

Двухкомпонентный метод автоматической кластеризации задач компоновки конструктивных узлов

В. М. Курейчик¹, И. Б. Сафроненкова²

Южный федеральный университет

¹ vmkureychik@sfedu.ru, ² safronenkova050788@yandex.ru

О. О. Варламов³

Московский автомобильно-дорожный государственный
технический университет

³ ovar@narod.ru

Аннотация. Настоящая работа посвящена разработке нового метода кластеризации задач компоновки как одного из этапов автоматизированного построения предметной онтологии. Предложена модель пространства признаков для задач компоновки, учитывающая специфику рассматриваемой предметной области. Проведен анализ известных формальных моделей целевых функций задачи компоновки, на основании которого предложена модель представления целевой функции в виде категорий. Разработан новый метод кластеризации, состоящий из двух последовательных этапов. Принципиальным отличием предложенного метода является его двухэтапность, позволяющая сгруппировать задачи компоновки конструктивных узлов по признаку формализованного описания ЦФ.

Ключевые слова: кластеризация; компоновка конструктивных узлов; целевая функция; категориальные признаки; поддержка принятия решений

I. ВВЕДЕНИЕ

Основные задачи конструкторского этапа проектирования, к которым относят компоновку узлов, их размещение, трассировку межсоединений на всех уровнях и получение конструкторско-технологической документации, обладают рядом особенностей, отличающих их от задач других этапов разработки электронной аппаратуры. Прежде всего сюда следует отнести прямую зависимость от критериев и ограничений, используемых при решении таких задач. Кроме того, важно подчеркнуть большую трудоемкость их решения [1].

Задача компоновки (ЗК) является первой на этапе конструкторского проектирования и служит источником информации для последующих этапов. Разнообразие и зачастую противоречивость требований, возникающих при решении ЗК, затрудняет разработку единой алгоритмической методики компоновки электронных узлов [2]. Таким образом, лицо, принимающее решение, (ЛПР) сталкивается с выбором оптимального метода решения ЗК. Данная задача представляет собой

нетривиальную процедуру принятия решения, для которой в рамках промышленных производств весьма обосновано применение интеллектуальных систем поддержки принятия решений (ИСППР). В связи с этим возникает необходимость формализации ЗК с целью формирования исходной базы знаний для принятия решений в конкретных ситуациях. Базой знаний в ИСППР, как правило, является онтология, ручная разработка которой представляет собой трудоемкий и дорогостоящий процесс [3, 4]. Более того, рассматриваемая предметная область обладает рядом специфических особенностей, накладывающих существенные ограничения на применение традиционных методов построения онтологий.

В данной работе будем рассматривать статические методы построения онтологий. Подход, основанный на статических методах, включает следующие этапы: предварительная обработка данных, кластеризация, определение классов онтологии, определение отношений между классами, оценка качества онтологии [5]. В рамках данной статьи рассмотрим предлагаемый подход для этапа кластеризации задач компоновки. Анализ открытых источников показал, что проблема автоматизации построения онтологии задач компоновки практически не освещена. Следовательно, проблема, освещенная в рамках данной работы, является актуальной и требует разработки новых методов решения.

II. ПРОСТРАНСТВО ПРИЗНАКОВ ДЛЯ ЗАДАЧИ КОМПОНОВКИ КОНСТРУКТИВНЫХ УЗЛОВ

Компоновкой коммутационной схемы на конструктивно законченные части называется процесс распределения элементов низшего конструктивного уровня в высший в соответствии с заданными критериями [6]. Таким образом, при решении ЗК, инженер сталкивается с решением оптимизационной задачи, зачастую многокритериальной. Известно, что формальная модель задачи оптимизации описывается набором: $(F(X), \text{Огр}, \text{Гр})$, где $F(X)$ – целевая функция, представляющая собой критерий или критерии оптимизации; Огр – ограничения, накладываемые на изменяемые параметры; Гр – граничные условия [7].

Работа выполнена при финансовой поддержке РФФИ, проект № 18-07-00050, ГЗ № 25537.2017/6.7.

Для формального описания ЗК необходимо формальное представление ее исходных данных. Один из известных способов заключается в представлении КС в виде взвешенного графа схемы, который может быть описан матрицей соединений $R=\|r_{ij}\|_{n \times n}$, строки и столбцы которой соответствуют элементам схемы, а r_{ij} – равен весу, приписанному соединению элементов e_i и e_j [8].

