, тогда граф $P=(V,E)$ является полным, и если кроме этого E обладает отношением порядка:

$$e_{12}\le e_{13}\le...\le e_{ij}\le...\le e_{nn-1},$$

то граф P становится упорядоченным графом.

Определение. Графом π близости называется упорядоченный граф $P=(V,E)$, у которого $V=\{v_1,v_2,...v_n\}$ - множество объектов (вершин), подвергаемых кластерингу, и $E=\{e_{12},e_{13},e_{ij}...e_{nn-1}\}$ - множество пар объектов (ребер), называемых связями, величина которых определяется данными о π близости между объектами v_i и v_j следующей формулой:

$$e_{ij}^{\pi}=\sum_{\substack{(i,\alpha)\in E_{index}\\ \alpha\ne j}}\rho_{i\alpha}+(n+k)\rho_{ij}+\sum_{\substack{(\beta,j)\in E_{index}\\ \beta\ne i}}\rho_{\beta j}, \tag{9}$$

где ρ_{ij} - расстояние между объектами v_i и v_j $(v_i,v_j\in V)$ из метрического пространства, k - положительное (неотрицательное) целое число $(k\ge 0)$, E_{index} - совокупность указателей (индексов) всех пар объектов из множества E.

Что касается упорядоченности графа, то под этим словом подразумевается следующая реберная цепь (линейный порядок):

$$e_{12}^{\pi}\le e_{13}^{\pi}\le...e_{ij}^{\pi}\le...e_{nn-1}^{\pi}$$

Отношение порядка определяется данными о π близости для пар объектов. Отношение $e_{ij}^{\pi}\le e_{2S}^{\pi}$ означает, что объекты v_i и v_j не менее сходны, чем v_r и v_s $\left(e_{ij}^{\pi}=v_iv_s,e_{rs}^{\pi}=v_rv_s\right)$. Тождественная запись $e_{ij}^{\pi}\equiv e_{rs}^{\pi}$ означает, что могущие быть различными связи e_{ij}^{π} и e_{rs}^{π} имеют тот же порядок, что и в равенстве $e_{ij}^{\pi}=e_{rs}^{\pi}$, означающем, что e_{ij}^{π} и e_{rs}^{π} одинаковые связи, т.е. $ij=rs$.

Уровнем стягивания[1] графа π близости $P=(V,E)$ называются уровни $S=0,S=m=|E^n|$, а также все S, $1\le S\le m-1$, для которых $\left(e_{ij}^{\pi}\right)_s<\left(e_{rq}^{\pi}\right)_{s+1}$. Для каждого

[1] Стягивание ребра *uv* означает отожествление вершин *u* и *v*. Двойственной к операции стягивания является операция расщепления (см. [1]).

уровня стягивания $m \geq S \geq 0$ с $E_s^{\pi} = \left\{ \left(e_{ij}^{\pi} \right)_k \right\}_{k=\overline{m,s}}$, упорядоченный граф $P = \left(V, E_S^{\pi} \right)$, где E_S^{π} обладает тем же отношением порядка, что и E^{π}, но и в пределах E_S^{π}, является подграфом π близости S–го порядка для P. Граф $T_S = \left(V, E_S^{\pi} \right)$, где E_S^{π} не предполагается упорядоченным, называется пороговым подграфом S-го порядка для P.

Обратимся теперь к теории монотонных систем. Обозначим через W множество парных связей объектов из конечного множества X. Пусть H подмножества $W(H \subseteq W, W = \{(i_1, j_1), (i_2, j_2), ..., (i_n, j_n)\})$. Сравнивая мощности множества W и её подмножества H, если $|E| = h$, то

$$h \leq n(n-1)/2.$$

Введем скалярную функцию π; каждой паре $((i, j), H)$, где $(i, j) \in H$, а H произвольное подмножество $W(H \subseteq W)$ поставлено в соответствие число $\pi((i, j), H)$. Семантически это число означает близость парного элемента (i, j) к подмножеству H, или важность, вес элемента (i, j) в подмножестве H.

Таким образом, мы получаем систему $\langle W, \pi \rangle$, составленную из конечного множества парных элементов W, с заданной на произвольном подмножестве $H(H \subseteq W)$ функцией $\pi((i, j), H)((i, j) \in H)$.

Если система $\langle W, \pi \rangle$ удовлетворяет условиям:

$$\pi((i, j), H \mid (r, s)) \leq \pi((i, j), H), \ \forall (i, j), (r, s) \in H, \ (i...i) \neq (r, s), \ \forall H \in W$$

или

$$\pi((i, j), H \mid (r, s)) \geq \pi((i, j), H), \ \forall (i, j), (r, s) \in H, \ (i...i) \neq (r, s), \ \forall H \in W,$$

то система $\langle W, \pi \rangle$ называется монотонной, соответственно $(-)$ или $(+)$ типа. В дальнейшем мы будем иметь дело с монотонной системой $(-)$ типа.

Функцию $\pi((i, j), H)$ можно задать аналитическим способом следующей формулой:

$$\pi((i, j), H) = \sum_{\substack{(i,j) \in H \\ \alpha \neq j}} \rho_{i\alpha} + (n+k)\rho_{ij} + \sum_{\substack{(\beta,j) \in H \\ \beta \neq i}} \rho_{\beta j}, \tag{10}$$

где ρ - расстояние между элементами i и j, а k неотрицательное целое число $(k \geq 0)$.

