

Desarrollo de un sistema inteligente para la clasificación de documentos ya digitalizados aplicando redes neuronales supervisadas

Doris Jimenez^a, Henry Paz-Arias^b, Andres Larco-A^b

^a Área de la Energía, Universidad Nacional de Loja, Ciudadela Universitaria Guillermo Falconí Espinosa La Argelia. Loja, Ecuador
twofastloja@hotmail.com

^b Facultad de Ingeniería en Sistemas, Escuela Politécnica Nacional, Ladrón de Guevara E11-25 y Andalucía.
henry.paz@epn.edu.ec, andres.larco@epn.edu.ec

Resumen. El aumento exponencial de la información disponible en formato digital durante los últimos años y las expectativas de crecimiento futuro, hacen necesaria la organización de la información con el fin de mejorar la búsqueda y acceso a la información. Por esta razón adquiere importancia la investigación e implementación de un sistema de clasificación automática de textos que permitan la organización y facilidad de categorizar documentos de acuerdo a su categoría correspondiente utilizando redes neuronales con aprendizaje supervisado, de tal manera que realice un proceso más rápido en un menor tiempo y costo. El criterio para realizar la clasificación de documentos está basada de acuerdo a categorías definidas.

Palabras Clave: categorización de textos, redes neuronales artificiales, Perceptrón multicapa.

1 Introducción

La clasificación automática de documentos ha ganado gran interés en los últimos tiempos, pues el aumento exponencial de la información disponible en formato digital durante los últimos años y las expectativas de crecimiento futuro, hacen necesaria la organización de todo este contenido con el fin de mejorar la búsqueda y acceso a la información, lo que se ha convertido en una difícil tarea la clasificación manual de documentos [1]. Con este fin adquiere importancia el desarrollo de un sistema inteligente para la clasificación automática de textos.

El desarrollo de la ciencia y tecnología viene avanzando aceleradamente la información en cada área de conocimiento se incrementa en forma exponencial, y su tratamiento así como su almacenamiento se hace más compleja. El explosivo crecimiento de la información disponible en documentos digitales en el área de informática y sistemas, ha hecho necesario desarrollar nuevos instrumentos y herramientas que faciliten la realización de procesos de búsquedas de forma eficiente y efectiva así como la administración de estos recursos. Es frecuente que para facilitar la búsqueda de información se proceda a la categorización de los documentos en un conjunto acotado

de clases o categorías. Estas clases permiten representar áreas específicas del conocimiento y son generalmente consolidadas por expertos [2].

El contexto del presente trabajo tiene como finalidad crear un sistema inteligente que permita categorizar automáticamente documentos utilizando sistemas expertos [3] de tal manera que realice un proceso más rápido en un menor tiempo y costo. Para el desarrollo de la investigación se utilizaron datos disponibles del repositorio de la Universidad Autónoma de la ciudad de México (Red de Revistas Científicas de América Latina y el Caribe, España y Portugal) en distintos formatos (Microsoft Word, PDF, texto plano) que han sido utilizados para entrenar y posteriormente evaluar los resultados obtenidos en cada grupo de entrenamiento. Cabe indicar que la categorización de los documentos van a estar categorizados en 14 categorías las cuales son física, matemáticas, ciencias sociales, ciencias naturales, arte, economía, educación, ingeniería, medio ambiente, medicina, jurídica, psicología, lenguaje y diverso.

2. REDES NEURONALES

2.1. Definición

Existen numerosas formas de definir a las redes neuronales; desde las definiciones cortas y genéricas hasta las que intentan explicar más detalladamente qué son las redes neuronales. Por ejemplo:

Una nueva forma de computación, inspirada en modelos biológicos.

Un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles [4].

Un sistema de computación compuesto por un gran número de elementos simples, elementos de procesos muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas [4].

Redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico. [4].

Definiciones según algunos autores:

- Haykin, S.: "Una red neuronal es un procesamiento distribuido masivamente paralelo que tiene una tendencia natural para almacenar conocimiento empírico y hacerlo disponible para el uso [5]. Recuerda al cerebro en dos aspectos:

1. Conocimiento se adquiere por la red a través de un proceso de aprendizaje.
2. Las conexiones interneurónicas se conocen como pesos sinápticos y se usan para almacenar el conocimiento."

