Tarea- Mapas auto-organizables

Oscar Andrés Rosas Hernández - Redes Neuronales.
Universidad Nacional Autónoma de México, Facultad de Ciencias

I. Mapas de Kohomen en la vida real (artículo)

Veamos una aplicación (artículo) para los SOMs (ej. Kohonen) en el mundo real, que están dando una solución a un problema específico y veamos cual es la idea general para dar una solución al problema planteado.
Este será: **Picking Mutual Funds with Self-Organizing Maps by Guido Deboeck** [1]

I-A. La idea

La idea es usar mapas auto organizados (en su versión original en la que tenemos un vector en una sola dimensión) que nos permitan traducir datos multidimensionales de varios mutual funds (herramientas de inversion de la cual la verdad desconozco una traduccion al español) en simples mapas de dos dimensiones.
Estos mapas proporcionan un mejora significativa sobre la información que tradicionalmente se publica en los fondos de inversión. Crean una mejor base para la selección de una cartera, para la comparación de el desempeño de los mutual funds y para la creación de puntos de referencia.
Para esto usaron datos publicados por Morningstar, utilizaron los mapas auto organizados para demostrar claramente distinciones y patrones entre los mutual funds mejor calificados.
Como entradas usamos el rendimiento de los mutual funds, las mediciones de riesgo de Morningstar, la "tenencia" de los gestores de inversiones, y las diversas relaciones de gastos.
Con este trabajo se puede ver que los SOM simplifican clasificación de fondos; pueden usarse para el apoyo en la toma de decisiones; y para proporcionar clasificaciones que tienen más significado que las listas ordenadas basadas en múltiples criterios pues no necesitan hacer presuposiciones.
Estos mapas se utiliza para buscar patrones entre mutual funds aparentemente similares. La principal razon para usar SOM son:

- Es un método de minería de datos numérico en lugar de simbólico
- Es un método no paramétrico, lo que significa que no hay suposiciones a priori sobre como debe hacerse la distribución de los datos.
- Es un método que puede detectar estructuras o patrones inesperados mediante el aprendizaje sin supervisión.

42 Guido Deboeck

Table 3.1. Sample binary data matrix of selected mutual funds

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	Fund name	Harbor International	T. Rowe Price International	Scudder International	Templeton Foreign	Vanguard International Growth	Fidelity World	Janus World	Pulham Growth	Scudder Global	Templeton Global Small Companies	Templeton Growth	Merrill Lynch Dev Cap	Templeton Developing Markets	Emerging Mkts Telecommunications	Latin Am Equity	Morgan Stanley Africa	Morgan Stanley Emerg Mkts	Templeton Emerging Markets
Attributes	Fund symbol	HBI	TRP	SCI	TFO	VIG	FWO	JWO	PUG	SCG	TGS	TGI	MLD	TDM	EMT	LAC	AFR	MSE	TEM
1 Type IS	foreign	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	world equity	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0
3	emerging market	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
4	closed-end	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
5 Total Return	= > EAFE	1	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
6 - 3 yr annualized	about EAFE	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	< = EAFE	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
8 MorningStar	average 3	0	0	1	0	0	1	0	1	1	0	0	0	0	1	1	0	1	0
9 Category Rating	high 4	0	1	0	1	1	0	0	0	0	1	1	1	0	0	0	0	0	1
10	highest 5	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
11 Risk	low	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0
12	medium	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	1
13	high	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0
14 Management	< = 2 years	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
15 Tenure	3–4 years	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
16	5 years +	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0
17	10 years	1	0	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0	1
18 Expense Ratio	low < 1.0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
19	medium 1–1.25	1	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0
20	high 1.25+	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1

Source: MorningStar Year-end Wrap January 1997

Figura 1. Figura 3.1: [1]

I-B. El dataset

Morningstar es una empresa privada fundada en 1984 para proporcionar a los inversores información útil para tomar decisiones de inversión inteligentes e informadas.
Morningstar publica Morningstar Mutual Funds, una revista analítica exhaustiva sobre mutual funds "Morningstar Investor", este es un compendio mensual de 450 fondos abiertos y 50 fondos cerrados seleccionados de un universo de más de 6000 mutual funds.
Los 500 representan, según Morningstar, los fondos más exitosos en la industria de fondos mutuos en la actualidad.
Como pueden ver se trata de un dataset binario y que tiene 20 caracteristicas.