Докажем, что система $\langle W, \pi \rangle$, где функция π определяется равенством (2), является монотонной.

Действительно:

$$\begin{aligned} \pi((i, j), H) &= \sum_{\substack{(i,j) \in H \\ \alpha \neq j}} \rho_{i\alpha} + (n+k)\rho_{ij} + \sum_{\substack{(\beta,j) \in H \\ \beta \neq i}} \rho_{\beta j} \geq \\ &\geq \sum_{\substack{(i,\alpha) \in H \setminus (r,s) \\ \alpha \neq j}} \rho_{i\alpha} + (n+k)\rho_{ij} + \sum_{\substack{(\beta,j) \in H \setminus (r,s) \\ \beta \neq i}} \rho_{\beta j} = \pi((i, j), H \setminus (r, s)). \end{aligned} \tag{11}$$

В каждом метрическом пространстве вместе с множеством элементов задается также расстояние. По определению расстояние – это однозначная, неотрицательная действительная функция ρ, определенная для любых x, y из заданного непустого множества и подчиненная следующим трем аксиомам:

1) $\rho(x, y) = 0$ тогда и только тогда, когда $x = y$;

2) $\rho(x, y) = \rho(y, x)$ - аксиоме симметрии:;

3) $\rho(x,z) \le \rho(x,y) + \rho(y,z)$ - аксиоме треугольника:.

Учитывая свойство неотрицательности функции ρ, становиться ясным, что удалением парного элемента (r,s) из подмножества H, сумма в выражении (3)-м становится не меньше, чем в начале.

В результате получим:

$$\pi\big((i,j), H \setminus (r,s)\big) \le \pi\big((i,j), H\big).$$

Монотонность системы доказана.

Наблюдение за вышеуказанными построениями по теории графов и по теории монотонных систем наводит на мысль некоторых ассоциаций, в частности, в определении монотонных систем фигурирует множество W - всех пар элементов из $X\,(X \neq \varnothing)$. По содержанию значение множества $E(.)$ в определении графа совпадает со значением W в определении монотонных систем. Сопоставляя правые части (1) и (2), сразу можно заметить тождественность формулировки с одной стороны величины π близости в графах близости, а с другой - аналитическим заданием весовой функции в теории монотонных систем. Исходя из этих смысловых сравнений, можно записать:

$$e_{ij}^{\pi} = \pi\big((i,j), W\big),\ (i,j) \in W,\ e_{ij}^{\pi} \in E^{\pi}. \tag{12}$$

Графическое преобразование монотонных систем, как будет показано в дальнейшем, является весьма весовым в разработке алгоритма кластеринга, представляя тем самым интересный пример использованная теории графов в области науки управления.

Рассмотрим монотонную систему $\langle W, \pi \rangle$ парных элементов связей. Каждому парному элементу $(i,j) \in W$ поставим в соответствие весовую функцию $\pi\big((i,j), W\big)$. Обозначим через S множество, состоящее из всех элементов

$$\pi\big((i,j), W\big)\Big(S = \{(i,j), W\}_{i,j=\overline{1,t}, i \neq j};$$
$$X = \{1,2,\ldots n\}, |X| = n, t = n(n-1)/2\Big)$$

где t мощность множества S.

Покажем, что S является алгебраической решеткой. Будет вполне удобным, если все элементы, принадлежащие множеству S заменим малыми латинскими буквами:

$$\pi\big((i_k, j_k), W\big) = a,\ \pi\big((i_e, j_e), W\big) = b\ldots$$

Введем на множестве S бинарную коммутативную операцию умножения, удовлетворяющую ассоциативному закону: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, так как S - непустое множество, то она становится мультипликативной полугруппой.

Если каждый элемент полугруппы является идемпотентным $(a \cdot a = a, \forall a \in S)$, то полугруппу S именуют как полугруппу идемпотентов, или связкой (см. [3]).

Во многих случаях полезную роль играет отношение естественного частичного порядка на S, задаваемое условием:

$$a \le b,\ \text{если}\ a \cdot b = b \cdot a = a. \tag{13}$$

Докажем справедливость (5).

Чтобы полугруппа была частично упорядоченной, она должна удовлетворять следующим трем условиям:

- рефлективности;
- антисимметричности;
- транзитивности.

Так как полугруппа S является связкой (полугруппа идемпотентов), поэтому свойство рефлективности очевидно.

Свойство антисимметричности:

$$\text{если } a \le b \text{ и } b \le a \Rightarrow a = b.$$

Действительно: $a \le b$ означает

$$a = a \cdot b; \tag{14}$$

аналогично $b \le a$ означает

$$b = b \cdot a. \tag{15}$$

Объединяя (6) и (7) получим следующую цепочку соотношений:

$$a = a \cdot b = b \cdot a = b$$

Наконец свойство транзитивности.

