

8. Оценка классификатора. [Электронный ресурс] // Личный блог. – Режим доступа: <http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html>, свободный.

УДК 004.4

Использование алгоритма k-means в обработке изображений

А.Р. Сыдыкпаева¹, С.А. Шаймерден^{1,2}

¹*ВКГУ им. С. Аманжолова, г. Усть-Каменогорск;*

²*АлтГУ, г. Барнаул*

Спутниковые снимки, снятые в различных спектральных диапазонах, содержат очень полезную информацию и хранятся в цифровом виде. Использование космических снимков в оперативном обновлении карт среднего и мелкомасштабного экономически выгодно. На основе трех спектрального канала дистанционного зондирования цветные изображения переносят больше информации, чем наземные или аэрофотоснимки, а стереопары изображений позволяют проводить трехмерный анализ пространственных объектов [1,2].

k-means-самый распространенный вид из методов кластеризации. В связи с тем, что алгоритм имеет простоту и высокую скорость выполнения, он имеет большое значение [3]. Алгоритм k-means выполняется итеративно, который разбивает заданный набор пикселей на точки кластера k, приближающиеся к их центрам, и в результате перемещения места этих центров, выполняется кластеризация.

Алгоритм классирования k-means, последовательный в языке C++. Алгоритм SK means (последовательный k-средний). Обнаружен в ал-

горитме SK means в виде функции $J = \sum_{n=1}^N \sum_{k=1}^k \|x_n - c_k\|^2$, чтобы минимизировать функцию SK means назначения, то есть сделать погрешность

функции квадратом. $J = \sum_{n=1}^N \sum_{k=1}^k \|x_n - c_k\|^2$, N – расстояние между дан-

ными, соответствующими центру J кластера, x_n ($1 \leq n \leq N$) обозначает заданные точки и c_k ($1 \leq k \leq K$) определяет тяжесть кластера

$\|x_n - c_k\|^2$ – x_n и c_k – определение расстояния между ними. В

$\mu_k = \frac{1}{N_k} \sum_{n \in c_k} x_n$, μ_k ($1 \leq k \leq K$) средняя точка сведения, подлежащие к k группе, и указывает количество объектов, подлежащих кластер N_k k [3,4].

В нашем алгоритме *SK means* количество K кластеров является параметром, заданным пользователем. Во-первых, отображаются N объекты из входного файла [5]. В первоначальном случае K центроид выбирает случайно, определяет в виде μ_k ($1 \leq k \leq K$). Во-вторых, алгоритм *SK means* рассчитывает расстояние и определяет точки, относящиеся к кластеру, итеративно к каждому объекту. До последнего значения, указанного пользователем, процесс повторяется.

SK means алгоритм:

Input: количество K кластеров, количество данных N объектов

Output: K центроиды

1) Найти объекты N в файле;

2) Случайно выбирает точку K как первичный центроид, образуется в виде μ_k ($1 \leq k \leq K$);

3) J вычисляет из формулы $J = \sum_{n=1}^N \sum_{k=1}^k \|x_n - c_k\|^2$, который определяет его через J' ;

4) Присваивает каждому объекту x_n ($1 \leq n \leq N$) до ближайшего кластера;

5) Рассчитывает новый кластер из каждого кластера μ_k по следующей формуле: $\mu_k = \frac{1}{N_k} \sum_{n \in c_k} x_n$;

6) Вычислят J из Формулы $J = \sum_{n=1}^N \sum_{k=1}^k \|x_n - c_k\|^2$;

7) Пересчитывает шаги 3-6, $J' - J$ до $<$ порог ($J' - J < \text{threshold}$) ;

8) Выводит значение Классификации: K центроиды [6,7].

Результаты алгоритма *k means* на языке C++. Встроенные точки:

[-42.07, 42.2], [-175.19, 179.3], [163.62, 2.27], [-41.53, -59.16], [-53.16, -90.43], [-163.99, 24.32], [184.87, -192.87], [163.36, -187.6], [-110.19, 55.71], [158.29, 10.98], [-186.48, 14.26], [-85.21, 14.94], [-61.03, -93.47], [166.61, -31.78], [94.31, -75.98], [-197.38, 179.82], [103.08, -73.08], [-115.14, 97.68], [-67.63, 27.99], [-180.65, 190.64], [92.14, -74.73], [-64.65, -

