Red neuronal artificial para orientación profesional "UDProfession"

Neural Network for Professional Orientation "UDProfession"

Diana Marcela Pulido Morales*

Ana Celmira Gualteros Gualteros**

Jorge Enrique Rodríguez Rodríguez***

Fecha de recepción: 20 de agosto del 2009
Fecha de aceptación: 28 de septiembre del 2009

Resumen

El artículo presenta el desarrollo de un software para orientación profesional, a través de la implementación de redes neuronales artificiales. Inicialmente, se da una introducción sobre el tema de la orientación profesional. En segundo lugar, se describe la forma como el psicólogo lleva a cabo el proceso vocacional, destacando la manera como éste razona. Luego se justifica la selección, el diseño e implementación de redes neuronales artificiales. Posteriormente, se aborda el diseño del software, basado en aspectos, tales como: definición de entradas y salidas para la red neuronal, modelo de la red, el preprocesamiento de los datos, método de entrenamiento, entre otros. Por último, se muestra el análisis de pruebas y resultados obtenidos junto con las conclusiones.

* Tecnóloga en Sistematización de Datos. Integrante del grupo de investigación en Inteligencia Artificial de la Universidad Distrital Francisco José de Caldas. Correo electrónico: dianapulido49@hotmail.com

** Tecnóloga en Sistematización de Datos. Integrante del grupo de investigación en Inteligencia Artificial de la Universidad Distrital Francisco José de Caldas. Correo electrónico: anymaniacs@hotmail.com

*** Magíster en Ingeniería de Sistemas. Especialista en Ingeniería de Software. Especialista en Diseño y Construcción de Soluciones Telemáticas. Ingeniero de Sistemas. Docente de la Universidad Distrital Francisco José de Caldas. Correo electrónico: jrodri@udistrital.edu.co

Palabras clave: orientación profesional, pruebas vocacionales, red neuronal artificial, entrenamiento, preprocesamiento de datos.

Abstract

This paper shows the development of software for Professional Orientation, through the implementation of neural networks. The vocational process allows orientating the young men that they are finishing your secondary to taking the decision more succeeded as for your future professional, realizing a series of vocational tests that measure your preferences and skills.

Key words: professional orientation, vocational tests, artificial neural network, training, data preprocessing.

Introducción

El concepto de orientación profesional, no sólo tiene que ver con los jóvenes bachilleres, que aspiran a ingresar a las instituciones de educación superior, sino también con aquellos estudiantes que tras iniciar un programa académico no se sienten satisfechos con él [7]. El proceso de orientación profesional conlleva al desarrollo de ciertas pruebas vocacionales, en las que se busca que el estudiante se conozca a sí mismo, explore sus aficiones y se relacione con las diversas áreas del conocimiento que le permitan definir un perfil profesional específico. Cuanto más temprana sea la etapa de la vida escolar en que se inicie este proceso, más adecuados serán los resultados.

Para ayudar a los jóvenes a tomar una decisión, los orientadores profesionales utilizan numerosas pruebas psicológicas y han diseñado diversidad de modelos formales y no formales que miden la capacidad de decisión del individuo y evalúan su habilidad en dicho proceso. No obstante, las medidas resultantes son exploratorias y limitadas, y no tienen la última palabra: sólo son instrumentos que ayudan a interpretar la relación entre el estilo de vida y el entorno de vida del estudiante [10].

Para la realización de "UDProfession", se seleccionó, diseñó e implementó una técnica de aprendizaje computacional (redes neuronales artificiales), que constituyen un buen método para resolver este tipo de problemas para aquellas tareas de clasificación, identificación, diagnóstico o predicción en las cuales el balance datos/conocimiento se inclina hacia los datos y en las que, adicionalmente, puede haber la necesidad de aprendizaje en tiempo de ejecución y de cierta tolerancia a fallos. En estos casos, las redes neuronales artificiales se adaptan dinámicamente reajustando constantemente los pesos de sus interconexiones [1].

La utilización de redes neuronales en "UDProfession" permite apoyar el proceso vocacional de un estudiante y facilita al psicólogo la realización –como se demuestra con

el análisis de pruebas y resultados–, calificación e interpretación de las pruebas vocacionales que son aplicadas al estudiante, cada prueba contiene una cantidad considerable de preguntas, esto genera un volumen extenso, aproximadamente 500 preguntas. Es importante señalar que la propiedad más relevante de las redes neuronales artificiales es su capacidad de aprender a partir de un conjunto de patrones de entrenamiento,[1] para finalmente clasificar e interpretar acertadamente cualquier respuesta dada por el estudiante a cada una de las pruebas vocacionales [1].

