

УДК 519.6

ИССЛЕДОВАНИЕ МНОГОКРИТЕРИАЛЬНОЙ ПОСТАНОВКИ ТЕОРЕТИКО-ГРАФОВОЙ ЗАДАЧИ СЕГМЕНТАЦИИ НА ДВУДОЛЬНОМ ГРАФЕ

© 2005 г. В.А. Перепелица, Ф.Б. Тебуева, Т.М. Шенкао

Market segmentation problem is stated as a problem of bipartite graph cover with stars. It was proved that this problem is intractable in multicriterion statement. Polynomially solvable subclass was found and corresponding algorithm was built.

В маркетинге термин «сегментация» означает выделение различных целевых групп (сегментов) потенциальных потребителей, требующих соответственно различных подходов в стратегии разработки новых видов продукции, организации товародвижения, рекламы и стимулирования сбыта [1]. В настоящей работе математическая модель формируется в предположении, что априори определена база для сегментации рынка [1], т.е. является известным множество потенциальных покупателей (потребителей и организаций), выбраны критерии (факторы) сегментации: географические, демографические, социально-экономические, психографические, поведенческие и др.

Оговорим заранее, что недостающие определения терминов и понятий теории графов можно найти в [2], сегментного анализа – в [1].

В качестве иллюстративного примера критериев можно назвать их перечень, используемый при сегментации рынка персональных компьютеров (ПК): F1 – технические характеристики; F2 – цена; F3 – особые качества; F4 – надежность; F5 – удобство в использовании; F6 – совместимость; F7 – периферийное оборудование; F8 – программное обеспечение. Кроме того, определено множество моделей однотипного товара, предъявляемого рынку.

Математическая постановка задачи сегментации рынка базируется на двудольном графе $G = (V_1, V_2, E)$, в котором вершины множества V_1 (т.е. первой доли) перенумерованы индексом $i = \overline{1, m}$, V_2 (второй доли) – $j = \overline{1, l}$.

Содержательно вершины $v_i \in V_1$ поставлены во взаимно-однозначное соответствие предъявленным типам моделей товара, которые также перенумерованы индексом $i = \overline{1, m}$. Вершины $v_j \in V_2$ поставлены во взаимно-однозначное соответствие группам потребителей, перенумерованных индексом $j = \overline{1, l}$. В упомянутом примере сегментации рынка персональных компьютеров в качестве иллюстративного примера можно назвать следующие шесть групп потребителей, выделенных по областям применения: дома, в школе, в вузе, в кабинете, в мелком бизнесе, в корпорации.

Через n_j будем обозначать прогнозируемое количество покупаемых единиц товара представителями j -й группы. Ребро $e = (v_i, v_j)$ принадлежит множеству E тогда и только тогда, когда i -я модель может оказаться приемлемой для покупателей j -й группы ($1 \leq j \leq l$).

Через k_i обозначим пороговое, т.е. минимально допустимое количество экземпляров товара i -й модели, производство которых может оказаться экономически выгодным ($i = \overline{1, m}$).

Допустимым решением рассматриваемой задачи сегментации является всякая такая часть $x = (V_1^x, V_2, E_x)$, $V_1^x \subseteq V_1$, $E_x \subseteq E$, $|E_x| = |V_2|$ графа $G = (V_1, V_2, E)$, каждая компонента связности которой представляет собой звезду, центром которой является некоторая вершина $v_i \in V_1$ и ребра которой образуют множество E_x , $i \in \{1, 2, \dots, m\}$. При этом висячие вершины этой звезды образуют подмножество $V_1^x \subseteq V_2$, удовлетворяющее неравенству

$$\sum_{v_j \in V_1^x} n_j \geq k_i \quad (1)$$

для каждого центра v_i непустой звезды.

Множество допустимых решений (МДР) на графе $G = (V_1, V_2, E)$ обозначим через $X = X(G) = \{x\}$.

