

MODELO DE IDENTIFICACIÓN DE INDICADORES DE GESTIÓN DE RIESGO FINANCIERO MEDIANTE LA REDUCCIÓN DE VARIABLES O RAZONES FINANCIERAS

Ruth Guillén, Alexis Melo

Resumen

Una opción en las técnicas multivariantes es la de ayudar a comprender el dominio de un problema real mediante características resaltantes en ese mismo dominio que consolidan la información dispersa en los fenómenos observados, bien sea individual o conjuntamente. Mediante Análisis de Componentes Principales (ACP) se puede disminuir el efecto de redundancia o dispersión de los fenómenos originales observados a través de nuevas características con poca pérdida de la información aportada originalmente. Esta reducción de variables además de crear nuevas características para el análisis, permite el desarrollo de modelos de pronóstico en los que el trabajo de cómputo se disminuye notablemente. Algunos trabajos han demostrado que produce excelente resultado y más aún, un excelente desempeño de pronóstico cuando estas nuevas características son los patrones de entrenamiento de los modelos de redes neuronales mediante el uso del algoritmo no lineal de retropropagación del error (RN-RP). Adicionalmente, y como aporte particular de esta investigación, se está incorporando como patrones de salida de este modelo híbrido (ACP y RN-RP), los valores obtenidos en una clasificación previa realizada mediante análisis discriminante (AD). El preprocesamiento de los datos se complementa con la redefinición del conjunto de entrada que son las razones financieras transformadas a unas características de origen totalmente nuevo y de variables totalmente independientes, y un conjunto de salida con valores cualitativos. El desempeño de este modelo está siendo evaluado mediante el uso del conjunto de datos construido para tales fines. El logro fundamental de esta iniciativa es la de vincular ambas técnicas

para obtener un mecanismo automático de pronóstico de riesgo y la eventual capacidad de generalizar el comportamiento de un banco en particular en un período dado

Palabras claves: Entrenamiento supervisado, análisis de componentes principales, retropropagación del error, riesgo financiero.

Introducción.

El reconocimiento de patrones en un dominio de problema real, desde un punto de vista estadístico, se refiere fundamentalmente a la selección o extracción de características en la fuente original de datos a través de las variables observadas. Esta selección se rota a un espacio donde el conjunto de los datos que describe todas las características observadas, es transformado a uno nuevo donde se conserva la misma dimensión del espacio original pero provee de una nueva información más vinculada al conjunto original de características descritas por los datos. Sin embargo, de acuerdo al tipo de medición que se haga de las variables observadas, puede variar la utilidad de las técnicas que permiten este tipo de transformación. En los métodos multivariantes, Análisis Discriminante (AD), Análisis de Correspondencia (AC) y Análisis de Componentes Principales (ACP) son las técnicas más comúnmente utilizadas, pero cada una de ellas requiere de un diferente tipo de medición de las características de las variables. Tal es el caso de AD y su fuente de datos lo conforman variables de tipo categórico o discreto, ACP y variables continuas o cuantitativas y AC con una mezcla de ambos tipos de medición. (Sharma, 1996).

En este caso particular, además de aprovechar las bondades de la técnica ACP para preparar nueva información mediante la interacción de las variables originales por una suerte de combinación lineal entre ellas, explicando en su contenido integral nueva información latente, se hace

énfasis en el significado de reducir significativamente las variables originales a esas variables latentes sin sacrificar mayor cantidad de información. En este contexto y en el caso de lo que se investiga, se realizó una buena contribución con un muy excelente resultado de pronóstico de riesgo de quiebra desde el punto de vista estadístico. La interpretación de las variables latentes en el nuevo espacio ortogonal permitió de manera descriptiva identificar individuos (banca comercial) propensos a la quiebra como consecuencia de analizar los resultados de los componentes principales obtenidos como variables latentes. Sin embargo, se corre con el riesgo de no poder explicar el dominio original del problema debido a la pérdida de porciones de la varianza explicada presente en los datos originales, como consecuencia de la presencia de variables originales con un alto grado de interdependencia y que no representan un aporte de información significativo. (Ayesterán, et al., 1996)

En este estudio se aprovecha otra utilidad práctica de la técnica ACP. La reducción de las variables originales a un nuevo conjunto tal que sean las variables explicativas o patrones de entrada, en los modelos de pronósticos, tales como Regresión Lineal Múltiple (RLM) y las redes neuronales. Hay varias experiencias que han demostrado un mejor y eficiente desempeño de los modelos usando estas variables reducidas al disminuir el tiempo de cómputo y el de mantener o mejorar la capacidad de pronóstico.

