

Algoritmo de estimación de compatibilidad entre parejas humanas

Quetzali Madera
Instituto Tecnológico de Tijuana
Maestría en ciencias computacionales
Tijuana, Baja California Mexico
quetz1@msn.com

Abstract

En ocasiones es difícil conseguir pareja porque se está muy ocupado y puede no haber tiempo para salir con diferentes personas con la intención de conocerlas. Una persona en promedio puede salir con 7 individuos antes de elegir uno para comenzar una relación. Este algoritmo predice posibles relaciones duraderas entre 2 personas utilizando una red neuronal Feed-forward con el algoritmo de aprendizaje Levenberg-Marquardt y una serie de datos de parejas con una relación mayor a 2 años, de esta forma se reduce el esfuerzo y el tiempo que se invierte en la búsqueda de una pareja ideal.

1. Introducción

Una persona puede salir con 7 individuos antes de elegir a uno para comenzar una relación y esta relación puede no ser necesariamente una relación de largo plazo o duradera. El algoritmo que se explicará en este artículo predice posibles relaciones duraderas entre 2 personas utilizando una red neuronal Feed-forward con el algoritmo de aprendizaje Levenberg-Marquardt y una serie de datos de parejas con una relación mayor a 2 años, de esta forma se reduce el esfuerzo y el tiempo que se invierte en la búsqueda de una pareja ideal.

1.1.. Red neuronal

Las redes neuronales son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas que colaboran entre sí para producir un estímulo de salida. En inteligencia artificial es frecuente referirse a ellas como redes de neuronas o redes neuronales. La estructura de una red neuronal se puede apreciar en la figura 1

Una red neuronal se compone de unidades llamadas neuronas. Cada neurona recibe una serie de entradas a través de interconexiones y emite una salida. Esta salida viene dada por tres funciones:

- Una función de propagación (también conocida como función de excitación), que por lo general consiste en el sumatorio de cada entrada multiplicada por el peso de su interconexión (valor neto). Si el peso es positivo, la conexión se denomina excitatoria; si es negativo, se denomina inhibitoria.
- Una función de activación, que modifica a la anterior. Puede no existir, siendo en este caso la salida la misma función de propagación.
- Una función de transferencia, que se aplica al valor devuelto por la función de activación. Se utiliza para acotar la salida de la neurona y generalmente viene dada por la interpretación que queramos darle a dichas salidas. Algunas de las más utilizadas son la función sigmoidea (para obtener valores en el intervalo $[0,1]$) y la tangente hiperbólica (para obtener valores en el intervalo $[-1,1]$).

Una red neurona es similar al cerebro humano en dos aspectos:

- El conocimiento es adquirido por la red a través de un proceso de aprendizaje.
- Los pesos sinápticos o fuerza con que están interconectadas las neuronas se utilizan para almacenar la información

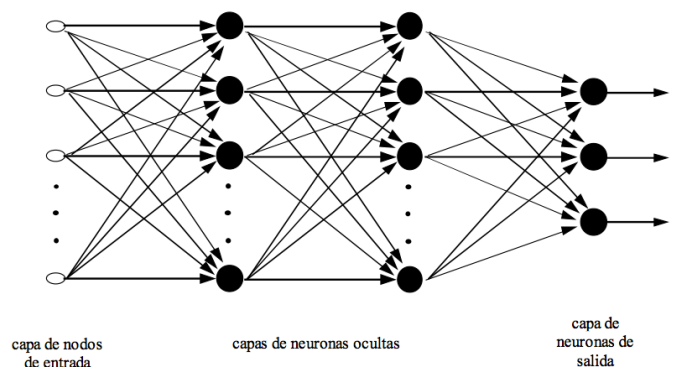


Figura 1. Estructura de una red neuronal

1.2.. Similitud por Coseno

Coseno similitud es una medida de similitud entre dos vectores de un espacio de producto interior que mide el coseno del ángulo entre ellos. El coseno de 0° es 1, y es inferior a 1 para cualquier otro ángulo. Es por lo tanto un juicio de orientación y no magnitud: dos vectores con la misma orientación tienen una similitud coseno de 1, dos vectores a 90° tener una similitud de 0, y dos vectores diametralmente opuestas tienen una similitud de -1, independiente de su magnitud. Coseno similitud se utiliza sobre todo en el espacio positivo, donde el resultado se limita de forma ordenada en $[0,1]$.

