

ГЕОМЕТРИЧЕСКИЕ МЕТОДЫ В ТЕОРИИ НЕЙРОННОЙ АССОЦИАТИВНОЙ ПАМЯТИ: ОПЫТ РАЗРАБОТКИ АЛГОРИТМА КЛАСТЕРИЗАЦИИ

Abstract: This paper is dedicated to the new algorithm for unsupervised learning and clustering. This algorithm is based on Hopfield-type pseudoinverse associative memory. Using methods of Riemannian geometry we establish the procedure of generalized averaging on the space of projective matrices of fixed rank: this space is isomorphic to the Grassmann manifold. This procedure enables us to endow the associative memory with ability of data generalization. In the paper we provide experimental testing for the algorithm using simulated random data and images from the MNIST database (handwritten digits).

Key words: neural networks, associative memory, clustering, optimization, Grassmann manifold.

Анотація: Робота присвячена розробці нового алгоритму для навчання без учителя й кластеризації. Запропонований алгоритм базується на псевдоінверсній нейронній асоціативній пам'яті. Використовуючи методи риманової геометрії, ми будемо процедуру узагальненого усереднення на просторі проєкційних матриць постійного рангу – цей простір ізоморфний багатовиду Грассмана. Така процедура дозволяє наділити неітеративні парадигми нейронної асоціативної пам'яті здатністю до узагальнення даних. У статті наведено експериментальні результати для модельних даних, а так само рукописних цифр із бази даних MNIST.

Ключові слова: нейронні мережі, асоціативна пам'ять, кластеризація, оптимізація, многовид Грассмана.

Аннотация: Работа посвящена разработке нового алгоритма для обучения без учителя и кластеризации. Предлагаемый алгоритм основан на псевдоинверсной нейронной ассоциативной памяти. Используя методы римановой геометрии, мы строим процедуру обобщенного усреднения на пространстве проекционных матриц постоянного ранга – это пространство изоморфно многообразию Грассмана. Данная процедура позволяет наделять неитеративные парадигмы нейронной ассоциативной памяти способностью к обобщению данных. В статье проводятся экспериментальные результаты для модельных данных, а так же рукописных цифр из базы данных MNIST.

Ключевые слова: нейронные сети, ассоциативная память, кластеризация, оптимизация, многообразие Грассмана.

1. Введение

Данная работа посвящена использованию методов римановой геометрии в теории ассоциативной памяти (АП). Оказывается, что римановы многообразия, возникающие в линейной алгебре (такие, как многообразия Штифеля и Грассмана), являются удобными объектами для представления синаптических матриц этих нейросетей. Использование многообразия Грассмана дает возможность построить на базе ассоциативной памяти алгоритм обучения без учителя и кластеризации.

Наш алгоритм основан на псевдоинверсной ассоциативной памяти [1]. Эта АП, как и другие сети типа Хопфилда, способны обучаться, в некотором смысле, «без учителя». Запоминаемые данные не снабжаются метками классов. Однако такая нейросеть не может использоваться для кластеризации, поскольку не способна обобщать данные: все обучающие образы запоминаются «как есть». Это значит, что такая система не сможет выделить центры кластеров в массиве данных.

Эта проблема частично решена в [2] и [3]. Авторы данных работ предлагают алгоритм адаптивной фильтрации. Этот алгоритм обладает некоторой способностью обобщать данные, однако матрица сети здесь не является проективной, что приводит к искажению и порче сети по мере накопления данных. После некоторого количества обучающих данных такая система перестает функционировать как ассоциативная память.

В отличие от [3, 4], по нашему методу всегда производятся проекционные матрицы. С помощью техники обобщенного усреднения на римановых многообразиях строится синаптическая матрица сети. АП с такой матрицей содержит образы, обобщающие обучающие данные. То есть эти векторы могут использоваться как центры кластеров, а их области притяжения выступают в роли самих кластеров.