- Zurada, J.M.: "Sistemas de redes neuronales artificiales, o redes neuronales son sistemas celulares físicos que puedan adquirir, almacenar y usar conocimiento empírico [5]."

- El concepto de Red Neuronal Artificial está inspirado en las Redes Neuronales Biológicas. Una Red Neuronal Biológica es un dispositivo no lineal altamente paralelo, caracterizado por su robustez y su tolerancia a fallos. Sus principales características son las siguientes:

- Aprendizaje mediante adaptación de sus pesos sinápticos a los cambios en el entorno.
- Manejo de imprecisión, ruido e información probabilística.
- Generalización a partir de ejemplos.

2.2. Elementos de una red neuronal

A continuación se puede ver en la figura 1, un esquema de una red neuronal:

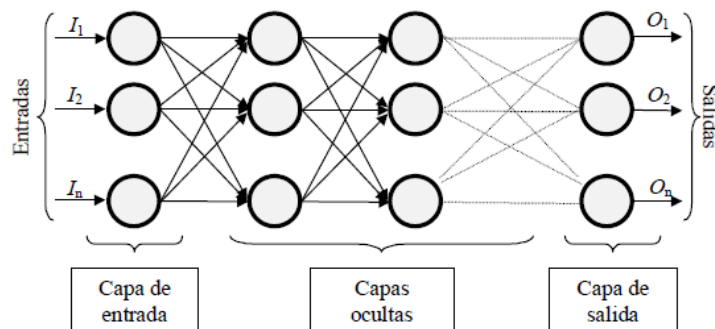


Fig. 1. Ejemplo de una red totalmente conectada

La misma está constituida por neuronas interconectadas y arregladas en tres capas (esto último puede variar). Los datos ingresan por medio de la “capa de entrada”, pasan a través de la “capa oculta” y salen por la “capa de salida”. Cabe mencionar que la capa oculta puede estar constituida por varias capas.

La distribución de neuronas dentro de la red se realiza formando niveles o capas, con un número determinado de dichas neuronas en cada una de ellas. A partir de su situación dentro de la red, se pueden distinguir tres tipos de capas:

- De entrada: es la capa que recibe directamente la información proveniente de las fuentes externas de la red.
- Ocultas: son internas a la red y no tienen contacto directo con el entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales, están encargadas de realizar el trabajo de la red.
- De salidas: transfieren información de la red hacia el exterior.

2.3. Redes Neuronales Multicapa con aprendizaje supervisado.

Las redes multicapa se forman por un conjunto de redes de capa simple en cascada unidas por pesos, donde la salida de una capa es la entrada de la siguiente capa. Generalmente son capaces de aprender funciones que una red de capa simple no puede aprender, por lo que ofrecen mejores capacidades computacionales. Para que este incremento en poder computacional sea tal, tiene que existir una función de activación no lineal entre las capas, por lo que generalmente se utilizará una función de activación sigmoidea en detrimento de la lineal o umbral [6].

El Perceptrón Multicapa es un tipo de Red que está compuesta por varios perceptrones y que permite clasificar adecuadamente más de dos clases. Estas neuronas están organizadas mediante capas, las cuales transmitan la información de capa en capa. La característica de este tipo de Red es que sus conexiones están hechas de atrás hacia adelante y que además las neuronas de la misma capa no se relacionan entre sí.

Las redes multicapas son aquellas que disponen de un conjunto de neuronas agrupadas en varios (2, 3, etc.) niveles o capas.

Para calcular la salida de una red multicapa se debe hacer de la misma manera que en las redes de capa simple, teniendo en cuenta que las salidas de una capa son las entradas de la siguiente capa.

2.3.1. Arquitectura

La arquitectura de este tipo de red se caracteriza porque tiene todas sus neuronas agrupadas en distintos niveles llamados capas. El primer nivel corresponde a la capa de entrada que se encarga únicamente de propagar por el resto de la red las entradas recibidas.

El último nivel es el de la capa de salida. Se encarga de proporcionar los valores de salida de la red. En las capas intermedias denominadas capas ocultas se realiza un procesamiento no lineal de los patrones recibidos.

