

К.Б. Чихачев

**МЕТОД КЛАСТЕРИЗАЦИИ ОЧАГОВ ПОЖАРОВ
ПО ДАННЫМ КОСМИЧЕСКОГО ЗОНДИРОВАНИЯ**

К.В. Tchikhatchev

**SEATS OF FIRE CLUSTERIZATION METHOD
WITH RESPECT TO COSMIC EXPLORATION DATA**

Предлагается метод выделения очагов пожаров по данным космического зондирования для последующего решения задач численного моделирования состояния атмосферы при лесных и торфяных пожарах. Дается краткий обзор основных применяемых методов кластеризации, применяемых метрик и критериев оптимальности. Приводится описание принципа работы двух алгоритмов кластеризации. Предложен метод нахождения оптимальных настроечных параметров и критерий оценки качества работы одного из алгоритмов применительно к поставленной задаче.

Ключевые слова: лесные пожары, численное моделирование атмосферы, методы кластеризации.

Seats of fire allocation method with respect to cosmic exploration data is proposed for future solving problems of numerical modeling of the atmosphere during forest and peat fires. A brief overview of the main applied clusterization methods, applied metrics and optimality criteria provides. A description is given for the principle of the two algorithms for clustering. Method is proposed for finding the optimal tuning parameters and criteria for assessing the quality of one of the algorithms applied to the task.

Key words: forest fires, numerical modeling of the atmosphere, clustering techniques.

Наиболее активные процессы выноса продуктов сгорания от лесных и торфяных пожаров в более высокие слои атмосферы реализуются возникающими в случаях интенсивного перегрева в зоне сгорания вертикальными струями от некоторой совокупности очагов горения, которые взаимодействуют как между собой, так и с внешним потоком [1].

На основе данных зондирования с искусственных спутников Земли (ИСЗ) в оперативном режиме оказывается возможным приближенно оценивать местоположение и яркостную температуру очагов горения, причем местоположение дается с точностью до размера проекции пикселя матрицы приемника излучения на ИСЗ (около 1 км на местности).

При обработке этих данных применительно к информационному обеспечению последующих расчетов с использованием моделей, возникает задача оптимальной группировки одиночных очагов пожаров. Общее требование здесь состоит в том, чтобы, с одной стороны, выделенные области (кластеры очагов горения) не были слишком велики и допускали последующее 3D моделирование с высоким пространственным разрешением, а с другой – не взаимодействовали динамически и термодинамически друг с другом.

Поскольку данные ИСЗ дают нам весьма разнообразные по яркостной температуре пиксели, как активному горению, так и угасающим участкам очагов пожара, то при кластеризации возникает необходимость использовать не только координаты точек, но и зафиксированную в них температуру, поскольку интерес представляют, в первую очередь, именно участки активного горения. Для этого при кластеризации целесообразно применять не обычную евклидову метрику расстояния между пикселями, а метрику, использующую помимо взаимного положения пикселей еще и разность зафиксированных в них температур. В рамках данной работы использовалась следующая метрика взаимного положения пикселей $p = (p_x, p_y, p_t)$ и $q = (q_x, q_y, q_t)$:

$$|p, q| = \left((p_x - q_x)^2 + (p_y - q_y)^2 \right)^{1/2} + k |p_t - q_t|, \quad (1)$$

где k – размерный коэффициент увязки влияния температуры точек, индексами x и y обозначены координаты пикселя (прямоугольные или географические), а индексом t – температура.

Существует значительное количество алгоритмов для решения задач такого типа кластеризации, которые ищут оптимальное разбиение точек на кластеры по разным нормам и разным показателям оптимальности (например, [2, 4, 5]). Некоторые из них накладывают дополнительные ограничения на применяемую норму. В большинстве случаев – это требование выполнения неравенства треугольника. Нетрудно видеть, что в нашем случае это условие выполнено.

В качестве критерия оптимальности может применяться диаметр кластера, средняя удаленность от центра кластера, среднее попарное расстояние, размеры окаймляющего кластер прямоугольника, количество кластеров. Для целей решаемой нами задачи критериями оптимальности будут отклонение от желаемого количества кластеров и размеры окаймляющего прямоугольника, поскольку основные ограничения задачи моделирования – это целостность очага пожара и ограниченность области моделирования.