Atiya hace un buen inventario de resultados evaluativos de la aplicación de este tipo de técnicas, e inclusive algunas no paramétricas, en el análisis del riesgo bancario, fundamentalmente en clasificación y predicción. (Atiya, 2001). Las técnicas multivariantes, en las mayorías de los estudios revisados, fueron aplicadas a grupos de variables asociadas a las razones financieras, y a su vez, comparadas con métodos numéricos emergentes demostrando su utilidad.

ACP, en este estudio, actuará como agente preprocesador de datos, al ser utilizado como método de reducción de variables para conformar el conjunto definitivo de patrones de entrada para el modelo de pronóstico. Algunos estudios realizados de reducción de variables, tales como Kramer, 1992; Tan et al., 1995, Colmenares, 2004, han demostrado que este mecanismo es de muy alta utilidad para la conformación de los patrones de entrenamiento y prueba usados en los modelos de redes neuronales, al reducir en un alto grado el nivel de complejidad de la arquitectura y construcción de estos modelos. Por ejemplo en Tan se puede observar que mediante un algoritmo no lineal donde combina ACP y redes neuronales, los ejemplos originales son reducidos a nuevas características que pueden ser valores de entrada para la preparación de modelos de redes neuronales. En Colmenares, por otro lado, mediante combinaciones de técnicas estadísticas de muestreo y ACP, se pueden lograr conjuntos muy reducidos de datos, altamente representativos del dominio del problema, que resultan ser eficientes ejemplos de entrada para los modelos de redes neuronales. Lo importante a destacar en estos casos, inclusive en Kramer, es el de incorporar la no linealidad inherente a los problemas reales introduciendo el concepto de redes autoasociativas para realizar reducción de variables en el mismo espíritu que lo hace ACP. De hecho, Kramer desarrolla un algoritmo totalmente no paramétrico en el que se reducen las variables originales a un nuevo conjunto, incorporando tácitamente el concepto de curvas principales. Este método es un proceso de transformación a nuevos atributos o componentes principales no lineales, que a su vez son replicados y transformados nuevamente a representaciones muy parecidas a las que los originaron, garantizando así una aceptable sumarización de las características originales en nuevos conjuntos sustancialmente reducidos, pero que logran describir el contenido informativo original.

Por otro lado, los patrones de salida, similar a los que conforman la entrada, son estimados mediante análisis discriminante, indicando en cada patrón, la característica cualitativa de riesgo dependiente del patrón de entrada que lo origina. Este preprocesamiento de los fuente original de datos permite además de construir un modelo de redes neuronales con funciones de base radial (Melo et al.), ser fuente de datos de salida para el conjunto de modelos de red neuronales con funciones logísticas que se aplicará en este estudio. Es decir, un modelo híbrido conformado por ACP y redes neuronales usando el algoritmo de retropropagación del error (RN-RP).

En las siguientes secciones se describen ambas técnicas, y el empleo de ellas como componentes del modelo que se construye para el pronóstico de riesgo bancario.

ACP aplicado al riesgo bancario.

La selección o extracción de características en la serie de datos está vinculada a procesos numéricos donde el espacio original sufre una transformación resaltando un nuevo conjunto *efectivo y reducido* de características que retienen la mayor cantidad de información contenida internamente en los serie de datos originales. Es decir, hay una *reducción de dimensionalidad* del espacio de los datos originales. Esto es básicamente el propósito de esta técnica donde la varianza total juega un papel importante en esta técnica.

Esta metodología de reducción de variables ha mostrado excelentes resultados. El beneficio alcanzado reutilizando fuentes históricas de datos, es el de permitir la disminución del tiempo de cómputo empleado en la construcción y uso de los modelos debido al empleo de las técnicas de preprocesamiento que seleccionan confiablemente la menor cantidad de observaciones y variables posible. Por ejemplo, el método Stratified/PCA (Colmenares, 2004), como técnica de preprocesamiento,

puede reducir una matriz original de observaciones y variables a un subconjunto apreciablemente menor que sus originales con un nivel de confiabilidad y consistencia bastante altos como para poder ser aplicados sobre fuentes históricas de datos. Adicionalmente, estos conjuntos reducidos de datos empleados en la construcción de modelos no lineales usando RNA, han mostrado tener similar capacidad de predicción y/o clasificación que los construidos con el conjunto total de los datos originales. De igual modo, mostraron ser mucho más consistente en sus resultados después de varias réplicas con diferentes conjuntos reducidos de datos. (Dong et al. 1995), (Tan et al. 1995).