¿Cómo se lleva a cabo el proceso de orientación profesional?

La orientación profesional se puede entender como un proceso estructurado de apoyo técnico dirigido a una persona o grupo en incertidumbre acerca de su conducta vocacional.[2] A través de éste se busca que el asesorado desarrolle un proceso de toma de decisiones funcionales con respecto a su conducta y motivación vocacionales [2].

El proceso se realiza a través de 8 pruebas vocacionales.

IAIM (Prueba de Inteligencias Múltiples)

La teoría de las inteligencias múltiples es un modelo propuesto por Howard Gardner en el que la inteligencia no es vista como algo unitario, que agrupa diferentes capacidades específicas con distinto nivel de generalidad, sino como un conjunto de inteligencias múltiples, distintas e independientes. Gardner define la inteligencia como "la capacidad de resolver problemas o elaborar productos que sean valiosos en una o más culturas". Howard Gardner añade que igual que hay muchos tipos de problemas que resolver, también hay muchos tipos de inteligencia. Howard Gardner y su equipo de la Universidad de Harvard han identificado ocho tipos distintos:

- *Inteligencia lingüística*, la que tienen los escritores, los poetas, los buenos redactores. Utiliza ambos hemisferios.
- *Inteligencia lógica-matemática*, la que se utiliza para resolver problemas de lógica y matemáticas. Es la inteligencia que tienen los científicos. Se corresponde con el modo de pensamiento del hemisferio lógico y con lo que la cultura occidental ha considerado siempre como la única inteligencia.
- *Inteligencia espacial,* consiste en formar un modelo mental del mundo en tres dimensiones; es la inteligencia que tienen los marineros, los ingenieros, los cirujanos, los escultores, los arquitectos o los decoradores.
- *Inteligencia musical*, es aquella que permite desenvolverse adecuadamente a cantantes, compositores, músicos y bailarines.
- *Inteligencia corporal-cinestésica*, o la capacidad de utilizar el propio cuerpo para realizar actividades o resolver problemas. Es la inteligencia de los deportistas, los artesanos, los cirujanos y los bailarines.
- *Inteligencia intrapersonal*, es la que permite entenderse a sí mismo. No está asociada a ninguna actividad concreta.
- *Inteligencia interpersonal*, la que permite entender a los demás; se la suele encontrar en los buenos vendedores, políticos, profesores o terapeutas [3].

1 Los patrones de entrenamiento, son una muestra de las respuestas dadas a las pruebas y su interpretación.

2 La conducta vocacional se refiere a un conjunto de procesos psicológicos (intereses, aptitudes, valores), que un individuo desarrolla a lo largo de su ciclo vital con el propósito final de lograr su autorrealización.

FIP (Formulario de Intereses Profesionales)

Esta prueba tiene como objetivo indicarle al estudiante el área profesional hacia la que apuntan sus intereses. No evalúa conocimientos y, por tanto, no hay respuestas buenas ni malas; la correcta será aquella que mejor refleje el tipo de actividades que prefiere desarrollar el estudiante, sus aficiones. De ahí que se requiera la mayor sinceridad posible a la hora de responder este cuestionario.

Esta prueba comprende ocho áreas profesionales, agrupadas de la siguiente forma: ciencias económicas, administrativas y del mercadeo, ciencias humanísticas y de investigación de la cultura, ciencias de la salud y del servicio social, ciencias exactas y naturales aplicadas a la investigación de la salud y de la alimentación, administración e investigación de las ciencias agropecuarias, ciencias exactas y naturales aplicadas a la tecnología, áreas artísticas e investigación en ciencias exactas.

DAT (Prueba de Aptitudes Diferenciales)

Ésta es una batería integrada de pruebas de orientación que fueron elaboradas a fin de ofrecer un procedimiento integrado, científico y bien estandarizado para medir las aptitudes de los alumnos de ambos sexos de los ciclos básico y especializado de la enseñanza secundaria, para fines de orientación educacional y vocacional.

La batería de pruebas de aptitudes diferenciales, incluye las siguientes: razonamiento verbal (VR), aptitud numérica (NA), razonamiento abstracto (AR), relaciones espaciales (SR), razonamiento mecánico (MR), y rapidez y precisión perceptivas (CSA).

