УДК 519.1

Н.Г. Мустафин, А.В. Пономарев, С.В. Савосин

МЕТОД ОПТИМАЛЬНОЙ ГРУППИРОВКИ ВЕКТОРНЫХ ОБЪЕКТОВ ОТНОСИТЕЛЬНО ЦЕНТРОВ

Мустафин Н.Г., Пономарев А.В., Савосин С.В. **Метод оптимальной группировки векторных объектов относительно центров.**
Аннотация. Рассматривается задача группировки векторных объектов относительно возможных центров с учетом ограничений на состав групп. Предлагается способ представления специфических векторных ограничений в виде ограничений эквивалентной задачи целочисленного программирования и полиномиальные алгоритмы для некоторых частных случаев.
Ключевые слова: алгоритм, группировка, целочисленное программирование, поток в сети, матроид.

Mustafin N.G., Ponomarev A.V., Savosin S.V. **A method for optimal grouping of vector objects around centers.**
Abstract. The paper addresses the problem of grouping vector objects around potential centers with respect to restrictions imposed on the group structure. A method for transforming vector restrictions into the restrictions of equivalent integer programming problem is proposed. Polynomial algorithms for some special cases are suggested.
Keywords: algorithm, grouping, integer programming, network flow, matroid.

1. Введение. Значительное количество задач, возникающих в научной и практической деятельности человека, имеют в своей основе общую идею: образование групп из заданных объектов при определенных условиях. Многие из этих задач успешно решаются с применением тех или иных методов и моделей – теории графов, кластерного анализа, математического программирования. Вместе с тем, с одной стороны, почти нет работ, в которых бы предпринимались попытки осознать и описать внутреннее родство таких задач, выделить общие идеи, подходы и методы, с другой — в литературе уделяется мало внимания специфическим вопросам группировки векторных объектов, описываемых набором разнородных характеристик.

Требования, предъявляемые к структуре формируемых групп, в значительной мере определяют методы, с помощью которых производится группировка. В данной статье рассматривается частный случай группировки — группировка относительно центров, то есть исходными являются два множества объектов C и S — объекты первого типа играют роль центров, относительно которых могут объединятся объекты второго типа, сателлиты.

2. Постановка задачи. Пусть каждый из объектов множества S описывается набором из k атрибутов: $s(a_1, \ldots, a_k) \in S$, $a_i \in \mathfrak{D}_i^S$. Таким образом, можно говорить об объектах множества S как о точках в k–мерном пространстве. Через $a_i(s)$ будем обозначать значение атрибута a_i объекта $s \in S$. Аналогично, каждый из объектов множества C описывается набором l атрибутов: $c(b_1, \ldots, b_l) \in C$, $b_i \in \mathfrak{D}_i^C$.

Дадим формальное определение задачи группировки относительно центров. Под группой G будем понимать пару множеств $\langle C_G, S_G \rangle$, таких, что $|C_G| = 1, C_G \subseteq C$, $S_G \subseteq S$. Для обозначения компонент группы будем пользоваться функциональной нотацией: $C(G) \equiv C_G$, $S(G) \equiv S_G$.

Группировку $\mathfrak{G}$ на семействе исходных множеств определим как множество попарно не пересекающихся групп. То есть,

$$\mathfrak{G} = \{G\},$$
$$\forall (G_1, G_2) | G_1, G_2 \in \mathfrak{G} \rightarrow C(G_1) \cap C(G_2) = \varnothing, S(G_1) \cap S(G_2) = \varnothing.$$

Введем характеристику *покрытия* группировкой $\mathfrak{G}$ исходных множеств. Группировка будет называться *полной по* C, если

$$\bigcup_{g \in \mathfrak{G}} C(g) = C.$$

Группировка будет называться *полной по* S, если

$$\bigcup_{g \in \mathfrak{G}} S(g) = S.$$

Группировка будет называться *полной*, если она является полной по C и полной по S. Наконец, группировка, не являющаяся ни полной по C, ни полной по S, называется *неполной*.

О поиске оптимальной группировки можно говорить только в том случае, если качество каждой группировки может быть оценено. Для этого используется система показателей, задающая отображение $val \colon \mathfrak{G} \mapsto \mathbb{R}$. Системы показателей могут очень различаться, и от структуры системы показателей в существенной мере зависит подход к решению задачи группировки. В этой статье рассматривается группировка с простейшей системой показателей.

