

Un sistema de ayuda a la toma de decisiones en series temporales fuzzy

Abel Rubio, José D. Bermúdez, Enriqueta Vercher arufor@alumni.uv.es,
jose.d.bermudez@uv.es, enriqueta.vercher@uv.es

Departamento de Estadística e Investigación Operativa, Universitat de València
C/ Doctor Moliner 50, Burjassot - Valencia

Resumen En este trabajo se proponen varios métodos de ponderación para realizar predicciones en series temporales fuzzy. Para gestionar eficazmente la diversidad de métodos de ponderación, se construye un árbol de toma de decisiones en el que las elecciones están basadas en el histórico de datos. Para establecer predicciones ex-post nuestro sistema clasifica las características de la tendencia mediante un operador fuzzy, y devuelve como predicción un número fuzzy trapezoidal. Para analizar el comportamiento del sistema se han realizado diferentes experimentos con datos de series temporales de cotizaciones de activos financieros.

1. Introducción

El comportamiento de los índices bursátiles ha sido analizado por una amplia cantidad de expertos en economía financiera desde diferentes perspectivas, para que mediante un razonamiento adecuado pueda aprenderse a dilucidar su comportamiento futuro. En este trabajo, nos vamos a centrar en la observación de 1 a cotización diaria de dos índices bursátiles de referencia (IBEX35 y NIKKEI255), con el propósito de aprender a decidir sobre cómo debería predicirse ese comportamiento futuro, asumiendo que podemos aprender sobre el futuro en base al comportamiento pasado de dichos índices, independientemente considerados.

Dada la ambigüedad de los datos financieros (generalmente no lineales) y la asimetría de la volatilidad a lo largo del tiempo, se pretende analizar las series temporales de cotización de los índices bursátiles desde un punto de vista fuzzy. En particular, seguiremos el esquema básico de las series temporales fuzzy (STF) que utilizan ponderaciones sobre la información extraída de las relaciones fuzzy entre variables lingüísticas y las relaciones fuzzy de grupos [Song and Chissom(1993),Chen(1996),Huang(2001b)]. Para poder decidir entre las distintas estrategias de ponderación desarrolladas para el análisis de series temporales con metodología fuzzy, proponemos aplicar un sistema de apoyo a la decisión (DSS) que, a partir del análisis exploratorio de los datos observados, discrimine entre los diferentes métodos de ponderación propuestos. Asimismo, hemos decidido utilizar minería de datos, para categorizar las distintas series temporales de cotizaciones de índices bursátiles según algunas de sus propiedades y/o parámetros.

La minería de datos trata de aportar conocimiento y estructura ante la ambigüedad de los datos almacenados en las grandes bases de datos, con el objeto de proporcionar una mejor interpretación, clasificación, segmentación y análisis de los mismos. Dada la complejidad y magnitud de los datos económico-financieros, esta disciplina está siendo ampliamente aplicada para el tratamiento y posterior análisis de las series financieras de datos.

Las series temporales fuzzy permiten establecer un nuevo enfoque para el tratamiento de series temporales no lineales, proporcionando estimaciones razonablemente ajustadas al valor real. Además, pueden aplicarse cuando el histórico de datos viene expresado a través de variables lingüísticas.

[Song and Chissom(1993)] introdujeron los conceptos fundamentales de series temporales fuzzy, y propusieron modelos de predicción invariantes y variantes temporales. Posteriormente, en [Chen(1996)] se desarrolló una simplificación de la metodología de las STF, y se establecieron los principios básicos de los modelos de predicción basados en series temporales fuzzy. Recientemente, en [Chen and Hsu(2004)] se introducen una serie de reglas heurísticas para la predicción que completan el trabajo original de Chen. Los pasos principales del modelo de Chen son: (i) definición y partición del universo del discurso; (ii) establecimiento de los valores lingüísticos representados mediante conjuntos fuzzy y clasificación de los datos; (iii) definición de las relaciones fuzzy entre conjuntos y de las relaciones fuzzy de grupos; y (iv) cálculo del valor crisp de la predicción, a partir de unas reglas de predicción previamente establecidas.

La ambigüedad conceptual contenida en cada una de las fases anteriores, abre la posibilidad de subdivisión y desarrollo en múltiples pasos. En particular, los principales pasos que han sido desarrollados por su influencia en la predicción de modelos de STF son: la partición inicial del universo del discurso y los criterios utilizados en las reglas de predicción.

