

Автоматизированный поиск устойчивых научных (исследовательских) групп в составе организаций – участников государственных программ



В. Б. Михайлец,
к. т. н., доцент,
с. н. с.

e-mail: mikhailets@fcntp.ru



И. В. Радин,
ведущий
специалист

e-mail: radin@fcntp.ru



И. С. Соцкова,
главный
специалист

e-mail: sis@fcntp.ru



А. В. Карапышев,
ведущий
специалист

e-mail: karapyshev@fcntp.ru

**Отдел информационно-аналитического и организационного обеспечения,
ФГБНУ «Дирекция научно-технических программ», Москва**

Предложен алгоритм, позволяющий выделять из состава организаций – участников государственных программ устойчивые научные группы, способные выполнять заказы на исследования и разработки, востребованные государством. Алгоритм основан на сортировке в определенной последовательности строк и столбцов матриц парного участия конкретных исполнителей научно-исследовательских и опытно-конструкторских работ. Проверка алгоритма в тестовом режиме показала его практическую применимость при выделении научных (исследовательских) групп в нескольких крупных научно-образовательных учреждениях.

Показано, что выделение из состава организаций – участников государственных программ устойчивых научных групп возможно также с помощью методов иерархической кластеризации.

Ключевые слова: государственные программы исследований и разработок, кластерный анализ, научная (исследовательская) группа.

1. Постановка задачи

В практике реализации государственных программ поддержки научных исследований и технологических разработок все более актуальной становится задача оперативного и объективного выявления научных коллективов или групп, которые могут выполнить конкретный государственный заказ. Знание таких групп способствует лучшей организации программно-целевого планирования развития отечественной науки и техники, в том числе «за счет концентрации ресурсов на работах, осуществляемых научными коллективами, способными на выполнение научных исследований на мировом уровне, правильной постановки задач, персональной мотивации научных кадров» [1].

Выявление научных коллективов в научных организациях возможно путем анкетирования их сотрудников, однако этот путь не лишен влияния различного рода субъективных факторов. С одной стороны, искажение информации о научных коллективах организации может происходить непреднамеренно из-за различного уровня квалификации респондентов, а с

другой – не исключено умышленное ложное информирование для достижения определенных целей. Кроме того, анкетирование может оказаться бесполезным в случае возникновения необходимости выявления научных групп, активно и успешно работающих в рамках одной или нескольких государственных программ. Может случиться так, что известные в научном сообществе группы исследователей вовсе не принимали участия в интересующей программе.

Значительно более объективный и ясный путь – это выделение научных (исследовательских) групп из состава организаций – участников государственных программ. Например, если организация в течение нескольких лет выполняла научно-исследовательские и опытные конструкторские работы в рамках какой-то программы или нескольких программ, то наверняка в составе этой организации есть сотрудники, неоднократно участвовавшие в этих работах. Основное препятствие, которое может возникнуть на этом пути, это – не структурированность информации о сотрудниках-исполнителях в отчетных документах, либо – недостаточность информации, или ее полное отсутствие.

Целью данной работы стало создание алгоритма автоматизированного поиска устойчивых научных (исследовательских) групп в составе организаций – участников государственных программ.

Под устойчивой научной группой в данной работе понимается группа специалистов численностью более S человек, участвовавших совместно в выполнении не менее W научно-исследовательских работ (контрактов) по одной или нескольким государственным программам развития науки и технологий в течение T последних лет.

2. Выделение научных (исследовательских) групп с помощью матриц парного участия

Пусть, например, некоторая организация участвовала в реализации одной из государственных программ и выполнила 5 контрактов по исследованиям и разработкам в течение последних нескольких лет (см. табл. 1).

В столбце 7 табл. 1 приведены количество выполненных контрактов (частоты участия) 10 сотрудников в выполнении НИОКР в рамках некоторой государственной программы. Если установить, что научной (исследовательской) группой можно считать группу, выполнившую хотя бы один контракт программы, то на первый взгляд кажется, что группа из 10 человек как раз и является такой группой. Однако, если проанализировать содержание столбца 8 табл. 1, то становится понятной ошибочность такого

предположения, поскольку в таком составе группа не выполнила ни одного контракта. Хотя этот факт кажется очевидным, при большем числе сотрудников-исполнителей и, особенно, при ужесточении требования к нижней границе признания группы устойчивой по числу контрактов, распознавание группы существенно затрудняется.