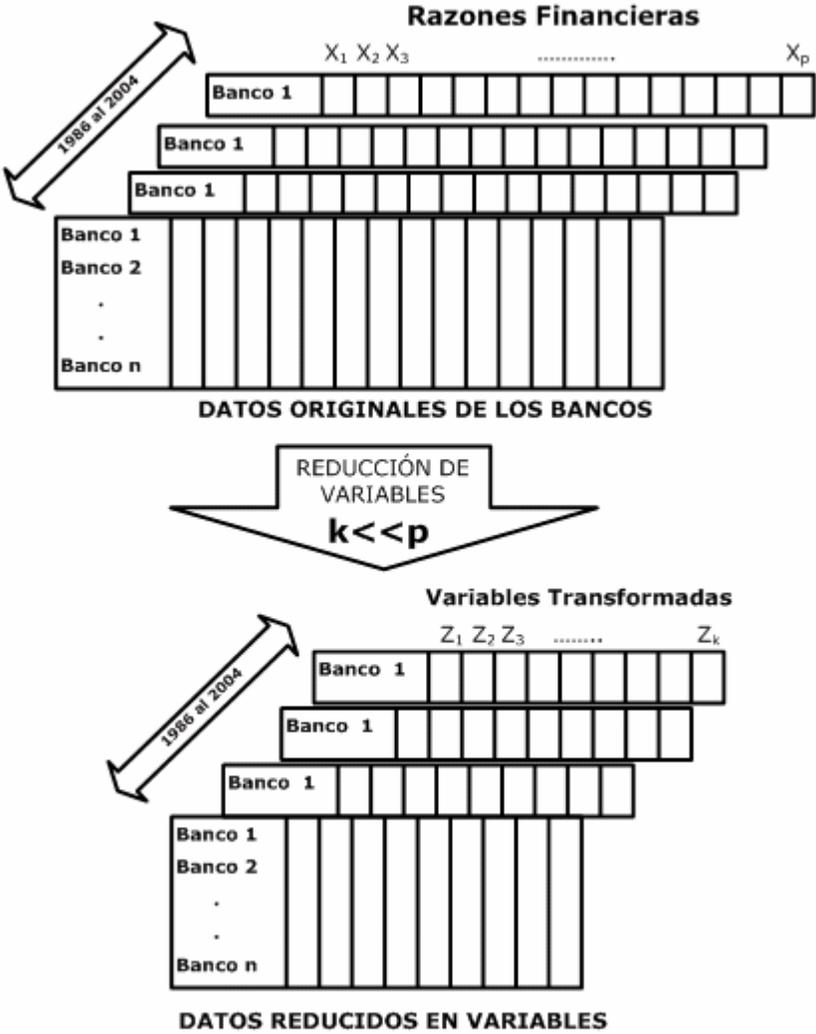


Figura 1. ACP como técnica de reducción de variables

Enfoque analítico de la reducción de las variables originales

Sea \mathbf{X} , una matriz de orden p -dimensional conformada por p variables con N observaciones para cada variable.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdot & \cdot & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdot & \cdot & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdot & \cdot & x_{3p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & x_{N3} & \cdot & \cdot & x_{Np} \end{bmatrix}$$

Mediante Componentes Principales (ACP), el conjunto total de p variables podría ser reducido a un nuevo conjunto *enteramente independiente* de nuevas variables (algunas veces conocidas como variables latentes) expresada en un matriz resultante \mathbf{Z} de orden k .

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & \cdot & \cdot & z_{1k} \\ z_{21} & z_{22} & z_{23} & \cdot & \cdot & z_{2k} \\ z_{31} & z_{32} & z_{33} & \cdot & \cdot & z_{3k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_{N1} & z_{N2} & z_{N3} & \cdot & \cdot & z_{Nk} \end{bmatrix}$$

El nuevo conjunto de variables define un espacio k -dimensional mucho más reducido que el original donde $k \ll p$, del cual se le puede hacer algunas consideraciones:

Los valores entre sí, de las nuevas variables presentes en la matriz \mathbf{Z} no están correlacionados. Es decir *las nuevas variables son completamente independientes*.

A pesar de que la dimensión de la matriz \mathbf{Z} es igual a la de la matriz \mathbf{X} , su utilidad práctica preferiblemente se reduce a una matriz de dimensión mucho menor ($k \ll p$).