Наиболее подходящим для нашей задачи следует признать алгоритм [6], основанный на выделении ровно m центров кластеров и дальнейшем отнесении оставшихся точек к ближайшему центру. Задача выделения m центров является NP -полной, т.е. построение точного алгоритма, который находил бы оптимальное решение за полиномиальное от количества точек время, является в настоящее время неразрешимой задачей. Таким образом, для значительных по размерам входных данных возможно использование лишь приближенных алгоритмов. Предложенный приближенный алгоритм находит решение не более чем в 2 раза худшее, чем оптимальное. Суть самого алгоритма заключается в последовательном нахождении очередной точки, которая будет добавлена последней при применении алгоритма Дейкстры [3] из уже найденных точек.

Распространение очага пожара зависит от многих факторов, поэтому является неравномерным по направлениям. Вследствие этого очаги пожара зача-

стью могут иметь неправильную форму. По этой причине применение описанного алгоритма кластеризации не дает желаемых результатов, поскольку строит выпуклые кластеры наименьшего диаметра, а не в форме очагов пожаров. Для решения этой проблемы предлагается алгоритм кластеризации на основе построения скелета (остовного дерева) точек с последующим его разделением на кластеры.

Рассмотрим полный граф на множестве всех точек зондирования с весами на ребрах, равными расстоянию между точками в метрике (1). Используя алгоритм Прима [3], выделим основное дерево, т.е. минимальный по сумме длин набор ребер при котором граф остается связным.



Рис. 1. Блок-схема алгоритма

Алгоритм заключается в следующем. Будем строить дерево, последовательно добавляя в него ребра. На каждом следующем шаге выбираем из вершин еще не принадлежащих дереву ту вершину, у которой длина кратчайшего ребра до вершин уже построенного дерева минимальна. Добавим ее в дерево с ее кратчайшим ребром. Поскольку изначальный граф связный (в нашем случае он полный), то данную операцию возможно проводить до тех пор пока в дерево не будут добавлены все вершины. Также следует заметить, что поиск вершины для добавления можно производить не перебором всех вершин и всех ребер из нее, а сохранив для каждой вершины, не принадлежащей дереву, кратчайшее ребро до вершин дерева. Блок-схема алгоритма показана на рис. 1.

При добавлении очередной вершины в дерево кратчайшее ребро для какой-то другой вершины может измениться только в том случае, если ребро до вновь добавленной вершины короче сохраненного. Время обновления массива кратчайших ребер пропорционально количеству вершин, в то время как непосредственное его вычисление – квадрату количества вершин. Работа алгоритма проиллюстрирована на рис. 2, а. Текущее дерево выделено черными ребрами, кратчайшие ребра для остальных вершин – серыми. На очередном шаге будет взято наименьшее серое ребро и добавлено в дерево.

Как показали проведенные эксперименты, построенный таким образом остов содержит в основном ребра, лежащие внутри очагов пожара, соединяя кроме этого лишь ближайшие очаги единичными ребрами большой длины. После исключения ребер с длиной большей порогового значения d_{max} , граф распадается на компоненты связности, которые и соответствуют очагам пожаров (или группам близкорасположенных очагов).

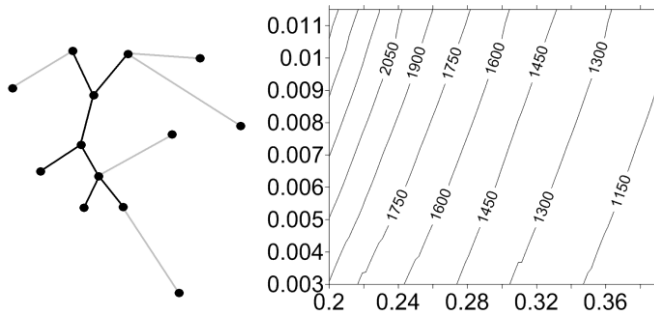


Рис. 2. *a* – пример построенного фрагмента остова дерева и кратчайшие ребра до него; *б* – расчетная зависимость количества кластеров от порогового значения d_{\max} и коэффициента k

Следует заметить, что поскольку данный алгоритм использует в метрике расстояние между точками на всем наборе точек зондирования, т.е. на поверхности всего земного шара, то логичнее было бы в формуле (1) использовать вместо выражения $[(p_x - q_x)^2 + (p_y - q_y)^2]^{1/2}$, которое соответствует расстоянию в проекции Меркатора, расстояние между точками на эллипсоиде. Однако в этом нет необходимости, поскольку результат работы алгоритма Прима существенно зависит только от соотношения расстояний близкорасположенных точек, которое несущественно различается из-за достаточной густоты расположения точек зондирования в районах пожаров.