Решение задачи выделения групп по заданным критериям из групп большего масштаба может быть в значительной степени облегчено за счет использования матриц парного участия сотрудников в выполнении работ (см. табл. 2).

В матрице парного участия на пересечении строк и столбцов отражено число контрактов организации, в которых принимали участие пары сотрудников. В ячейках матрицы парного участия содержится число контрактов, выполненных совместно сотрудниками i -й строки и j -го столбца (n_{ij}).

Например, из матрицы парного участия для рассматриваемой условной организации следует, что 10 сотрудников {Волков, Егоров, Ежов, Зайцев, Иванов, Козлов, Медведев, Орлов, Петров и Сидоров} не могут быть признаны группой, потому что совместно они не выполнили ни одного контракта, хотя каждый из них участвовал в выполнении не менее одного контракта. Но, возможно, эти 10 сотрудников составляют не одну, а несколько групп?

Для ответа на этот вопрос, была проведена работа по изучению возможности применения теории графов, методов комбинаторной оптимизации, кластерного

Таблица 1

Сотрудники, выполнявшие контракты по исследованиям и разработкам в рамках одной из государственных программ

№ п/п	Исполнители контрактов	№ контракта	Порядковый номер исполнителя в контракте	№ п/п	Перечень исполнителей контрактов	Количество выполненных контрактов	Номера выполненных контрактов
1	2	3	4	5	6	7	8
1	Иванов	1	1	1	Волков	3	3 4 5
2	Петров	1	2	2	Егоров	3	3 4 5
3	Зайцев	1	3	3	Ежов	2	3 4
4	Медведев	1	4	4	Зайцев	2	1 3
5	Сидоров	1	5	5	Иванов	2	1 2
6	Иванов	2	1	6	Козлов	3	3 4 5
7	Петров	2	2	7	Медведев	2	1 3
8	Орлов	2	3	8	Орлов	3	2 4 5
9	Козлов	3	1	9	Петров	2	1 2
10	Егоров	3	2	10	Сидоров	2	1 3
11	Зайцев	3	3				
12	Волков	3	4				
13	Ежов	3	5				
14	Медведев	3	6				
15	Сидоров	3	7				
16	Козлов	4	1				
17	Егоров	4	2				
18	Орлов	4	3				
19	Волков	4	4				
20	Ежов	4	5				
21	Козлов	5	1				
22	Егоров	5	2				
23	Орлов	5	3				
24	Волков	5	4				

Таблица 2

Матрица парного участия сотрудников в выполнении работ

	Волков	Егоров	Ежов	Зайцев	Иванов	Козлов	Медведев	Орлов	Петров	Сидоров
Волков	3	3	2	1	0	3	1	2	0	1
Егоров	3	3	2	1	0	3	1	2	0	1
Ежов	2	2	2	1	0	2	1	1	0	1
Зайцев	1	1	1	2	1	1	2	0	1	2
Иванов	0	0	0	1	2	0	1	1	2	1
Козлов	3	3	2	1	0	3	1	2	0	1
Медведев	1	1	1	2	1	1	2	0	1	2
Орлов	2	2	1	0	1	2	0	3	1	0
Петров	0	0	0	1	2	0	1	1	2	1
Сидоров	1	1	1	2	1	1	2	0	1	2

Таблица 3

Матрица парного участия после сортировки главной диагонали

	Волков	Егоров	Козлов	Орлов	Ежов	Зайцев	Медведев	Сидоров	Иванов	Петров
Волков	3	3	3	2	2	1	1	1	0	0
Егоров	3	3	3	2	2	1	1	1	0	0
Козлов	3	3	3	2	2	1	1	1	0	0
Орлов	2	2	2	3	1	0	0	0	1	1
Ежов	2	2	2	1	2	1	1	1	0	0
Зайцев	1	1	1	0	1	2	2	2	1	1
Медведев	1	1	1	0	1	2	2	2	1	1
Сидоров	1	1	1	0	1	2	2	2	1	1
Иванов	0	0	0	1	0	1	1	1	2	2
Петров	0	0	0	1	0	1	1	1	2	2

анализа и др. [2–10] для решения задачи выделения устойчивых научных групп в составе организаций. В результате, был разработан метод выделения групп, в основу которого положена последовательная перестановка строк и столбцов матрицы парного участия для их ранжирования по числу совместно выполненных контрактов i -м и j -м сотрудниками.