Estas pruebas constan de un número determinado de preguntas (aproximadamente sesenta), con las cuáles se miden los gustos y las aptitudes que tiene un estudiante hacia una determinada área profesional. Del grado de coherencia y sinceridad con que el estudiante resuelva estas pruebas, depende el éxito del proceso, porque esto le permite al psicólogo calificar y determinar un buen resultado. Cada prueba consta de preguntas enfocadas a un perfil específico, por tanto, se busca que en cada una de ellas, las respuestas estén dirigidas hacia un mismo perfil, esto facilitaría la interpretación de éstas; en el caso contrario, si los resultados obtenidos divergen entre sí de una sola área profesional, el psicólogo debe preparar una entrevista personal, en la que pueda analizar más a profundidad los gustos y las preferencias del estudiante por un ámbito profesional en específico y así lograr orientar al estudiante en un área profesional acorde con sus intereses profesionales y habilidades para desempeñarse en una determinada área.

Una correcta elección profesional debe tener en cuenta la satisfacción y realización profesionales, el éxito y el bienestar económico y personal que de ella puedan derivar. Estas metas sólo se logran mediante un trabajo sistemático y disciplinado, cuyo resultado es el autoconocimiento, un proceso que se debe cultivar durante toda la existencia [7].

El proceso de orientación profesional busca lograr que los estudiantes se conozcan a sí mismos, exploren sus aficiones y destrezas y se relacionen con las diversas áreas del conocimiento, para que, llegado el momento de elegir, lo hagan con conocimiento de causa [7].

¿Por qué redes neuronales artificiales?

El estudio de las redes neuronales fue liderado por psicólogos quienes se dedicaron a desarrollar y evaluar el comportamiento de las neuronas, esto implicaba intentar sistematizar el modelo neuronal humano, bajo el cual las neuronas aprenden a través de conexiones entre sí mismas, cada una de estas conexiones tiene un peso asociado, la red realiza su proceso de aprendizaje, ajustando los pesos de tal forma que se logre dar un alto grado de predicción y exactitud a la clase a la que pertenece cada patrón [2].

Inicialmente, se contempló la posibilidad de emplear sistemas expertos, que se definen como un sistema informático que simula el proceso de aprendizaje [13], de memorización, de razonamiento, de comunicación y de acción de un experto humano en una determinada rama de la ciencia; de esta forma se suministra un consultor que puede sustituirlo o apoyarlo con unas ciertas garantías de éxito [8]. Estos sistemas se basan en reglas, cuya función principal es simular el razonamiento sin ningún tipo de aprendizaje, por tanto, cualquier condición dada que se salga de las reglas implementadas, el sistema experto no arrojaría ninguna respuesta; debido a esto, fue necesario buscar otra técnica capaz de aprender a partir de unos patrones dados como ejemplo, de este modo se decide implementar una red neuronal artificial que tiene la capacidad de aprendizaje para disminuir el margen de error entre la respuesta deseada y la obtenida [11], así responde a la necesidad de garantizar la veracidad y precisión en las respuestas que el software debe ofrecer durante el desarrollo del proceso.

Una característica importante de las redes neuronales artificiales es que son altamente tolerantes al ruido y robustas frente a fallos estructurales: la eliminación o mal funcionamiento de un porcentaje importante de unidades no provoca un "colapso", sino una disminución progresiva en el rendimiento de la red, características que, desde luego, posee el cerebro humano. Sin embargo, la característica más valorada de las redes neuronales artificiales es su capacidad de aprendizaje [5].

Durante el desarrollo del proceso de orientación profesional, el psicólogo debe tener un criterio amplio, preciso y oportuno para aconsejar de forma acertada a un estudiante, ya que de esto depende su futuro profesional. En muchas ocasiones este proceso se ve seriamente afectado por el corto tiempo del que se dispone para dedicar a la interpretación de cada una de las pruebas, asimismo, por la divergencia que puede existir entre cada prueba, cuando lo ideal es que todas las respuestas de las pruebas estuvieran enfocadas a hacia uno o máximo dos perfiles; pero este caso no es común, por esto el análisis que se requiere para cada prueba debe ser minucioso y de ser necesario, el psicólogo debe recurrir a una entrevista personal para lograr enfocar un área profesional específica.

Con la implementación de "UDProfession", los psicólogos, disponen de un software que les permite apoyar su labor de orientar a los estudiantes de bachiller. A continuación, se exponen algunos criterios que justifican el desarrollo "UDProfession":

- Unificar el conocimiento y criterio de varios expertos en psicología, a través de la creación de una base de conocimiento sobre la cual la red neuronal aprenderá.
- Ofrecer a los psicólogos un software que les ayude a llevar su proceso de orientación profesional de una forma más interactiva con los estudiantes de bachiller.
- Las redes neuronales artificiales constan de múltiples unidades elementales de

procesamiento o neuronas interconectadas y son capaces de aprender [4]; este factor es vital para el proceso de orientación vocacional que se implementó, ya que el objetivo fundamental es que la red aprenda y logre por sí misma interpretar las respuestas que un estudiante da a cada prueba y posterior a ello clasificar y arrojar el resultado deseado.