La entrada principal al ACP es la matriz de covarianza Σ . Sin embargo, las variables originales podrían ser estandarizadas (media

cero y varianza uno) para eliminar el efecto de la varianza relativa de las variables originales. En este caso, se puede sustituir como entrada principal a ACP a la matriz de correlación \mathbf{R} por la matriz de covarianza. Esta sustitución es útil para eliminar las altas varianzas generadas por las variables involucradas con diferente escala en las unidades de medida.

La traza de la matriz de correlación es igual a la varianza total de las variables transformadas.

Los autovectores de la matriz de covarianza o de correlación definen los nuevos ejes en el espacio k -dimensional.

La cantidad de componentes principales es igual a la cantidad de variables originales consideradas en el ACP. Es decir, son p componentes.

Cada z_j es una variable transformada de las $x_i(s)$ variables originales. Siendo $i=1,2,3,\dots,p$. Estas nuevas variables contienen los valores de las variables transformadas.

ACP captura solamente linealidad entre las variables. Por eso se le conoce como una técnica que transforma las variables originales mediante un método de combinación lineal. Es decir las z_i son combinación lineal de los componentes y las variables originales x_i .

Los primeros componentes obtenidos mediante ACP explican la mayor cantidad de la varianza total de las variables originales. Es decir, agrupa la mayor cantidad de información que puedan suministrar las variables originales.

Habiendo definido \mathbf{X} y \mathbf{Z} , con p variables (originales y transformadas) y N observaciones, consideremos adicionalmente la matriz de covarianza Σ de la matriz \mathbf{X} ; entonces ACP permite calcular mediante la matriz Σ un nuevo conjunto de p variables no correlacionadas (z) tal que ellas sean combinación lineal de las variables originales.

De este modo, para cualquier observación i en \mathbf{X} dado por $[x_1 \ x_2 \ x_3 \ \dots \ x_p]$, existe una función lineal $\mathbf{Z} = \mathbf{a}'_i \mathbf{x}$, para todo $i=1, \dots, N$. Es decir,

$$z_j = a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \dots + a_{ip}x_p, \quad j=1, \dots, P, \quad i=1, \dots, N, \text{ donde:}$$

\mathbf{a}'_i es un vector transpuesto de pesos o parámetros del i -ésimo componente principal.

\mathbf{a}_j es un vector de pesos o parámetros del j -ésimo componente principal o *autovector* de Σ para formar la combinación lineal de las p variables originales.

\mathbf{a}_j tiene que ser ortogonal y ortonormal. Es decir $\mathbf{a}'_i \mathbf{a}_j = 0$ y $\mathbf{a}'_i \mathbf{a}_i = 1$.

z_j se corresponde con los valores nuevos de las variables originales.

La varianza del j -ésimo componente es

$$V(z_j) = \text{var}(\mathbf{a}'_j \mathbf{x}) = \mathbf{a}'_j \Sigma \mathbf{a}_j, \text{ para todo } j=1, \dots, P, \quad i=1, \dots, N.$$

Las varianzas resultantes para cada componente son decrecientes y en estricto orden: $V(z_1) > V(z_2) > V(z_3) > \dots > V(z_p)$. Su suma representa la variación total de las variables originales.

En general, el objetivo es entonces el de encontrar los componentes principales \mathbf{a}_j tal que $\mathbf{a}'_j \Sigma \mathbf{a}_j$ es un máximo sujeto a que $\mathbf{a}'_j \mathbf{a}_j = 1$, donde $j=1, \dots, P$. Usando Lagrange para el j -ésimo componente, se tiene:

$$L = \mathbf{a}'_j \Sigma \mathbf{a}_j - \lambda_j (\mathbf{a}'_j \mathbf{a}_j - 1),$$

λ_j es el multiplicador y la parcial con respecto al componente es

$$\frac{\partial L}{\partial \mathbf{a}_j} = 2 \Sigma \mathbf{a}_j - 2 \lambda_j \mathbf{a}_j.$$

Igualamos a cero la expresión anterior y podemos calcular los valores de λ_j mediante $(\Sigma - \lambda_j \mathbf{I}) \mathbf{a}_j = 0$, donde $|\Sigma - \lambda_j \mathbf{I}| = 0$ para que $\mathbf{a}_j \neq 0$. Existen p raíces para el polinomio. Estas raíces son los autovalores de Σ y son $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$. Cada valor de λ_j permite el cálculo de su correspondiente autovector \mathbf{a}_j mediante $(\Sigma - \lambda_j \mathbf{I}) \mathbf{a}_j = 0$.