Суть метода решения задачи выделения научных (исследовательских) групп показана в табл. 3–7. Базовый алгоритм для построения общего комплекса процедур выделения групп основан на использовании матриц парного участия сотрудников организации в государственной программе и включает следующие основные шаги:

Шаг А. Сортировка строк и столбцов матрицы парного участия относительно содержимого ячеек главной диагонали (см. табл. 3).

Шаг Б. Сортировка строк и столбцов матрицы парного участия по содержимому первого столбца и первой строки (см. табл. 4).

Шаг В. Сортировка строк и столбцов матрицы парного участия по содержимому i -го столбца и i -й строки (см. табл. 5).

Шаг Г. Процедура сортировки строк и столбцов продолжается до $N-2$ строки матрицы парного участия (см. табл. 6).

Шаг Д. Выделение научных (исследовательских) групп путем очерчивания ячеек, расположенных симметрично главной диагонали, число контрактов в которых не равно нулю или превышает требуемое значение.

В табл. 6 можно выделить две научные (исследовательские) группы по 5 человек в каждой, выполнившие хотя бы один контракт в рамках рассматриваемой программы.

Таблица 4

Матрица парного участия после сортировки строк и столбцов по содержимому первого столбца и первой строки

	Волков	Егоров	Козлов	Орлов	Ежов	Зайцев	Иванов	Медведев	Петров	Сидоров
Волков	3	3	3	2	2	1	0	1	0	1
Егоров	3	3	3	2	2	1	0	1	0	1
Козлов	3	3	3	2	2	1	0	1	0	1
Орлов	2	2	2	3	1	0	1	0	1	0
Ежов	2	2	2	1	2	1	0	1	0	1
Зайцев	1	1	1	0	1	2	1	2	1	2
Иванов	0	0	0	1	0	1	2	1	2	1
Медведев	1	1	1	0	1	2	1	2	1	2
Петров	0	0	0	1	0	1	2	1	2	1
Сидоров	1	1	1	0	1	2	1	2	1	2

Таблица 5

*Матрица парного участия после сортировки строк и столбцов по содержимому i-го столбца и i-й строки
(для примера показан результат сортировки четвертых столбца и строки)*

	Волков	Егоров	Козлов	Орлов	Ежов	Иванов	Петров	Зайцев	Медведев	Сидоров
Волков	3	3	3	2	2	0	0	1	1	1
Егоров	3	3	3	2	2	0	0	1	1	1
Козлов	3	3	3	2	2	0	0	1	1	1
Орлов	2	2	2	3	1	1	1	0	0	0
Ежов	2	2	2	1	2	0	0	1	1	1
Иванов	0	0	0	1	0	2	2	1	1	1
Петров	0	0	0	1	0	2	2	1	1	1
Зайцев	1	1	1	0	1	1	1	2	2	2
Медведев	1	1	1	0	1	1	1	2	2	2
Сидоров	1	1	1	0	1	1	1	2	2	2

Таблица 6

*Матрица парного участия после сортировки N–2 строк и столбцов при требовании
выполнения группой хотя бы одного контракта*

	Волков	Егоров	Козлов	Орлов	Ежов	Зайцев	Медведев	Сидоров	Иванов	Петров
Волков	3	3	3	2	2	1	1	1	0	0
Егоров	3	3	3	2	2	1	1	1	0	0
Козлов	3	3	3	2	2	1	1	1	0	0
Орлов	2	2	2	3	1	0	0	0	1	1
Ежов	2	2	2	1	2	1	1	1	0	0
Зайцев	1	1	1	0	1	2	2	2	1	1
Медведев	1	1	1	0	1	2	2	2	1	1
Сидоров	1	1	1	0	1	2	2	2	1	1
Иванов	0	0	0	1	0	1	1	1	2	2
Петров	0	0	0	1	0	1	1	1	2	2

Если требование к числу выполненных группой контрактов ужесточить (к примеру, более 1 контракта), то количественный состав групп или их количество, естественно, уменьшится. В приведенном примере при таком требовании остается две группы (см. табл. 7, 8: одна группа из четырех человек {Волков, Егоров, Козлов, Орлов}, вторая группа из трех человек {Зайцев, Медведев, Сидоров}).

Шаг Е. Уточнение состава научных (исследовательских) групп по результатам анализа дополнительной информации.