Докажем, что если $a \le b$ и $b \le c$, то $a \le c$. Из отношения $a \le b$ вытекает

$$a = a \cdot b. \tag{16}$$

Из отношения $b \le c$ выводим

$$b = b \cdot c. \tag{17}$$

Подставляя значение b из (9) в равенстве (8):

$$a = a \cdot b = a \cdot b \cdot c = b \cdot c. \tag{18}$$

Но (10) в свою очередь означает $a \le c$, что и требовалось доказать.

Определим операцию умножения по правилу:

$$a \cdot b = a \wedge b = \inf\{a, b\},$$

тогда коммутативная связка превращается в нижнюю полурешетку и если S является аддитивной полугруппой и примем, что

$$a + b = a \vee b = \sup\{a, b\},$$

то S будет верхней полурешеткой. Нижняя и верхняя полурешетки вместе составляют решетку S.

Из последнего логического вывода можно сделать общее заключение, что полугруппа S, а следовательно заданное множество S, содержащее элементы $\pi\big((i,j),W\big),(i,j)\in W$, преобразуется в решетку.

Теперь приведем некоторые необходимые сведения из теории решеток. В теореме о гомоморфизме для решеток** фигурирует отображение $\varphi_\theta : x \to [x]\theta$.

Лемма 10. Отображение $\varphi_\theta : x \to [x]\theta \ (x \in L)$ является гомоморфизмом из решетки L на L/θ (см. [4]).

Постараемся интерпретировать эту информацию для нашего случая (для решетки S).

Из следствия 2*** можно сделать заключение, что полуоткрытые интервалы:

$$\left[\Gamma_q^W, \Gamma_q^W / \Gamma_{q+1}^W\right)_{\substack{q=\overline{1,P_R}\\ P_R \le p}}$$

являются смежными классами факторрешетки S/θ. Эти классы содержат значения функций $\pi\big((i,j),W\big)$ для всех пар $(i,j)\in W$ (напомним, что $\langle \Gamma_1, \Gamma_2, ..., \Gamma_p \rangle$ - подпоследовательность последовательности $H = \langle H_1, ..., H_\wedge \rangle$, вложенных подмножеств множества W, где $\Gamma_1 = H_1 = W$, $H_2 = H_1 \setminus (i_1, j_1)$, $...H_{K+1} = H_K \setminus (i_{K+1}, j_{K+1})$, $...H_N = (i_N, j_N)$ соответствующий максимальной определяющей последовательности парных элементов заданного множества

** Теорема о гомоморфизме. Любой гомоморфный образ решетки L изоморфен подходящей фактор-решетки L. В действительности, если $\varphi : L \to L_1$, гомоморфизм из L на L_1 и θ конгруэнция на L, такая что $x \equiv y(\theta)$, тогда и только тогда, когда $x\varphi = y\varphi$, то $L/\theta \cong L_1$ и изоморфизм может быть задан следующим образом: $\psi : [x]\theta \to x\varphi(x \in L)$ (знак $\equiv$ означает эквивалентность, а знак $\cong$ - изоморфизм) (см. [4]).

*** Весовое значение каждого элемента множества S содержится в одной из последовательности интервалов следующего вида: $\left[\Gamma_q^W, \Gamma_q^W / \Gamma_{q+1}^W\right)_{\substack{q=\overline{1,P_R}\\ P_R \le p}}$.

(X) - $\{(i_1, j_1)...(i_N, j_N)\}$. Слова "максимальный" означает максимальный относительно включения элементов Γ_i, т.е. не содержащихся в последовательности $\langle \Gamma_1, \Gamma_2, ..., \Gamma_p \rangle$ с большим числом элементов, удовлетворяющих условию:

$$\pi\left(\left(i_k, j_i\right), H_K\right) > F\left(\Gamma_{j+1}\right),\ \forall\left(i_k, j_k\right) \in \Gamma_j \setminus \Gamma_{j+1},\ N = \frac{n(n-1)}{2}. \qquad (19)$$

Вследствие этого, теорема о гомоморфизма в нашем частном случае можно изобразить графически (в виде диаграммы).

Здесь (рис. 1) каждый смежный класс в факторрешетке представлен как $\pi\left[\left(\left(i_k, j_i\right), W\right)\theta\right.$, где $(i, j) \in \left[\Gamma_q^W, \Gamma_q^W / \Gamma_{q+1}^W\right)$, а $q = \overline{1, P_R}$.

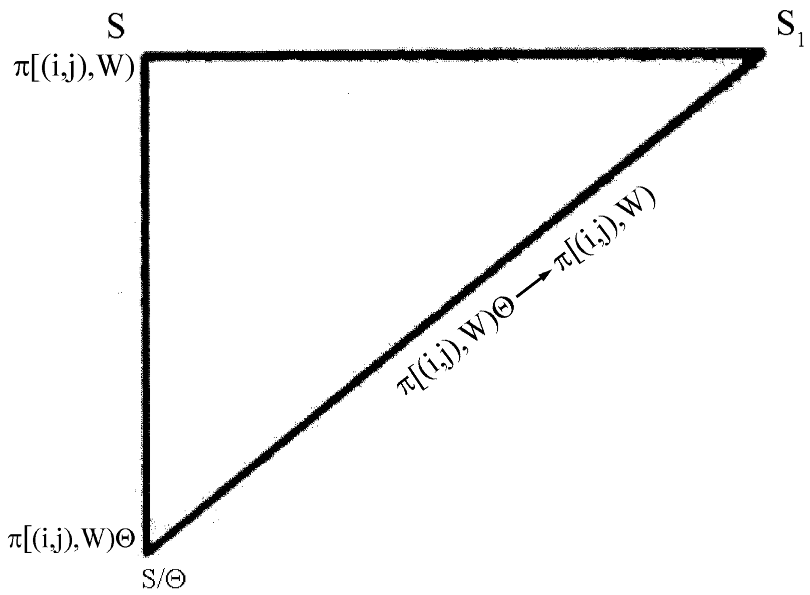

Рис. 1.