- El tema de la orientación profesional no se constituye una práctica obligatoria en los colegios del sector, en los cuales cada día más jóvenes terminan su bachillerato sin saber qué carrera profesional pueden empezar. Al desarrollar este software, se brinda la posibilidad a todos los estudiantes de realizar pruebas vocacionales y en poco tiempo recibir una orientación oportuna y confiable de la carrera profesional que pueden estudiar.
- Las redes neuronales van aprendiendo a partir de ejemplos específicos, para luego tomar decisiones por sí mismas. En el caso de "UDprofession" se presentan una variedad de posibles respuestas a cada una de las pruebas y esto conlleva a diversas interpretaciones de éstas; por tanto, se seleccionó una parte de estas respuestas para que se tomarán como patrones de entrenamiento, así la red estará en la capacidad de dar una interpretación satisfactoria a cualquier tipo de respuesta.
- La psicología es un área de estudio que se relaciona estrechamente con el comportamiento del cerebro humano, en cuanto a su capacidad de toma de decisiones; es por esto que la implementación de una técnica de aprendizaje computacional apoyaría en gran medida el proceso de orientación que debe desempeñar el psicólogo cuando empieza el proceso vocacional con los estudiantes, dada la similitud para representar le conocimiento y aprender de casos registrados.

Diseño del modelo neuronal

"UDProfession" cuenta con un modelo de red neuronal tipo feedforward[3] compuesto de varias capas de neuronas entre la entrada y la salida. El algoritmo de backpropagation[4] es implementado sobre la red, éste es una regla de aprendizaje que se puede utilizar en modelos de redes con más de dos capas de neuronas. El funcionamiento de la red consiste en un aprendizaje de un conjunto predefinido de pares de entradas-salidas, dados como ejemplo. El aprendizaje involucra un ajuste de los pesos comparando la salida deseada con la respuesta de la red e iterando hasta encontrar un mínimo error. La importancia de este algoritmo consiste en su capacidad de autoadaptar los pesos de las neuronas de las capas intermedias para aprender la relación que existe entre un conjunto de patrones dados como ejemplo y sus salidas correspondientes. Una característica importante que se exige a los sistemas de aprendizaje es la capacidad de generalización que se entiende como la facilidad de dar salidas satisfactorias a las entradas que el sistema no ha visto nunca en su fase de entrenamiento [6].

El algoritmo backpropagation requiere el uso de neuronas cuya función de activación sea continua y, por tanto, diferenciable. Generalmente, la función utilizada será de tipo sigmoidal.[5] El objetivo de este algoritmo consiste

3 Una red neuronal feedforward consiste en un conjunto de neuronas organizadas en capas compuestas por una capa de entrada, una o más capas ocultas y una de salida, donde la salida de la neurona de una capa, suministra información para todas las neuronas de la capa siguiente.

4 El algoritmo se basa en $w_{ji}(t+1) = w_{ji}(t) + [\Delta w_{ji}(t+1)]$, donde wji(t+1) es el nuevo peso, y wji(t) es el peso actual.