Tenga en cuenta que estos límites se aplican para cualquier número de dimensiones, y coseno similitud se utilizan más comúnmente en los espacios positivos de alta dimensión. Por ejemplo, en la recuperación de la información, cada término se asigna teóricamente una dimensión diferente y un documento se caracteriza por un vector, donde el valor de cada dimensión se corresponde con el número de veces que término aparece en el documento. Coseno similitud se da una medida útil de la similitud de ambos documentos es probable que sean en cuanto a su objeto.

La técnica también se utiliza para comparar los documentos en la minería de texto. Además, se utiliza para medir la cohesión dentro de las agrupaciones en el ámbito de la minería de datos. Una de las razones de la popularidad de coseno similitud es que es muy eficiente para evaluar, sobre todo para los vectores de escaso, ya que sólo las dimensiones no-cero necesitan ser consideradas.

Dados dos vectores de atributos, A y B, la similitud del coseno, $\cos(\theta)$, se representa mediante un producto de punto y magnitud como:

$$\text{similitud} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

1.3.. MATLAB

MATLAB (abreviatura de MATrix LABoratory, "laboratorio de matrices") es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M). Está disponible para las plataformas Unix, Windows y Mac OS X. Entre sus prestaciones básicas se hallan: la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes y con otros dispositivos hardware.

El paquete MATLAB dispone de dos herramientas adicionales que expanden sus prestaciones, a saber, Simulink

(plataforma de simulación multidominio) y GUIDE (editor de interfaces de usuario - GUI). Además, se pueden ampliar las capacidades de MATLAB con las cajas de herramientas (toolboxes); y las de Simulink con los paquetes de bloques (blocksets). Es un software muy usado en universidades y centros de investigación y desarrollo. En los últimos años ha aumentado el número de prestaciones, como la de programar directamente procesadores digitales de señal o crear código VHDL.

2.. Objetivo

Desarrollar un algoritmo capaz de sugerir una persona a un usuario para entablar una relación de pareja que pueda ser duradera con los datos aportados por la red social Facebook respecto a este usuario. Una pareja duradera se considera posible si es capaz de permanecer más de 2 años en una relación afectuosa.

3.. Aplicaciones

Este algoritmo puede introducirse en aplicaciones para Facebook para generar posibles relaciones de compatibilidad entre usuarios, estas relaciones pueden ser sentimentales y con algunas modificaciones se puede sugerir otra clase de relaciones como sociedades de negocios o amistades.

4.. Desarrollo

Los datos de la entrada a la red neuronal son un vector de 6 posiciones que representan los valores de cada uno de los datos del usuario (Nombre, edad, fotos, amigos, post, me gusta, comentarios). Dado que el rango de los números que representan los datos es muy grande, los datos se normalizarán antes de introducirlos al vector de la red neuronal. La red neuronal se entrena 20 veces con esos datos utilizando el algoritmo de aprendizaje Levenberg-Marquardt. La red neuronal tiene 6 neuronas de entrada, 2 capas ocultas: una con 10 neuronas y otra con 6 neuronas y 6 neuronas en la capa de salida, la función de activación es tangente sigmoideal en la capa oculta y en la capa de salida una función lineal.

Para comparar el vector de salida con las personas almacenadas y encontrar la similitud entre los datos se utilizó similitud por coseno, una vez obtenida la similitud de los vectores se busca el nombre de la persona a la que corresponda ese vector y de esa forma se puede imprimir en pantalla el nombre de la persona que la red neurona considera ideal para el usuario según los datos proporcionados.