Для осуществления процесса кластеризации, необходимо векторное представление кластеризуемого объекта (ЗК). Тогда, ее формальное описание сводится к представлению данной задачи в виде вектора признаков, компонентами которого будут являться матрица соединений исходной схемы и формализованная постановка задачи оптимизации:

$$W = (F(X), \text{Огр}, \text{Гр}, R), \quad (1)$$

где $F(X)$ – целевая функция; Огр – ограничения, накладываемые на изменяемые параметры; Гр – граничные условия; R – матрица соединений схемы.

Остановимся более подробно на каждом компоненте вектора признаков [9-12]. $F(X)$ – целевая вектор-функция, определенная в области D_X со значениями в целевом пространстве $\{F\}=\mathbb{R}^{|F|}$. В общем случае имеет вид: $F(X)=(f_1(X), f_2(X), \dots, f_n(X))$, где $f_1(X), f_2(X), \dots, f_n(X)$ – частные целевые функции; $\{X\}$ – ограниченное и замкнутое множество допустимых значений вектора варьируемых параметров $D_X=\{X|G(X) \geq 0\} \subset \{X\}=\mathbb{R}^{|X|}$. Ограничения в задачах оптимизации задаются системой уравнений и неравенств: $f_i(x) \leq b_i, i=1, \dots, k; f_i(x)=b_i, i=k+1, \dots, m$, где $x \in \mathbb{R}^n, f_i(x): \mathbb{R}^n \rightarrow \mathbb{R} (i=0, \dots, m)$. Граничные условия записываются неравенствами типа $x_i^{(u)} \leq x_i \leq x_i^{(l)}, i=1, \dots, n$.

Как можно заметить, приведенные выше выражения представляют собой различные типы данных, что ведет к неоднозначности при определении схожести между векторами признаков, описанными выражением (1). В рамках данной работы рассмотрим процесс кластеризации целевых функций, входящих в состав вектора признаков.

III. АНАЛИЗ ФОРМАЛЬНЫХ МОДЕЛЕЙ ЦФ ЗАДАЧИ КОМПОНОВКИ

Целевая функция (ЦФ) представляет собой зависимость критерия (критериев) оптимальности от параметров, влияющих на ее значение. Вид ЦФ определяется конкретной задачей оптимизации. Проведенный анализ ЗК позволил выделить основные критерии оптимальности, используемые при их решении: минимум числа межузловых соединений (минимум числа выводов на узлах) (F_1); минимум числа узлов (F_2); минимум задержек в распространении сигналов (F_3).

Ниже приведены формальные постановки вышеперечисленных критериев оптимальности для задач компоновки функциональных узлов [13, 14].

A. Минимум числа межузловых соединений

Пусть S_j – j -й маршрут от одного из начальных ребер гиперграфа H к одному из конечных ребер. l_j – условная длина j -го маршрута, которая определяется числом частей гиперграфа H , через которые проходит маршрут S_j .

$$l_j = \sum_{i=1}^m \sum_{k=2}^t \alpha_{ki}^{(j)},$$

где $t=|E_j|, E_j=\{e_j|e_j \in S_j\}$,

$$\alpha_{ki}^{(j)} = \begin{cases} 1, \text{если } X_{k-1} \subset P_i, \text{ но } \exists x \in X_k : x \notin P_i \\ 0, \text{в противном случае} \end{cases}.$$

Таким образом, некоторое фиксированное разбиение f характеризуется величиной: $\alpha_f = l_{j_{\max}}$, которая является максимальным числом частей гиперграфа H , через которые проходят маршруты от начальных ребер гиперграфа к конечным при данном разрезании.

Задача компоновки состоит в отыскании такого разрезания f из множества возможных F гиперграфа H , при котором минимизируется значение α_f при заданных ограничениях: $\Phi = \min \alpha_f$.

B. Минимум числа узлов

Будем считать, что схема представлена взвешенным графом $G=(E,U)$ с матрицей соединений $R=\|r_{ij}\|_{n \times n}$. Пусть каждому элементу $e_i \in E$ приписан некоторый вес $\rho_i > 0 (i=1, 2, \dots, n)$. Заданы ограничения на формирование узлов (k – максимальное число элементов в узле, v – максимальное число выводов на узле). Требуется осуществить при этих условиях компоновку элементов E в узлы $T_1 (l=1, 2, \dots, \gamma)$, таким образом, чтоб количество узлов было минимальным. Решение задачи компоновки сводится к минимизации значения γ при заданных

ограничениях: $\Phi = \min \gamma$, где $\gamma = \sum_{s=1}^p y_s, y_s \in \{0, 1\}, s=1, 2, \dots, p$.