5 La función sigmoidal está dada a través de la siguiente ecuación $f(x) = \frac{1}{1+e^{-x}}$

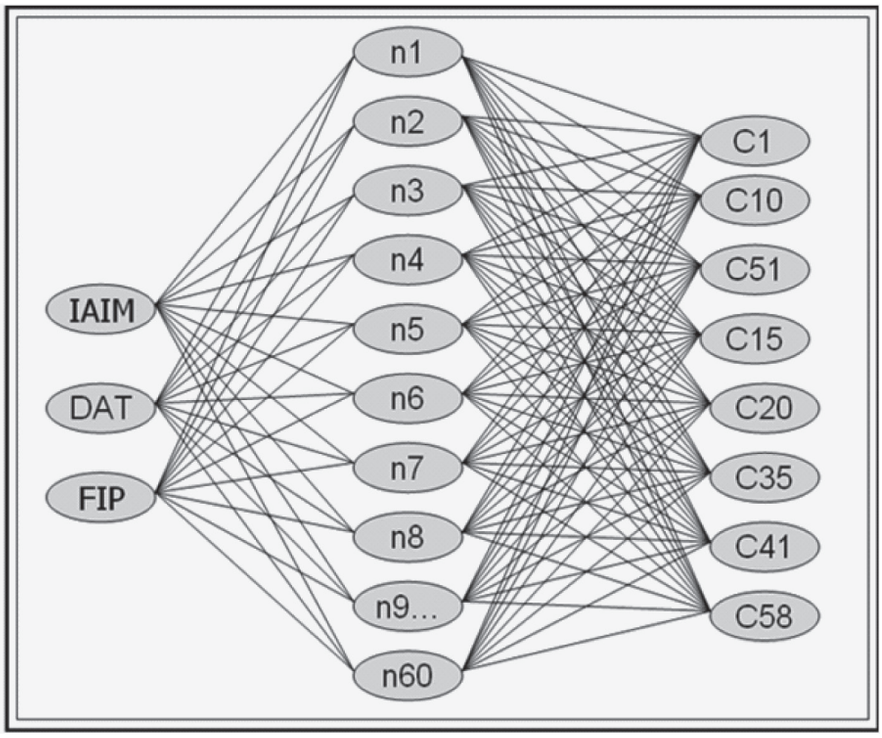
en hallar un conjunto de pesos que minimice el error cometido al clasificar los patrones; para esto, en primer lugar, se realiza una inicialización aleatoria de todos los pesos de la red; a continuación, se entra en la fase iterativa en la que se elige un ejemplo aleatoriamente del conjunto de ejemplos y se suministra a la red; el resultado es el cálculo de las salidas y posterior a esto, se comparan las salidas obtenidas con las salidas deseadas [9].

Un aspecto importante con respecto al aprendizaje en las redes neuronales es conocer cómo se modifican los valores de los pesos; es decir, cuáles son los criterios que se siguen para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información. Estos criterios determinan lo que se conoce como la regla de aprendizaje de la red. De forma general, se suelen considerar dos tipos de reglas: las que responden a lo que se conoce como aprendizaje supervisado y las de aprendizaje no supervisado [10].

"UDProfession" es una red neuronal con aprendizaje supervisado que se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor comprueba la salida de la red y en el caso de que ésta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, a fin de conseguir que la salida obtenida se aproxime a la deseada [10]. Para determinar el área profesional se utilizó una red neuronal, con las siguientes especificaciones: las entradas son los resultados de cada una de las ocho pruebas vocacionales: IAIM (Prueba de Inteligencias Múltiples), FIP (Formulario de Intereses Profesionales, DAT (razonamiento abstracto, razonamiento mecánico, aptitud numérica, rapidez y precisión perceptivas, relaciones espaciales y aptitud verbal). Con los datos de las pruebas anteriores se genera una red con tres capas: la capa de entrada contiene tres neuronas, que corresponden a la respuesta final de cada una de las pruebas (DAT, FIP, IAIM), la capa de salida cuenta con 58 neuronas y en la capa oculta se definió 60 neuronas (a través de la experimentación), véase figura 1.

Las salidas generadas, identificadas como clases: (C1, C2, C3,….C58) corresponden a todas las áreas profesionales: C1 cinestésica corporal, C2 cinestésica corporal e intrapersonal, C3 cinestésica corporal y naturista, C4 espacial, C5 espacial y cinestésica corporal, C6 espacial y intrapersonal, C7 espacial musical, C8 espacial y naturista, C9 indeterminado, C10 interpersonal, C11 interpersonal y cinestésica corporal, C12 interpersonal y intrapersonal, C13 interpersonal y naturalista, C14 intrapersonal, C15 intrapersonal y naturista, C16 lingüística, C17 lingüística y cinestésica corporal, C18 lingüística y espacial, C19 lingüística e interpersonal, C20 lingüística e intrapersonal, C21 lingüística y lógica matemática, C22 lingüística y musical, C23 lingüística y naturista, C24 lógica matemática, C25 lógica matemática y cinestésica corporal, C26 lógica matemática y espacial, C27 lógica matemática e interpersonal, C28 lógica matemática e intrapersonal, C29 lógica matemática y musical, C30 lógica matemática y naturalista, C31 musical, C32 musical y cinestésica corporal, C33 musical e interpersonal, C34 musical e intrapersonal, C35 musical y naturalista, C36 naturalista, C37 ciencias humanística y investigación de cultura, C38 razonamiento mecánico, C39 razonamiento abstracto, C40 relaciones espaciales, C41 aptitud numérica, C42 rapidez y precisión perceptivas, C43 aptitud numérica y razonamiento abstracto, C44 relaciones espaciales y rapidez y precisión perceptivas, C45 aptitud numérica y razonamiento

Figura 1. Red neuronal para el proceso de orientación profesional



mecánico, C46 aptitud numérica y rapidez y precisión perceptivas, C47 razonamiento verbal y rapidez y precisión perceptivas, C48 razonamiento mecánico y rapidez y precisión perceptivas, C49 razonamiento mecánico y razonamiento abstracto, C50 indeterminado, C51 ciencias económicas administrativas y del mercadeo, C52 ciencias humanística y investigación de cultura, C53 ciencias de la salud y servicio social, C54 ciencias exactas aplicadas salud y alimentación, C55 administración e investigación de ciencias agropecuarias, C56 ciencias exactas aplicadas tecnología, C57 áreas artísticas, C58 investigación en ciencias exactas.