Пусть η множество всех факторрешеток, а τ множество графов близости.

Введем отображение $f: \eta \to \tau$ из множества факторрешеток в множество графов близости. По определению отображения, каждому элементу x множества η (факторрешеток) сопоставляется однозначно определенный y - элемент другого множества τ (графа близости). Такое соотношение между элементами η и τ записывается в виде $y = f(x)$.

Из дедуктивного правила можно записать:

$$f^*\left(\left[\left[\pi(i, j)_{\Gamma_q^W}, W\right)\theta\right) = \bigvee_{g=0}^{P_R - q}\left(\neg\left(\bigwedge_{t=g+p}^{P_R}\left[\pi(i, j)_{\Gamma_q^W}, W\right)\theta\right)\right). \qquad (20)$$

Конкретное отображение соответствует конкретному подмножеству множества факторрешеток – факторрешетки решетки, инициированные монотонными системами парных элементов связей (область определения отображения) и, продолжая начатую мысль, соответствует также конкретному подмножеству множества графов близости – графов π близости (область значения отображения). Так как факторрешетка сама является сложным понятием, состоящим из смежных классов, а графы π близости сами содержат подграфы π близости, то в итоге получим отображение из множества смежных классов факторрешетки вышеуказанного типа (область определения отображения) множества подграфов π близости графа π близости. Функциональная зависимость подграфа $T_{\left|E^\pi\right| - \left|\Gamma_q^W\right|}$ графа π близости от смежного класса $\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta$ факторрешетки S/θ решетки S принимает следующий аналитический вид:

$$f^*\left(\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)=\bigvee_{g=0}^{P_R-q}\left(\dot{-}\left[\bigwedge_{t=p+g}^{P_R}\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)=T_{\left|E^N\right|-\left|\Gamma_q^W\right|} \qquad (21)$$

где $\left|E^\pi\right|$ и $\left|\Gamma_q^W\right|$ мощности соответственно множества E^π и элемента из подпоследовательности $\left\langle \Gamma_1^W,\Gamma_2^W,...,\Gamma_P^W\right\rangle$, $-\cdot$ операция удаления, $q=\overline{1,P_R}$, T_q есть q-ый подграф близости.

Покажем истинность соотношения (21).

После преобразования (21) имеем:

$$\begin{aligned}
&f^*\left(\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)=\bigvee_{g=0}^{P_R-q}\left(\dot{-}\left(\bigwedge_{t=p+g}^{P_R}\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)\right)=\left(\left(\dot{-}\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)\wedge\right.\\
&\wedge\left(\dot{-}\left[\pi(i,j)_{\Gamma_{q+1}^W},W\right)\theta\right)\wedge...\wedge=\left(\dot{-}\left[\pi(i,j)_{\Gamma_{P_R}^W},W\right)\theta\right)\Bigg)\wedge\left(\dot{-}\left[\pi(i,j)_{\Gamma_{q+1}^W},W\right)\theta\right)\wedge \qquad (22)\\
&\wedge\left(\dot{-}\left[\pi(i,j)_{\Gamma_{q+2}^W},W\right)\theta\right)\wedge...\wedge=\left(\dot{-}\left[\pi(i,j)_{\Gamma_{P_R}^W},W\right)\theta\right)\Bigg)\wedge...\wedge\left(\dot{-}\left[\pi(i,j)_{\Gamma_{P_R}^W},W\right)\theta\right)=
\end{aligned}$$

В силу того, что множества смежных классов разбиении решетки S индуцированной конгруэнцией $\theta-S/\theta$, определена операция $\wedge$ и имеет место равенство:

$$\left[\pi(i,j)_{\Gamma_k^W},W\right)\theta\wedge\left[\pi(i,j)_{\Gamma_e^W},W\right)\theta=\left[\pi\left((i,j)_{\Gamma_k^W},W\right)\wedge\left[\pi(i,j)_{\Gamma_e^W},W\right)\right)\theta.$$

С учетом закона ассоциативности, в продолжении (22) получим:

$$\begin{aligned}
&=\left[\dot{-}\left(\pi(i,j)_{\Gamma_q^W},W\right)\wedge^{\dot{-}}\left(\pi(i,j)_{\Gamma_{q+1}^W},W\right)\wedge...\wedge^{\dot{-}}\left(\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta\vee\\
&\vee\left[\dot{-}\left(\pi(i,j)_{\Gamma_{q+1}^W},W\right)\wedge^{\dot{-}}\left(\pi(i,j)_{\Gamma_{q+2}^W},W\right)\wedge...\wedge^{\dot{-}}\left(\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta\vee \qquad (23)\\
&\vee...\vee\left(\dot{-}\left[\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta
\end{aligned}$$