I-C. Arquitectura; fases de entrenamiento y clasificación

SOM de una matriz de datos de una muestra binaria, se utilizo un tamaño de mapa de 10 por 1, donde la longitud inicial del entrenamiento fue de 1000 epochs.
La tasa de aprendizaje inicial fue de 0.05 con un radio inicial de 5;
La segunda fase del entrenamiento fue de 10,000 epochs, con una tasa de aprendizaje de 0.02 y un radio de 1.
Todos los pesos fueron inicializados con un número aleatorio al inicio y se uso la clásica distancia euleriana.

Figura 2. Figura 3.3: [1]

Table 3.2. Key to mutual fund names and symbols

Fund name	*Symbol*	*Fund name*	*Symbol*
Harbor International	HBI	Templeton Global Small Companies	TGS
T. Row Price International	TRP	Templeton Growth	TGI
Scudder International	SCI	Merrill Lynch Dev Cap	MLD
Templeton Foreign	TFO	Templeton Developing Markets	TDM
Vanguard International Growth	VIG	Emerging Mkts Telecommunications	EMT
Fidelity World	FWO	Latin Am Equity	LAC
Janus World	JWO	Morgan Stanley Africa	AFR
Putham Growth	PUG	Morgan Stanley Emerg Mkts	MSE
Scudder Global	SCG	Templeton Emerging Markets	TEM

Figura 3. Figura 3.2: [1]

I-D. Resultados

El primer ejemplo en el articulo es un SOM de 1 por 10 neuronas. La figura 3.1 muestra que varios fondos de inversión se agrupan.
Por ejemplo, Fidelity World (FWO), Janus World OWO), Scudder Global (SCG) y Templeton Global Small Companies (TGS) se agrupan en una neurona;

Las telecomunicaciones de mercados emergentes (EMT), la renta variable de América Latina (LAC), Morgan Stanley África (AFR) y los mercados emergentes de Morgan Stanley (MSE) se agrupan en otra neurona.
Hay otras tres neuronas que contienen dos mutual funds.
Dado que los vectores de entrada que son similares se agrupan en las mismas neuronas o en las cercanas, las agrupaciones que se muestran en la Figura 3.1 indican la similitud de los atributos de entrada entre los fondos que están en las mismas neuronas o en las cercanas.
Para verificar esto, buscamos en la Tabla 3.1 el primer grupo listado y lo comparamos con el segundo grupo principal.
El primer grupo de fondos está más diversificado, mientras que el segundo grupo contiene fondos que están más enfocados en sectores específicos o regiones geográficas del mundo. EMT, por ejemplo, es un fondo que solo invierte en empresas de telecomunicaciones; ALC, AFR y MSE son fondos que se especializan en inversiones en América Latina, África o mercados emergentes, respectivamente.
El sombreado gris de las neuronas en este mapa indica que esas neuronas con sombreado similar contienen mutual funds que están más cercanos entre sí, mientras que las neuronas que están sombreadas de manera oscura crean brechas entre los grupos.

II. Ideas para usar mapas de Kohomen en la vida real

II-A. Mi idea, arquitectura y demás

Mi idea era desarrollar un mapa que le permitiera a un robot aprender como estaban relacionados los movimientos de los que era capaz, por ejemplo supongamos que tenemos un robot con 5 o 6 grados de libertad, en ese caso podemos usar un mapa autorganizado para poder visualizar de manera sencilla que patrones existen dentro de los movimientos que realiza el robot, investigando un poco mas me encontre con que ya existe estas ideas, asi que ahora no puedo decir que es mi dia porque me vi influido por otros, en especial por una red desarrollada por Saxon y Mukerjee que aprende el mapa de movimiento de un brazo robotico de dos grados de libertad utilizando un mapa de características autoorganizadas.
Este mapa se puede utilizar para planificar una ruta con el fin de evitar obstáculos. El sistema utiliza un mapa de características en 2D del como las propuestas por Kohonen.
La idea seria tener un mapa que recibe cuatro entradas: dos que representan la posición en el espacio cartesiano del efector final y dos que representan la posición del espacio articular (angulos) del efector final.
El sistema se entrena moviendo el brazo a un punto aleatorio en el espacio articular, y con esto registrando la posición cartesiana del efector final y luego alimentando la información cartesiana y articular a la red.