"UDProfession" cuenta con dos ambientes de trabajo: un ambiente está desarrollado en java (eclipse), en la cual se implementa el modelo de la red neuronal para su entrenamiento y el segundo ambiente, está desarrollado en PHP (acrónimo de *Hypertext Preprocessor*), en el cual se implementan las pruebas vocacionales, así mismo, se realiza la calificación de éstas, para que dichos resultados sean las entradas a la red neuronal.

Para empezar a configurar la red neuronal, primero se realizó un preprocesamiento de los datos,[6] el cual se encarga de eliminar los datos erróneos, inconsistentes, faltantes (limpieza), etc., para presentarlos de una mejor forma al iniciar el entrenamiento**.** Las bases de datos manejan una cantidad considerable de información, que está llena de datos erróneos, datos faltantes, diversos formatos, etc., lo que se convierte en un problema para rea-

6 El propósito fundamental del preprocesamiento de datos eliminar ruido de los datos.

lizar un análisis adecuado. De este problema se deriva el concepto de "calidad de los datos", y gran parte de la responsabilidad del análisis exhaustivo y el aprendizaje de los datos dependen de este concepto.

Preprocesamiento de los datos

El propósito fundamental del preprocesamiento de datos es manipular y transformar cada conjunto de datos haciendo que la información contenida dentro de ellos sea más accesible y coherente [12]. En el preprocesamiento una parte muy importante dentro del proceso de minería de datos, se estima que se lleva el 60% de los esfuerzos y las tareas fundamentales radican en la limpieza e integración de datos [15].

La recopilación de los datos de diferentes fuentes y aun de una sola implican la ejecución de una limpieza exhaustiva de los datos para un excelente análisis, que en ocasiones se convierte en una tarea bastante tediosa, debido a que se pueden tener muchas inconsistencias en los datos que impide un excelente aprendizaje de éstos. Estas inconsistencias se verán reflejadas a la hora de tomar decisiones.

El software inicialmente se pensaba para desarrollar cuatro pruebas psicológicas que eran: cuestionario de hábitos y técnicas de estudio (CHTE), inventario de autoeficacia para inteligencias múltiples (IAIM), test de aptitudes diferenciales (DAT), formulario de intereses profesionales (FIP). Pero al analizar la primera prueba en mención la cual evalúa algunos aspectos de los hábitos y técnicas de estudio, se consideró que no se realizaría porque su resultado no encaminaba a la salida deseada. Se definieron tres entradas que hacen referencia a las pruebas; para éstas se realizaron el proceso de numerización y la normalización.

Numerización

Este proceso se utilizó porque en las pruebas vocacionales, la mayoría de las respuestas se dan de forma nominal [14]. La numerización es utilizada para la conversión del tipo de datos de un valor categórico a un valor numérico, empleando la técnica denominada: variables Dummy[7].

Normalización

La normalización se refiere al escalado o centrado de los datos, en el que se representan los datos numéricos en una escala predeterminada, en este caso de 0 a 1. La fórmula utilizada es la siguiente, ver ecuación 1:

$$v = \left(\frac{(dato - \text{var_min})}{(\text{var_max} - \text{var_min})} \right) \times (nuevo_\text{max} - nue_\text{min}) + nuevo_\text{min}$$

(Ecuación 1)

var_max : valor máximo
var_min : valor mínimo
nuevo_max : nuevo valor máximo
nuevo_min : nuevo valor mínimo

Aquí se pretende normalizar el rango de un valor numérico para que los atributos estén en función de la misma medida [14].

Análisis de pruebas y resultados

Para el entrenamiento de la red, se recopilaron 4.212 patrones, de éstos el 70% se asignó al entrenamiento de la red y el 30% restante para pruebas. Este porcentaje se selecciona aleatoriamente para que la red pueda tomar la mayor combinación posible. A continuación, se presentan algunas pruebas realizadas a la red,

7 Estas variables toman dos valores, usualmente cero y uno. Los dos valores significan que la observación pertenece a una de dos categorías. Las variables dummy o indicadoras sirven para identificar categorías o clases a las que pertenecen las observaciones..

en las que se modificó la tasa de aprendizaje, el momentum, el umbral y la cantidad de capas ocultas, hasta encontrar la configuración más adecuada y que ofreciera los mejores resultados, en cuanto a efectividad de refiere.