Под простейшей системой показателей понимается то, что определено одно отображение $D\colon S \times C \mapsto \mathbb{R}$, задающее численный эффект от присоединения заданного сателлита к группе с заданным центром. При этом оценка группы формируется как простая сумма соответствующих значений сателлитов:

$$val(G) = \sum_{s \in S(G)} d(s, C(G)),$$

а оценка всей группировки — как сумма оценок групп:

$$val(\mathfrak{G}) = \sum_{G \in \mathfrak{G}} val(G).$$

Предполагается, что наилучшим решением задачи является группировка с наибольшей оценкой.

3. Переход от векторного описания объектов к описанию взаимосвязей.

Ограничение на разброс значений атрибута в группе. В ряде случаев при постановке задачи векторной группировки необходимо наложить ограничение, заключающееся в том, что разброс значений какого-то атрибута (a_r) для всех объектов, входящих в любую группу, не может превышать определенного значения (Δ_r).

$$\forall G \in \mathfrak{G} : \max_{s \in G} a_r(s) - \min_{s \in G} a_r(s) \leqslant \Delta_r$$

Ограничения подобного вида можно представить в виде дополнительных линейных ограничений к рассмотренной выше базовой постановке задачи группировки как задачи линейного программирования, если свести их к бинарному отношению, задающему запрет на одновременное вхождение объектов в группу. Введем на множестве $S \times S$ бинарное отношение R^{Δ_r}, определяемое следующим образом:

$$\mathrm{R}^{\Delta_r}(x, y) \equiv \{\langle x, y\rangle \colon (|a_r(x) - a_r(y)| > \Delta_r), x, y \in S\} \qquad (1)$$

Это отношение задает множество всех пар объектов, которые не могут находиться в одной группе.

Ограничение на состав группы. Определение группы может содержать в себе ограничения, связанные со структурой группы, например:

- требование того, чтобы в группе было более m_r элементов, имеющих значение атрибута a_r более заданного и ровно m_s элементов, имеющих значение атрибута a_s равное заданному;
- требование того, чтобы элементов группы, имеющих определенное значение атрибута a_r было больше, чем элементов, имеющих определенное значение атрибута a_s;
- требование того, чтобы элементов группы, имеющих определенное значение атрибута a_r, было больше половины всех элементов группы;
- и т.д.

Во всех приведенных ограничениях есть два общих момента:

- с помощью наложения условия на значение атрибутов задается один или несколько классов (подмножеств) исходных элементов;
- устанавливается ограничение на количественное соотношение внутри группы элементов, принаджлежащих различным классам.

Задание класса элементов есть ни что иное, как определение унарного отношения на S (отношения принадлежности классу). Так, требование определенного значения (v_r) атрибута a_r эквивалентно определению отношения:

$$\mathrm{R}^{=v_r}(s) \equiv \{s\colon a_r(s) = v_r,\ s \in S\}.$$

Принципы формирования отображения D. Отображение D в простейшей системе показателей задает эффект в целевой функции от присоединения сателлита к узлу с определенным центром. Если рассматриваются векторные объекты, то, в общем случае, D — это какая-то функция над $k+l$ переменными. Смысл этой функции полностью зависит от интерпретации задачи группировки, однако при ее задании могут особым образом обрабатываться запреты на присоединение s к c.

На практике часто возникает случай, когда S и C являются разными типами объектов, принадлежащих одному метрическому пространству (например, находящихся на поверхности Земли). Тогда существуют такие проекции s_p и c_p, представляющие собой координаты объектов в данном пространстве. В этом случае отображение D может являться некоторой функцией от ρ — меры данного пространства.

4. Постановка в виде задачи целочисленного программирования. Для того, чтобы сформулировать задачу группировки как задачу математического программирования, нужно показать способ перехода от терминологии описания задачи группировки (класс ОГ, система показателей) к трем составляющим задачи МП: набору переменных, набору ограничений на значения переменных и целевой функции, оптимальное значение которой должно достигаться искомой программой.