Сделаем паузу для комментариев. В правой части (20), кроме решеточно-алгебраических операций $\vee$ и $\wedge$, фигурирует дополнительная операция удаления -·. Поскольку указанная операция представляет собой n-арную алгебраическую операцию, которая отображает множества A^n в A (в нашем конкретном случае в (23) $n=0$, получается нульарная операция удаления, которая отображает A^0 в нейтральном элементе 0 или $\varnothing$ - множество A).****

Отметим, что под словом “нейтральный элемент” подразумевается нейтральный элемент аддитивной универсальной алгебры. По причине того, что для всех членов (23), которые связаны между собой операциями пересечения $\wedge$ и объединения – $\vee$, записана операция удаления -·, результат которой зафиксирован и ее значение во всех случаях не меняет свою величину – $\varnothing$, и нарушение порядка выполнения алгебраических операций – вынос за скобки операции удаления не отразится на результате:

**** Здесь $A^n = AxAx\cdot A$ есть n-ая прямая (декартовая) степень, т.е. множества всевозможных упорядоченных последовательностей $(a_1,a_2,a_3,...,a_n,)$ $a_i\in A$, которою алгебраическая операция переводит в однозначно определяющий элемент A

$$=^{-}\left[\left(\pi(i,j)_{\Gamma_q^W},W\right)\wedge\left(\pi(i,j)_{\Gamma_{q+1}^W},W\right)\wedge\ldots\wedge\left(\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta\vee$$
$$\vee^{-}\left[\left(\pi(i,j)_{\Gamma_{q+1}^W},W\right)\wedge\left(\pi(i,j)_{\Gamma_{q+2}^W},W\right)\wedge\ldots\wedge\left(\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta\vee \qquad (24)$$
$$\vee\ldots\vee^{-}\left[\left(\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta=$$

Используя закон ассоциативности, а также следующее решеточное отношение, которое позволяет выбрать *inf* алгебраическо-решеточного многочлена (см. [3]):

$$(\ldots(a_0\vee a_1)\wedge a_2)\ldots)a_{n-1}=\inf\{a_0,a_1,\ldots,a_{n-1}\}$$

где $a_0,a_1,\ldots,a_{n-1}$ элементы некоторой решетки H, получим:

$$=^{-}\left(\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)\wedge^{-}\left(\left[\pi(i,j)_{\Gamma_{q+1}^W},W\right)\theta\right)\vee\ldots\vee$$
$$\ldots\vee^{-}\left(\left[\pi(i,j)_{\Gamma_{P_R}^W},W\right)\theta\right)=^{-}\left(\left(\left[\pi(i,j)_{\Gamma_q^W},W\right)\theta\right)\vee \qquad (25)$$
$$\vee\left(\left[\pi(i,j)_{\Gamma_{q+1}^W},W\right)\theta\right)\wedge\ldots\wedge\left(\left[\pi(i,j)_{\Gamma_{P_R}^W},W\right)\right)\theta=$$

Теперь посмотрим на поставленную задачу с другой стороны.

Рассмотрим W последовательности квазядер $\overline{\Gamma}^W=\left\langle \Gamma_1^W,\Gamma_2^W,\ldots,\Gamma_i^W,\ldots,\Gamma_{P_R}^W\right\rangle$, которая является последовательностью вложенных множеств:

$$W=\Gamma_1^W\supset\Gamma_2^W\supset\ldots\supset\Gamma_i^W\supset\ldots\supset\Gamma_{P_R}^W.$$

Обозначим через G множества подмножеств W. Определим отношение на G следующим образом:

$$\Gamma_k^W\le\Gamma_e^W,\ \text{если}\ \Gamma_k^W\subseteq\Gamma_e^W \qquad (26)$$

Отношение (26) является порядком.

Действительно, имеют место следующие признаки порядка – отношение на множестве G:

1. Рефлексивное: $\Gamma_k^W\le\Gamma_k^W$;
2. Антисимметричное: если $\Gamma_k^W\le\Gamma_e^W$ и $\Gamma_e^W\le\Gamma_k^W$ то $\Gamma_k^W=\Gamma_e^W$;
3. Транзитивное: если $\Gamma_k^W\le\Gamma_e^W$ и $\Gamma_e^W\le\Gamma_Q^W$, то $\Gamma_k^W\le\Gamma_Q^W$.

Для примера покажем справедливость 1.

$$\Gamma_k^W\subseteq\Gamma_k^W,\ \text{поэтому}\ \Gamma_k^W\le\Gamma_k^W.$$

Можно показать справедливость и остальных двух.

Положим:

$$\Gamma_k^W\le\Gamma_e^W,\ \text{если}\ \Gamma_k^W=\Gamma_k^W\cap\Gamma_e^W. \qquad (27)$$

При этом

$$\Gamma_k^W\cap\Gamma_e^W=\Gamma_k^W\wedge\Gamma_e^W=\inf\left\{\Gamma_k^W,\Gamma_e^W\right\},$$
$$\Gamma_k^W\cup\Gamma_e^W=\Gamma_k^W\vee\Gamma_e^W=\sup\left\{\Gamma_k^W,\Gamma_e^W\right\}.$$

Тогда множество подмножеств G оказывается решеткой.