Si se desarrollara una aplicación para Facebook la salida final se podría mantener en forma de vector sin utilizar similitud por coseno. y de esa forma presentar las características de la persona ideal para después comparar con la base de datos de las amistades de ese usuario. La

similitud por coseno nos da la ventaja de no necesitar que los datos de la persona sugerida sean estrictamente iguales a la salida del vector.

En la figura 2 se puede observar la estructura de la red neuronal, y datos de entrenamiento mientras que en la figura 3 se muestra la interfaz que se desarrolló para capturar los datos de entrada a la red neurona y presentar el nombre de la persona seleccionada como posible pareja duradera.

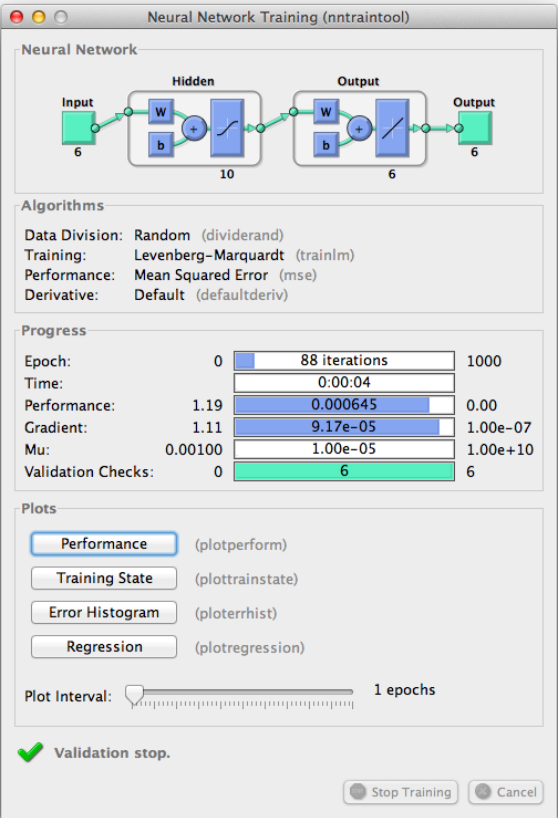


Figura 2. Entrenamiento de la red neuronal

4.1.. Datos

Se proporcionarán datos de la red social Facebook de 30 personas: 15 parejas con relación a largo plazo (más de 2 años).

Los datos proporcionados son los siguientes:

- Nombre.
- Edad.
- Fotos.
- Amigos.
- Cantidad de "Post"(anual).

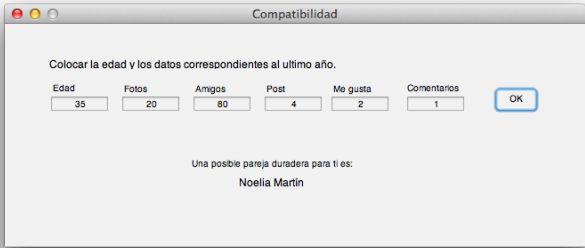


Figura 3. Interfaz de usuario para ingresar los datos

- Cantidad de "Me gusta"(anual).
- Comentarios de los post (anual).

Los datos se normalizaron para que la entrada a la red neuronal estuviera en un rango de 0 a 1: Cada dato se dividió entre el dato mas grande de su conjunto. En la figura 4 se muestran los datos del entrenamiento: de lado izquierdo los datos están sin normalizar y los datos de lado derecho son los datos normalizados.

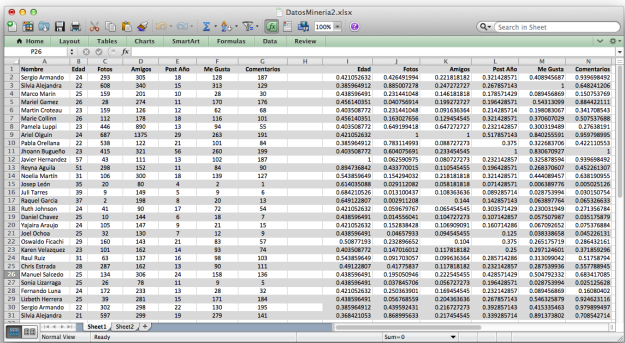


Figura 4. Tabla de datos

5.. Resultados

Se desarrollo el algoritmo Compatibilidad.m para el entorno MATLAB 2012b. En él se ha programado una red neurona con 6 neuronas de entrada y 6 neuronas de salida. También dispone de dos capas ocultas una con 10 neuronas y otra con 6 neuronas. Se a utilizando el algoritmo de aprendizaje Levenberg-Marquardt. En la figura 5 se muestran los resultados de las pruebas con los datos de Facebook.