De ahí, el primer autovector \mathbf{a}_1 le corresponde el primer autovalor λ_1 , el cual se obtiene mediante $(\Sigma - \lambda_1 \mathbf{I})\mathbf{a}_1 = 0$ y la condición de ortonormalidad $\mathbf{a}'_1 \mathbf{a}_1 = 1$; a su vez \mathbf{a}_1 es el autovector correspondiente a la varianza más grande de Σ , $\text{var}(\mathbf{a}'_1 \mathbf{x}) = \mathbf{a}'_1 \Sigma \mathbf{a}_1 = \lambda_1$, y $z_1 = \mathbf{a}'_1 \mathbf{x}$. El segundo autovector \mathbf{a}_2 es la segunda varianza más grande de Σ y así sucesivamente. Además, \mathbf{a}_2 es ortonormal, $\mathbf{a}'_2 \mathbf{a}_2 = 1$, \mathbf{a}_2 ortogonal a \mathbf{a}_1 , $\mathbf{a}'_2 \mathbf{a}_1 = 0$, y la varianza de z_2 es $\text{var}(\mathbf{a}'_2 \mathbf{x}) = \mathbf{a}'_2 \Sigma \mathbf{a}_2 = \lambda_2$. En general, El j -ésimo \mathbf{a}_j es ortonormal. $\mathbf{a}'_j \mathbf{a}_j = 1$, ortogonal al resto de los autovectores previos ($\mathbf{a}'_j \mathbf{a}_k = 0$, $k = 1, 2, 3, \dots, j-1$) y la varianza de z_j es $\text{var}(\mathbf{a}'_j \mathbf{x}) = \mathbf{a}'_j \Sigma \mathbf{a}_j = \lambda_j$. Del mismo modo, las varianzas para los p autovectores es decreciente tal que $V(z_1) > V(z_2) > \dots > V(z_p)$.

Finalmente, los autovectores son los componentes principales y los autovalores la varianza de las nuevas variables. Así, las nuevas variables son $z_j = \mathbf{a}'_j \mathbf{x}$, o en notación matricial sería, $\mathbf{Z} = \mathbf{X} \mathbf{A}$. Más detalles en (Jolliffe, 1986), (Sharma, 1996).

En conclusión, contrario a \mathbf{X} , \mathbf{Z} está formado de vectores columnas totalmente ortogonales, es decir por variables no correlacionadas. El procedimiento general para calcular el ACP es como sigue:

La primera variable transformada z_1 es la combinación lineal de \mathbf{X} describiendo la mayor cantidad de variabilidad medida en Σ mediante su autovalor asociado λ_1 dado por $z_1 = \mathbf{X} \mathbf{a}_1$, donde $\lambda_1 = \max(\mathbf{a}'_1 \Sigma \mathbf{a}_1)$.

La segunda variable transformada z_2 , es entonces la combinación lineal $z_2 = \mathbf{X} \mathbf{a}_2$, la cual describe la segunda varianza más grande de Σ . Esto es, maximiza la varianza para $\mathbf{a}'_2 \Sigma \mathbf{a}_2$, sujeta a que z_2 no debe estar correlacionado con z_1 . Es decir, $\mathbf{z}_1 \mathbf{z}_2 = 0$.

En general, la i -ésima variable transformada z_i es $z_i = \mathbf{X} \mathbf{a}_i$ y corresponde con la i -ésima varianza más grande de Σ , dada por

$a_i \sum a_i$, sujeta a que z_i no debe estar correlacionado con el resto de las variables en Z . Es decir, $z_i z_j = 0, j=1, \dots, p$ con $i \neq j$ y $z_i z_i = 1$.

Existe un método matricial directo conocido como Valor de Descomposición Simple (VDS) que puede ayudar a resolver ACP. Supóngase que se tiene la misma matriz original X de N observaciones (filas) y p variables (columnas). La descomposición espectral de esta matriz X puede ser escrita de la siguiente manera:

$$X = QSE'$$

Donde E es una matriz ortogonal $p \times p$ cuyos autovectores son $X'X$, S es una matriz diagonal $p \times p$ formada por los autovalores de $X'X$, Q es una matriz ortogonal $N \times p$. Además, E y Q son ortonormales, es decir $EE' = 1$ y $QQ' = 1$. De igual modo, el producto $X'X$ forma una matriz simétrica y puede ser escrito mediante la descomposición espectral de la siguiente manera,

$$X'X = (QSE')'(QSE') = ES'SE' = E\Lambda E'$$

Nuevamente, sea Σ la matriz de covarianza de X ; puesto que Σ es una matriz cuadrada y simétrica, entonces la podemos expresar mediante la descomposición espectral como $\Sigma = E\Lambda E'$, donde la diagonal de Λ está formada por el cuadrado de los p autovalores de Σ . E es una matriz ortogonal $p \times p$ donde la i -ésima columna corresponde al i -ésimo autovalor.