Определение. Дистрибутивная решетка с дополнениями, т.е. дистрибутивная решетка с 0 и 1, в которой каждый элемент имеет дополнение, называется булевой решеткой (см. [3]).

Очевидно, что (27) (множество G) является булевой решеткой с теоретико-множественными операциями объединения и пересечения. Интересно показать свойство дистрибутивности в нашем случае (27).

Пусть $\Gamma_e^W, \Gamma_m^W, \Gamma_k^W$ подмножества множеств W (элементы множества G), докажем что:

$$\left(\Gamma_e^W \cup \Gamma_m^W\right) \cap \Gamma_k^W = \left(\Gamma_e^W \cap \Gamma_k^W\right) \cup \left(\Gamma_m^W \cap \Gamma_k^W\right).$$

Если $\pi(i,j) \in \left(\Gamma_e^W \cup \Gamma_m^W\right) \cap \Gamma_k^W$, то $\pi(i,j) \in \Gamma_k^W$ и кроме этого $\pi(i,j) \in \Gamma_e^W$ или $\pi(i,j) \in \Gamma_m^W$. Отсюда вытекает, что $\pi(i,j) \in \Gamma_e^W \cap \Gamma_k^W$ или $\pi(i,j) \in \Gamma_m^W \cap \Gamma_k^W$, а значит $\pi(i,j) \in \left(\Gamma_e^W \cap \Gamma_k^W\right) \cup \left(\Gamma_m^W \cap \Gamma_k^W\right)$. Пусть сейчас $\pi(i,j) \in \left(\Gamma_e^W \cap \Gamma_k^W\right) \cup \left(\Gamma_m^W \cap \Gamma_k^W\right)$, тогда $\pi(i,j) \in \Gamma_e^W \cap \Gamma_k^W$ или $\Gamma_m^W \cap \Gamma_k^W$.

Если $\pi(i,j) \in \Gamma_e^W \cap \Gamma_k^W$, тогда $\pi(i,j) \in \Gamma_e^W$ и $\pi(i,j) \in \Gamma_k^W$, отсюда выводим $\pi(i,j) \in \Gamma_e^W \cup \Gamma_k^W$ и в результате $\pi(i,j) \in \left(\Gamma_e^W \cup \Gamma_m^W\right) \cap \Gamma_k^W$. Аналогично доказывается, если $\pi(i,j) \in \Gamma_m^W \cap \Gamma_k^W$. Итак, G является булевой алгеброй, в которой роль 0 играет $\varnothing$, а роль 1 само множество W. Булева решетка сигнатуры $\mathbf{(+,\cdot,0,1)}$ называется булевой алгеброй. В нашем конкретном случае булевой алгеброй будет булева решетка с сигнатурой $(G, \cup, \cap, ', \varnothing, W)$.

Определение. Непустая совокупность подмножеств некоторого множества X называется алгеброй (полем) множеств, если оно замкнуто относительно теоретико-множественных операций $\cup, \cap$ и $'$ (см. [5]).

Очевидно, что множество G (булева алгебра) является алгеброй множеств. В алгебре множеств вводятся некоторые дополнительные операции, среди них симметрическая разность, которая обозначается символом Δ и равна:

$$A \Delta B = \left(A \cap B'\right) \cup \left(A' \cap B\right),$$

где A и B подмножества некоторого непустого множества; буква B со знаком B' означает дополнение множества B.

А теперь представим графическое выражение элементов множества G со своими функциональными значениями:

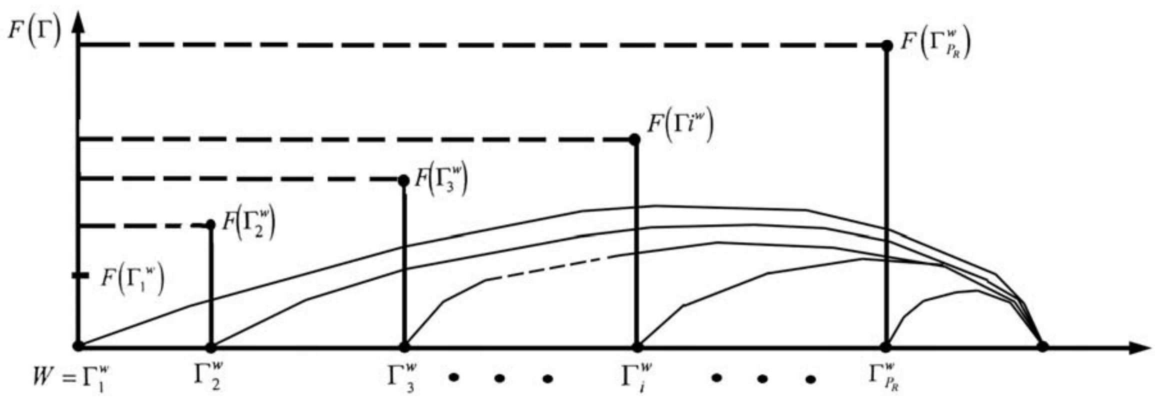

Рис. 2.