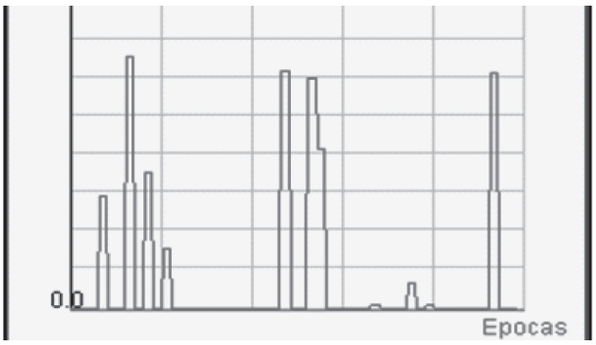
En la tabla 1 se pueden apreciar las pruebas realizadas, variando los datos de entrenamiento; tasa de aprendizaje, momentum, épocas, capas ocultas, porcentaje de entrenamiento y de prueba. Según los datos resultantes, la segunda corrida se considera como el mejor entrenamiento, porque el tiempo es bajo en comparación con otras como la quinta, en la que se obtiene el mismo resultado y el error total es menor. Como se puede apreciar en la tabla 1, la mejor corrida es la número 2, ya que en ésta se puede apreciar un margen de error mucho menor que el de las demás corridas.

En la figura 2 se observa la grafica de error de entrenamiento, que consiste en el error total que se presentó en cada época y así analizar el comportamiento y aprendizaje de la red. Ésta fue la corrida que mostró un mejor desempeño, ya que el tiempo de entrenamiento fue mínimo y la cantidad de épocas fue menor que en los otros entrenamientos, se probó con distintos valores en la tasa de aprendizaje, el momentum, el número de épocas y el número de capas ocultas.

Tabla 1. Muestra de corridas de entrenamiento y pruebas

Datos de Entrenamiento										**Resultados**			
Corrida	Tasa Aprendizaje	Momentum	Épocas	Capas Ocultas	% de Entrenamiento	% de Prueba	Patrones Totales	Patrones Entrenamiento	Patrones Prueba	Error Total	Éxito Entrenamiento	Éxito Prueba	Tiempo
1	0.75	1.0	50	30	70	30	4212	2948	1264	$118x10^{-7}$	93%	96%	29s
2	0.75	1.0	50	60	70	30	4212	2948	1264	$6{,}37x10^{-9}$	93%	95%	39s
3	0.9	1.0	100	60	80	20	4212	3369	843	0.2554	99%	100%	94s
4	0.9	0.05	100	65	70	30	4212	3369	843	$6{,}732x10^{-5}$	94%	96%	82s
5	0.75	1.0	500	60	67	33	4212	2822	1390	$7{,}62x10^{-9}$	99%	100%	646s
6	0.75	1.0	100	160	70	30	4212	3369	843	$1{,}124x10^{-7}$	100%	100%	152s

Figura 2. Grafica de error de entrenamiento



Algunas de las salidas no se clasificaron bien, debido a que las respuestas generadas de los datos no ofrecen una interpretación clara, porque las entradas de los patrones no se relacionan entre sí y esto implicaría la aplicación de otro proceso realizado por el experto.

Para mejorar el entrenamiento, se tomó aleatoriamente los patrones; de esta manera, la red recorrerá una buena cantidad de patrones y no se concentrará tan sólo en unos cuantos, evitando así el sobreentrenamiento. La matriz de confusión generada del entrenamiento de la red neuronal es de tamaño 58*58, que representa las 58 salidas de la red neuronal; a través de ésta se podrá observar la efectividad del entrenamiento y la prueba. La matriz de confusión hace la comparación entre las salidas obtenidas y las salidas deseadas, de este modo clasificará la clase resultante, es decir, las clases que quedaron en la diagonal son las clasificadas correctamente y las que están fuera de la diagonal son las erróneas. En la tabla 2 se puede apreciar un ejemplo de la matriz de confusión.

En la tabla 3, se aprecia una muestra de los patrones que se recopilaron para realizar el entrenamiento de la red neuronal, cada número corresponde al resultado de cada una de las pruebas realizadas por "UDProfession".

Tabla 2. Ejemplo de una matriz de confusión

	C1	C2	C3	C4	C5
C1	1	0	0	0	0
C2	0	1	0	0	0
C3	0	0	1	0	0
C4	0	0	0	1	0
C5	0	0	0	0	1

Tabla 3. Muestra de patrones trabajados

Pruebas / Personas	IAIM	DAT	FIP	SALIDAS
1	1	1	4	1
2	31	27	8	9
3	16	1	3	36
4	36	25	6	27
5	1	2	2	15

En la tabla 4, se puede apreciar la interpretación (área profesional) de los patrones relacionados en la tabla 2.