Выделенные с помощью указанных процедур группы требуют проверки и уточнения. Так, например, наряду с признаками устойчивости совместного сотрудничества по числу выполненных контрактов возможно наличие обстоятельств, заставляющих изменять состав групп. Такими обстоятельствами могут быть, например, сведения о реальной роли и личном вкладе того или иного сотрудника в вы-

полнение научно-исследовательских и опытно-конструкторских работ. Такие сведения в отчетных документах, содержащих в том или ином виде списки сотрудников-исполнителей, как правило, отсутствуют. Тогда в составе групп могут оказываться лица, выполняющие контролирующие, административные и другие функции (менеджеры, нормоконтролеры, планировщики, бухгалтера и т. п.), не связанные с научным творчеством.

Для исключения ошибок состав полученных научных (исследовательских) групп должен быть изучен, как с использованием сведений во всех доступных отчетных документах в базах данных государственных программ, так и с использованием других доступных (открытых) источников информации, в том числе интернета.

Таким образом, разработанный алгоритм выделения научных (исследовательских) групп из состава организаций оказался вполне пригодным

Таблица 7

*Матрицы парного участия после сортировки всех строк
и столбцов при требовании выполнения группой
более одного контракта*

	Волков	Егоров	Козлов	Орлов
Волков	3	3	3	2
Егоров	3	3	3	2
Козлов	3	3	3	2
Орлов	2	2	2	3

Таблица 8

*Матрицы парного участия после сортировки всех строк
и столбцов при требовании выполнения группой
более одного контракта*

	Зайцев	Медведев	Сидоров
Зайцев	2	2	2
Медведев	2	2	2
Сидоров	2	2	2

для практического применения. С его помощью в тестовом режиме были выделены научные (исследовательские) группы в нескольких крупных научно-образовательных учреждениях.

3. Выделение научных (исследовательских) групп средствами иерархической кластеризации

Кластерный анализ (англ. cluster analysis) — многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.

В литературе [4–10] описывается множество различных методов кластеризации, основанных на использовании матриц сходства, оценивании функций плотности статистического распределения, эвристических алгоритмах перебора, идеях математического программирования и др. Выяснилось, однако, что особенность задачи, состоящая в необходимости проверки соблюдения условия совместности выполнения научных исследований и разработок кандидатами в члены научной группы в более, чем в определенном числе контрактов, не позволяет напрямую использовать известные методы кластеризации. Причиной тому является нацеленность кластерного анализа только на декомпозицию объектов на группы по формальным признакам самих объектов. Признак совместности функционирования рассматриваемых объектов — столь важный в решаемой задаче — не может быть признаком (атрибутом) этих объектов, так как характеризует не объекты, а группы объектов. Несмотря на это, как будет показано ниже, методы кластерного анализа могут быть весьма полезны для решения задачи выделения научных (исследовательских) групп из состава организаций — участников государственных программ.

Общепринятой классификации методов кластеризации не существует, но можно выделить ряд групп подходов.

1. Вероятностный подход: предполагается, что каждый рассматриваемый объект относится к одному из k классов.
2. Подходы на основе систем искусственного интеллекта.
3. Логический подход: разбиение на группы осуществляется с помощью дерева решений.
4. Теоретико-графовый подход: основан на графовых алгоритмах кластеризации.
5. Иерархический подход: предполагается наличие вложенных групп (кластеров различного порядка).

Наиболее применимы для разработки научно-методического инструментария формализованной идентификации научных (исследовательских) групп, по мнению авторов данной статьи, являются методы иерархической кластеризации.

Иерархическая кластеризация — это совокупность алгоритмов упорядочивания данных. Алгоритмы упорядочивания данных указанного типа исходят из того, что некое множество объектов характеризуется определенной степенью связности. Предполагается

наличие вложенных групп (кластеров различного порядка). Для визуализации результатов иерархической кластеризации, наряду с другими средствами, широко используется дендрограмма.

Под дендрограммой обычно понимается дерево, то есть граф без циклов, построенный по матрице мер близости. Дендрограмма позволяет изобразить взаимные связи между объектами из заданного множества. Для создания дендрограммы требуется матрица сходства (или различия), которая определяет уровень сходства между парами объектов.

Наиболее употребительными алгоритмами иерархической кластеризации являются:

- метод одиночной связи (англ. single linkage) — «метод ближайшего соседа»;
- метод полной связи (англ. complete linkage) — «метод дальнего соседа»;
- метод средней связи (англ. pair-group method using arithmetic averages);
- центроидный метод (англ. pair-group method using the centroid average);
- метод Уорда (англ. Ward's method) — пошаговый алгоритм [9, 10].