Las conexiones del Perceptrón multicapa son hacia adelante. Generalmente todas las neuronas de un nivel se conectan con todas las neuronas de la capa inmediatamente

posterior. A veces dependiendo de la red se encuentran conexiones de neuronas que no están en niveles consecutivos, o alguna de las conexiones entre dos neuronas de niveles consecutivos no existe, es decir el peso asociado a dicha conexión es constante e igual a cero. Además, todas las neuronas de la red tienen un valor umbral asociado.

El Perceptrón multicapa es una red formada por una capa de entrada, al menos una capa oculta y una de salida su estructura se muestra en la Figura 2

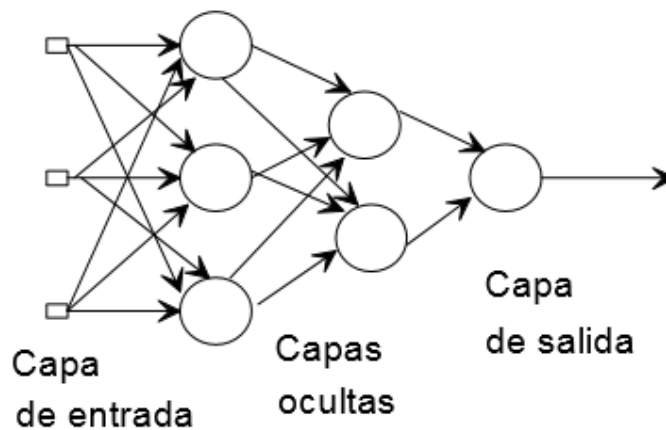


Fig. 2. Estructura de un Perceptrón multicapa

Este modelo se compone de la siguiente manera:

- o Capa de entrada: sólo se encarga de recibir las señales de entrada y propagarla a la siguiente capa.
- o Capa de salida: proporciona al exterior la respuesta de la red para cada patrón de entrada.
- o Capas ocultas: realizan un procesamiento no lineal de los datos de entrada.

3. RESULTADOS

3.1. FASE DE SELECCIÓN O ANÁLISIS

3.1.1. Selección del repositorio de estudio

Para la utilización del conjunto de documentos se ha hecho un estudio de varios repositorios como se muestra en la Tabla 1, luego de haber analizados sus ventajas y desventajas de cada uno de ellos se escogió el repositorio que más se acopló con el tema de investigación, llegando a la conclusión que el repositorio para ser utilizado es el de la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal de la

Universidad autónoma de México [7]. Este repositorio cuenta con una gran cantidad de documentos digitalizados como es 338 816 documentos compuestos por: artículos científicos a texto completos, revistas científicas y fascículos.

Tabla 1. Análisis del corpus de datos.

| Repositorios | Tipos de documentos | Ventajas | Desventajas |
|---|--|---|--|
| Repositorio de Ciencia - Science Library | Cuenta con libros de divulgación científica y artículos | Tiene un enlace de descarga total de documentos digitalizados | Es un repositorio muy pequeño de documentos Metadatos incompletos |
| Biblioteca Virtual Universal | - Cuenta con obras clásicas y estudios científicos. - Trabajos de divulgación científica y de investigación | Contiene un índice medio de documentos a requerir | La mayoría de los metadatos son incompletos |
| Biblioteca Utpl | Tesis doctorales, revistas y libros | Solo trabaja con documentos de su propia biblioteca | La estructura de sus metadatos tienen muy poca información |
| Repositorio Red de Revistas Científicas de América Latina y el Caribe, | Contiene artículos científicos a texto completos, revistas | Existe una gran numero y variedad de documentos digitalizados | Dos de cada 20 documentos no contienen metadatos completos |

| | | | |
|--------------------------|--------------------------|--|--|
| España y Portugal | científicas y fascículos | | |
|--------------------------|--------------------------|--|--|

Como se ha expuesto anteriormente una de las razones para el uso de este repositorio es que sus metadatos tiene una estructura completa como son autor, título, palabras claves, publicación, url entre otros. Cada uno de estos datos permiten dar mayor información sobre el documento, es por ello que al seleccionar el repositorio se tuvo en cuenta los metadatos del corpus ya que posteriormente estos serán de gran ayuda para dicha clasificación.