Figura 4. [4]

Después de muchos miles, y miles de estos movimientos, el sistema debería aprender un mapeo topológico del espacio de entrada.
En esencia, cada neurona representa un punto en el espacio cartesiano y un punto en el espacio articular.
Una vez que se hubiera aprendido el mapa de movimiento, entonces tenemos la base para un medio bastante elegante de planificar trayectorias alrededor de obstáculos como lo describen Saxon y Mukerjee.

La idea básica es que cada neurona en la red 2-D corresponde tanto a un espacio articular como a una posición de espacio cartesiano.
Por lo tanto realizar la planificación de la trayectoria utilizando esta red permite la incorporación de información tanto de espacios articulares como cartesianos, mientras que solo requiere que se busque una matriz 2D de posibles caminos. (lo cual puede potencialmente mejorar en gran medida la complejidad algoritmia de tiempo en la busqueda).
Se pueden incorporar obstáculos desactivando las entradas del espacio articular a la red, aplicando la posición del espacio cartesiano de los objetos a la red y luego desactivando las neuronas que produjeron respuestas máximas.
Luego, se puede determinar una posible trayectoria que evite estos obstáculos eligiendo un camino a través de la matriz neural 2-D que comienza en la neurona que representa más parecido a la posición actual del espacio articular del brazo, evitando las neuronas que se han desactivado y terminando en el punto que representa la posicion más cercana a la posición deseada del efector final (en espacio articular o cartesiano).
Dadas las caracteristicas de este problema podriamos sin ningun problema empezar en pesos aleatorios y usar la clasica distancia euleriana pues justo estamos hablando de problemas en lo que estamos mapeando al mundo real, por lo que resulta una forma de medir distancias bastante prometedora a priori.

II-B. ¿Porque es diferente a las demas redes?

Esta red es claramente diferente a todas las demas que hemos visto a lo largo del curso primeramente porque se trata de aprendizaje no supervisado, por lo que a finalidad es tomar un monton de datos y usar estos mapa para poder extraer caracteristicas de la informacion.
En otras palabras: Los Mapas autorganizados de Kohonen son un tipo de red neuronal utilizado en problemas de reconocimiento de patrones.
Es especialmente capaz de agrupar y visualizar datos complejos de altas dimensiones y puede aplicarse potencialmente para resolver muchos problemas complejos del mundo real.
Si bien podemos argumentar que por ejemplo las redes convolucionales tambien son redes en las que ellas mismas aprenden los patrones necesarios para reconocer clases, esto es diferente y es justo en lo que acabo de decir, pues en el caso de las convolucionales ya tenemos unas clases dadas mientras que con estas podemos generar clases automaticamente.
Ahora el problema que acabamos de ver seria bastante dificil que hacer usando una red neuronal como las anteriores, de mismo modo que la idea que acabo de contar, pues por ejemplo para el brazo robotico tendriamos que darle a una red neuronal la relación (y encontrar una forma de cuantificar la relación) que existe entre las dimensiones cartesianas y las de articulaciones, pero en que en mi ejemplo justo lo que no tenemos son dichas relaciones y usamos los mapas autorganizados para encontrarlas.

II-C. Desafios

Uno de los principales problemas a los que me podria enfrentar es que la gente me dijera que para sistemas bien definidos, como este caso, a menudo será mucho más rápido calcular una respuesta exacta en lugar de utilizar un mapa autorganizados para obtener una solución aproximada.
Esto es particularmente cierto cuando se emplea un procesador secuencial.
Tambien habria que ver que al ser aprendizaje no supervisado necesitamos grandes cantidades de datos, mismos que quiza podrian contener una gran cantidad de errores (sobretodo con el uso de sensores) y sin conocer la relacion per ser nos seria bastante dificil para nosotros saber si la informacion que le estamos ingresando a la red es 100 % correcta.
Tambien elegir correctamente el espacio de salida y el radio de “vecindad”.

