

A Fuzzy Random Forest

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

International Journal of Approximate Reasoning, 2010

January 26, 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.

Introducción 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 imcompletos y con mucho ruido, y entre ello 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”.

FRF

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 estategias que los autores han implementado para respaldar su trabajo.

Tenemos dos estrategias Strategy1 y Strategy2. Se aplican Ambas estategas para cada combinacion de metodos.

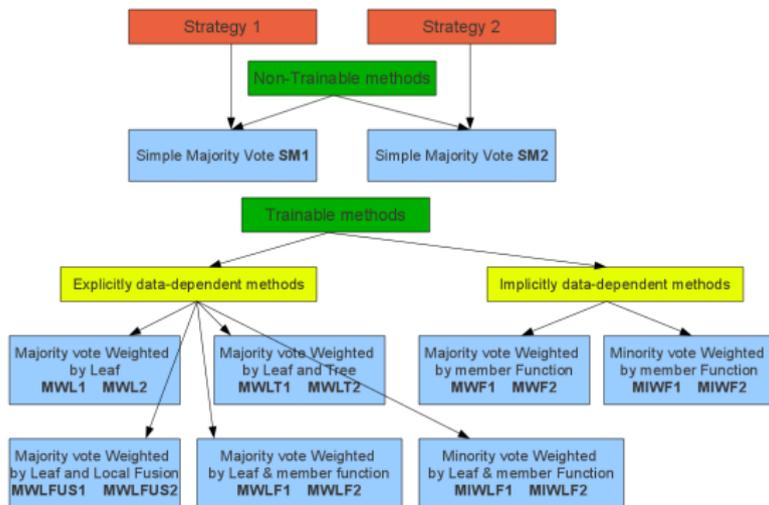
Tenemos metodos que no necesitan entrenamiento y metodos que si necesitan entrenamiento.

Entre los que si necesitan entrenamiento. Tenemos los que necesitan entranmiento sobre los datos a clasificar y los que no necesitan entrenar sobre los datos a clasificar.

Las clasificaciones se evaluan mediante mayoria o minoria en votaciones con pesos en los clasificadores.

Metodos en el FRF

Figure: Esquema de los metodos



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 porteriori para comparar multiples metodos.

Table: Datasets. E:ejemplos, M:atributos, l:clases

Dataset	Abbr	E	M	l
Appendicitis	APE	106	7	2
Wisconsin breast C.	BCW	683	9	2
BUPA liver disorders	BLD	345	6	2
Contraceptive method	CMC	1473	9	3
German credit	GER	1000	24	2
Glass	GLA	214	9	7
Statlog heart disease	HEA	270	13	2
Ionosphere	ION	351	34	2
Iris plants	IRP	150	4	3
Pima indian diabetes	PID	532	7	2
Pima indian diabetes	PIM	768	8	2
Image segmentation	SEG	2130	19	7
Attitude smoking	SMO	2855	8	3
Thyroid disease	THY	7200	21	3
Vehicle	VEH	846	18	4
Wine	WIN	178	13	3

Resultados 1

Table: Test accuracy of FRF diferente porcentajes of missing values

Dataset	Introducing missing values			
	Without	5%	15%	30%
	% Decrease average accuracy			
APE	91.13 _(8,70) MWFI,MWLF1	0.82 _{MWFI,MWLF1}	1.03 _{MWLT2,MWIF2}	0.21 _{MWFI,MWLF1}
BCW	97.31 _(1,76) MWLT1,MWLT2	0.12 _{MWLT2,MWLT2}	0.79 _{MWLT2}	2.92 _{MWLT2}
GER	76.68 _(3,97) MWIF2	0.70 _{MWLT2}	3.86 _{MWIFUS2}	5.16 _{MWIFUS2}
GLA	77.66 _(7,36) MWLT2	6.62 _{MWIFUS2}	10.95 _{MWLT2}	17.20 _{MWLT2}
ION	96.41 _(2,89) MWLF1	0.94 _{MWFI,MWLF1}	2.66 _{MWIFUS2}	6.09 _{MWFI,MWLF1}
IRP	97.33 _(3,93) **	1.23 _{M1(M2),MWFI,MWIF2}	4.11 _{MWFI}	16.71 _{MWLT2}
PIM	77.14 _(4,88) MWFI	0.82 _{MWFI}	2.57 _{MWFI}	7.47 _{MWLT2}
WIN	97.87 _(2,85) MWLF1	4.41 _{MWFI,MWLF1}	6.27 _{MWLT2}	14.21 _{MWFI,MWLF1}

Table: Test accuracy of FRF different percentages of fuzzy values

Dataset	Introducing fuzzy values			
	Without	5%	15%	30%
	% Decrease average accuracy			
APE	91.13 _(8,70) MWFI,MWLF1	-0.21 _{MWFI,MWLF1}	0.82 _{MWFI,MWLF1,MWLT1,MWIF2}	0.21 _{MWFI,MWLF1}
BCW	97.31 _(1,76) MWLT1,MWLT2	0.42 _{MWLT1,MWLT2,MWIFUS1,MWIFUS2}	0.28 _{MWLT1,MWLT2}	0.12 _{MWLT1,MWLT2}
GER	76.68 _(3,97) MWIF2	0.08 _{MWFI,MWLF1}	-0.29 _{MWFI}	-0.26 _{MWIF2}
GLA	77.66 _(7,36) MWLT2	-1.08 _{MWIF2}	0.00 _{M2}	1.31 _{MWIF2}
ION	96.41 _(2,89) MWLF1	0.53 _{MWIF1}	0.46 _{MWIF2}	0.94 _{MWFI,MWLF1}
IRP	97.33 _(3,93) **	0.00 **	0.00 **	0.00 **
PIM	77.14 _(4,88) MWFI	0.52 _{MWFI}	0.17 _{MWFI}	1.49 _{M1}
WIN	97.87 _(2,85) MWLF1	-0.29 _{MWLT2,MWLT2,MWIFUS2,MWIF2}	-0.51 _{MWFI,MWIF2}	0.75 _{MWIF2}