Для определения наилучших значений порогового значения d_{\max} и коэффициента увязки k был проведен комплекс расчетов кластеров по архиву данных космического зондирования за лето 2011 г. для широкого диапазона значений параметров. Данные включают в себя следующие поля: географические координаты точки, размеры прямоугольника, на котором зафиксировано превышение температуры, дата, температура и выброс. Исследование показало, что наилучший результат в смысле решения поставленной в начале данной работы задачи достигается при количестве кластеров близком к 1700. Зависимость количества кластеров от порогового значения d_{\max} и коэффициента k показана на рис. 2, *б*.

Из рисунка видно, что значение 1700 задает нам лишь изолинию взаимозависимости параметров, однако по смыслу понятно, что ни расстояние, ни разность температур не могут превалировать в определении метрики, поэтому следует использовать такое значение k , при котором расстояние и разность температур вносят одинаковый вклад в определение метрики. Взяв в качестве такого значения $0,0075^\circ/\text{K}$ получает значение d_{\max} равным $0,15^\circ$.

В случае если на каком-либо наборе данных эти значения дают неудовлетворительный результат, то возможно их определение с использованием экспертной оценки. При этом настроечным параметром будет количество кластеров, а оцениваться будет качество разбиения на кластеры при вычисленных значениях d_{\max} и k .

Пример использования предложенного алгоритма применительно к данным на 7 октября 2011 г. приведен на рис. 3. Очаги представлены черными квадратами и обрамлены прямоугольниками кластеров.

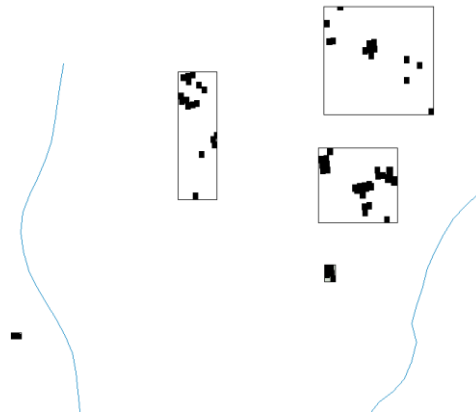


Рис. 3. Очаги пожара, выделенные при помощи алгоритма

Отдельные точки, выделенные алгоритмом в самостоятельные кластеры, являются точками с низкой температурой и интереса для моделирования не представляют.

Таким образом, предложенный алгоритм оказывается весьма надежным инструментом для автоматического выделения локальных областей территории с группами очагов лесных или торфяных пожаров для последующего численно-го моделирования переноса и рассеяния в атмосфере продуктов сгорания.

Автор признателен Р.Е. Ванкевичу за предоставление архива спутниковых наблюдений и А.С. Гаврилову за руководство работой.

Работа выполнена в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» по теме: «Проведение проблемно-ориентированных поисковых исследований в области технологий мониторинга и прогнозирования состояния атмосферы при лесных и торфяных пожарах» (контракт № 16.515.11.5029 от 12 мая 2011 г.).

Литература

1. Гаврилов А.С. Численная модель подъема тепловой струи от лесного пожара с учетом ее взаимодействия с внешним потоком // Уч. зап. РГГМУ, 2012, № 26, с. 130-142.
2. Rui Xu, Donald C. Wunsch, *Clustering*, IEEE Press, 2009.
3. Кормен, Т., Лейзерсон, Ч., Ривест, Р. Алгоритмы: построение и анализ / Пер. с англ. / Под ред. А. Шеня. – М.: МЦНМО, 2000. – 960 с.
4. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
5. Дюран Б., Оделл П. Кластерный анализ. – М.: Статистика, 1977. – 128 с.
6. Dasgupta S., Papadimitriou C. H., Vazirani U. V. *Algorithms*, 2006.