Обозначим мощность множества S через n, а множества C через m и введем набор переменных $x_{ij}, i \in \{1, \ldots, n\},\ j \in \{1, \ldots, m\}$. Переменная x_{ij} соответствует факту вхождения элемента множества S с индексом i в группу, центром которой является элемент множества C с индексом j, и может принимать значения из множества $\{0, 1\}$. Отметим, что множества S и C здесь упоминаются и как множества, и как последовательности, допускающие обращение к элементу по порядковому номеру. Эта двойственность выбрана в качестве компромисса, дабы не вводить дополнительные обозначения для последовательностей, поскольку сама последовательность может формироваться из множества произвольным образом и самостоятельного смысла не имеет. Просто предполагается, что в процессе перехода к задаче МП элементы перечислимых множеств S и C были пронумерованы натуральными числами. Обозначать элемент множества S с индексом i будем s_i, а элемент множества C с индексом j — c_j.

Введем также переменные $y_j, j \in \{1, \ldots, m\}$, принимающие значения из множества $\{0, 1\}$ и означающие то, что c_j — потенциальный центр группы — в данном варианте решения является действительным центром какой-либо группы.

Базовая постановка задачи МП для двухуровневых централизованных группировок будет записываться следующим образом:

$$\text{Максимизировать } z = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} d(s_i, c_j) \qquad (2)$$

при ограничениях

$$\sum_{j=1}^{m} x_{ij} \leqslant 1 \qquad i \in \{1, \ldots, n\} \qquad \text{(3-а)}$$

$$x_{ij} \leqslant y_j \qquad i \in \{1, \ldots, n\},\ j \in \{1, \ldots, m\} \qquad \text{(3-б)}$$

$$x_{ij}, y_j \in \{0, 1\} \qquad i \in \{1, \ldots, n\},\ j \in \{1, \ldots, m\}.$$

Ограничения вида 3-б назовем ограничениями *целостности*, поскольку они выражают требование внутренней согласованности и непротиворечивости значений переменных x_{ij} и y_j.

Ограничения вида 3-а выражают отношение объектов множества S к группировке, поэтому назовем их ограничениями *покрытия*. Смысл и разновидности ограничений покрытия будет раскрыт ниже в ходе рассмотрения возможных дополнений к этой базовой постановке.

Полные, частично полные и неполные задачи. Покажем, как требование полноты включается в математическую постановку задачи.

Применительно к двухуровневой централизованной группировке может существовать четыре разновидности полноты: полная по S, полная по C, полная и неполная.

Полнота по S определяется видом ограничения покрытия (3-а) базовой постановки. Действительно, сумма в правой части этого ограничения - это количество объектов множества C, с которыми в текущей группировке связан данный объект S. Количество это не может быть больше 1 по определению этого вида группировки, но, варьируя отношение в этом ограничении, мы можем либо допустить нулевое значение для этой суммы, позволив, тем самым, объекту S не быть связанным ни с одним объектом C, а значит, не войти ни в одну группу, либо потребовать строгого равенства суммы единице, обязав каждый объект C быть связанным с одним центром и, следовательно, войти ровно в одну группу. Таким образом, для полных по S группировок ограничение покрытия записывается следующим образом:

$$\sum_{j=1}^{m} x_{ij} = 1 \qquad i \in \{1, \ldots, n\}.$$

Полнота по C определяется тем, какое количество переменных y_j принимают значение единицы. Для полной по C группировки, очевидно, все y_j должны быть равны единице, а это означает, что базовая задача должна быть дополнена ограничением:

$$y_j = 1 \qquad j \in \{1, \ldots, n\}.$$

При рассмотрении окончательной задачи (с учетом всех дополнений) подобное ограничение может позволить упростить систему ограничений подстановкой во все уравнения единицы вместо y_j и сокращением количества переменных задачи.

Запреты на одновременное вхождение. Пусть задано бинарное отношение $\mathrm{R}_{sc} \subseteq S \times C$, задающее такой набор пар $\langle s, c\rangle$, которые не могут входить в одну группу. В задаче математического программирования переменная x_{ij} соответствует факту вхождения сателлита s_i в группу с центром c_j, а поскольку это единственный вариант, когда сателлит и центр оказываются в одной группе, то запрет на вхождение, выражаемый отношением R_{sc}, может быть учтен в задаче математического программирования добавлением ограничений вида:

$$x_{ij} = 0, \qquad \forall i \forall j (s_i \mathrm{R}_{sc} c_j).$$

Другой способ учесть отношение R_{sc} в математической постановке задачи группировки — вместо D использовать в целевой функции отображение D', определяемое следующим образом:

$$D'(s_i, c_i) = \begin{cases} D(s_i, c_i), & \langle s_i, c_i\rangle \notin \mathrm{R}_{sc} \\ M, & \langle s_i, c_i\rangle \in \mathrm{R}_{sc}. \end{cases}$$

Здесь M — большое по модулю отрицательное число, выполняющее роль штрафа.