C. Минимум задержек в распространении сигналов

Пусть каждой вершине и ребру ориентированного гиперграфа ставится в соответствие значение условной функциональной задержки (УФЗ). В том случае, если отсутствуют данные о результатах функционально-логического моделирования схемы, то значение УФЗ ставит в соответствие вершине или ребру определенный шаг волны, распространяющейся в гиперграфе от его начальных ребер (для которых УФЗ=1) к конечным. Обозначим УФЗ вершины X_i через r_{xi} , а ребра e_j через r_{ej} . УФЗ ребра e_j , для которого вершина X_i интерпретирует элемент схемы, являющийся источником сигнала в цепи j определяется как $r_{ej}=r_{xi}+1$. УФЗ вершины X_i определяется как максимальное значение УФЗ инцидентных этой вершине ребер $E'=\{e_s, e_t, \dots, e_f\}$ для интерпретируемых которыми цепей схемы, элемент взаимно-однозначно соответствующий вершине X_i не является источником сигнала в цепях $s, t, \dots, f \in I$ схемы, т.е. $r_{xi} = \max_{e_j \in E} r_{ej}$.

Если известны результаты функционально-логического моделирования схемы, то начальным ребрам графа присваиваются значения, равные нулю. УФЗ ребра e_j определяется как $r_{ej} = r_{xi} + \Delta r_{xi}$, где Δr_{xi} – оценка задержки элемента, соответствующего вершине X_i , полученная в результате моделирования схемы.

Каждая часть гиперграфа P_γ обладает такой характеристикой, как интервал значений УФЗ принадлежащих ей вершин: $\Delta R_\gamma = r_{\gamma\max} - r_{\gamma\min}$; где ΔR_γ – интервал значений УФЗ части P_γ гиперграфа H ; $r_{\gamma\max}$ – максимальное значение УФЗ вершин, принадлежащих P_γ ; $r_{\gamma\min}$ – минимальное значение УФЗ вершин, принадлежащих P_γ . Произвольное фиксированное разбиение f гиперграфа H характеризуется максимальной величиной интервала при данном разбиении $\Delta R_f = \Delta R_{\gamma\max}$. При компоновке схемы ставится задача разрезать гиперграф так, чтобы получить минимальное из возможных значений ΔR_f при заданных ограничениях: $\phi = \min \Delta R_{\gamma\max}$.

IV. МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ЦФ В ВИДЕ КАТЕГОРИЙ

Важным этапом в кластеризации является определение признаков, по которым объекты будут распределяться по кластерам. Для того чтобы произвести автоматическую кластеризацию по признаку ЦФ, авторы данной работы предлагают использовать категориальные признаки для описания ЦФ.

Предлагается следующее формальное категориальное описание ЦФ задач компоновки: $F = (O, Op, el)$, где O – искомый тип экстремума; Op – тип операции (сумма, произведение, определение интервала значений); el – тип элементов, над которыми совершается операция Op из множества допустимых при описании ЦФ задач компоновки. Например, описанные выше ЦФ, могут быть представлены в виде: $F_1 = (\min, \sum, \alpha_f)$; $F_2 = (\min, \sum, \gamma)$; $F_3 = (\min, \Delta, R_{\gamma\max})$.

Учтем тот факт, что ЦФ может включать n -частных целевых функций. Тогда в общем случае формальное описание ЦФ имеет вид: $F = (F_1, F_2, \dots, F_n)$, где n – количество частных целевых функций.

V. ПРЕДЛАГАЕМЫЙ МЕТОД

Поскольку, ЦФ задачи компоновки может включать в себя более одного критерия оптимальности, задача оптимизации становится многокритериальной. Стоит отметить, что методы, используемые при решении задачи с одним критерием оптимальности, существенно отличаются от методов решения многокритериальных задач. Поэтому на первом этапе кластеризации авторы предлагают произвести начальное разбиение на кластеры множества задач по признаку количества критериев оптимальности. Такое разбиение представляет собой простую группировку по количественному признаку (количество частных ЦФ). В таком случае число групп определяют по формуле Стерджесса:

$$k = 1 + 3,3221 \lg n = \log_2 + 1 \quad (2),$$

где k – число групп; n – число единиц совокупности [15].

Вторым этапом предлагаемого метода является кластеризация, производимая внутри первоначально образованных кластеров, по признаку формального описания ЦФ. Здесь необходимо использовать такой алгоритм кластерного анализа, который позволяет произвести кластеризацию по категориальным признакам.