86.24], [110.68, -84.65], [-92.61, 22.92], [190.92, -177.26], [-62.99, -86.89], [-71.47, -67.25], [95.08, -83.19], [-59.51, -85.96], [110.07, -78.02], [178.45, -162.32], [-63.1, -84.14], [-63.19, -98.54], [148.74, -25.85], [-47.03, -86.63], [199.73, -160.16], [194.1, -160.22], [177.97, -58.1], [-74.22, 42.39], [-61.98, -87.73], [194.65, -194.63], [-160.78, 174.02], [-179.91, 184.81], [188.65, -176.59], [-118.73, -35.8], [188.49, -182.72], [-46.24, 19.95], [191.35, -175.82], [-180.02, 198.89], [-181.88, 154.86]

Центр кластеризации:

[129.14, -52.01], [187.46, -177.02], [-74.02, 32.3], [-155.21, 45.42], [-179.4, 180.34], [-64.03, -80.19]

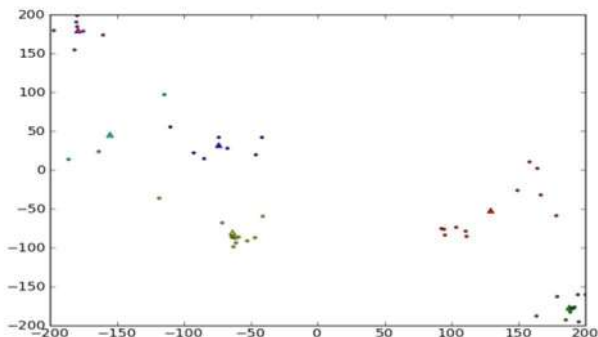


Рисунок 1 – Результат работы алгоритма k-means

Заключение. В настоящее время на гиперспектральных рисунках очень много алгоритмов осуществления распознавания объектов. Один из лидирующих алгоритмов из них – алгоритм кластеризации k-means. В магистерской диссертации был изучен алгоритм кластеризации k-means и рассмотрен последовательный вариант алгоритма на языке программы C++. В дальнейшем буду использовать в реальном распознавании снимков с космоса, с использованием алгоритма кластеризации k-means.

Библиографический список

1. Дж. Ту, Р. Гонсалес Принципы распознавания образов. – С.110–112. Москва, 1978г.
2. Кашкин В.Б., Сухинин А.И. Дистанционное зондирование Земли из космоса. Цифровая обработка изображений: Учебное пособие. – М.: Логос, 264 с., 2001г.
3. В.В. Сергеев Анализ и обработка изображений, получаемых при наблюдениях земли из космоса // Стенограмма научного сообщения на совместном семинаре ИСОИ РАН и Института компьютерных исследований СГАУ 18 апреля 2006 года.

4. R.C. Dubes and A.K. Jain. Algorithms for Clustering Data. Prentice Hall, 1988.

5. M. Armbrust, I. Stoica, M. Zaharia, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, April 2010.

6. W. Zhao, H. Ma, Q. He, "Parallel K-Means Clustering Based on MapReduce," Cloud Computing, vol. 5931, pp. 674-679, 2009.

S. Kantabutra, A. Couch, "Parallel K-means Clustering Algorithm on NOWs," Technical Journal, vol. 1, no. 6, 2000.

УДК 004

Разработка статистического пакета «Корреляционно-регрессионный анализ данных»

А.С. Шевченко, Н.Д. Химочкина

Рубцовский институт (филиал) АлтГУ, г. Рубцовск

Изучение теории корреляционно-регрессионного анализа является неотъемлемой частью процесса обучения любого экономического или технического направления подготовки высшего образования. Для достижения этой цели в рамках образовательного процесса очень важно обеспечить проведение эффективных практических и лабораторных занятий.

Большое внимание этому вопросу уделяется в Рубцовском институте (филиале) Алтайского государственного университета, так как корреляционно-регрессионный анализ является отдельным разделом большинства дисциплин различных образовательных программ: «Эконометрика», «Интеллектуальный анализ данных», «Математические методы в юриспруденции», «Макроэкономическое планирование и прогнозирование», «Статистика», «Правовая статистика».

Таким образом, в связи с повышением требований к качеству современного образования возникает необходимость в разработке статистического пакета для корреляционно-регрессионного анализа данных, который позволит автоматизировать учебный процесс разработки различных заданий для фонда оценочных средств, а также выполнения и проверки заданий в рамках изучаемых дисциплин.

Объектом исследования является организация учебного процесса по дисциплинам, связанным с математическим моделированием.

Предметом исследования является корреляционно-регрессионный анализ данных.

Целью исследования является разработка статистического пакета «Корреляционно-регрессионный анализ данных».