Поскольку наш алгоритм основан на неитеративной нейропарадигме, он имеет преимущество по скорости. Действительно, даже для больших массивов данных требуется малое количество эпох. Эта особенность делает алгоритм ассоциативной кластеризации привлекательным по сравнению с другими нейропарадигмами, используемыми для этих целей, например, с самоорганизующимися картами Кохонена [4]. Действительно, обучение карт Кохонена – процесс медленный и зачастую требуются миллионы эпох.

Предлагаемые алгоритмы кластеризации на базе ассоциативной памяти были исследованы экспериментально. Проведены эксперименты с использованием модельных данных, а также рукописных цифр из базы данных MNIST [5].

2. Необходимые сведения

Псевдоинверсная ассоциативная память

Наш алгоритм основан на автоассоциативной памяти с псевдоинверсным правилом обучения [1]. В сетях этого типа запоминаются биполярные векторы: $v_k \in \{-1, 1\}^n$, $k = 1 \dots m$. Пусть эти векторы образуют столбцы матрицы \mathbf{V} размером $m \times n$. Синаптическая матрица \mathbf{C} дается соотношением:

$$\mathbf{C} = \mathbf{V}\mathbf{V}^+, \quad (1)$$

где \mathbf{V}^+ – матрица, псевдообратная к \mathbf{V} по Муру-Пенроузу. Ее можно вычислить напрямую по формуле $\mathbf{V}^+ = (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T$ или по формулам Гревилля [7].

Ассоциативный поиск осуществляется с помощью процедуры экзамена: входной вектор x_0 служит начальной точкой итераций вида

$$\mathbf{x}_{t+1} = f(\mathbf{C}\mathbf{x}_t), \quad (2)$$

где f – монотонная нечетная функция, такая, что $\lim_{s \rightarrow \pm\infty} f(s) = \pm 1$. К векторному аргументу она применяется покомпонентно. Устойчивую неподвижную точку этого отображения будем называть аттрактором. Расстояние по Хэммингу между входным вектором x_0 и запомненным образом v_k такое, что процедура экзамена все еще сходится к v_k и называется *аттракторным радиусом*.

Многообразие Грассмана

Существует известная математическая модель, позволяющая описать множество всех линейных подпространств размерности m в R^n и наделить его метрической структурой. Это риманово многообразие Грассмана (МГ). Риманова структура этого многообразия позволяет определить

касательные пространства, метрику, геодезические расстояния. В данной работе рассматриваются только вещественные многообразия Грассмана.

Определение 1. Множество всех матриц Y размера $n \times m$, таких, что $Y^T Y = I$ с римановой метрикой, индуцированной стандартным вложением в $M_n(\mathbb{R})$, называется *многообразием Штифеля* и обозначается $V_{n,m}$.

Определение 2. Многообразие всех классов эквивалентности матриц Y из многообразия Штифеля по отношению к $Y \cong YU$, где U – произвольная ортогональная матрица размером $m \times m$, называется *многообразием Грассмана* и обозначается $G_{n,m}$.

Это многообразие является многообразием всех m -мерных подпространств \mathbb{R}^n . В самом деле, каждый элемент многообразия Штифеля задает ортонормированный базис в таком подпространстве, а эквивалентность в определении 2 обеспечивает независимость от выбора базиса. Таким образом, многообразие Грассмана есть фактор многообразия Штифеля относительно действия ортогональной группы $G_{n,m} = V_{n,m} / O_m(\mathbb{R})$ [0].

Из линейной алгебры известно, что существует взаимно-однозначное соответствие между самосопряженными проекционными операторами ранга m , m -мерными линейными подпространствами и элементами многообразия Грассмана $G_{n,m}$. Это соответствие задает

Лемма 1. Каждому m -мерному подпространству $L_m \subset \mathbb{R}^n$ взаимно-однозначно соответствует симметричный проекционный оператор C ранга m и точка c на многообразии Грассмана $G_{n,m}$.