3.1.2. Identificación de los metadatos a utilizar

Los metadatos son datos estructurados y codificados que describen características de instancias conteniendo informaciones para ayudar a identificar, descubrir, valorar y administrar. En otras palabras los metadatos son datos sobre los datos, es por ello que luego del análisis de los repositorios se extraerán los metadatos que se van a emplear para la identificación de los documentos, los cuales estos datos son los siguientes como se muestra en la tabla 2.

Tabla 2. Metadatos utilizados para un documento

| |
|----------------------|
| Autores |
| Título |
| Descripción |
| Fecha de publicación |
| Palabras Clave |
| Lenguaje |
| Url |

Desarrollo del algoritmo para la lectura de los metadatos.

El siguiente código como se muestra en la Figura 3 el método para la lectura de metadatos el cual permite extraer los metadatos de título, autor, palabras claves, año, url, descripción, de un cierto documento del repositorio Red de Revistas Científicas de América Latina y el Caribe, España y Portugal. Los metadatos como título, palabras claves y descripción servirán como entrada para la red, a partir de la unión de estos datos de entrada se obtiene un texto lematizado [8], y el resultado determinará cuál es la salida o a que categoría pertenece el documento.

```
public boolean cargarLibroWeb(String urls){

    Document doc;
    try {
        URL url = new URL(urls);
        URLConnection conexion = url.openConnection();
        conexion.connect();
        InputStream is = conexion.getInputStream();
        BufferedReader br = new BufferedReader(new
InputStreamReader(is));
        char[] buffer = new char[1000];
        int leido;
        List <String> cont = new ArrayList<String>();
        String aux = null;
        while ((leido = br.read(buffer)) > 0) {
            aux = aux + new String(buffer, 0, leido);
        }
        doc = Jsoup.parse(aux);
        String title = doc.title();
        List<Element> autor = new ArrayList<>();
        try {
            autor = doc.select("meta[name=citation_author]");
        } catch (Exception e) {
        }

        String desc =
doc.select("meta[name=description]").first().attr("content");

        String keybor="";
        try {
            keybor =
doc.select("meta[name=keywords]").first().attr("content");
        } catch (Exception e) {
            keybor =
doc.select("meta[name=citation_keywords]").first().attr("content");
        }

        String urlPag =
doc.select("meta[name=citation_pdf_url]").first().attr("content");
        String autores = "";
        if (!autor.isEmpty()) {
```



```
for (Element element : autor) {
    autores = autores + " - "+element.attr("content");
}
}
String ano = "";
try {
    ano =
doc.select("meta[name=citation_publication_date]").first().attr("content");
} catch (Exception e) {
    ano = "";
}
Libro l=new Libro();
l.setTitulo(title);
l.setDescripcion(desc);
l.setFechaPublicacion(ano);
l.setAutor(autores);
l.setPalabrasClave(keybor);
l.setUrl(urlPag);
}
}
```

Fig. 3. Algoritmo para la lectura de metadatos

3.2. FASE DE DISEÑO

3.2.1. Diseño de la arquitectura de la red

Debido a que se va a construir una red multicapa supervisada de antemano se debe tener en claro cuáles son las salidas de la red, esta red está aplicada a la clasificación de documentos en 14 categorías las cuales son: física, matemáticas, ciencias sociales, ciencias naturales, arte, economía, educación, ingeniería, medio ambiente, medicina, jurídica, psicología, lenguaje y diverso, estas categorías se las menciona luego de hacer un estudio y una revisión exhaustiva de cada uno de los documentos del repositorio .

Para determinar que un documento pertenezca a una categoría se tuvo que asignar a cada una de ellas un vocabulario, es decir cada categoría cuenta con subcategorías que describen de la mejor manera a la categoría. Para ellos usamos el sistema de clasificación denominado Sistema de Clasificación Dewey [9], el cual constituye listas estructuradas de términos (conceptos) que representan de forma univoca el contenido conceptual de los documentos, además es un sistema que cuantifica la relevancia de un término para describir a una categoría.