Пусть задано бинарное отношение $\mathrm{R}_{\mathrm{ss}} \subseteq S \times S$, задающее такой набор пар $\langle s_i, s_j\rangle$, которые не могут входить в одну группу.

В рассматриваемой базовой постановке задачи группировки как задачи ЛП сумма вида $x_{i_1,j} + x_{i_2,j} + \cdots + x_{i_l,j}$ позволяет оценить, какое количество из заданных l объектов входят в группу j

в текущем решении. Тогда ограничение на совместное вхождение объектов s_{i_1} и s_{i_2} в группу j записывается как $x_{i_1,j} + x_{i_2,j} \leqslant 1$, а полный набор ограничений, покрывающий все отношение $\mathrm{R_{ss}}$ и все возможные группы:

$$x_{i_1,j} + x_{i_2,j} \leqslant 1 \qquad j \in \{1, \ldots, m\},\ i_1, i_2 \in \{1, \ldots, n\},\ s_{i_1}\mathrm{R_{ss}}s_{i_2}.$$

Количественные ограничения. Пусть задан некоторый одноместный предикат $P(s), s \in S$, образующий в сочетании с исходным множеством S подкласс $S_p = \{s | s \in S, P(s)\}$. В ряде случаев необходимо задать ограничение на количество элементов S_p, входящих в группу, или размер группы.

В ограничениях задачи ЛП для группировки количество элементов в группе, как уже было сказано выше, задается как сумма значений переменных с определенными индексами, тогда размер группы j вычисляется как:

$$\sum_{i \in n} x_{ij},$$

а количество элементов в группе j, принадлежащих классу, будет задаваться следующей суммой:

$$\sum_{s_i \in S_p} x_{ij}.$$

Например, требование того, чтобы количество элементов класса S_p в каждой группе было больше b, записывается следующим образом:

$$\sum_i x_{ij} \geqslant (b+1)y_j \qquad j \in \{1, \ldots, m\}, s_i \in S_p.$$

Произведение в правой части неравенства необходимо, чтобы не накладывать ограничение на несуществующие группы. Вместе с тем, для несуществующих групп ограничение должно выполняться всегда, а поскольку левая и правая части в этом случае равны нулю, то само ограничение должно записываться как нестрогое неравенство. Преобразование из строгого неравенства в нестрогое для натуральных b тривиально.

5. Подходы к решению. Методы решения задачи централизованной группировки с простейшей системой показателей существенно зависят от характеристик покрытия искомой группировки, от вида отношений R_{ss} и R_{sc} и от наличия количественных ограничений на состав групп.

В общем, и наиболее сложном, случае эта задача решается традиционными для задач целочисленного программирования способами — например, вариантом метода ветвей и границ. Однако в ряде случаев возможно построение эффективных полиномиальных алгоритмов. Рассмотрим такие случаи.

Полная или полная по C группировка без количественных ограничений при $R_{ss} = \varnothing$. Подход к решению задачи группировки в этом частном случае основан на представлении ее в виде эквивалентной $m \times n$ задачи о назначении, решение которой по [1] сводится к поиску потока минимальной стоимости в специальным образом построеной сети.

Принцип построения сети $M'(V, E, a, f)$ следующий. Пусть множество вершин $V = C \cup S \cup \{s\} \cup \{t\} \cup \{q\}$. То есть, в множестве вершин графа есть по одной вершине для каждого из элементов множества $C \cup S$, вершина–источник s, вершина–сток t и дополнительная фиктивная вершина q. Обозначение множества вершин V мы будем употреблять также и в функциональном контексте в качестве обозначения отображения между множествами исходных объектов и множеством вершин графа. При этом запись $V(s_i)$ следует интерпретировать как «вершина графа, соответствующая объекту s_i». Множество дуг E, вместе с отображениями a и f, задающими пропускные способности дуг и стоимости использования единицы пропускной способности, является объединением следующих множеств:

- множество дуг вида $\langle V(c_j), V(s_i)\rangle$, где $s_i \in S, c_j \in C, \langle s_i, c_j\rangle \notin R_{sc}$. Для таких дуг принимается

$$a(V(c_j), V(s_i)) = \infty, \quad f(V(c_j), V(s_i)) = -d(s_i, c_j);$$

- множество дуг вида $\langle s, V(c_j)\rangle$, где $c_j \in C$. Для таких дуг принимается $a(s, V(c_j)) = 1$, $f(s, V(c_j)) = 0$;

- множество дуг вида $\langle V(s_i), t\rangle$, где $s_i \in S$. Для таких дуг принимается $a(V(s_i), t) = 1$, $f(V(s_i), t) = 0$;

- дуга $\langle s, q\rangle$, для которой принимается

$$a(s,q) = |S| - |C|, f(s,q) = 0;$$

- множество дуг вида $\langle s, V(c_j)\rangle$, где $c_j \in C$. Для таких дуг принимается $a(s, V(c_j)) = \infty$, $f(s, V(c_j)) = 0$.

Легко показать, что поток величины m в такой сети соответствует полной группировке исходных множеств, а поток величины m минимальной стоимости — оптимальной полной группировке.

Для поиска полных по C группировок можно рассмотреть семейство сетей $M'[x]$, которые строятся так же, как и сеть M', но имеют пропускную способность дуги $\langle s, q\rangle$ равную x. Решением задачи полной по C группировке является максимальный поток минимальной стоимости в сети $M'[x]|_{0 \leqslant x \leqslant |S|-|C|}$. В самом простом случае, можно рассматривать последовательно все сети $M'[0], \ldots, M'[|S| - |C|]$ и решать для каждой задачу поиска максимального потока минимальной стоимости.

Неполная или полная по S группировка без количественных ограничений при $R_{ss} = \varnothing$. При анализе подобного вида группировок плодотворным также оказывается графовое представление задачи. Рассмотрим взвешенный неориентированный граф $M(V, E, f)$, у которого $V = C \cup S$ (по аналогии с предыдущей задачей мы будем употреблять V и в функциональном контексте как обозначение соответствующего отображения из $C \cup S$ в множество вершин графа). Множество же ребер задается так: $E = V(S) \times V(C) \setminus R_{sc}$, а вес ребра $f(V(c_j), V(s_i)) = d(s_i, c_j), \langle V(s_i), V(c_j)\rangle \in E$.

Тогда нахождение полной по S оптимальной группировки равносильно нахождению такого множества ребер графа, что каждой вершине $s \in V(S)$ инцидентно одно и только одно ребро из этого множества, а суммарный вес ребер максимален.

Рассмотрим пару множеств (E, P), где E — множество всех ребер графа M, а P — множество таких подмножеств E, где ни одна пара ребер не имеет общей вершины из множества $V(S)$.

Утверждение 1. *Пара множеств (E, P) является матроидом, в качестве носителя в котором выступает E — множество ребер графа M, а независимым множеством является такое множество ребер графа M, ни одна пара из которых не имеет общей вершины в S.*

Доказательство. По определению, матроидом называется такая пара (X, I), где X - конечное множество, а I - множество подмножеств X, для которых выполняются следующие условия [2]:

1. $\varnothing \in I$;
2. Если $A \in I$ и $B \in A$, то $B \in I$;
3. Если $A, B \in I$ и мощность A больше мощности B, то существует $x \in A \setminus B$ такой, что $B \cup \{x\} \in I$.

Первое и второе условия очевидны, докажем, что для пары (E, P) выполняется третье. Итак, пусть есть два $P_1, P_2 \in P$ и $|P_1| > |P_2|$. Оценим величину $q^s(I_p)$ - количество $s \in V(S)$, имеющих инцидентные ребра в некотором независимом множестве $I_p \in P$. Поскольку каждое ребро инцидентно только одной вершине из $V(S)$, искомое количество не может быть больше количества ребер в I_p, то есть, $q^s(I_p) \leqslant |I_p|$. С другой стороны, из способа задания P следует, что никакая вершина из $V(S)$ не может быть инцидентна более чем одному ребру, что исключает случай $q^s(I_p) < |I_p|$. Следовательно, $q^s(I_p) = |I_p|$. Значит, $q^s(P_1) > q^s(P_2)$ и в P_1 найдется такой элемент $\langle s_a, c_b \rangle$, что $s_a \notin \{s_i | \langle s_i, c_j \rangle \in P_2\}$. Найденный элемент можно добавить к P_2. Расширенное таким образом множество P_2 все равно останется независимым, а значит, выполняется и третье условие определения матроида.