Напомним правило определения функционала F. На множестве всех подмножеств W монотонной системы определяется скалярная функция F, ставящая число $F(H)$ в соответствие каждому подмножеству H, по следующему правилу:

$$F(H) = \min_{(i,j) \in H} \pi((i,j), H)$$

Цель задачи - выделение экстремальной подсистемы монотонной системы – наибольшего ядра (наибольшее по значению F).[****].

Исходя из содержания класса факторрешетки по конгруэнции $\theta - \left[\pi(i,j)_{\Gamma_k^W}, W\right)\theta$ и проведении операции симметрической разности, в алгебре множеств G, для произвольного индекса из $k=\overline{1,P_R}$ можно записать:

$$\Gamma_k^W \Delta \Gamma_{k+1}^W = \left(\Gamma_k^W \cap \left(\Gamma_{k+1}^W\right)'\right) \cup \left(\left(\Gamma_k^W\right)' \cap \Gamma_{k+1}^W\right) =$$

$$= \left(\Gamma_k^W \cap \left(\Gamma_{k+1}^W\right)'\right) \cup \varnothing = \Gamma_k^W \cap \left(\Gamma_{k+1}^W\right)' = \left[\pi(i,j)_{\Gamma_k^W}, W\right)\theta.$$

Введем соотношение $\Gamma_{P_{R+1}}^W = \varnothing$, и в силу вложенности множества подмножеств $\left\{\Gamma_k^W, k=\overline{1,P_{R+1}}\right\}$, а также, с учетом того, что множество можно обозначить не только прописными буквами, но и перечнем или фиксированием его элементов, соотношение (25) преобразуется следующим образом:

$$\begin{aligned} &=^{-}\left(\left(\Gamma_q^W \cap \left(\Gamma_{q+1}^W\right)\right)' \cup \left(\Gamma_{q+1}^W \cap \Gamma_{q+2}^W\right)'\right) \cup \ldots \cup \left(\Gamma_{P_R}^W \cap \left(\Gamma_{P_R+1}^W\right)'\right)\Big) = \\ &=^{-}\left(\Gamma_q^W\right) =^{-} \left\{\left(\pi(i,j),W\right) \middle| \pi(i,j) \in \Gamma_q^W\right\} = \end{aligned} \qquad (28)$$

Учитывая (4), а также определение уровня стягивания графа π близости и наконец изоморфизм множеств E^π и W, в заключении имеем:

$$=^{-}\left\{e_{ij}^\pi \middle| e_{ij}^\pi \notin E_{\left|\Gamma_q^W\right|}^\pi\right\} = T_{|W|-\left|\Gamma_q^W\right|} = T_{\left|E^\pi\right|-\left|\Gamma_q^W\right|}, \qquad (29)$$

где T подграф π близости соответствующего уровня, а $|W|$ и $\left|\Gamma_q^W\right|$ мощности указанных множеств.

Чтобы облегчить детальное ознакомление достаточно сложной функции (20), а также разложение аналитической формулы отображения, приведем промежуточные результаты, взятые из программной реализации алгоритма. Мы надеемся, что указанный материал вместе с графическими изображениями создаст читателю наглядное представление о рассматриваемом методе кластеринга.

Пусть X множество объектов $X=\{1,2,3,4,5,6\}$, W множество связей между этими объектами:

$$W = \{(1,2),(1,3),\ldots,(5,6)\}, |W| = \frac{6 \cdot 5}{2} = 15.$$

Значение функции F на ключевых множествах монотонной системы[*****] – квазиядрах $\left\langle \Gamma_1^W, \Gamma_2^W, \ldots, \Gamma_{P_R}^W \right\rangle$ будет соответственно:

$$\{(\pi(1,2),W)=116, (\pi(1,3),W)=196, \ldots, (\pi(5,6),W)=305\}.$$

Вследствие этого и определенных вычислительных процедур, получим последовательность смежных классов факторрешетки, инициированной конгруэнцией θ:

[****] Ядрами называют такие подмножества множества W, на которых достигается максимум функции $F(H)$.

[*****] Заданная весовая функция имеет следующий вид: $\pi((i,j),H) = \sum_{\substack{(i,\alpha)\in H \\ \alpha \neq j}} \rho_{i\alpha} + (6+14)_{\rho_{ij}} + \sum_{\substack{(\beta,j)\in H \\ \beta \neq i}} \rho_{\beta j}.$

$$\left[\pi(i,j)_{\Gamma_9^w},W\right)\theta;\left[\pi(i,j)_{\Gamma_8^w},W\right)\theta;\left[\pi(i,j)_{\Gamma_7^w},W\right)\theta;$$
$$\left[\pi(i,j)_{\Gamma_6^w},W\right)\theta;\left[\pi(i,j)_{\Gamma_5^w},W\right)\theta;\left[\pi(i,j)_{\Gamma_4^w},W\right)\theta; \qquad (30)$$
$$\left[\pi(i,j)_{\Gamma_3^w},W\right)\theta;\left[\pi(i,j)_{\Gamma_2^w},W\right)\theta;\left[\pi(i,j)_{\Gamma_1^w},W\right)\theta.$$

Графическое изображение смотрите на рис. 3, где в качестве области значений соответствующей последовательности смежных классов факторрешетки S/θ (30) (область определения функции f^*) фигурирует последовательность подграфов близости:

T_{15-4}; T_{15-5}; T_{15-6}; T_{15-9}; T_{15-10}; T_{15-11}; T_{15-13}; T_{15-14}; T_{15-15} или

T_{11}; T_{10}; T_9; T_6; T_5; T_4; T_2; T_1; T_0.

