

Combinación de clasificadores mediante el método Boosting. Una aplicación a la predicción del fracaso empresarial en España

DATOS BÁSICOS DE TESIS DOCTORAL

Autor: Esteban Alfaro Cortés

Director: Matías Gámez Martínez

Universidad y fecha de lectura: Universidad de Castilla-La Mancha, 30 de noviembre de 2005.

Palabras clave: Predicción del fracaso empresarial, combinación de clasificadores, método Boosting

Clasificación JEL: 0272, 0463, 0984

Acceso al documento completo: ProQuest Information and Learning Company (<http://il.proquest.com/brand/umi.shtml>)

Publicaciones:

Alfaro, E., Gamez, M. and Garcia, N. (2006): "Multiclass corporate failure prediction by Adaboost.M1", *International Advances in Economic Research*, próxima publicación.

RESUMEN

La presente tesis doctoral está estructurada en tres partes.

La primera parte comprende del capítulo I al capítulo IV. Tras exponer algunos aspectos generales de los problemas de clasificación, se analizan algunos de los métodos de clasificación individuales más utilizados en la actualidad, destacando sus principales ventajas e inconvenientes.

En la segunda parte (capítulos V al VII) se analizan algunos aspectos relacionados con el comportamiento y las propiedades de los clasificadores individuales. En concreto, se plantean las dificultades que pueden surgir debido al uso de los clasificadores individuales, como son la precisión y la estabilidad de los mismos. A continuación, en el capítulo VI, se aborda el estudio de la combinación de clasificadores prestando especial atención al método boosting. Además, se recoge una taxonomía de los métodos de combinación y se introducen también el método bagging y el bosque aleatorio. Por último, se estudian los primeros algoritmos que han dado lugar al desarrollo posterior del método boosting. También se exponen algunas de las modificaciones que se han propuesto al algoritmo Adaboost, incluyendo las que sirven para afrontar la existencia de más de dos clases y, para acabar, se analiza cuál debe ser el tamaño adecuado de los árboles utilizados en la combinación.

En la tercera parte (capítulos 8-10) se proporciona una visión general de la predicción del fallo empresarial, sus antecedentes y estado actual. Además, se elaborará un listado con los ratios financieros que han resultado

de mayor utilidad para el pronóstico del fracaso. También se lleva a cabo una descripción de la evolución que han seguido en España las empresas fracasadas. El capítulo IX se centra en la aplicación práctica. Después de recoger brevemente algunas consideraciones teóricas sobre el tratamiento de la información, se realiza un análisis exploratorio de los datos. Además de catorce ratios financieros, se utilizan otras tres variables menos habituales que intentan recoger el tamaño de la empresa, la actividad a la que se dedica y la forma jurídica que presenta. Se coteja el método boosting con los árboles de clasificación, tanto para el caso dicotómico, como cuando se distingue entre tres clases. A continuación se realiza una comparación, algo menos detallada, con otros cinco métodos de clasificación. Posteriormente, se examina la capacidad de los modelos establecidos anteriormente para predecir el fracaso empresarial cuando aumenta la distancia temporal al período en que se hace efectivo el fallo. Finalmente, se concluye que boosting mejora los resultados de los árboles de clasificación individuales. Entre las principales aportaciones de este trabajo destacan el uso de una técnica novedosa, el método boosting y la consideración de un concepto de fracaso empresarial más amplio del habitual.

ÍNDICE

Capítulo 0. Planteamiento y objetivos

0.1. Planteamiento

0.2. Objetivos

Primera Parte

Capítulo 1. Introducción

1.1. Introducción

1.2. Definición de clasificación

1.3. Ejemplos de clasificadores

Capítulo 2. Clasificación

2.1. Introducción

2.2. Estructura general de los problemas de clasificación

2.3. La regla de Bayes a posteriori

2.4. Clasificación de los procedimientos de clasificación

2.5. La selección de variables

2.6. Estimación de la precisión de un sistema de clasificación

Capítulo 3. Métodos de Clasificación Individuales: Análisis Discriminante y Vecino Más Próximo