Para llegar a concluir la categoría se usa la relación que existen categoría y glosario, es decir la entrada de la red será comparada con cada glosario o vocabulario de cada categoría, de esta forma se irán asignado pesos a cada categoría llegando a determinar la categoría a la que corresponde.

```

public String[] neuralNetworkInput(List<String> words){
    String[] entrada = {"0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0"};

    for (String word : words) {
        Glosario gosa =glosarioList.obtenerGlosarioPorPalabra(word);
        if (gosa!=null) {
            List<Categoria> cats = gosa.getCategorias();
            for (Categoria categoria : cats) {
                int indice = Integer.parseInt(""+categoria.getId()-1);
                float aux = Float.parseFloat(entrada[indice]);
                if (aux!=0) {
                    aux = (float) (aux + 0.1 - 0.01);
                }else{
                    aux = (float) (aux + 0.1);
                }
                entrada[indice]= ""+aux;
            }
        }
    }
    return entrada;
}

```

Fig. 4. Algoritmo para asignación de pesos

En la Figura 5 muestra la estructura de la red y sus conexiones, la capa de entradas estarán representadas por los datos de título, palabras claves y descripción y la capa de salida estará constituida por las 14 categorías. El modelo final corresponde a una red con 14 entradas y 14 salidas.

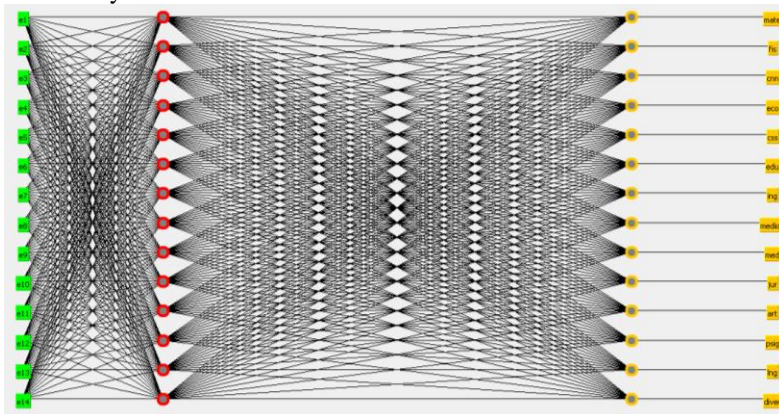


Fig. 5. Resultado de la red en Weka

3.2.2. Aprendizaje de la red

El aprendizaje supervisado se caracteriza por conocer la respuesta que debería tener la red frente a una determinada entrada. De esta manera se compara la salida deseada con la salida de la red y si existen discrepancias se ajusta iterativamente los pesos. Así la etapa de aprendizaje tiene por objeto hacer mínimo el error entre la salida brindada por la red y la salida deseada o verdadera

3.3. FASE DE DESARROLLO

3.3.1. FASE DE ENTRENAMIENTO

Estos resultados están basados en los dos conjuntos de entrenamiento de 1000 y 3000 documentos, a continuación se presentan y se discuten los resultados de cada uno de ellos. Para desarrollar las pruebas correspondientes de entrenamiento de clasificación de los textos según las categorías se empleó Weka (Waikato Environment for Knowledge Analysis), lo cual permite la verificación y constancia de los resultados de la correcta y mala clasificación que se pueda estar dando.

Muestra 1

En la Figura 6 se observa el primer conjunto de datos consta de 1000 documentos, se observa que sus resultados tienen un alto porcentaje de clasificación correcta, lo cual indica que hay una buena clasificación, a continuación se muestran sus resultados en porcentajes de la versatilidad de clasificación y margen de error.