Доказательство. Выберем в L_m какой-нибудь ортонормированный базис. Пусть векторы этого базиса суть столбцы матрицы Y . Тогда $C = YY^T$ – самосопряженный проекционный оператор, образ которого совпадает с L_m . Обратно, для каждого симметричного проекционного оператора C $L_m = \text{im} C$. Соответствие линейных подпространств и элементов $G_{n,m}$ было показано выше. ■

Существуют различные представления точек многообразия Грассмана. Одно из таких представлений базируется на использовании ортогональных $n \times m$ -матриц. Один элемент МГ есть класс эквивалентности таких матриц (см. выше). Устранить эту неоднозначность можно, избрав представление с помощью проекционных матриц: лемма 2.1 гласит, что каждая такая матрица (взаимно) однозначно определяет элемент МГ. Однако это представление избыточно: используется n^2 (или $n(n-1)/2$ при хранении симметричной матрицы в треугольном виде) вещественных чисел, в то время как размерность МГ составляет всего лишь $m(n-m)$.

Кроме того, представление Грассмана можно использовать для сравнения наборов собственных векторов синаптических матриц, получаемых при различных отклонениях от проекционного алгоритма. Действительно, почти любое искажение исходной матрицы снимает

вырождение: m -кратное собственное значение 1 распадается на m различных СЗ. Получившиеся при этом собственные векторы могут сильно отличаться от запоминаемых в матрице. Чтобы корректно сравнить две такие сети, надо определить расстояние между инвариантными подпространствами в целом. В этом случае метрика на многообразии Грассмана служит хорошей мерой близости.

3. Алгоритм

Постановка задачи

Пусть у нас имеется выборка, содержащая K образов, $x_1 \dots x_k \in R^n$. Ассоциативную память, которая содержит образы, обобщающие данные выборки, будем строить так.

Сначала разобьем их случайным образом на группы по m штук. Число $m < n$ не должно превышать количество кластеров, которые мы хотим получить. Затем построим по каждой группе матрицу проекционной АП C_k , $k = 1 \dots N$. Для того, чтобы объединить все экземпляры ассоциативной памяти в одной матрице, прибегнем к обобщенному усреднению на римановом многообразии.

Техника обобщенного усреднения на многообразии

Рассмотрим конечное множество точек некоторого метрического пространства M с расстоянием

$\rho(x,y): \{x_i\}_{i=1}^N \subset M$. Обобщенным средним данного множества точек будем называть элемент (3)

$$\bar{x} = \min_{x \in M} \sum_{i=1}^N (\rho(x, x_i))^2. \quad (3)$$

Аналогично обобщенной медианой конечного множества точек называется точка

$$x_m = \min_{x \in M} \sum_{i=1}^N \rho(x, x_i). \quad (4)$$

Нетрудно видеть, что если M – евклидово пространство, то обобщенное среднее и медиана совпадают с обычным средним и медианой соответственно.

Как вычислить обобщенное среднее на многообразии Грассмана?

Существует несколько представлений элементов многообразия Грассмана $G_{n,m}$. Остановимся на представлении в виде проекционных матриц ранга m – такое представление наиболее естественно для задач ассоциативной памяти. В качестве расстояния выберем норму Фробениуса разности матриц. Тогда задача об обобщенном среднем превращается в следующую задачу минимизации:

$$\min \varphi(X) = \sum_{k=1}^N \|X - C_k\|^2;$$

$$X^2 = X; \text{ rank } X = m. \quad (5)$$

Преобразуем целевую функцию следующим образом:

$$\begin{aligned}
\varphi(\mathbf{X}) &= \sum_{k=1}^N \sum_{i,j=1}^n (x_{ij} - c_{k,ij})^2 = \sum_{k=1}^N \sum_{i,j=1}^n (x_{ij}^2 - 2x_{ij}c_{k,ij} + c_{k,ij}^2) = \\
&= \sum_{i,j=1}^N \left(Nx_{ij}^2 - 2x_{ij} \sum_{k=1}^N c_{ij,k} + \sum_{k=1}^N c_{ij,k}^2 \right) = \sum_{i,j=1}^n N \left(x_{ij} - \frac{1}{N} \sum_{k=1}^N c_{k,ij} \right)^2 + \text{const} = \\
&= N \left\| X - \frac{1}{N} \sum_{k=1}^N C_k \right\|^2 + \text{const} = N \left\| X - \bar{C} \right\|^2 + \text{const} .
\end{aligned}$$

Таким образом, задача (5) об обобщенном среднем свелась к отысканию проекционной матрицы ранга m , ближайшей к арифметическому среднему набору матриц C_k .