III. ¿Porqué la tasa de aprendizaje baja?

La selección de un factor de ganancia adecuado (learning rate), parece ser un compromiso entre la tasa de aprendizaje y la precisión del aprendizaje.
Si se elige un factor de alta ganancia, el aprendizaje continuará rápidamente, pero la red cambiará constantemente con cada nuevo patrón de entrada y, por lo tanto, es posible que no se clasifique de manera uniforme todo el espacio de entrada.
Por otro lado, si eligimos un factor de baja ganancia, entonces el aprendizaje será muy lento pero eventualmente debería alcanzar un resultado muy preciso. Por lo tanto, el factor de ganancia generalmente se elige para ser inicialmente bastante grande (para asegurar un aprendizaje inicial rápido) y luego se reduce para aumentar la precisión del resultado final.
Kohonen ha demostrado que una red unidimensional de neuronas con una sola entrada formará un mapeo ordenado siempre que alfa sea menor que uno. Sin embargo, no está claro que esta prueba se escalara para los casos en que la entrada o la red fueran de mayor dimensionalidad.
En particular, es poco probable que se aplique para el caso donde la entrada tenía una dimensión mucho más alta que la red.
La decisión sobre qué tamaño de vecindario se debe usar también parece determinarse de una manera un tanto ad hoc.

IV. Ideas biológicas

Parece claro que, al menos que los mamíferos se refieren, el cerebro contiene un registro de la información sensorial previa. Además, dado que el número de entradas posibles supera mucho el número de neuronas en el cerebro parece claro que se produce una forma de compresión (o abstracción).
El mecanismo por el cual el cerebro almacena estas representaciones comprimidas no está bien entendida, sin embargo, se hace las hipótesis que se registran en grandes hojas de neuronas en dos dimensiones y organizadas topologicamente". (me costo mucho entender y traducir este termino de la literatura, originalmente se llaman: "2-D metrically and topologically organized"sheets of neural cells)
En otras palabras el cerebro forma un mapeo entre el espacio de entrada multidimensional y el de dos dimensiones de las neuronas.
Hay unas evidencias para esta idea. Se han encontrado encontrado en la parte del cerebro donde se procesa el sonido en los mamíferos las neuronas responden selectivamente a los sonidos de diferentes frecuencias.
De manera que la ubicación de una neurona parece estar directamente relacionada con la frecuencia de del sonido de entrada.
Se han encontrado otras representaciones topográficas en la zona de procesamiento visual, sensorial y motor del cerebro.
Las redes de Kohonen son un intento, mas bien modesto de modelar el funcionamento del cerebro en esta regiones.

V. Desventajas de este tipo de red

Otra desventaja ademas de toda las que ya hemos dicho antes es que que el comportamiento de estas redes no está bien definido matemáticamente.
Esto significa que la selección de parámetros de red (como factores de ganancia y tamaños de vecindad) debe realizarse de una manera un tanto ad hoc.
En particular existe la preocupación de una falta de garantía de que una red, que tiene permitido adaptarse en el mundo real, siempre mantendrá un mapeo ordenado consistente.

VI. Cosas curiosas

Resulta que investigando todo esto encontre con que hay alguien (Eklavya Sarkar) que como tesis hizo unas visualizaciones de estos mapas, anexo la tesis y varios screenshots del resultado.
Porque resulta que el codigo que uso puede ser visto facilmente montando un servidor basado en sus codigos en github. [6]

University of Liverpool

Bachelor Thesis

Artificial Neural Networks: Kohonen Self-Organising Maps

Author:
Eklavya Sarkar

Supervisors:
Dr. Irina Biktasheva
Dr. Rida Laraki

A thesis submitted in the partial fulfillment of the requirements for the degree of Bachelor of Science

in the

Department of Computer Science

DESCRIPTION

CHALLENGES

IMPLEMENTATION

OUTCOMES

KOHONEN SELF-ORGANISING MAPS

2017-18

Description

This project contains the implementation of a Kohonen Network as an application from scratch without any ML library, in order to demonstrate its usefulness and explain in depth the concepts of Machine Learning, by means of a converging Self-Organising Map. It provides an in depth study of Kohonen's algorithm, and present insights of its properties, by designing and implementing three functional model on different datasets.