Применительно к решаемой задаче выделения научных (исследовательских) групп алгоритмы иерархической кластеризации могут быть использованы для упорядочивания списков фамилий исполнителей контрактов по критериям близости их по числу совместно выполненных контрактов. Однако нужно всегда учитывать, что любое решение задачи кластеризации принципиально неоднозначно. Причинами неоднозначности являются:

1. Не существует однозначно наилучшего критерия качества кластеризации. Известен целый ряд алгоритмов, осуществляющих достаточно разумную кластеризацию «по построению». Все они могут давать разные результаты.
2. Число кластеров, как правило, неизвестно заранее и устанавливается в соответствии с некоторым субъективным критерием.
3. Результат кластеризации существенно зависит от метрики и начальных условий, которые, к тому же, определяются или задаются экспертным путем.

Постановка задачи иерархической кластеризации для выделения научных (исследовательских) групп из состава организаций — исполнителей государственных контрактов заключается в следующем.

Имеется n сотрудников-исполнителей m контрактов. Каждый i -й сотрудник мог участвовать или не участвовать в выполнении k -го контракта — K_{ik} ($K_{ik} \in \{0, 1\}$, $i \in \{1, 2, \dots, n\}$, $k \in \{1, 2, \dots, m\}$).

Кроме того, каждый i -й сотрудник может характеризоваться некоторыми свойствами H . Свойства H могут быть, как измеряемы, так и не измеряемы, но те и другие должны быть кодируемы и, при необходимости, нормируемы. Свойствами H могут быть, например, вклад сотрудника в выполнение работы по контракту, специализация и т. п. H_{ir} — значения свойств ($H_{ir} \in \{-\infty, +\infty\}$, $i \in \{1, 2, \dots, n\}$, $r \in \{m+1, m+2, \dots, M\}$).

Таким образом, для задачи выделения научных (исследовательских) групп так называемая матрица наблюдений $X (X_{ij})$ размерами $n \times M$ (n —

число исполнителей контрактов, $i \in \{1, 2, \dots, n\}$; M — число признаков: номеров контрактов, свойств, $j \in \{1, 2, \dots, m, m+1, m+2, \dots, M\}$, являющаяся входом для алгоритмов иерархической кластеризации, имеет следующий вид:

$$[X] = [[K][H]] = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1M} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{iM} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{nj} & \dots & X_{nM} \end{bmatrix} = \quad (1)$$

$$= \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1k} & \dots & K_{1m} & H_{1,m+1} & H_{1,m+2} & \dots & H_{1,m+r} & \dots & H_{1M} \\ K_{21} & K_{22} & \dots & K_{2k} & \dots & K_{2m} & H_{2,m+1} & H_{2,m+2} & \dots & H_{2,m+r} & \dots & H_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_{i1} & K_{i2} & \dots & K_{ik} & \dots & K_{im} & H_{i,m+1} & H_{i,m+2} & \dots & H_{i,m+r} & \dots & H_{iM} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & \dots & K_{nk} & \dots & K_{nm} & H_{n,m+1} & H_{n,m+2} & \dots & H_{n,m+r} & \dots & H_{nM} \end{bmatrix},$$

где [] — обозначение матрицы.

Требуется разбить n исполнителей $\{1, 2, 3, \dots, i, \dots, n\}$ на непересекающиеся подмножества кластеры $\{S_1, S_2, \dots, S_k\}$ так, чтобы обеспечить минимум (экстремум) некоторого критерия (функционала качества), т. е.:

$$S = \{S_1, S_2, \dots, S_k\}; F(S) \rightarrow \min (\max) \quad (2)$$

Если эти данные понимать как точки в признаком пространстве, то задача кластерного анализа формулируется как выделение «сгущений точек» и разбиение исходной совокупности на однородные подмножества объектов. Кластерный анализ можно рассматривать также как метод редукции (сжатия) некоторого множества данных в более компактную классификацию объектов. В этом и состоит наиболее общая постановка задачи иерархической кластеризации.

В табл. 9, для примера, приведена матрица наблюдений для иерархической кластеризации только по пяти признакам (свойствам) — номерам контрактов, чтобы обеспечить сопоставимость с результатами

выделения групп путем сортировки столбцов и строк матриц парного участия (см. раздел 2 выше).