Заключение

В работе вводится понятие графов близости; строится весовая функция элементов монотонных систем парных связей; задается отображение из множества факторрешеток в множество графов близости. Затем отображение принимает конкретный вид - . К работе прилагается наглядный пример, который проверяет и подтверждает правильность теоретических выводов.

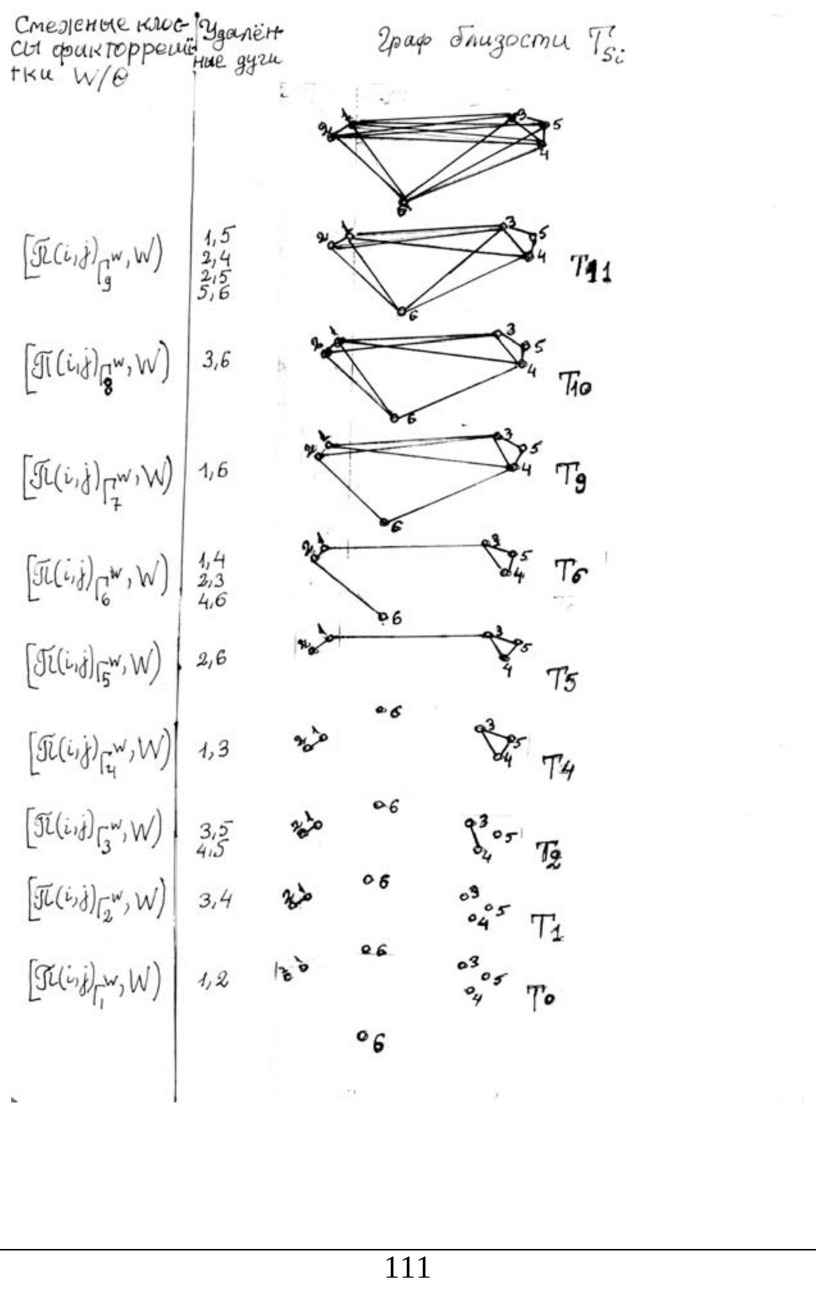

Рис. 3

Цитированная литература

1. Журавлев Ю.И. Распознание – Классификация – Прогноз. М.:Наука, 1988 г.

2. Артамонов В.А., Салий В.Н., Скорняков Л.А., Шеврин Л.Н., Шульгейфер Е.Г. Общая алгебра. М.: Наука, т. 2. 1991 г.

3. Гретцер Г. Общая теория решеток. М.: Наука, 1982.

4. Емеличев В.А., Мельников О.И., Сарванов О.И., Тишкевич Р.И. Лекции по теории графов. М.: Наука, Физматлит. 1990 г.

5. Математический энциклопедический словарь. М.: Советская энциклопедия, 1988.

Статья получена: 2008-08-09