Для численной оценки качества допустимых решений $x \in X$, ребрам $e = (v_i, v_j) \in E$ приписаны веса $w_v(e)$, $v = \overline{1, N}$, где $w_v(e)$ представляют собой экспертно определенную степень (коэффициент) потребительской пригодности i -й марки для покупателей из группы j , $0 < w_v(e) < 1$, $v = \overline{1, N}$, $e \in E$.

На МДР X определена векторная целевая функция (ВЦФ)

$$F(x) = (F_1(x), F_2(x), \dots, F_N(x)), \quad (2)$$

состоящая из максимизируемых критериев

$$F_v(x) = \sum_{e \in E_x} w_v(e) \rightarrow \max, \quad v = \overline{1, N}. \quad (3)$$

ВЦФ (2)–(3) определяет собой в МДР X паретовское множество (ПМ) \tilde{X} , состоящее из всех паретовских оптимумов (ПО) $\tilde{x} \in \tilde{X}$ [3, 4].

Всякая пара ПО $\tilde{x}_1, \tilde{x}_2 \in \tilde{X}$ считается эквивалентной, если выполняется равенство значений ВЦФ: $F(\tilde{x}_1) = F(\tilde{x}_2)$. Поэтому в настоящей работе рассматриваем алгоритмическую проблему нахождения так называемого полного множества альтернатив (ПМА) [4].

Подмножество $X^0 \in \tilde{X}$ называется ПМА, если его мощность $|X^0|$

минимальна при выполнении равенства $F(X^0) = F(\tilde{X})$, где $F(X^*) = \{F(x) : x \in X^*\}$, $\forall X^* \subseteq X^0$.

Максимальная мощность МДР для задачи сегментации рынка растёт экспоненциально с ростом размерности входа задачи $(m + n)$ -количества вершин графа $G = (V_1, V_2, E)$.

В настоящей работе строго доказывается, что в терминах вычислительной сложности [5] алгоритмов является справедливой следующая оценка вычислительной сложности рассматриваемой задачи: при $N \geq 2$ проблема нахождения ПМА задачи сегментации с ВЦФ (2)–(3) является труднорешаемой, т.е. нижняя оценка ее вычислительной сложности растёт экспоненциально от размерности $(m + n)$.

Сформулированная на конкретном N -взвешенном графе G задача сегментации с ВЦФ (2)–(3) является полной, если существуют такие веса

$w_v(e)$, $v = \overline{1, N}$, $e \in E$, при которых выполняются равенства $X = \tilde{X} = X(G)$.

Массовая задача называется полной, если существует ее полная индивидуальная задача для каждого $(m + n)$ – вершинного графа.

Лемма 1. Для всякого $N \geq 2$ задача сегментации с ВЦФ (2)–(3) является полной.

Доказательство. Выберем произвольное $x \in X$, которое определяется данным n -вершинным графом $G = (V_1, V_2, E)$. Для тривиальных случаев $X = \emptyset$ или $X = 1$ утверждение очевидно. Пусть мощность МДР $|X| > 2$. Рассмотрим вначале случай $N = 2$, когда ВЦФ (2) имеет вид

$$F(x) = (F_1(x), F_2(x)) = (w_1(x), w_2(x)). \tag{4}$$

В данном графе G ребра $e \in E$ перенумеруем числами $t = t(e) = \overline{1, q}$, $q = |E|$, а их веса определим следующим образом:

$$w_1(t) = 2^t, w_2(t) = r_0 - w_1(t), t = \overline{1, q}, t = \overline{1, q}, r_0 = 2^q + 1. \tag{5}$$

Из (4), (5) получаем

$$F_1(x) + F_2(x) = c_0^* r_0, \tag{6}$$

где константа $c_0 = |V_2|$.