4. Экспериментальные результаты

Модельные данные

Данная серия экспериментов проводилась с целью проверки способности нейросети работать с данными, имеющими заранее известную кластерную структуру. Обучающая выборка может быть разбита на части, сгруппированные вокруг известных центров.

Мы использовали случайно сгенерированные биполярные данные. Сначала было образовано $p \times n$ -мерных векторов, будущих центров кластеров. Векторы данных как таковые были построены путем добавления к центрам биполярного шума интенсивности h . Это означает, что для получения вектора данных мы случайным образом выбирали h компонент одного из центров и изменяли их знак.

Затем были построены N проекционных матриц ассоциативной памяти, каждая содержала m обучающих векторов. Общий размер обучающей выборки равнялся, следовательно, Nm . Эти матрицы были осреднены с помощью вышеописанного алгоритма, и получена результирующая матрица X .

Чтобы измерить аттракторный радиус такой сети, мы провели процедуру экзамена: в качестве входных векторов использовались центры кластеров, искаженные биполярным шумом интенсивности H . Аттракторный радиус H_{attr} полагался равным максимальному значению H , при котором итерационный процесс все еще сходил к соответствующему центру.

Для всех экспериментов использовались следующие значения параметров: $N = 1000$, $n = 256$; m и p варьировались. Рис. 1 соответствует ситуации, когда объем памяти равнялся числу кластеров, $m = p$ при различных m .

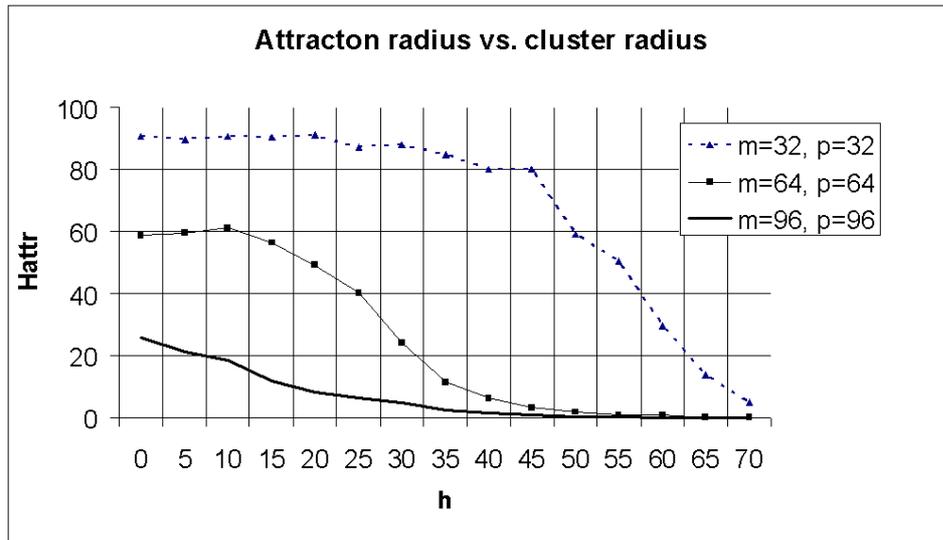


Рис. 1. Аттракторный радиус в зависимости от размера кластера, $m = p$

На рис. 2 показаны результаты для $p < m$ при постоянном $m = 64$. Везде значения H_{attr} при $h = 0$ соответствуют аттракторному радиусу обычной проекционной АП, запомнившей центры кластеров.