Постановка задачи: Пусть имеется множество задач компоновки D , состоящее из множества ЗК $\{d_1, d_2, \dots, d_n\}$. Каждая ЗК есть набор объектов $\{O, Op, el\}$. Множество кластеров $\{C_1, \dots, C_k\}$ есть разбиение множества $\{d_1, \dots, d_n\}$, такое, что $C_1 \dots C_k = \{d_1, d_2, \dots, d_n\}$ $C_i \neq \emptyset \wedge C_i \cap C_j = \emptyset$ для $i \geq 1, j \leq k$. Каждый элемент C_i называется кластером, n, k – количество ЗК и число кластеров соответственно.

Для решения ЗК формализованных ЦФ, необходимо использовать метод, позволяющий группировать категориальные признаки. Мера процент несогласия вычисляется по формуле:

$$\mu(x, y) = \frac{q(x_i \neq y_i)}{i}, \quad (3)$$

где μ – расстояние между двумя объектами x и y ; q – количество неравных пар; i – общее число пар [16].

Далее для кластеризации предлагается использовать любую иерархическую процедуру. Результатом второго этапа кластеризации будет являться распределение ЗК по группам, произведенным в соответствии с формализованным описанием ЦФ конкретной задачи.

Таким образом, результатом первого этапа предлагаемого метода является группировка на первоначальные кластеры по признаку количества частных ЦФ. Далее производим кластеризацию внутри каждого кластера, полученного в результате группировки на первом этапе, любым методом кластерного анализа, позволяющего группировать категориальные данные (рис. 1).

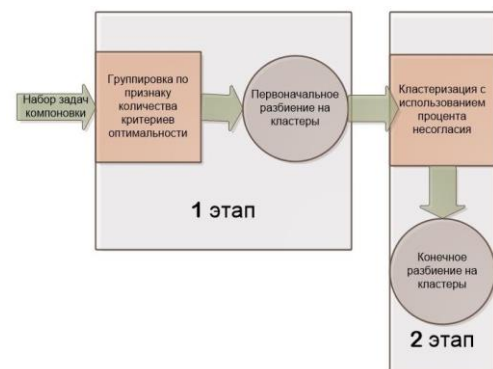


Рис. 1. Схема двухкомпонентного метода автоматической кластеризации

Пример. Пусть имеется множество задач компоновки, состоящее из 5 задач: $\{d_1, d_2, d_3, d_4, d_5\}$. Каждая ЗК категориально описывается некоторой ЦФ F : $F_1 = (F_1, F_1)$, $F_2 = (F_2, F_2, F_2)$, $F_3, F_4, F_5 = (F_5, F_5, F_5)$. В свою очередь: $F_1 = (\min, \sum, \alpha_f)$, $F_1 = (\min, \sum, \gamma)$; $F_2 = (\min, \sum, \gamma)$, $F_2 = (\min, \sum, \gamma)$, $F_2 = (\min, \Delta, R_{\gamma\max})$;

$$F_3 = (\min, \Delta, R_{\gamma_{\max}});$$

$$F_4 = (\min, \sum, \alpha_i);$$

$$F_5 = (\min, \sum, \gamma), F_5'' = (\min, \sum, \gamma), F_5''' = (\min, \Delta, R_{\gamma_{\max}}).$$

Первый этап: группировка по признаку количества критериев оптимальности. В рамках данного примера использование формулы 2 не имеет смысла, поскольку группировка очевидна. Итак, первоначальное разбиение на кластеры имеет вид: 1 кластер: F_1 ; 2 кластер: F_2, F_5 ; 3 кластер: F_3, F_4 (рис. 2).

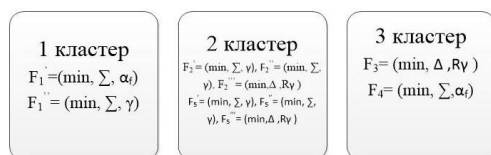


Рис. 2. Результат первоначального разбиения на кластеры

Далее необходимо провести кластеризацию внутри каждого из трех образовавшихся кластеров, используя меру процент несогласия и любой известный алгоритм кластерного анализа. Результаты вычисления процента несогласия сведены в таблицу (формула 3).