Обозначим разность $R_{s_1, s_2} = E_{s_1} \setminus E_{s_2}$ для пары $x_{s_1}, x_{s_2} \in X$. Тогда для всяких $x_1, x_2 \in X$

$$R_{1,2} \cap R_{2,1} = \emptyset, |R_{1,2}| = |R_{2,1}|. \tag{7}$$

Пусть среди элементов множества $R_{1,2} \cup R_{2,1}$ ребро с наибольшим номером $t = t(e)$ принадлежит $R_{1,2}$. Тогда из (4)–(7) вытекают неравенства

$F_1(x_1) > F_1(x_2)$, $F_2(x_1) < F_2(x_2)$, которые означают, что любая пара $x_1, x_2 \in X$ является векторно несравнимой по ВЦФ (4). Последнее с учетом

необходимого равенства $F(X^0) = F(\tilde{X})$ означает справедливость

$X = \tilde{X} = X^0$. Для $N = 2$ лемма 1 доказана.

Так как при любом $N > 2$ для всякой индивидуальной задачи с ВЦФ (2)–(3) добавление новых критериев к этой ВЦФ либо оставляет ПМ \tilde{X} и ПМА X^0 неизменными, либо пополняет их новыми альтернативами, то

равенство мощностей $|X| = |\tilde{X}| = |X^0|$ выполняется и при $N > 3$, если для $v = 1, 2$ критерии (3) определим согласно (4)–(5), а для $v = \overline{3, N}$ – произвольным образом. Лемма 1 доказана.

Двудольный граф $G = (V_1, V_2, E)$ называем нетривиальным, если мощности его долей строго больше 1. Частным случаем рассматриваемой задачи сегментации рынка является задача о совершенном сочетании, у которой, как известно [6], мощность МДР растет экспоненциально с ростом числа вершин $(m + n)$. Отсюда с учетом леммы 1 является справедливой.

Лемма 2. Максимальная $r(m, l)$ мощность МДР задачи сегментации для случая $m = 2$ определяется точной формулой $|X| = r(m, l) = r(2, l) = 2^l$, где через m и l обозначены мощности долей V_1 и V_2 , $m \leq l$, соответственно полного двудольного графа $G = (V_1, V_2, E)$.

Доказательство. Согласно определению допустимое решение $x = (V_1^x, V_2, E_x)$, $V_1^x \in V_1$, $E_x \subseteq E$ содержит все вершины $v_j \in V_2$, $j = \overline{1, l}$ и подмножество вершин $v_i \in V_1^x$, $i = 1, 2$. Поэтому выбор типа компоненты связности решения x для одной из двух вершин $v_i \in V_1^x$, $i = 1, 2$, однозначно определяет тип компоненты связности для второй вершины множества V_1^x . С учетом этих замечаний рассмотрим возможные варианты покрытия полного двудольного графа $G = (V_1, V_2, E)$ $(h + l)$ -вершинными звездами, называемыми h -звездами. Пусть одна из вершин, для определенности первая v_1 , является центром 1-звезды, тогда вторая вершина v_2 множества V_1^x является изолированной. Такое покрытие возможно в C_l^1 (одном) случае. Если v_1 является центром $(l - 1)$ -звезды, тогда решение x в качестве второй компоненты связности содержит ребро $e \in E_x$, инцидентное второй вершине v_2 множества V_1^x . Такое покрытие возможно построить C_l^{l-1} способами. Если v_1 является центром $(l - z)$ -звезды, тогда

вторая вершина v_2 множества V_1^x есть центр z -звезды. Такое покрытие возможно построить C_l^{l-z} способами. Если v_1 – изолированная вершина, то вершина v_2 множества V_1^x является центром 1-звезды. Такое покрытие возможно в C_l^0 (одном) случае.

Таким образом, мощность МДР есть сумма всех рассмотренных вариантов покрытий полного двудольного графа $G = (V_1, V_2, E)$ для $m = 2$ h -звездами и определяется формулой $|X| = \sum_{z=0}^l C_l^z = 2^l$.

Лемма 3. Для случая $m \geq 3$ максимальная мощность МДР задачи сегментации определяется рекуррентной формулой

$$|X| = m \sum_{z=0}^l C_l^z r(m-1, l-z), \text{ где } |V_1| = m \text{ и } |V_2| = l, \quad m \leq l, \text{ для полного}$$

двудольного графа $G = (V_1, V_2, E)$.