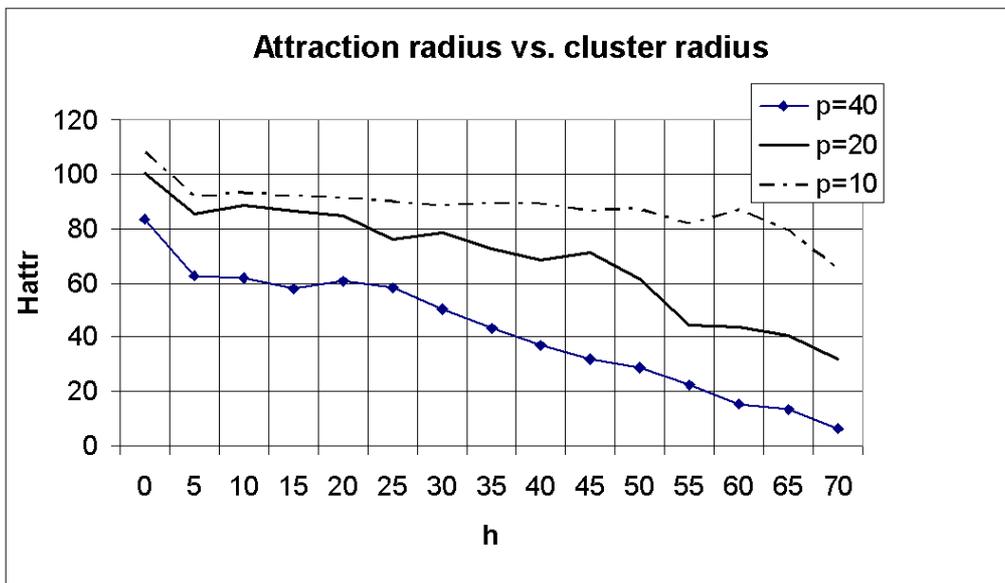


Рис. 2. Аттракторный радиус в зависимости от размера кластеров при $m = 64$ и $p < m$

Видно, что аттракторные свойства сети улучшаются при малом числе кластеров, даже если при этом емкость памяти достаточно велика. Конвергенция происходит примерно до тех пор, пока диаметр кластеров (он равен $2h$) не становится близок к АР проекционной сети, запоминающей их центры.

Рукописные цифры из базы данных MNIST

Для проверки работоспособности наших алгоритмов на реальных данных мы провели эксперименты с использованием образов из известной базы данных MNIST [5]. Она содержит 10 классов изображений (соответствующих цифрам от 0 до 9) в серой шкале с разрешением 28x28 пикселей. Образы снабжены метками классов. Имеются обучающая и тестовая выборки по 10000 и 60000 образов соответственно. Число образов в классе составляет от 5843 до 6472 в обучающей выборке и около 1000 – в тестовой.

Каждый образ представлялся в виде n -мерного вектора $n = 28 \cdot 28 = 784$. Векторы приводились к биполярному виду с помощью порогового фильтра.