En [Huang(2001a)] se prueba que la partición inicial del universo U afecta a las predicciones, y a partir del análisis de los datos de la serie temporal se proponen dos algoritmos (distribution-based length y average-based length) que determinan la longitud intervalar de la partición inicial del universo, consiguiendo una disminución notable de los errores de ajuste. Aunque sus modelos STF siguen el esquema básico propuesto por Chen, sus propuestas de partición de U permiten alcanzar una disminución notable de los errores de ajuste. Por su parte, [Yu(2005)] introduce las series temporales fuzzy ponderadas y toma en consideración las recurrencias en las relaciones fuzzy de grupos, proponiendo una ponderación cronológica. Recientemente, se han propuesto otras particiones con intervalos de diferente longitud basándose en procedimientos de información granular (véase, por ejemplo, [Wang et al.(2014)]).

En lo que respecta a los criterios para establecer predicciones ex-post, algunos modelos de STF se centran en la mejora de los errores de ajuste más que en la precisión de la predicción futura. En tanto que algunos modelos no permiten una adaptación eficiente frente a series de naturaleza errática, de modo que la ponderación no consigue captar los cambios de tendencias actuales y/o

fluctuaciones súbitas, pues están concebidos para series más segmentadas y/o lingüísticas. Ahora bien, el comportamiento que pueda tener una serie financiera en cierto segmento de su cotización, no explica o permite deducir su estado actual ni sus alteraciones consecuentes, por lo que para explicar su comportamiento actual se necesita trabajar con propiedades globales.

Nuestra propuesta para la predicción de series temporales financieras mediante conjuntos difusos, puede considerarse un método ponderado para la predicción del valor futuro de un índice bursátil de referencia o de la cotización de un activo. Para mejorar la precisión y robustez de la predicción hemos introducido algunos cambios que afectan tanto al esquema de relaciones de las variables lingüísticas de los modelos clásicos de STF como a las estrategias de ponderación, que estarían basadas en el análisis de patrones. En el presente trabajo se sigue el esquema siguiente: en la sección 2 presentamos el modelo ponderado que se propone para el análisis de series temporales, y la estrategia de ayuda a la decisión entre las diferentes propuestas de ponderación, y en la sección 3 se incluyen los experimentos numéricos realizados con cotizaciones diarias de los dos índices bursátiles, IBEX35 y NIKKEI225.

2. Modelo ponderado de serie temporal fuzzy

El marco de referencia de nuestro trabajo es la metodología básica introducida en [Chen(1996)], aunque se han introducido variaciones en diferentes etapas del procedimiento. En un trabajo previo, habíamos propuesto un operador de ponderación cronológico basado en las relaciones fuzzy de la serie temporal, y que utilizaba dichos pesos en la relación fuzzy de grupos [Rubio et al(2016)]. Nuestro enfoque permite utilizar números fuzzy trapezoidales como representantes de las variables lingüísticas de cada elemento de la partición del universo del discurso, y obtenerlos como predicciones fuzzy ex-post (la idea de trabajar con números trapezoidales simétricos había sido sugerida en [Liu(2007)]).

A lo largo de todo este artículo utilizamos diferentes técnicas de minería de datos, empezamos haciendo un análisis exploratorio de los datos y realizamos una clasificación de dichos datos basada en la desviación típica de los mismos. Posteriormente construimos un conjunto de números trapezoidales fuzzy y etiquetamos los datos utilizando las diferentes variables lingüísticas. A partir de aquí se extraen las relaciones fuzzy existentes, de donde se deriva una de las propiedades de la serie que consideramos más relevante los 'saltos', analizamos dichos saltos utilizando técnicas de reconocimiento de patrones y finalmente acabamos con unas sencillas reglas heurísticas para la predicción. Notamos que el sistema de ayuda a la decisión afecta tanto al análisis de los patrones de saltos como a las reglas heurísticas de la predicción.

2.1. Operadores de ponderación

Para definir la partición del universo del discurso U en intervalos de idéntica longitud tenemos en cuenta la dispersión de los datos observados; es decir,

utilizamos la desviación típica de los datos para determinar la longitud de los intervalos en la partición de U . La importancia de la partición (en intervalos de idéntica amplitud o no idénticos) y del número de intervalos sobre los que definimos los números fuzzy trapezoidales ya se ha analizado en multitud de trabajos anteriores.