Las variables transformadas, quedan definida como por Z y $Z = XE$, que no son más que los valores de X proyectados sobre la nueva base ortonormal definida por los autovectores en E . La matriz de covarianza de estas variables transformadas en Z , puede estar expresada por

$$E(Z'Z) = (XE)'(XE) = E'X'XE = E'\Sigma E.$$

De este modo, al sustituir a Σ , la matriz de covarianza será

$$E(Z'Z) = E'E\Lambda E'E = \Lambda.$$

Esto es, los autovalores $\lambda_1, \lambda_2 \dots \lambda_p$ correspondientes a la diagonal de Λ , representan la varianza para cada una de las variables transformadas en \mathbf{Z} , siendo no correlacionadas y en consecuencia, \mathbf{Z} es una matriz no correlacionada.

Otra particularidad de los ACP es que la varianza total de \mathbf{Z} es igual a la varianza total de \mathbf{X} . La varianza total de Σ corresponde a la suma de la varianza de cada una de las variables originales en \mathbf{X} y está dada por su

propia traza, $\text{tr}(\Sigma) = \sum_{j=1}^p \sigma_{jj}^2$. Con la ayuda de la propiedad que indica que

$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, se puede mostrar que la varianzas de \mathbf{X} y \mathbf{Z} son iguales como sigue:

$$\text{tr}(\Sigma) = \text{tr}(\mathbf{E}\Lambda\mathbf{E}') = \text{tr}(\mathbf{E}'\mathbf{E}\Lambda) = \text{tr}(\Lambda).$$

Si las variables originales son estandarizadas, entonces la matriz de correlación \mathbf{R} sería usada en sustitución de Σ . En este caso, la descomposición espectral para \mathbf{R} sería $\mathbf{R} = \mathbf{E}\Lambda\mathbf{E}'$. De igual manera que para la matriz de covarianza, $\mathbf{Z} = \mathbf{X}\mathbf{E}$ y $E(\mathbf{Z}'\mathbf{Z}) = \Lambda$. Adicionalmente, para este caso en particular, la suma de los autovalores en Λ es igual a p y es igual a la suma de de los elementos de la diagonal de \mathbf{R} .

Lo que es lo mismo $\text{tr}(\mathbf{R}) = \text{rango}(\mathbf{R}) = \text{tr}(\Lambda) = p$. Es importante tener en cuenta que los autovectores y los autovalores son sensibles a los cambios de escala. Σ podría mostrar severos cambios por esta misma razón. \mathbf{R} , por el contrario, omite este problema y de ahí que los autovalores y autovectores originados por \mathbf{R} o Σ son diferentes.

Las Redes Neuronales Multicapas con funciones logísticas como función de activación y el algoritmo de retropropagación del error.

En el entendido de que es un procesador paralelo distribuido en los que se almacena el conocimiento, emula su símil biológico en dos aspectos fundamentalmente: a) el conocimiento es adquirido a través de un

proceso de aprendizaje y b) la fortaleza de la interconexión de las neuronas artificiales está descrita en los pesos sinápticos.