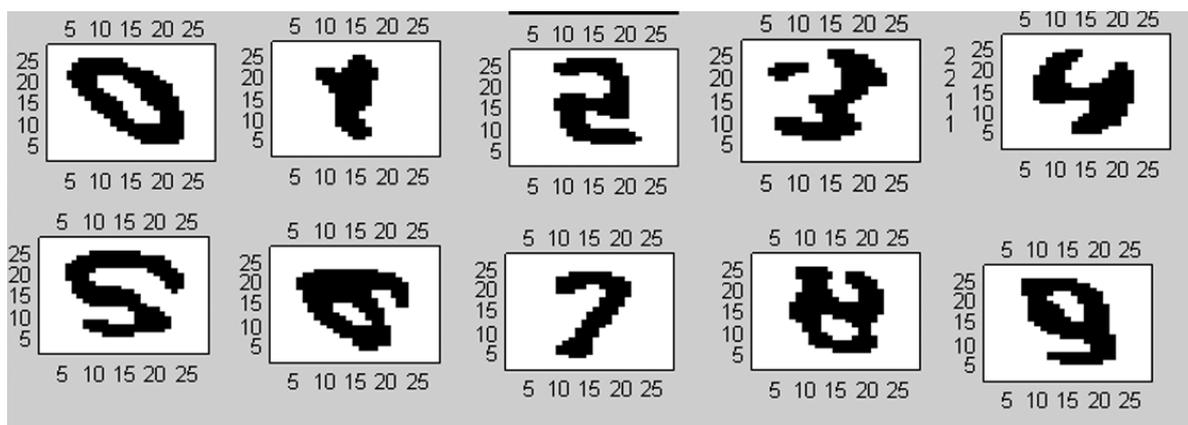


Рис. 3. Некоторые аттракторы АП, полученные на основе данных MNIST

В этих экспериментах синаптическая матрица сети получена обобщенным усреднением 3750 матриц C_k . Каждая из них – это матрица псевдоинверсной АП, содержащая 16 случайно выбранных образов. Такая сеть имеет множество аттракторов, по несколько аттракторов на каждый класс. Поэтому трудно оценить качество распознавания такой сети. Примеры аттракторов, выглядящих как цифры, изображены на рис. 3. Эти изображения – результат обобщения, они показывают, что сеть «знает» о данных. Кроме аттракторов такого рода есть и некоторое количество паразитных. На вид – это бессмысленные фигуры.

Качество распознавания в данном эксперименте можно оценить следующим образом. Сначала процедура экзамена выполняется на всех образах из обучающей выборки. При этом выявляются аттракторы. Каждый аттрактор приписывается тому классу, из которого наибольшее число образов привели к данному аттрактору. Затем при тестировании образ относят к тому классу, к которому «приписан» соответствующий аттрактор. При таком методе тестирования количество правильно распознанных образов составило около 92 %.

5. Выводы

Проведенные эксперименты показывают, что можно построить нейронные сети ассоциативной памяти, способные обобщать данные. Это открывает возможность создания систем кластеризации

на их основе. Неитеративная природа ассоциативной памяти делает ее привлекательной по сравнению с общепринятыми нейропарадигмами, предназначенными для обучения без учителя.

К сожалению, текущий вариант алгоритма ассоциативной кластеризации имеет очень мало параметров, поддающихся регулировке. Например, мы не можем явно задать желаемое число кластеров: оно лишь ограничено сверху объемом памяти m . Этот недостаток можно преодолеть, расширив класс многообразий, используемых в основном алгоритме.

Важно также, что в нашем подходе используются методы оптимизации на римановых многообразиях. Геометрические методы уже доказали свою эффективность в задачах вычислительной линейной алгебры, адаптивной фильтрации и др. (например, [8]). Есть основания полагать, что их применение перспективно и в области искусственных нейронных сетей, распознавания образов. В данной работе используются многообразия одного типа (Грассмана). Обобщение предлагаемого метода на произвольные римановы многообразия с использованием соответствующих алгоритмов (например, описанных в [9]) позволит решить более широкий класс задач.

СПИСОК ЛИТЕРАТУРЫ

1. Personnaz L., Guyon I., Dreyfus G. Collective computational properties of neural networks: New learning mechanisms // *Phys. Rev. A.* – 1986. – Vol. 34 (5). – P. 4217–4228.
2. Reznik A.M. Non-Iterative Learning for Neural Networks // *Proc. of the International Joint Conference on Neural Networks.* – Washington DC. – 1999. – July 10 – 16.
3. Сичов О.С. Дослідження методів удосконалення нейронної асоціативної пам'яті та її застосування в гібридних модульних нейромережах: Дис. канд. техн. наук: 01.05.03. – Київ, 2003. – 133 с.
4. Kohonen Teuvo Self-organizing maps. Third edition. Springer Series in Information Sciences, 30. – Springer-Verlag Berlin, 2001.
5. The MNISTdatabase. <http://yann.lecun.com/exdb/mnist/index.html>
6. Edelman A., Arias T. Smith S. The Geometry of Algorithms with Orthogonality Constraints // *Siam J. Matrix Anal. Appl.* – Vol. 20, N 2. – P. 303–353.
7. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание: Пер. с англ. – М.: Наука, 1977.
9. Smith S.T. Geometric Optimization Methods for Adaptive Filtering // Ph. D. Thesis.– Harvard University, Cambridge MA, 1993.
10. Dedieu J.-P., NOVITSKY D. Symplectic Methods for the Approximation of the Exponential Map and the Newton Iteration on Riemannian Submanifolds. Submitted to the *Journal of Complexity.* – 2004.