Nuestra aportación más importante radica en la introducción de nuevos operadores de ponderación que permiten analizar la tendencia de la serie, y predecir fuera del dominio de definición del universo de discurso. Para ello se hace necesario introducir el concepto de *salto* en la serie temporal fuzzy: para dos conjuntos fuzzy consecutivos o relacionadas, el salto se corresponde con la diferencia (temporal) entre sus subíndices, lo que deriva en la obtención de una nueva serie temporal de *saltos*. Los *saltos* son la manifestación de los movimientos que infiere la lógica fuzzy implícita en las relaciones fuzzy entre los conjuntos (números fuzzy o variables lingüísticas) definidos mediante la partición del universo del discurso. Los *saltos* pueden considerarse como una característica intrínseca de la serie temporal fuzzy que intentan captar la tendencia de la serie, al tiempo que persigue ligar la ponderación de las relaciones fuzzy a toda la serie, en lugar de a las relaciones fuzzy de grupo (como suele hacerse habitualmente).

Finalmente, introducimos un nuevo operador basado en la secuencia de *saltos* asociados a las relaciones lógicas fuzzy. A partir del análisis de dichas secuencias, se buscan patrones mediante la identificación de recurrencias en las secuencias de *saltos*. En resumen, proponemos utilizar alternativa o conjuntamente tres diferentes operadores de ponderación, como sigue:

1. Operador Cronológico: opera sobre las relaciones fuzzy entre los intervalos del universo U , y normaliza dichos pesos a partir de las relaciones fuzzy de grupos.
2. Operador de Saltos: opera sobre la serie temporal de saltos, y normaliza los pesos a partir de las relaciones entre los saltos.
3. Operador de Patrones: opera sobre las secuencias de saltos en la serie temporal de saltos, buscando patrones.

Las modificaciones introducidas para la composición y asignación de los pesos obvian el uso de las relaciones fuzzy de grupo (salvo en el caso del operador cronológico), y afectan de manera importante al proceso de predicción fuzzy, proporcionando como predicción del comportamiento futuro un número fuzzy trapezoidal, no necesariamente simétrico.

La obtención de dicho número trapezoidal fuzzy permite calcular el valor y la ambigüedad del comportamiento futuro de la variable de interés, así como algunos momentos posibilísticos a partir del concepto de intervalo de medias [Dubois and Prade(1987),Fuller and Majlender(2003)].

2.2. Sistema de ayuda a la decisión

Nuestra propuesta para elegir qué operador u operadores deben utilizarse para una serie temporal dada, toma en consideración la información

proporcionada por el histórico de datos (tamaño de la muestra y dispersión de los datos observados), y las características específicas de las series temporales fuzzy que hemos creado al analizar las relaciones entre los conjuntos fuzzy que componen el universo del discurso (i.e. tendencia fuzzy de la serie temporal de *saltos*). La decisión es una elección entre las alternativas basadas en las estimaciones de los valores de esas alternativas, y para categorizar las distintas series temporales de datos originales utilizaremos ambos tipos de información.

En primer lugar, definimos *tendencia fuzzy* de una serie temporal como la suma de los *saltos* obtenidos en la serie temporal fuzzy. Por construcción, los saltos son números enteros, cuya suma será otro número entero que puede ser:

- Positivo, se entenderá como tendencia fuzzy creciente.
- Negativo, se entenderá como tendencia fuzzy decreciente.
- Cero, se entenderá como tendencia fuzzy nula.

La tendencia fuzzy nula se considera como un movimiento horizontal o bucle de la serie temporal alrededor de cierto valor observado. Cabe destacar que también se consideran o se les asigna una tendencia nula a aquellas series cuya tendencia fuzzy tome valores cercanos a cero. Aunque el hecho de estar alrededor del cero dependerá del número de saltos que intervengan en el sumatorio y del rango de los saltos. Por ejemplo, en el caso de las series de cotizaciones, con las que trabajaremos en la sección siguiente, si se pretenden calcular tendencias fuzzy mensuales, se agruparan los saltos de 20 en 20, si fueran semanales de 5 en 5, y así en función de la periodicidad de los datos. Todo dependerá del tipo de predicciones que se pretendan realizar (diarias, semanales, mensuales, etc). Notar que la clasificación de la tendencia fuzzy subdivide al operador de ponderación basado en patrones en tres categorías:

- Operador creciente: donde se almacenan todos los patrones que han estado en un periodo determinado con tendencia creciente.
- Operador decreciente: donde se almacenan todos los patrones que han estado en un periodo determinado con tendencia decreciente.
- Operador neutro: donde se almacenan todos los patrones que han estado en un periodo determinado con tendencia nula.