Кластеризация «методом дальнего соседа» при евклидовой метрике расстояний между 10 сотрудниками-исполнителями дает следующую дендрограмму кластеров (см. рис. 1):

- 2 группы (*{Иванов, Петров, Зайцев, Медведев, Сидоров}* и *{Волков, Козлов, Егоров, Ежов, Орлов}*);
- 3 группы (*{Иванов, Петров, Зайцев, Медведев, Сидоров}*, *{Волков, Козлов, Егоров, Ежов}* и *{Орлов}*);
- 4 группы (*{Иванов, Петров}*, *{Зайцев, Медведев, Сидоров}*, *{Волков, Козлов, Егоров, Ежов}* и *{Орлов}*);
- 5 групп (*{Иванов, Петров}*, *{Зайцев, Медведев, Сидоров}*, *{Волков, Козлов, Егоров}*, *{Ежов}* и *{Орлов}*).

(Следует отметить, что расчеты были выполнены с помощью пакета Statistics Toolbox программы MATLAB.)

Такое представление разбивки сотрудников на группы имеет недостаток, состоящий в отсутствии наглядности отражения причин, приведших к полученным результатам. Не видно также, соблюдается ли условие — «любая пара сотрудников должна была участвовать в более чем в заданном числе контрактов». Указанный недостаток устраняется использованием матриц парного участия, ячейки которых содержат не число выполненных каждой парой сотрудников контрактов, а их номера (см. табл. 10).

Теперь в зависимости от условий задачи выделения научных (исследовательских) групп можно формировать группы, как по количеству контрактов, так и по их тематике.

Например, если задано условие, что научные группы должны быть численностью более трех человек, выполнивших вместе хотя бы один контракт, то такому условию удовлетворяют 2 группы: *{Иванов, Петров, Зайцев, Медведев, Сидоров}* и *{Волков, Козлов, Егоров, Ежов, Орлов}*. Следует заметить, что такие же группы были выделены предложенным методом сортировки всех строк и столбцов в матрице парного участия (см. табл. 6).

Если задано условие, что научные группы должны быть численностью не менее двух человек, выполнивших вместе хотя бы два контракта, то такому условию удовлетворяют 3 группы: *{Иванов, Петров}*, *{Зайцев, Медведев, Сидоров}* и *{Волков, Козлов, Егоров, Ежов, Орлов}*.

Таблица 9

Матрица наблюдений для иерархической кластеризации

Исполнители	Признаки (свойства)				
	Конт-тракт 1	Конт-тракт 2	Конт-тракт 3	Конт-тракт 4	Конт-тракт 5
Волков	0	0	1	1	1
Егоров	0	0	1	1	1
Ежов	0	0	1	1	0
Зайцев	1	0	1	0	0
Иванов	1	1	0	0	0
Козлов	0	0	1	1	1
Медведев	1	0	1	0	0
Орлов	0	1	0	1	1
Петров	1	1	0	0	0
Сидоров	1	0	1	0	0

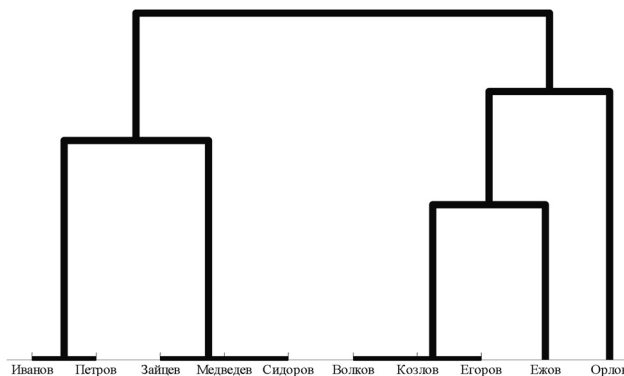


Рис. 1. Дендрограмма иерархической кластеризации

Матрица парного участия по номерам выполненных контрактов

	Иванов	Петров	Зайцев	Медведев	Сидоров	Волков	Козлов	Егоров	Ежов	Орлов
Иванов	1 2	1 2	1	1	1					2
Петров	1 2	1 2	1	1	1					2
Зайцев	1	1	1 3	1 3	1 3	3	3	3	3	
Медведев	1	1	1 3	1 3	1 3	3	3	3	3	
Сидоров	1	1	1 3	1 3	1 3	3	3	3	3	
Волков			3	3	3	3 4 5	3 4 5	3 4 5	3 4	4 5
Козлов			3	3	3	3 4 5	3 4 5	3 4 5	3 4	4 5
Егоров			3	3	3	3 4 5	3 4 5	3 4 5	3 4	4 5
Ежов			3	3	3	3 4	3 4	3 4	3 4	4
Орлов	2	2				4 5	4 5	4 5	4	2 4 5

Медведев, Сидоров и *{Волков, Козлов, Егоров, Ежов}* и т. д.