Доказательство. Пусть дан полный двудольный граф $G^* = (V_1^*, V_2, E)$, $|V_1^*| = m = 2$ и $|V_2| = l$, $m \leq l$. Согласно лемме 2, максимальная мощность МДР для этого случая определяется точной формулой $|X| = r(m, l) = r(2, l) = 2^l$. Добавим к первой доле V_1^* этого графа вершину v_3 . Получим полный двудольный граф $G = (V_1, V_2, E)$. Рассмотрим возможные варианты покрытия полученного графа $G = (V_1, V_2, E)$ h -звездами. Пусть одна из вершин, для определенности v_3 , является центром 1-звезды, тогда вершины v_1 и v_2 множества V_1^* являются изолированными. Такое покрытие возможно в C_l^l (одном) случае. Если v_3 является центром $(l-1)$ -звезды, тогда решение x в качестве второй компоненты связности содержит ребро $e \in E_x$, инцидентное одной из вершин v_1 или v_2 множества V_1^* . Выбор $(l-1)$ -звезды на графе $G = (V_1, V_2, E)$ можно осуществить C_l^{l-1} способами. Рассмотрим количество вариантов покрытий для подграфа $G^* = (V_1^*, V_2^1, E)$, где мощность множества $|V_2^z| = z = l-h$, $z = \overline{0, l}$ определяется количеством вершин h -звезды с центром в вершине v_3 . Для рассматриваемого случая $z=1$, а $|V_1^*| = 2$. Согласно лемме 2, количество вариантов покрытий для подграфа $G^* = (V_1^*, V_2^1, E)$ определяется как $r(2, l) = 2^l = 2$. Такое покрытие можно построить $C_l^{l-1} r(2, l)$ вариантами. Если v_3 является центром $(l-z)$ -звезды, тогда выбор $(l-z)$ -звезды на графе $G = (V_1, V_2, E)$ можно осуществить C_l^{l-z} способами. Такое покрытие можно построить $C_l^{l-z} r(2, z)$ вариантами, где $r(2, z)$ определяет количе-

ство вариантов покрытий для подграфа $G^* = (V_1^*, V_2^1, E)$. Аналогично, если v_3 – изолированная вершина, то существует $C_l^0 r(2, l)$ возможностей построения такого покрытия.

Итак, добавление одной вершины ведет к $\sum_{z=0}^l C_l^z r(2, l-z)$ описанным вариантам покрытий. Так как подобные построения покрытий можно провести и для других двух вершин, то окончательно приходим к формуле для случая $m=3$: $r(3, l) = 3 * \sum_{z=0}^l C_l^z r(2, l-z)$.

Проведя аналогичные построения покрытий, приходим к обобщенной рекуррентной формуле для подсчета мощности МДР $r(m, l) = |X| = m \sum_{z=0}^l C_l^z r(m-1, l-z)$. Лемма 3 доказана.

Из леммы 2 и 3 очевидна справедливость следствия 1.

Следствие. Максимальная мощность МДР для задачи сегментации рынка растет экспоненциально с ростом размерности входа задачи $(m+n)$ -количества вершин графа $G = (V_1, V_2, E)$.

Из следствия в терминах вычислительной сложности алгоритмов [5] является справедливой следующая теорема 1.

Теорема 1. При $N \geq 2$ проблема нахождения ПМА задачи сегментации с ВЦФ (2) – (3) является труднорешаемой.

Доказательство. Напомним, что, согласно [5], дискретная задача распознавания или оптимизационная называется труднорешаемой, если нижняя оценка вычислительной сложности нахождения искомого решения растет экспоненциально с ростом размерности задачи, точнее, если эта оценка не ограничена сверху никаким полиномом от размерности задачи.

В силу леммы 1 рассматриваемая задача является полной, т.е. мощность искомого ПМА может совпадать с мощностью МДР. Последнее, согласно следствию 1, растет экспоненциально с ростом размерности задачи, тем самым теорема 1 доказана.