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      1152           97.4619 %
Incorrectly Classified Instances     30            2.5381 %
Kappa statistic                     0.9717
Mean absolute error                  0.011
Root mean squared error              0.0629
Relative absolute error               8.5963 %
Root relative squared error          24.8401 %

```

Fig. 6. Resultado muestra 1

Al analizar la muestra 1 las instancias de clasificación correcta muestran un 97% de documentos bien clasificados y un 3% una mínima parte de margen de error, indicando un mayor rango de clasificación correcta de documentos como se muestra en la Figura 7.

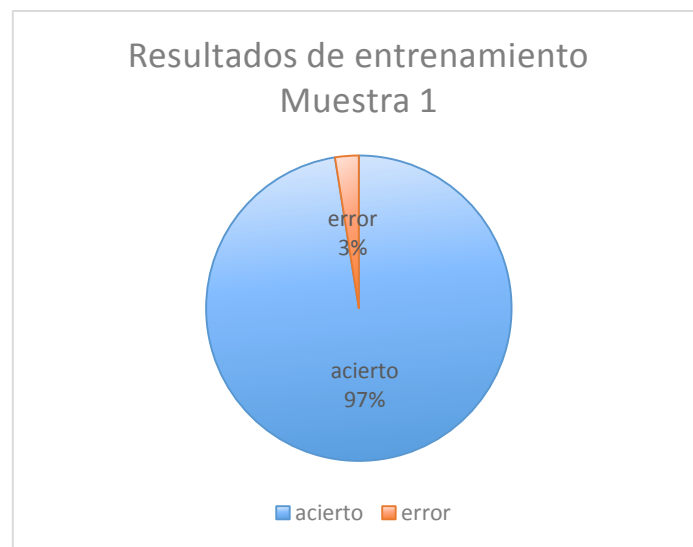


Fig. 7. Resultados de entrenamiento muestra 1

La matriz de confusión muestra en la Figura 8, el tipo de las predicciones correctas e incorrectas sobre el conjunto de documentos. La misma permite comprender en qué sentido se equivoca la red al intentar clasificar el nuevo conjunto de documentos. En el gráfico de esta matriz las predicciones correctas están representadas sobre la diagonal.

```

=== Confusion Matrix ===
      a  b  c  d  e  f  g  h  i  j  k  l  m  n  <-- classified as
164  0  0  0  0  0  0  0  0  0  0  0  0  0 | a = mate
  0 198  0  0  0  0  0  0  0  0  0  0  0  0 | b = fis
  0  0  98  0  0  0  0  0  0  0  0  0  0  0 | c = cnn
  0  0  0 120  0  0  0  0  0  0  0  0  0  1 | d = eco
  0  0  0  1 141  0  0  0  0  0  0  0  0  0 | e = css
  0  1  0  0  0 109  0  0  0  0  0  0  0  0 | f = edu
  0  2  1  0  0  0 13  0  0  0  0  1  0  0 | g = ing
  1  0  1  0  0  0  0 69  0  0  0  0  1  1 | h = medioA
  0  1  1  0  0  0  0  0 24  0  2  0  0  0 | i = med
  0  0  0  0  1  1  0  0  0 51  0  0  0  2 | j = jur
  0  0  1  0  2  0  0  0  0  0 20  0  0  0 | k = art
  2  2  2  0  0  0  0  0  1  0  0 62  0  0 | l = psig
  0  0  0  0  0  0  0  0  0  0  0  0 21  0 | m = lng
  1  0  0  0  0  0  0  0  0  0  0  0  0 62 | n = diver

```

Fig. 8. Matriz de confusión muestra 1

Muestra 2

En la Figura 9 se muestra el segundo conjunto de datos se llevó a cabo con 3000 documentos, al incrementar la cantidad de documentos se creó conveniente agregar nuevas palabras al vocabulario con la finalidad de que el margen de asertividad no decaiga.

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      2223      89.3489 %
Incorrectly Classified Instances    265      10.6511 %
Kappa statistic                    0.8819
Mean absolute error                 0.023
Root mean squared error             0.1109
Relative absolute error             17.8743 %
Root relative squared error         43.7063 %

```

Fig. 9. Resultados muestra 2

En la Figura 10 se presenta el resultado de este nuevo grupo de entrenamiento muestra un valor del 90% de instancias correctas y con un margen de error del 10%, a diferencia del primer grupo se redujo un mínimo porcentaje el margen de asertividad debido a la gran cantidad de documentos a ser analizadas, es decir que a medida que el aumento el número de documentos a ser analizados se deberá incrementar el vocabulario.