Таким образом, совместное применение методов иерархической кластеризации для упорядочивания сотрудников по числу совместно выполненных контрактов и сортировки в определенной последовательности строк и столбцов матриц парного участия конкретных сотрудников-исполнителей завершённых работ приводит к возможности автоматизированного поиска устойчивых научных (исследовательских) групп в составе организаций – участников государственных программ.

Дополнительными показателями личного вклада исполнителей в научные исследования и разработки в рамках государственных программ [11], предложенный алгоритм позволит использовать полученные результаты для принятия управленческих решений в научной сфере, особенно при оценке эффективности (успешности) работы отдельных научных групп и коллективов, а также выявлять вновь возникающие научные (исследовательские) группы.

Заключение

Предложенный алгоритм выделения научных (исследовательских) групп из состава организаций – участников государственных программ, опробованный в тестовом режиме, оказался вполне пригодным для практического применения. Сопоставление результатов выделения научных групп с помощью разработанного алгоритма в нескольких крупных научно-образовательных учреждениях со сведениями, собранными из других независимых источников, показывает достаточно высокое соответствие.

Предложенный алгоритм может быть использован в любых областях знаний, где требуется автоматизированное выявление из большого массива элементов устойчивых их совокупностей. При этом, устойчивость определяется по нескольким задаваемым базовым признакам.

Список использованных источников

1 Государственная программа Российской Федерации «Развитие науки и технологий» на 2013–2020 гг. Утверждена распоряжением Правительства Российской Федерации № 2433-р от 20 декабря 2012 г.

2. М. Свами, К. Тхуласираман. Графы, сети и алгоритмы. М: Мир, 1984.
 3. А. В. Леоненков. Решение задач оптимизации в среде MS Excel. СПб.: БХВ-Петербург, 2005.
 4. С. А. Айвазян, З. И. Бежаева, О. В. Староверов. Классификация многомерных наблюдений. М.: Статистика, 1974.
 5. Б. Дюран, П. Оделл. Кластерный анализ. М.: Статистика, 1977.
 6. Классификация и кластер/Под ред. Дж. Вэн-Райзина. М.: Мир, 1980.
 7. М. Жамбю. Иерархический кластер-анализ и соответствия. М.: Финансы и статистика, 1988.
 8. И. Д. Мандель. Кластерный анализ. М.: Финансы и статистика, 1988.
 9. http://ru.wikipedia.org/wiki/Иерархическая_кластеризация.
 10. <http://matlab.exponenta.ru/statist/book2/14/linkage.php>.
 11. А. В. Карапышев, А. А. Мазнев, В. Б. Михайлец, А. Н. Петров, И. В. Радин, И. С. Соцкова, К. В. Шуртаков. Измерение вклада исполнителей в научные исследования и разработки, проводимые в рамках государственных программ//Иновации, № 7, 2013.

Automated search for sustainable scientific (research) groups of the organizations participating in the state programs

V. B. Mikhailets, PhD (in Technical Sciences), lecturer, Deputy head, analytical and organizational support, FGFR «Directorate science and technology programs».

I. V. Radin, Leading specialist of the information, analytical and organizational support, FGFR «Directorate science and technology programs».

A. V. Karapyshev, Leading specialist of the information, analytical and organizational support, FGFR «Directorate science and technology programs».

I. S. Sotskova, chief specialist of the information, analytical and organizational support, FGFR «Directorate science and technology programs».

An algorithm is proposed, which allows to allocate from the participating institutions of the state programs steady scientific groups capable to carry out orders for research and development needed by the state. The algorithm is based on sorting in a certain sequence of rows and columns of matrices pair participation of specific performers of research and development work. Testing of the algorithm in test mode has shown its practical applicability of the allocation of the scientific (research) groups in several major scientific and educational institutions.

It is shown that the selection of sustainable scientific groups from the composition of the organizations participating in the state programs is also possible using the methods of hierarchical clustering.

Keywords: state research and development programs, cluster analysis, scientific (research) group.