Максимальным по включению независимым множеством (базой) матроида (E, P) является множество, в котором у каждой вершины графа $s \in S$ есть ровно одно инцидентное ребро. Сопоставив это с графовой постановкой задачи полной по S группировке, получаем, что данный вариант группировки сводится к поиску базы максимального веса в описанном матроиде.

Поиск базы максимального веса в матроиде эффективно осуществляется жадным алгоритмом Радо—Эдмондса [3].

6. Заключение.

В работе предложена схема преобразования некоторых видов векторных ограничений, а именно — ограничений на разброс значений атрибута в группе и на состав группы — в бинарные отношения и учета этих бинарных отношений в постановке задачи целочисленного программирования.

Проанализированы способы решения получающейся задачи целочисленного программирования и показано, что при отсутствии структурных ограничений и ограничений на одновременное вхождение в группу двух сателлитов. Для данной задачи существуют эффективные полиномиальные алгоритмы: задачи полной и полной по C группировки сводятся к поиску потока минимальной стоимости, а неполной и полной по S решаются жадным алгоритмом. В иных случаях можно применять общую схему решения задач целочисленного программирования.

Дальнейшую работу представляется целесообразным вести по следующим направлениям: рассмотрение других классов векторных ограничений, поиск эффективных алгоритмов для других частных случаев, исследование вопроса о существовании более быстрых способов нахождения оптимальных решений для семейства сетей $M'[x]|_{0\leqslant x\leqslant |S|-|C|}$, чем последовательный перебор элементов этого семейства.

Литература

1. *Bertsekas D.* Linear Network Optimization: Algorithms and Codes. The MIT Press, 1991.

2. *Новиков Ф.А.* Дискретная математика для программистов. СПб.: Питер, 2000.

3. *Edmonds J.* Matroids and the Greedy Algorithm. Mathematical Programming, volume 1, p.125–136. 1971.

Мустафин Николай Габдрахманович — к.т.н., доцент; ст. научный сотрудник лаборатории интегрированных систем автоматизации СПИИРАН, доцент кафедры автоматизированных систем обработки информации и управления С.-Петербургского государственного электротехнического университета (СПбГЭТУ «ЛЭТИ»). СПбГЭТУ, ул. профессора Попова, д. 5, г. Санкт-Петербург, 197376, РФ; р.т. +7(812)234-2773, факс +7(812)234-2773.

Mustafin Nikolai Gabdrahmanovich — PhD in Computer Science, Associate Professor; Computer-Aided Integrated Systems, SPIIRAS, Associate Professor of Automated Systems of Information Processing and Control department, SPbSETU. SPbSETU, Prof. Popova st., 5, St. Petersburg, 197376, Russia; office phone +7(812)234-2773, fax +7(812)234-2773.

Пономарев Андрей Васильевич — научный сотрудник лаборатории интегрированных систем автоматизации СПИИРАН, ассистент кафедры автоматизированных систем обработки информации и управления С.-Петербургского государственного электротехнического университета (СПбГЭТУ «ЛЭТИ»).

ponomarev.a.v@gmail.com; СПбГЭТУ, ул. профессора Попова, д. 5, г. Санкт-Петербург, 197376, РФ; р.т. +7(812)234-2773, факс +7(812)234-2773. Научный руководитель — Мустафин Н.Г.

Ponomarev Andrei Vasilyevich — Computer-Aided Integrated Systems, SPIIRAS, assistant at Automated Systems of Information Processing and Control department, SPbSETU. ponomarev.a.v@gmail.com; SPbSETU, Prof. Popova st., 5, St. Petersburg, 197376, Russia; office phone +7(812)234-2773, fax +7(812)234-2773. Scientific advisor — N.A. Mustafin.

Савосин Сергей Валентинович — к.т.н., доцент; ст. научный сотрудник лаборатории интегрированных систем автоматизации СПИИРАН, доцент кафедры автоматизированных систем обработки информации и управления С.-Петербургского государственного электротехнического университета (СПбГЭТУ «ЛЭТИ»). svsavosin@mail.ru; СПбГЭТУ, ул. профессора Попова, д. 5, г. Санкт-Петербург, 197376, РФ; р.т. +7(812)234-2773, факс +7(812)234-2773.

