

A Fuzzy Random Forest

Piero Bonissone, José M. Cadenas, M. Carmen Garrido, R.
Andrés Díaz-Valladares

International Journal of Approximate Reasoning, 2010

January 25, 2012

Outline

- 1 Introducción
- 2 Multiple Classifier systems and fuzzy logic
- 3 Fuzzy Random Forest
- 4 Combinar metodos en el FRF ensemble
- 5 Experimentos y Resultados
- 6 Conclusión

Introducción

- 1 Este articulo muestra una forma de clasificacion agrupando varios clasificadores.
- 2 La agrupacion es mediante la tecnica de Breiman conocida como Random Forest.
- 3 Esta demostrado que la combinacion de varios claificadores de forma adecuada obtiene mejores resultados que la aplicación de los clasificadores de forma individual.
- 4 Cuando entre los datos hay inconsistencias, faltan datos o hay mucho ruido la logica fuzzy ayuda a gestionar estos datos de forma adecuada.

Introduccion II

Este articulo pretende mostrar el uso de fuzzy logic combinado con Random forest para generar un clasificador multiple que:

- Utilice la robustez de los multclasificadores.
- La aleatoriedad de los Random Forest.
- La flexibilidad de la fuzzy logic contra datos imperfectos.

Multiple Classifier systems

Existen muchos sistemas de clasificación múltiple, entre todos ellos nos vamos a centrar en los basados en árboles de decisión.

- 1 Bagging: combina la decisión de los clasificadores usados dándoles a todos el mismo peso.
- 2 Boosting Algorithm: añade clasificadores uno tras otro de forma que clasifican sobre datos previamente clasificados.
- 3 Ho's random subspaces technique: usa subconjuntos de datos para entrenar los clasificadores.
- 4 Randomization: elige uno entre los 20 mejores atributos cada vez que se divide una rama del árbol.
- 5 Breiman: algoritmo Random Forest, que mezcla el algoritmo de Bagging con la randomización.
- 6 Banfield: comparan estas técnicas y calculan el promedio de ir aplicando cada una de las técnicas anteriores.

Metodos combinados

La combinacion de metodos para formar el multclasificador puede ser:

- No entrenable: Los clasificadores se entrenan por separado, al unirlos no necesitan mas entrenamiento.
- Entrenables: necesitan entrenamiento tras su unión. Pueden ser:
 - implícitos: se entrenan con datos de test.
 - explícitos: se entrenan con datos del problema.

Fuzzy Logic

- Los arboles de decisión son inestables cuando hay pequeñas perturbaciones en los datos. Ahí es donde entra la logica fuzzy.
- Se ha construido un multclasificador con Fuzzy logic y Random Forest, ya que se ha demostrado que la fuzzy logic y los arboles de decision son muy buenos con datos incompletos y con mucho ruido, y entre ellos los Random Forest porque son mas resistentes al ruido.

Fuzzy Random Forest (FRF)

- Los random forest de Breiman, cada arbol tienen el tamaño máximo¿?
- Durante el proceso de construcción cuando un nodo debe dividirse solo se elige un subconjunto de atributos de forma aleatoria. Para cada nodo.
- El tamaño de cada subconjunto es el único parámetro importante.
- El mejor atributo puede no ser seleccionado para un “split”, pero quizás si para otro.
- Hay dos elementos aleatorios:
 - La selección del conjunto de datos para cada arbol.
 - Los atributos seleccionados en cada “split”.

Para construir los arboles tenemos dos algoritmos:

- 1 Generar Random Forest con arboles de decision fuzzy.
- 2 Generar arboles de decision con una parte aleatoria de los atributos.

Los arboles de decision del algoritmo 1 se construyen mediante el algoritmo 2.

FRF strategy

Tenemos dos estrategias para clasificar:

- 1 Agrupar las decisiones de las hojas de cada uno de los arboles y despues agrupar las decisiones de los arboles usando el mismo u otro método de clasificación.
- 2 Combinar la información de todas las hojas de todos los arboles, generando una decisión global.

Lo que da pie a generar dos algoritmos, uno para la estrategia 1 o otro para la estrategia 2.

FRF ensemble

En esta sección se enumeran gran cantidad de metodos y estartegias que los autores han implementado para respaldar su trabajo.

A pesar de que la idea general del articulo esta bastante clara, me pierdo entre tanto metodo. Si fuera a usar FRF (que nunca se sabe) podria ser interesante sumergirse dentro de ellos.

Experimentos

- Se han elaborado experimentos para medir el comportamiento y estabilidad del FRF con:
 - datos imperfectos
 - Valores que faltan
 - Valores difusos
 - ruido
 - Ruido
 - Valores atípicos
- Comparación con otros clasificadores y ensembles.

Datos y parametros

- Se han usado conjuntos de datos de la UCI
- Se han usado test noparametricos para validar los resultados:
 - Wilcoxon signed-rank test para comparaciones entre dos metodos.
 - Friedman test y Benjamin-Hochberger procedure como test a posteriori para comparar multiples metodos.

Resultados

A continuación el artículo enumera los experimentos realizados y los resultados obtenidos. Al igual que el apartado 4, la información es abundante y requiere un estudio en profundidad para una mejor comprensión de la misma.

Según el abstract el método FRF es el mejor en todos los casos y esta sección lo corrobora, pero también se dice que en escenarios fuzzy con falta de datos y mucho ruido este método no mejora a otros métodos de clasificación múltiple, algo que parece corroborar las tablas de resultados pero que no queda bien explicado en el texto.

Conclusiones

En este apartado no explico las conclusiones de los autores sino las mias propias al leer este artículo.

- El articulo tiene algun error (una frase esta en castellano), partes se repiten dos incluso tres veces y es algo que se supone que cuidan mucho los revisores.
- Exponen un nuevo metodo combinando Fuzzy logic con Random Forest, pero ellos no han creado Ramdon Forest, ni Fuzzy logic y ademas dicen que fuzzy logic con random forest ya ha sido usado, referencias 21,22,23,26,29.
- Como parte positiva han generado muchas combinaciones de metodos en el ensemble FRF y han realizado muchos test.
- El uso de Fuzzy es para datos inconsistentes o con ruido pero experimentalmente es precisamente ahi donde este metodo no demuestra que sea mejor que los demas.