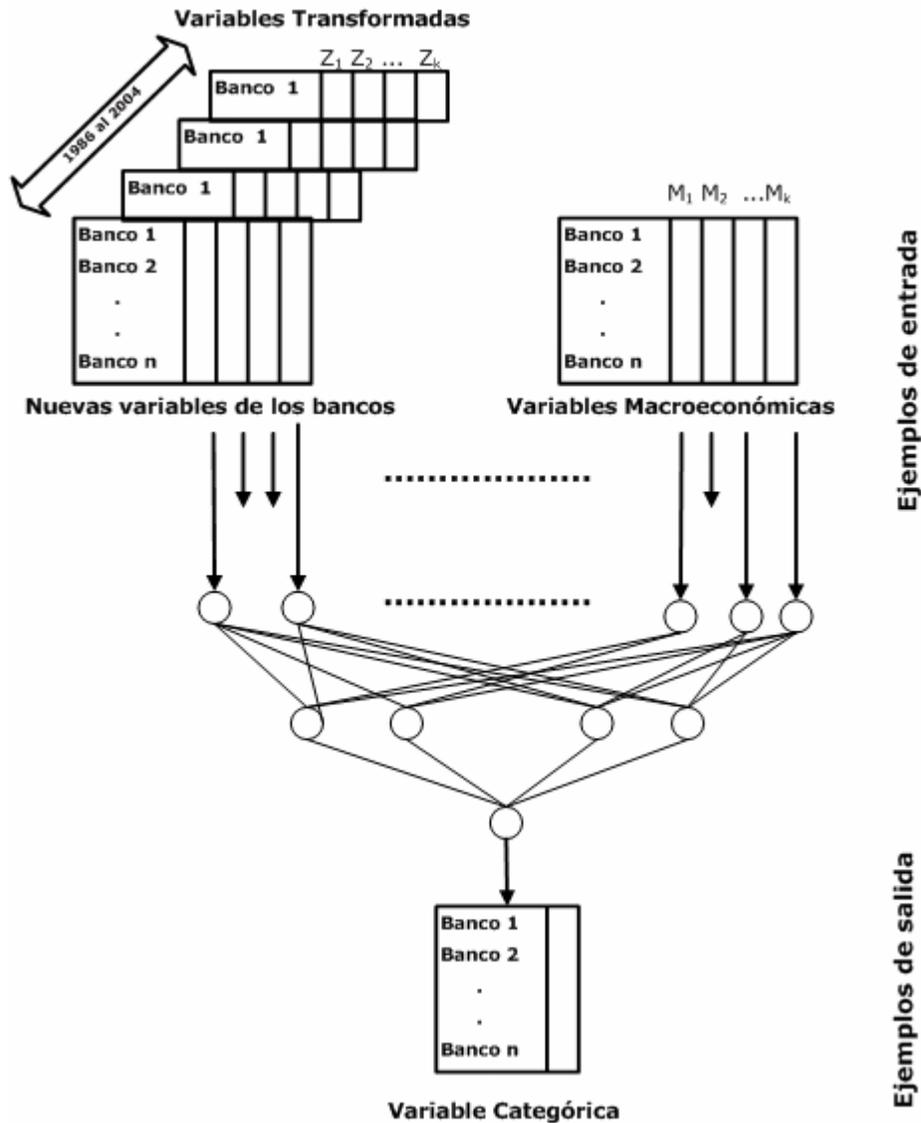


Figura 3. ACP y RN-RP para la construcción del modelo híbrido de pronóstico

Bajo estas consideraciones esenciales, toda red artificial ajusta permanentemente sus parámetros libres (pesos y umbrales) mediante mecanismos de aprendizaje que podrían ser supervisados o no. El caso

particular que se aplica a este modelo híbrido es un modelo de red neuronal multicapa que activa el cambio en las interconexiones de los pesos sinápticos a través de una función de excitación no lineal (logística) y un proceso de corrección de estos pesos mediante la constante evaluación de su convergencia a un umbral dado. Este proceso no supervisado se le conoce con el nombre de red neuronal con un algoritmo de retropropagación del error (backpropagation neural network) (RN-RP).

Tal como lo muestra la figura 3, la topología de una red neuronal RN-RP, es similar a la mostrada. Una capa de nodos que capturan los componentes de los patrones de entrada uno a la vez y del mismo un patrón de salida capturado por el o los nodos de salida. En el intermedio se incluye una capa de nodos que se le conoce como capa oculta y en cada uno de sus nodos se incluye la función de excitación. Esta función puede variar de acuerdo a las características del dominio del problema real. Puede ser funciones asimétricas como la hiperbólica tangencial, la lineal, y la función no simétrica conocida como logística. El proceso de aprendizaje en este tipo de redes es básicamente en dos fases (Werbos 1974; Rumelhart et al. 1986):

Fase hacia delante: Los parámetros libres de la red son establecidos y el patrón de entrada i es propagado a través de la red capa por capa. Esta fase finaliza con el cálculo del error e_i entre la salida calculada s_c y el patrón de salida dado s_d , $e_i = s_d - s_c$.

Fase hacia atrás: El error e_i es propagado a través de la red en dirección hacia atrás. De ahí el nombre del algoritmo. Es en esta fase que se realizan los ajustes a los parámetros libres de la red, así que se minimice el error en un sentido estadístico, mediante el error cuadrático medio.