Savosin Sergei Valentinovich — PhD in Computer Science, Associate Professor; Computer-Aided Integrated Systems, SPIIRAS, Associate Professor of Automated Systems of Information Processing and Control department, SPbSETU. svsavosin@mail.ru; SPbSETU, Prof. Popova st., 5, St. Petersburg, 197376, Russia; office phone +7(812)234-2773, fax +7(812)234-2773.

Рекомендовано ЛИСА СПИИРАН, зав. лаб. д.т.н, профессор Смирнов А.В.

Статья поступила в редакцию 17.02.2012.

РЕФЕРАТ

Мустафин Н.Г., Пономарев А.В., Савосин С.В. **Метод оптимальной группировки векторных объектов относительно центров.**

В статье рассматриваются вопросы, связанные с оптимальной группировкой объектов, каждый из которых описывается набором характеристик, относительно потенциальных центров групп. Целью работы является исследование подходов к представлению и решению задач такого рода.

Сложность задачи группировки существенным образом зависит от деталей ее постановки: от используемой системы показателей, через которую определяется оптимальность группировки, от требуемого покрытия исходных множеств и от дополнительных ограничений. В статье используется простейшая система показателей, в которой оценка группировки формируется только на основании бинарной функции, задающей эффект от присоединения объекта к определенному центру. С точки зрения покрытия анализируются 4 вида задач: полная, полная по S, полная по C и неполная.

Рассматриваются два вида специфических векторных ограничений задачи группировки: ограничение на разброс параметра в группе и ограничение на состав группы. Показано, каким образом эти ограничения могут быть учтены в постановке задачи группировки как задачи целочисленного программирования. Ограничение на разброс параметра в группе представляется в виде бинарного отношения запрета на одновременное вхождение, которое непосредственно преобразуется в множество попарных ограничений, ограничение же на состав реализуется введением предикатов принадлежности классам и составлением ограничений на основе сумм, образуемых с учетом этих предикатов.

Проанализированы способы решения получающейся задачи целочисленного программирования и показано, что при отсутствии структурных ограничений и ограничений на одновременное вхождение в группу двух сателлитов. Для данной задачи существуют эффективные полиномиальные алгоритмы: задачи полной и полной по C группировки сводятся к поиску потока минимальной стоимости, а неполной и полной по S решаются жадным алгоритмом. В иных случаях можно применять общую схему решения задач целочисленного программирования.

Дальнейшую работу представляется целесообразным вести по следующим направлениям: рассмотрение других классов векторных ограничений и поиск эффективных алгоритмов для других частных случаев.

SUMMARY

Mustafin N.G., Ponomarev A.V., Savosin S.V. **A method for optimal grouping of vector objects around centers.**

The paper discusses issues related to the optimal grouping of objects, each of which is described by a set of characteristics around potential centers of groups. The possible interpretations of this problem include optimal team building, various location and regional analysis tasks. The aim is to study the approaches to the representation and solution of problems of this kind.

The input for the grouping problem being discussed consists of two sets: potential centres (C) and satellites (S). Each group must contain one potential centre and may contain several satellites. The quality of solution is measured by a system of indicators. The complexity of the grouping task depends significantly on its details: the system of indicators, in which terms the grouping optimality is defined, the required coverage of the original sets and the additional restrictions. The paper uses a simple system of indicators, which estimate groupings only on the basis of binary function that defines the effect of the assigning an object to a particular center. From the standpoint of coverage four types of tasks are discussed: the complete, S–complete, C–complete, and incomplete.

Two types of vector-specific constraints of the grouping problem are reviewed: limiting of the spread parameter in the group and structural restrictions on the group. It is shown how these limitations can be addressed in the grouping problem as integer programming problem. Spread limit in the group is represented as a binary relation of mutual exclusion, which is directly converted into a set of pairwise constraints. Structural restrictions are based on introducing class predicates and composing integer problem constraints as sums with respect to these predicates.

The different methods of solving the resulted integer programming problem are discussed. It is shown that in case of absence of structural restrictions and mutual exclusion restrictions on pairs of satellites, there are efficient polynomial algorithms: the complete and C–complete grouping problems are reduced to finding the minimum cost flow in a network, built in a special way, on the other hand, the incomplete and S–complete grouping problems are solved by the greedy algorithm, because their structure is proven to be a matroid. In other cases, the general scheme of solving integer programming problems can be applied.

The directions of further work on this topic are: the analysis of other classes of vector constraints and the search for efficient algorithms for other special cases.