ТАБЛИЦА 1 РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЯ МЕРЫ НЕСОГЛАСИЯ

Номер кластера	1	2	3
$\mu(x,y)$	-	0	0,33

Исходя из значений, приведенных в таблице, получим следующее. Кластер 1 содержит одну ЦФ, состоящую из двух частных критериев оптимальности, поэтому мера несогласия в данном случае не вычисляется. В кластер 2 входят две ЦФ, имеющих одинаковое формализованное представление (три частных критерия оптимальности). Мера несогласия в данном случае равна 0. Следовательно, кластер 2 не подлежит дальнейшему разбиению. При вычислении меры несогласия в кластере 3 получили значение 0,33. Следовательно, кластер 3 подлежит дальнейшему разбиению на более мелкие кластеры. Результат 2 этапа кластеризации приведен на рис. 3.

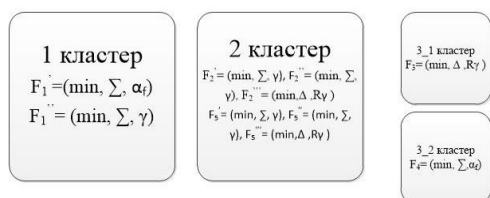


Рис. 3. Результат второго этапа кластеризации

VI. ЗАКЛЮЧЕНИЕ

В работе предложен новый двухкомпонентный метод кластеризации задач компоновки конструктивных узлов.

На основе проведенного анализа задач компоновки предложено новое формальное описание таких задач по признаку ЦФ с использованием категорий. Принципиальное отличие разработанного метода заключается в самой его структуре, представляющей собой комбинацию двух этапов разбиения на кластеры, результатом которого является группировка по признаку формализованного описания ЦФ. В работе рассмотрен пример кластеризации задач компоновки с использованием нового метода. Целью предлагаемого метода является автоматизация процесса кластеризации как одного из этапов построения онтологии. Это позволяет сократить ресурсы, затрачиваемые на разработку онтологий.

СПИСОК ЛИТЕРАТУРЫ

- [1] Курейчик В.М. Математическое обеспечение конструкторского и технологического проектирования с применением САПР: Учебник для вузов. М.: Радио и связь, 1990. 352 с.
- [2] Курейчик В.В., Курейчик В.М. Эволюционные методы компоновки блоков ЭВА. Монография. Таганрог: Изд-во ТРТУ, 2006. 120 с.
- [3] Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. Инженерия знаний. Модели и методы: Учебник. СПб.: Издательство «Лань», 2016. 324 с.
- [4] Курейчик В.М., Сафроненкова И.Б. Разработка архитектуры СППР по выбору методов решения задач компоновки // Информационные технологии. 2017. Т. 23, № 10, С. 736-741.
- [5] Платонов А.В., Полешук В.А. Методы автоматического построения онтологий // Программные продукты и системы. 2016. №2(114). С. 47-52.
- [6] Kureichik, V., Zaruba, D., Kureichik V. Jr.: Hybrid Approach for Graph Partitioning. In: Silhavy, R., Senkerik, R., Kominkova, O.Z., Prokopova, Z., Silhavy, P. (eds.) Artificial Intelligence Trends in Intelligent Systems. AISC, vol. 573. Springer, Cham, 2017, pp. 64-73.
- [7] Пиявский С.А. Методы оптимизации и принятия решений /С.А. Пиявский. Самара: СГАСУ, 2004.
- [8] Баринов С.В., Курейчик В.М., Гладков Л.А. Компоновка МЭС на основе итерационной кластеризации с учетом временных задержек. Известия ТРТУ. Таганрог: Изд-во ТРТУ, 2006. №8(63). С. 120-127.
- [9] Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учебное пособие / А.П. Карпенко. Москва : Издательство МГТУ им. Н.Э. Баумана, 2014. 446, [2] с.
- [10] Евгеньев Е.Б. Интеллектуальные системы проектирования: учеб.пособие / Г.Б. Евгеньев. М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. 334, [2] с.
- [11] Тихонов А.Н., Цветков В.Я. Методы и системы поддержки принятия решений. М.: МАКС Пресс, 2001. 312 с.
- [12] Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. М.: ФИЗМАТЛИТ, 2003. 423 с.
- [13] Таберкин П.А. Компоновка и размещение элементов схем ЭВА с минимизацией длительности задержки сигналов: Автореф. Дис. канд. тех. наук. Таганрог, 1988. 16 с.
- [14] Курейчик В.В., Сороколетов П.В. Композитные методы разбиения графовых моделей. Монография. Таганрог: Изд-во ТРТУ, 2006. 140 с.
- [15] Sturges H.A. The choice of a class interval // Journal of the American Statistical Association. 1926. Vol. 21, no. 153, P. 65-66.
- [16] Сегаран Т. Программируем коллективный разум. / Пер. с англ. СПб: Символ-Плюс, 2015. 368 с.