Fig. 10. Resultados de entrenamiento muestra 2

La matriz de confusión como se observa en la Figura 11 identifica el tipo de las predicciones correctas e incorrectas sobre el conjunto de datos. La misma permite comprender en qué sentido se equivoca la red al intentar clasificar los nuevos textos.

```
== Confusion Matrix ==  
  
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  <-- classified as  
153 0  4  1  2  4  0  0  0  0  1  0  0  0 | a = mate  
  1 198 2  4  0  3  6  1  0  1  5  0  2  0 | b = fis  
  2  0 121 3  2  5  1  8  0  0  2  1  1  0 | c = cnn  
  1  0  1 113 2  4  0  0  1  2  2  1  0  0 | d = eco  
  3  0  0  2 146 8  1  2  0  3  7  1  1  0 | e = css  
  0  0  1  1  1 357 2  1  1  3  4  0  2  0 | f = edu  
  4  4  1  1  2  9 354 3  1  2  7  1  5  0 | g = ing  
  2  2  3  2  3  4  4 267 3  0  2  0  6  1 | h = medioA  
  6  0  3  1  0  3  2  2 165 0  1  0  0  0 | i = med  
  4  0  0  2  2  4  1  0  0 104 4  1  1  0 | j = jur  
  3  1  0  0  0  2  1  2  0  0 101 0  2  0 | k = art  
  5  1  3  0  0  2  1  1  0  2  1 22  1  0 | l = psig  
  0  0  1  1  0  1  0  0  0  0  0 120 0  0 | m = lng  
  1  0  1  1  0  1  0  0  0  0  1  0  0  2 | n = diver
```

Fig. 11. Matriz de confusión muestra 2

3.4. CONCLUSIONES

La utilización de la herramienta Weka permite la ejecución de pruebas de entrenamiento de Redes Neuronales, con la finalidad de predecir el área de pertenencia de un texto.

Para garantizar una buena clasificación de documentos se creyó necesario en la muestra 2 se incrementa un mayor número de palabras al vocabulario con la finalidad de dar una mejor exactitud en la clasificación de los documentos.

Definir correctamente el vocabulario de cada una de las categorías hace que la clasificación tenga un buen porcentaje de aciertos logrando que el documento sea asignado correctamente a su categoría, de tal medida que el desempeño de la exactitud de la red mejora de acuerdo al tamaño del vocabulario.

La utilización de metadatos fue de gran ayuda para lograr mejores resultados a la hora de la representación, localización y recuperación de recursos electrónicos.

Referencias

- [1] Cortez Vasquez,A y Rojas Lazo,O y Calmet Agnelli,R ., “Categorización de Textos mediante Máquinas de Soporte Vectorial,” Revistas Signos, pp. 1-24, 2011
- [2] Mendoza, M y Ortiz, I y Rojas, V., “Categorización de texto en bases documentales a partir de modelos computacionales liviano,” Revista de investigación de Sistemas e Informática,vol 10,N. 1,pp. 2-12,Enero-Junio 2013.
- [3] Martínez Pérez, P. y Colarte, J. (feb. 2007) Multimedia para discapacitados. Presentada en: Congreso y Feria Internacional Informática 2007 [en línea]. Disponible en: <http://www.informaticabana.cu/eventovirtual/educacion/discapacitados.pdf>
- [4] Matich, D.J., “Redes Neuronales: Conceptos Básicos y Aplicaciones”, Facultad Regional Rosario, Dept. Ingeniería Química, Universidad Tecnológica Nacional, marzo 2001, [En línea] Link http://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monografias/matich-redesneuronales.pdf
- [5] “Redes Neuronales”, [En línea] LINK <http://www.usmp.edu.pe/publicaciones/boletin/fia/info32/pag4.htm>
- [6] Cárdenas Almeida, R., “Inteligencia Artificial”, [En línea] Link http://www2.ulpgc.es/hege/almacen/download/38/38584/practica_ia_2.pdf

[7] Red de Revistas Científicas de América Latina y el Caribe, España y Portugal,2013.

[En línea]. Disponible en:

<http://www.redalyc.org>

[8] “Snowball“; Link: <http://snowball.tartarus.org/>

[9] “Introducción al sistema de clasificación decimal Dewey” [En línea] Link

<http://bibliotecamachala.wikispaces.com/file/view/Clasificacion+decimal+dewey.pdf>