Рассмотрим оптимизационную, т.е. однокритериальную постановку сформулированной задачи сегментации. Она также представляет нетривиальную алгоритмическую проблему, поскольку при линейных критериях (3) и линейном ограничении (1) она не представляет собой задачу линейного программирования. Действительно, всякий критерий вида (3) можно представить следующим образом в виде ЦФ транспортной задачи. Допустимое решение определяется в виде матрицы $X = \|x_{ij}\|$, где $i = \overline{1, m}$, $j = \overline{1, l}$, x_{ij} означает количество покупателей, которые, являясь предста-

вителями групп j , приобретают товар марки i . Тогда, представляя вес $w_v(e)$ ребра $e = (v_i, v_j) \in E$ в виде коэффициента $w_{ij}^v = w(v_i, v_j) = w_v(e)$, можем записать

$$F_v(x) = \sum_{i=1}^m \sum_{j=1}^l w_{ij}^{v*} x_{ij} \rightarrow \max, \quad 1 \leq v \leq N. \quad (8)$$

При $N = 1$ задача с ЦФ вида (8) является классической транспортной задачей в том случае, если с учетом (1) допустимое решение определяется следующей системой ограничений:

$$\sum_{i=1}^n x_{ij} = n_j, \quad j = \overline{1, l}, \quad (9)$$

$$x_{ij} \in \{0, 1, 2, \dots, n_0\}, \quad n_0 = \max_{1 \leq j \leq l} n_j, \quad i = \overline{1, m}, \quad j = \overline{1, l}. \quad (10)$$

Для представления условия (1) в терминах и обозначениях транспортной задачи введем булеву переменную $y_i \in \{0; 1\}$. Условие (1) в терминах переменных представляется следующими неравенствами:

$$\sum_{j=1}^l x_{ij} \geq k_i y_i \quad y_i \in \{0; 1\}, \quad i = \overline{1, m}. \quad (11)$$

Содержательный смысл всякого неравенства системы (11) заключается в следующем: товар марки i либо может не производиться, либо в случае его производства количество закупаемого товара этой марки должно быть не меньше пороговой величины k_i , $i = \overline{1, m}$.

Таким образом, оптимизационная задача (8)–(11) не относится к классу задач линейного программирования. Актуальный вопрос: принадлежит ли эта задача к классу NP -трудных задач [5] остается открытым.

Особого внимания заслуживает тот факт, что задача с ЦФ (8) и МДР $X = \{x\}$, определяемом условиями (9)–(10), оказывается не только полиномиально разрешимой, но и разрешимой с помощью простого алгоритма α_1 , вычислительная сложность которого линейна относительно длины записи информации на его входе [5]. Эта информация представляется двудольным графом $G = (V_1, V_2, E)$ и весами ребер этого графа.

Алгоритм α_1 состоит из этапов $j = \overline{1, l}$. На этапе $j = j_0$ для рассматриваемой группы j_0 определяется такой индекс i_0 , для которого выполняется равенство $w_{i_0 j_0}^v = \max_{1 \leq i \leq m} w_{ij_0}^v$, после чего фиксируется значение переменных $x_{i_0 j_0} = n_{j_0}$ и $x_{ij_0} = 0$ для всех $i \neq j_0$. Вычисленные таким образом зна-

чения переменных x_{ij}^0 представляют полученное допустимое решение в виде матрицы $x = \left| x_{ij}^0 \right|$ размерности $m \times l$.

Если учесть, что в процессе получения решения x каждое ребро графа $G = (V_1, V_2, E)$ рассматривается не более одного раза, то из приведенных рассуждений с очевидностью вытекает, что является справедливой

Лемма 4. Оптимизационная задача сегментации с ЦФ (8) и МДР X , определяемым условиями (9)–(10), разрешима с вычислительной сложностью $\tau(\alpha_1) \leq O(|V_1| + |V_2| + |E|) \leq O(ml)$.

Примечание. Алгоритм α_1 можно использовать для представленной в лемме 4 задачи, которая рассматривается в N -критериальной постановке. В этом случае он представляет собой алгоритм линейной свертки критериев (ЛСК).