3.1. Introducción

3.2. Discriminante Lineal

3.3. Discriminante Cuadrático

3.4. Discriminante Logístico

3.5. Método del vecino más próximo

3.6. Cotas de error de las reglas 1-NN y k-NN

3.7. Clases de rechazo

3.8. Comentarios sobre las reglas k-NN

3.9. Reducción en las reglas del vecino más próximo

Capítulo 4. Métodos de Clasificación Individuales II: Redes Neuronales Artificiales y Árboles de Clasificación

- 4.1. Introducción
- 4.2. Motivación biológica de las redes neuronales artificiales
- 4.3. El perceptrón simple
- 4.4. Redes Multicapa
- 4.5. La aplicación práctica del algoritmo de retropropagación
- 4.6. Árboles de clasificación
- 4.7. Construcción del árbol de clasificación
- 4.8. La regla de corte o división
- 4.9. Criterio de parada
- 4.10. Sobreajuste en los árboles de clasificación
- 4.11. Técnicas de podado
- 4.12. Valores omitidos

Segunda Parte

Capítulo 5. Problemática de los clasificadores individuales

- 5.1. Introducción
- 5.2. Descomposición del error. Estudio del sesgo y de la varianza
- 5.3. Estudio de la inestabilidad de los clasificadores
- 5.4. ¿Por qué utilizar combinaciones de clasificadores?
- 5.5. La perspectiva bayesiana sobre la agregación de clasificadores

Capítulo 6. Métodos de combinación de clasificadores

- 6.1. Introducción
- 6.2. Clasificación de los métodos de combinación
- 6.3. Bagging
- 6.4. Boosting
- 6.5. Bosques aleatorios

Capítulo 7. El Método Boosting y sus variaciones

- 7.1. Introducción
- 7.2. Los orígenes del Boosting
- 7.3. Relación del método Boosting con los modelos aditivos generalizados
- 7.4. Versiones de Adaboost
- 7.5. Generalización del Boosting al caso de q clases
- 7.6. El tamaño de los árboles en Boosting

Tercera Parte

Capítulo 8. Predicción del fracaso empresarial

- 8.1. Introducción
- 8.2. Estado actual de la previsión del fracaso empresarial
- 8.3. Críticas y ventajas de los modelos de predicción del fracaso empresarial
- 8.4. Análisis descriptivo de las estadísticas sobre suspensiones de pagos, quiebras, fusiones y disoluciones

Capítulo 9. Predicción del fracaso empresarial mediante el método Bosting

- 9.1. Introducción
- 9.2. Algunas consideraciones teóricas sobre el tratamiento de la información

9.3. Recogida y tratamiento de la información

9.4. Aplicación del método Bosting a la previsión del fracaso empresarial

Capítulo 10. Conclusiones y líneas de investigación futura

10.1. Conclusiones

10.2. Líneas de investigación futura

Bibliografía

Apéndice A: Algoritmos de clasificación

Apéndice B: Salida de la función adaboost.M1

BIBLIOGRAFÍA RECOMENDADA

BREIMAN, L. [1998]: “Arcing classifiers”. *The Annals of Statistics*, vol. 26, 3, pp. 801-849.

FREUND, Y. Y SCHAPIRE, R.E. [1996]: “*Experiments with a new boosting algorithm*”. En Proc. 13th International Conference on Machine Learning, pp. 148-146. Morgan Kaufmann.

FREUND, Y. Y SCHAPIRE, R.E. [1997]: “A decision-theoretic generalization of on-line learning and an application to boosting”. *Journal of Computer and System Sciences*, 55[1], pp.119-139.

FRIEDMAN, J.; HASTIE, T. Y TIBSHIRANI, R. [2000]: “Additive logistic regression: a statistical view of boosting”. *Annals of Statistics*, 28, 2, pp. 337-407.

FRYDMAN, H.; ALTMAN, E. Y KAO, D. [1985]: “Introducing recursive partitioning for financial classification: the case of financial distress”. *Journal of Finance*, pp. 269-291.

GABÁS, F. [1997]: “Predicción de la insolvencia empresarial”. En Calvo-Flores, A. y García, D. [Coord.]: *Predicción de la insolvencia empresarial*, pp. 11-31. AECA.