Finalmente, en el procedimiento de automatización de los operadores que se van a utilizar para calcular la predicción fuzzy se sigue una heurística sencilla basada en el número de observaciones de la serie (o conjunto de datos de entrenamiento). De modo que para series con $n < 20$ observaciones se utiliza solamente la ponderación cronológica, pues el sistema no ha tenido posibilidad de aprender ni de verificar si hay tendencia; por motivos análogos, para $20 \leq n < 50$, utilizaríamos el operador de saltos (con el mismo peso que el cronológico), pero no el operador de patrones; para $50 \leq n < 100$, podemos encontrar series con tendencia nula (aplicaríamos el mismo esquema que en el caso anterior) o series tendencia positiva o negativa para las que utilizaríamos el operador de patrones de longitud 2. Conforme vaya creciendo el tamaño del conjunto de entrenamiento puede ir creciendo también la longitud del patrón.

3. Experimentos numéricos

Para analizar el comportamiento del sistema de decisión presentado en el apartado anterior, utilizaremos dos índices de referencia (IBEX35 y NIKKEI225) de los mercados bursátiles español y japonés, respectivamente.

Toda la información sobre las cotizaciones de dichos índices, han sido obtenidas de la web 'www.finance.yahoo.com'. Tanto el algoritmo básico como los distintos métodos de ponderación han sido implementados en un ordenador personal usando el entorno de programación de R (www.r-project.org). Para analizar la precisión alcanzada con el DSS hemos calculado el RMSE (root of the mean square error) y el MAD (mean absolute deviation) de los errores de predicción de una etapa.

Hemos considerado cotizaciones diarias de los años 2013 al 2015. Los datos observados se han dividido en dos conjuntos: datos de entrenamiento (ex-ante) y datos de testeo (ex-post). El análisis del conjunto de entrenamiento mediante minería de datos nos ha permitido discriminar entre los distintos métodos de ponderación, para aplicar del DSS propuesto.

3.1. Predicción del IBEX35

En el primer experimento hemos trabajado con datos diarios del IBEX35 del año 2013, las 100 primeras cotizaciones nos servirán como datos de entrenamiento para las primeras 5 predicciones (una hipotética semana de cinco días). Posteriormente, se van incorporando esas observaciones al conjunto de entrenamiento, en un esquema de horizonte rodante, se trabaja con esa nueva serie temporal desde un punto de vista de las STF y se calculan las matrices de peso asociadas a los realaciones fuzzy, las relaciones fuzzy de grupo, se redefinen los saltos y patrones y, se realizan 5 nuevas predicciones, y así sucesivamente de cinco en cinco. Se ha repetido este esquema de predicción semanal 40 veces, obteniéndose un total de 200 predicciones diarias.

La Figura 1 muestra las observaciones de esos 200 días del año 2013 junto con las predicciones obtenidas por el DSS para esos mismos días. Cabe destacar que inicialmente, debido al tamaño del conjunto de entrenamiento, el sistema no encontraba patrones, pero conforme ha ido creciendo dicho tamaño el sistema ha ido mejorando en las predicciones crisp realizadas.

Para analizar el comportamiento que tiene el sistema de predicción que hemos implementado, hemos escogido trabajar con patrones de longitud 4. Una vez concluido el experimento, hemos calculado el RMSE semanal y su valor medio para las 40 semanas consideradas, que ha resultado ser 90.97. En tanto que la valor medio del MAD para las 40 semanas ha sido de 69.44, con un sesgo medio de -5.57.