ЛСК, образующая ЦФ для решаемой задачи, определяется выражением

$$F^\lambda(x) = \sum_{v=1}^N \lambda_v * F^v(x) \rightarrow \max, \quad (12)$$

где вектор

$$\lambda \in \Lambda_N = \{ \lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) : \sum_{v=1}^N \lambda_v = 1, \lambda_v > 0, v = 1, 2, \dots, N \}.$$

Как известно [3], при всяком фиксированном $\lambda \in \Lambda$ решение x^λ , оптимальное по значению ЛСК (12), представляет собой ПО, соответствующий N -критериальной задаче, т.е. $x^{(\lambda)} \in \tilde{X}$.

Таким образом, варьируя векторы $\lambda \in \Lambda$, т.е. используя определенное подмножество $\Lambda_N^0 \subset \Lambda_N$, можно с помощью алгоритма ЛСК α_1 находить некоторое подмножество ПМ \tilde{X} и выделять из него подмножество ПМА $X^0(\Lambda_n^0) \subseteq X^0$. Подмножество $X^0(\Lambda_n^0)$ назовем нетривиальным, если его мощность $|X^0(\Lambda_n^0)| \geq 2$. Если существует алгоритм нахождения нетривиального подмножества $X^0(\Lambda_n^0)$, то говорим о частичной разрешимости рассматриваемой N -критериальной задачи с верхней оценкой $\tau(X^0(\Lambda_n^0))$ вычислительной сложности.

Из примечания и леммы 4 вытекает, что является справедливой следующая

Теорема 2. Многокритериальная задача сегментации с ВЦФ (2)–(3) и МДР X , которое определяется условиями (9), (10), частично разрешима с помощью алгоритма ЛСК α_1 . При этом вычислительная сложность находж-

дения нетривиального подмножества $X^0(\Lambda_n^0) \subset X^0$ имеет следующую верхнюю оценку: $\tau(X^0(\Lambda_n^0)) = O(mlrN)$, где $r = \left| \Lambda_N^0 \right|$.

Литература

1. Макдоналд М., Данбар Я. Сегментирование рынка: практическое руководство. М., 2002.
2. Емеличев В.А., Мельников О.И., Сарванов В.И., Тышкевич Р.И. Лекции по теории графов. М., 1990.
3. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. М., 1982.
4. Емеличев В.А., Перепелица В.А. // Дискретная математика. 1994. Т. 6. № 1. С. 3–33.
5. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М., 1982.
6. Кориунов А.Д. // Успехи математических наук. 1985. Т. 40. № 1 (241). С. 23–34.

Карачаево-Черкесская государственная технологическая академия

6 июля 2005 г.

УДК 681.518.54

ВИЗУАЛИЗАЦИЯ ТРЕХМЕРНОГО ТЕЧЕНИЯ С ПОМОЩЬЮ СИСТЕМЫ ЧАСТИЦ НА ОСНОВЕ ГРАФИЧЕСКОГО ПРОЦЕССОРА

© 2005 г. *О.А. Потий*

In this paper a fast and effective technique for 3D unsteady flow visualization based on particle system animation is presented. Particle advancement in flow is performed directly by means of a graphic processing unit (GPU).

The suggested method derives the advantages of texture visualization techniques - continuous cover of displayed domain and the most intuitive way of flow imagery. On the other hand, this technique allows performing 3D flow animation with interactive frame rate and sufficient quality. Tracking particles injected in the velocity field do not require expensive, from the computation point of view, 3D volume rendering procedure, in contrast with texture 3D visualization techniques. Employing of GPU hardware capacities and stream computation approach makes possible to relieve the stress from a central processing unit.

Развитие вычислительной математической физики в последние годы, частично обусловленное стремительным ростом компьютерных мощностей, дало возможность проводить сложные вычислительные эксперименты математического моделирования природных объектов и явлений. Весьма распространенным частным случаем такого моделирования в экологических исследованиях и задачах гидрогазодинамики является расчет двумерного или трехмерного поля скоростей движения среды в ограниченном объеме на регулярной или нерегулярной сетке. Результатами та-