Conclusiones

- La implementación de un software que permita desarrollar un proceso de orientación profesional es de gran importancia y utilidad en el ámbito académico, ya que le permite al estudiante ir desarrollando cada prueba vocacional a través de un entorno más interactivo y ameno. Con este desarrollo se ofrecen herramientas alternas para apoyar la selección

Tabla 4. Muestra de interpretación

Pruebas / Personas	IAIM	DAT	FIP	INTERPRETACIÓN
1	Cinestésica-corporal	Razonamiento verbal	Ciencias de la salud y servicio social	Ciencias de la salud y servicio social
2	Musical	Rapidez y precisión perceptivas	Áreas artísticas	Áreas artísticas
3	Lingüística	Razonamiento verbal	Ciencias humanísticas y investigación de cultura	Ciencias humanísticas investigación de cultura
4	Naturalista	Razonamiento mecánico	Admón. E investigación de ciencias agropecuarias	Admón. e investigación de ciencias agropecuarias
5	Cinestésica corporal	Razonamiento verbal-aptitud numérica	Ciencias económicas administrativas del mercado	Ciencias económicas administrativas del mercado

de una carrera profesional a un estudiante de bachillerato.

- Se seleccionó dentro del área de inteligencia artificial, una técnica de aprendizaje computacional: redes neuronales, que de acuerdo con sus características presenta un mejor desempeño y confiabilidad de respuesta al proceso de orientación profesional que se implementó. La red neuronal encuentra patrones que permiten orientar a una persona para ingresar a la universidad.
- Se logró sistematizar todas las pruebas de orientación profesional, a través de una interfaz diseñada en PHP, bajo un entorno de selección múltiple con única respuesta.

Trabajos futuros

- "UDProfession" podría integrar más información vocacional y alguna actividad interactiva que impulse el proceso y, asimismo, que pueda ampliar el conocimiento del estudiante acerca de su perfil profesional.
- Implementar más pruebas de orientación profesional, a fin de ampliar la base de conocimiento y permitir que la red pueda aprender con base en más patrones.
- Implementar otras técnicas de aprendizaje computacional, tales como: métodos bayesianos, algoritmos evolutivos, algoritmos basados en aprendizaje incremental, entre otros, con el objeto de com-

parar con el algoritmo backpropagation aquí implementado, en cuanto a efectividad se refiere.

Referencias bibliográficas

[1] W. S Mcculloch y W. Pitts. "A logical calculus of ideas immanent in nervous activity". *Bulletin of Mathematical Biophysics*, 1943.

[2] M. Gallego. Guías de Trabajo para "Conserjería y Orientación Profesional". Corporación Universitaria Iberoamericana. Bogotá: el autor, 2005.

[3] H. Gardner. ***Estructuras de la mente: la teoría de las inteligencias múltiples*****. México:** Fondo de Cultura Económica. **1999.**

[4] L. Álvarez Munárriz. *Fundamentos de Inteligencia Artificial*. España: Universidad de Murcia. 1994.

[5] S. Barro y J. Mira. *Computación neuronal*. España: Universidad de Santiago de Compostela. 1995.

[6] J. Hilera y V. Martínez. *Redes neuronales artificiales "fundamentos, modelos y aplicaciones"*. España: Alfaomega. 2000.

[7] M. A. Mejía. *Escoja bien su carrera*. Bogotá: Intermedio. Cáp.1, 2000.

[8] M. Corredor. *Introducción a la Inteligencia Artificial*. Ed. UIS. 1998.

[9] F. Escolano y M. A. Cazorla. *Inteligencia artificial: modelos, técnicas y áreas de aplicación*. España: Thompson Editores. 2003.

[10] M. L. Rodríguez Moreno. *Orientación e intervención psicopedagógica*. Barcelona: Barcanova. 1995.

[11] M. J. De la Fuente. *Redes Neuronales Artificiales*. España: Universidad de Valladolid. 2002.

[12] D. Pyle. *Data Preparation for Data Mining*. United States: Morgan Kaufmann. 1999.

[13] http://es.wikipedia.org/wiki/Sistema_experto

[14] http://bdigital.eafit.edu.co/bdigital/PROYECTO/P005.74C268U/fulltext.pdf

[15] http://svo.laeff.inta.es/Uploads/svo/svo-datamining/preprocesado.pdf