3.2. Predicción del índice NIKKEI225

En este segundo experimento hemos utilizado cotizaciones diarias del índice NIKKEI225 de los años 2013-2015. Las 500 primeras cotizaciones nos sirven

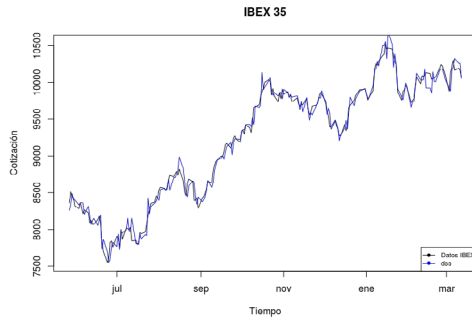


Figura 1. Evolución de las cotizaciones diarias del IBEX35 (en negro) y de las predicciones (en azul) para 200 días de 2013.

como datos de entrenamiento del DSS para realizar las primeras 5 predicciones, y, posteriormente se aplica el mismo esquema de horizonte rodante anterior, se van incorporando a los datos de entrenamiento las últimas 5 cotizaciones y se recalculan todos los elementos del modelo de STF antes de realizar las siguientes 5 predicciones, y así sucesivamente. Este procedimiento se repite 40 veces, con un total de 200 predicciones diarias. En la Figura 2 puede verse la representación de las cotizaciones y de las predicciones para los 200 días que se han utilizado para el testeo del DSS.

Análogamente al experimento anterior, hemos calculado las medias de las medidas de error de predicción de estas 40 aplicaciones del DSS, obteniéndose que $RSME=213.54$ y $MAD=140.25$. Notar que se ha trabajado con un conjunto de entrenamiento mucho mayor, lo que ha permitido utilizar patrones de mayor longitud, aunque la volatilidad en 2015 ha sido mucho mayor que la observada en 2013 para el IBEX35.

4. Conclusiones

En este trabajo se han introducido nuevas estrategias de análisis de datos de series temporales fuzzy, que incluyen la definición de la tendencia fuzzy de la serie, basada en el concepto de salto entre relaciones fuzzy en una STF. Este nuevo concepto ha permitido también la introducción de un operador de ponderación basado en patrones de saltos.

Con el propósito de automatizar el procedimiento de ponderación, que permite devolver las predicciones ex-post, se ha desarrollado un sistema de ayuda

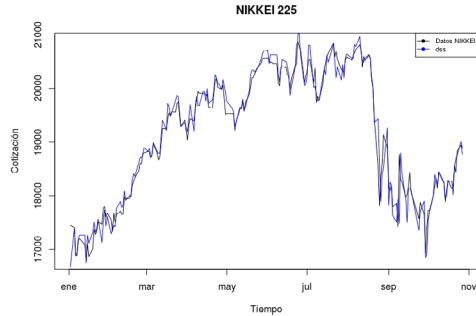


Figura 2. Evolución de las cotizaciones diarias (en negro) del NIKKEI225 y de las predicciones (en azul) para 200 días de 2015.

a la decisión en el que intervienen todos los parámetros fuzzy introducidos para el análisis de STF.

Hemos contrastado el funcionamiento del sistema con series de cotizaciones diarias de índices bursátiles con resultados prometedores.

Agradecimientos

Investigación parcialmente subvencionada por el Ministerio de Economía y Competitividad, España (Proyecto de Investigación MTM2014-56233-P)

Referencias

- [Song and Chissom(1993)] Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54, 269-277 (1993).
- [Chen(1996)] Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*, 81, 311-319 (1996).
- [Huang(2001b)] Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy sets and Systems*, 123, 387-394 (2001b)
- [Chen and Hsu(2004)] A new method to forecast enrollments using fuzzy time series.: *International Journal of Applied Science and...*, 1, 234-244 (2004).
- [Huang(2001a)] Heuristic models of fuzzy time series for forecasting. *Fuzzy sets and systems*, 123, 369-386, (2001a).
- [Yu(2005)] Weighted fuzzy time series models for TAIEX forecasting. *Physica A*, 349, 609-624 (2005).

- [Wang et al.(2014)] Determination of temporal information granules to improve forecasting in fuzzy time series. *Expert Systems with Applications*, 41, 3134-3142 (2014).
- [Rubio et al(2016)] Forecasting portfolio returns using weighted fuzzy time series methods. *International Journal of Approximate Reasoning*. doi: 10.1016/j.ijar.2016.03.007, (2016).
- [Liu(2007)] An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimization and Decision Making*, 6, 63-80, (2007).
- [Dubois and Prade(1987)] Dubois, D. and Prade, H.: The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24, 279-300 (1987).
- [Fullér and Majlender(2003)] Fullér, R. and Majlender, P.: On weighted possibilistic mean and variance of fuzzy numbers. *Fuzzy Sets and Systems*, 136, 363-374 (2003).