Para probar el algoritmo con datos que no estuvieran en las pruebas, se buscaban los datos de alguna otra pareja de personas con una relación a largo plazo con la que no se hubiese entrenado antes y estos datos se colocaban en el vector

de entrada por medio de la interfaz gráfica, una vez que sugería a una persona como pareja duradera se revisaban los datos de la sugerencia, se comparaban con los datos de la pareja de la persona la cual se introdujo como vector y en todas nuestras pruebas los datos del vector resultante fueron muy similares a los datos de la pareja introducida, por lo que concluimos que los resultados son satisfactorios.

Compatibilidad

Colocar la edad y los datos correspondientes al último año.

Edad	Fotos	Amigos	Post	Me gusta	Comentarios
23	159	126	12	62	68

OK

Una posible pareja duradera para ti es:
Marie Collinn

Compatibilidad

Colocar la edad y los datos correspondientes al último año.

Edad	Fotos	Amigos	Post	Me gusta	Comentarios
23	446	890	13	94	55

OK

Una posible pareja duradera para ti es:
Ariel Olguin

Compatibilidad

Colocar la edad y los datos correspondientes al último año.

Edad	Fotos	Amigos	Post	Me gusta	Comentarios
24	293	305	18	128	187

OK

Una posible pareja duradera para ti es:
Silvia Alejandra

Compatibilidad

Colocar la edad y los datos correspondientes al último año.

Edad	Fotos	Amigos	Post	Me gusta	Comentarios
25	159	201	10	28	30

OK

Una posible pareja duradera para ti es:
Mariel Gamez

Figura 5. Resultados de la Compatibilidad

6.. Conclusiones

Los resultados fueron muy satisfactorios y la red neuronal trabaja estupendamente. Las redes neuronales son una excelente opción cuando se trata de problemas de búsqueda de características o patrones entre individuos. En este artículo se desarrolló un proyecto relacionado con la redes sociales y la compatibilidad de individuos, pero este mismo método se puede utilizar en otra clase de problemas semejantes.

El desarrollo fue un poco más complicado de lo que se había predicho porque se ignoraban algunos conceptos como las posiciones del vector de entrada, y el manejo de la similaridad entre vectores. Una vez que se comprendió, el desarrollo fue más fluido.

En general el desarrollo del proyecto fue interesante y provechoso dado que se comprendieron muchos detalles previamente estudiados. En general los resultados fueron muy satisfactorios pero aun hay que hacer algunas modificaciones para optimizar el algoritmo y utilización como una aplicación.

7.. Referencias

- Singhal, Amit (2001). Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.
- P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining , Addison-Wesley (2005), ISBN.
- Perceptrn multicapa, Redes de Neuronas Artificiales, UC3M, RAI 2012.
- Haykin, Simon (1998). Neural Networks: A Comprehensive Foundation (2 edición). Prentice Hall. ISBN 0132733501
- Emiliano Aldabas-Rubira; Introduccion al reconocimiento de patrones mediante redes neuronales. JCEE, 2002.

8.. Agradecimientos

En esta sección quiero agradecer a mi maestro el Dr. Mario Garcia Valdez por todo el conocimiento que me ha dado, a mi novio Amaury Hernández Águila por su aportación explicándome la similitud por coseno y porque es el mejor novio del mundo, y a mis padres porque me brindan techo y comida para que yo avance con mis estudios de maestra.

Un agradecimiento especial a CONACYT porque me ha otorgado una beca y no hay mejor motivación que eso para hacer investigaciones.