El algoritmo de retropropagación se procesa en el modo lote, regularmente. Los pesos se ajustan sobre la base de ciclo a ciclo, donde

cada ciclo consiste de el conjunto completo de patrones de entrada y salida disponible para el entrenamiento. Este algoritmo es sencillo de implementar y eficiente computacionalmente. La arquitectura de los modelos neuronales seguirá una topología fija en cada uno de los experimentos conducentes a medir la consistencia en el entrenamiento y capacidad de generalización. El número de nodos en la capa oculta estará directamente vinculado con la cantidad de parámetros libres que incluiría la red neuronal y el nivel de convergencia deseado. De este modo, existirán una fase inicial de entonación de la red, ajustando nodos ocultos, función de activación, separación de los conjuntos de datos de entrenamiento y verificación, ajuste de la tasa de aprendizaje, selección del momento ideal para escapar lo más de las soluciones locales, nivel del error de convergencia, selección de un conjunto de pesos iniciales, y una segunda fase que es totalmente experimental comprendiendo solamente la construcción y prueba de los modelos. El criterio de parada empleado es el de la verificación cruzada (Haykin, 1999) que consiste en: a) Dividir el conjunto de entrenamiento en dos partes: subconjunto de estimación usado para el entrenamiento y subconjunto de verificación para evaluar el desempeño del modelo previamente entrenado; b) Entonación final de la red usando el conjunto entero de entrenamiento y luego probar la generalización con patrones nunca conocidos por el modelo.

Sección experimental

Haciendo referencia al diagrama funcional mostrado en la figura 3 y para el caso particular de los datos de la banca comercial compuesto por las razones financieras y la agrupación obtenida en el modelo de análisis discriminante en una variable categórica, se dividirá en dos subconjuntos al conjunto original, antes de iniciar la construcción de los modelos. El conjunto original de datos de aproximadamente 8000

observaciones, se dividirá en uno para realizar el entrenamiento, verificación y puesta a tono del modelo, de aproximadamente 6000 ejemplos, y el otro, formado por un subconjunto con aproximadamente 2000 ejemplos nunca usados por el modelo serán usados para verificar el nivel de desempeño o generalización. En la construcción de la red neuronal y como un primer intento de estimación del número de nodos presentes en la capa oculta, se empleará el criterio heurístico que determina el orden de acuerdo a una relación entre el nivel de error deseado y el número limitado de parámetros libres involucrados (pesos y umbrales).

Los modelos serán replicados una cantidad aceptable de veces con diferentes grupos de datos de entrenamiento y manteniendo constante las consideraciones iniciales asumidas para la topología y arquitectura de la red neuronal. El resultado de obtener varios modelos es para que permita medir la consistencia de los valores en las tres fases fundamentales de la construcción de los modelos: entrenamiento, verificación y generalización. Las medidas de evaluación que garantizarían resultados aceptables y consistentes serán el error medio cuadrático calculado entre el los valores deseados y calculados en la fase de entrenamiento y verificación, el error medio cuadrático (RMSE) obtenido durante la fase de generalización. Ambos resultados serán medidos para cada modelo y luego de su tabulación, serán observadas las variaciones en los resultados del RMSE obtenidos en las dos fases para determinar el grado de consistencia conseguido en la construcción de los modelos bajo condiciones uniformes de diseño. Una alta consistencia conseguida entre los modelos entrenados y probados garantiza una aceptable robustez en los procesos de pronóstico bajo las condiciones de construcción que se siguieron al inicio.

Referencias.

- Amir F. Atiya, *Senior Member, IEEE*. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 12, NO. 4, JULY 2001
- Ayesterán, José R., Ramoni, Josefa y Orlandoni Giampaolo. (1996). Crisis y Poder: El caso del Sistema Financiero. Economía Nueva Etapa. No. 11.
- Colmenares Gerardo, (2004). Reducing Archives to Build Non Linear Models Using Neural Networks. AMSE Periodicals. Lion. France
- Dong D. and McAvoy T. (1996). Nonlinear Principal Component Analysis -Based on Principal Curves and Neural Networks. *Computer Chem. Engng.*, vol 20, no. 1, pp. 65-78.
- Haykin Simon. (1995). Neural Networks. A comprehensive foundation. Macmillan College Publishing Company, Inc.
- Haykin Simon. (1998). Feedforward: Neural Networks: An Introduction. Chapter 1. Manuscrito
- Kramer M. (1992). Autoassociative Neural Networks. *Computer Chem. Engng.*, vol. 16, no. 4, pp. 313-328.
- Sharma Subash. (1996). Applied Multivariate Techniques. John Wiley. USA
- Tan S. and Mavrovouniotis M. (1995). Reducing data dimensionality through optimizing neural network inputs. *AIChE Journal*, vol 41, no. 6, pp. 1471-1480.