Kohonen's Self-Organising Maps is an algorithm used to improve a machine's performance in pattern recognition problems. The algorithm is especially capable of clustering and visualising complex high-dimensional data and can potentially be applied to solve many complex real-world problems.

The aim of this thesis is to provide an in-depth study of Kohonen's algorithm, and present insights of its properties, by implementing a complete and functional model.

As part of this project, an extensive literature review on Kohonen

EklavyaFCB / EMNIST-Kohonen-SOM
Watch 1
Star 3
Fork 1
Code
Issues 0
Pull requests 0
Projects 0
Wiki
Security
Insights
Using Kohonen's algorithm to make a Self Organising Map using the EMNIST database on handwritten alphabets.
24 commits
1 branch
0 releases
1 contributor
Branch: master
New pull request
Create new file
Upload files
Find file
Clone or download
EklavyaFCB Update README.md
Latest commit 6d2e95b 8 days ago
static Final 2 years ago
templates Final 2 years ago
.gitignore updates 2 years ago
AllLetters.png Add files via upload 8 days ago
README.md Update README.md 8 days ago
RGB.py Final 2 years ago
after.png Add files via upload 8 days ago
app.py Final 2 years ago
before.png Add files via upload 8 days ago
iris.py Final 2 years ago
som.py Final 2 years ago
sort.py Final 2 years ago
viewInput.py Final 2 years ago

TareaSOM.pdf
Kohomen.pdf
Kohonen SOM
localhost:5000
Kohonen's
Self-Organising Maps
Playing time: 20 mins • By Eklavya Sarkar

TareaSOM.pdf
Kohonen SOM
localhost:5000/cards1
Home
Kohonen Self-Organising Maps
About
Choose an option:
RGB
Select
Low dimension inputs
High volume data
Iris
Select
Low dimension inputs
Low volume data
Character Recognition
Select
High dimension inputs
High volume data

TareaSOM.pdf
Kohonen SOM
localhost:5000/1_4
Home
Kohonen Self-Organising Maps
About
Each of the three inputs is within the range of 0 to 255, representing the amount of red, blue and green light added together.
R: 128, G:128, B:128
Next

TareaSOM.pdf
Kohonen SOM
localhost:5000/1_4
Home
Kohonen Self-Organising Maps
About
If we make a map of 100 such 3D vectors, with each vector representing a slightly different colour.
10 by 10
R: 128, G:128, B:128
Next

TareaSOM.pdf
Kohonen SOM
localhost:5000/1_5
Home
Kohonen Self-Organising Maps
About
Graphs
At 0 iterations without noise
At n iterations without noise
At 0 iterations with noise
At n iterations with noise

TareaSOM.pdf
Kohonen SOM
localhost:5000/3
Home
Kohonen Self-Organising Maps
About
Graphs
Unsorted training nodes without noise
Sorted training nodes without noise
Unsorted testing nodes without noise
Sorted testing nodes without noise
Unsorted training and testing nodes without noise
Sorted training and testing nodes without noise
Unsorted training nodes with noise
Sorted training nodes with noise
Unsorted testing nodes with noise
Sorted testing nodes with noise
Unsorted training and testing nodes with noise
Sorted training and testing nodes with noise
Next

VII. Memes

Referencias

[1] Deboeck, G. and Kohonen, T. (1998). Visual explorations in finance with self-organising maps. London: Springer. Pueden descargarlo de aquí: (https://www.springer.com/gp/book/9783540762669)
[2] Guido J Deboeck Neural Network World, May 2000 Financial Applications of Self-Organizing Maps Pueden descargarlo de aquí: (https://pdfs.semanticscholar.org/e46d/77829d20ef37762d16baccc5d91eb114e25e.pdf)
[3] Self-Organizing Map: ScienceDirect https://www.sciencedirect.com/topics/engineering/self-organizing-map
[4] Craig Sayers, May 1991, University of Pennsylvania Self Organizing Feature Maps and Their Applications to Robotics
[5] Kohonen Self-Organizing Maps https://towardsdatascience.com/kohonen-self-organizing-maps-a29040d688da
[6] Eklavya Sarkar, University of Liverpool Artificial Neural Networks: Kohonen Self-Organising Maps Link: https://eklavyafcb.github.io/som.html