Resultados 2

Table: Test accuracy of FRF missing and fuzzy values

Dataset	Introducing missing values and fuzzy values			
	Without	5%	15%	30%
		% Decrease average accuracy		
APE	91.13 _(6.76) MW1/MW2	0.00 _(0.00) MW1	1.03 _(0.01) MW1	0.21 _(0.02)
BCW	97.31 _(1.76) MW1/MW2	0.76 _(0.01) MW1	0.31 _(0.01) MW1	0.67 _(0.01)
GER	76.68 _(1.07) MW2	0.76 _(0.01)	1.10 _(0.01) MW1	2.50 _(0.01) MW1
GLA	77.66 _(7.36) MW2	0.59 _(0.02)	5.90 _(0.02)	10.11 _(0.01) MW1
ION	96.41 _(2.89) MW1	1.18 _(0.02)	2.83 _(0.01) MW1	3.25 _(0.01) MW1
IRP	97.33 _(2.14) -	1.91 ₍₋₎	4.11 _(0.01) MW1/MW2/MW1	6.99 _(0.02)
IRM	77.14 _(4.88) MW1	1.11 _(0.01)	1.19 _(0.01)	1.36 _(0.01)
WIN	97.87 _(2.89) MW1	0.29 _(0.01)	4.89 _(0.02)	7.07 _(0.01) MW1

Table: Effect for different types of outliers on FRF

Dataset	Outliers			
	Without	Obtained with k_1	Obtained with k_2	Obtained with k_3
		% Increase average error		
BCW	97.30 _(1.48) MW1/MW1/MW2	-1.11 _(0.01) MW1	-5.19 _(0.01) MW1/MW2	-3.70 _(0.01) MW1/MW2
BLD	72.97 _(1.91) MW2	-3.22 _(0.02) MW2	-2.70 _(0.02)	1.59 _(0.02)
CMC	53.62 _(2.00) MW1	0.19 _(0.01)	0.30 _(0.01)	-0.11 _(0.01)
GLA	78.38 _(6.31) MW2	1.57 _(0.02)	3.75 _(0.02)	1.06 _(0.02)
HEA	82.87 _(3.30) MW1	-4.32 _(0.01)	-1.11 _(0.02)	-4.32 _(0.01)
ION	94.66 _(2.19) MW1	6.93 _(0.01)	6.74 _(0.01)	5.43 _(0.01)
IRP	97.33 _(2.14) -	0.00 ₍₋₎	0.00 ₍₋₎	0.00 ₍₋₎
IRD	79.61 _(3.27) MW2	1.62 _(0.02)	0.49 _(0.01)	2.11 _(0.02)
IRM	76.53 _(1.86) MW1	3.07 _(0.01)	1.79 _(0.01)	-0.13 _(0.01)
SEG	97.19 _(1.97) MW1/MW1	2.83 _(0.01)	3.91 _(0.01)	3.20 _(0.01)
SMO	69.54 _(1.97) MW1/MW1	0.03 _(0.01) MW1	0.03 _(0.01) MW1	0.03 _(0.01) MW1
VEH	75.18 _(1.81) MW2	0.60 _(0.01)	1.93 _(0.02)	1.33 _(0.02)
WIN	97.48 _(3.23) MW1	17.06 _(0.01)	11.51 _(0.01)	11.51 _(0.01)

Resultados 3

Table: Comparison of tree-based ensembles and classifiers with noise data

Dataset	% Increase error Hamza et al. [18]	Best classification method Hamza et al. [18]	% Increase error (FRF)	Best combination method (FRF)
BCW	54.84	ST-NLC-G	4.54	MWLFUS2
BLD	16.98	BA-NLC-G	6.78	MWLFUS2
CMC	-6.23	ST-NLC-G	0.57	MWLFUS2
HEA	-2.08	RF	-2.44	MWFF1
PD	5.74	RF	4.42	MWLFUS2
SEG	115.56	RF	81.36	MWFF2
SMD	3.07	BA-NLC-G	0.02	MWFF2,MWLFUS2
THY	23.08	RF	18.19	MWFF2
VEH	-0.76	ST-NLC-G	-0.95	MWFF1

Table: Test average accuracy of FRF with ensembles same base classifier

Dataset	Size RF	RF	fuzzy tree (FT)	Size ensembles	Boosting with FT	Bagging with FT	FRF ensemble
BCW	125	97.07(1.80)	95.50(1.75)	125	94.51(1.51)	95.68(1.60)	97.30(1.40)
BLD	200	72.68(6.32)	65.14(6.75)	200	65.79(6.23)	71.88(5.80)	72.97(5.19)
CMC	120	51.41(2.01)	47.17(2.62)	120	49.08(3.00)	51.49(2.04)	53.62(3.00)
GLA	120	78.85(6.05)	72.43(6.34)	50	74.89(6.26)	76.74(5.72)	78.38(6.11)
HEA	120	81.48(4.13)	74.26(6.58)	120	77.13(4.03)	81.02(5.13)	82.87(5.29)
ION	175	93.45(2.35)	92.50(3.28)	175	94.09(3.89)	93.25(3.12)	94.66(3.19)
RP	120	95.33(1.74)	97.00(2.00)	120	96.67(2.53)	96.67(2.53)	97.33(2.14)
PD	125	76.41(2.12)	71.85(3.60)	50	70.54(3.60)	78.05(3.19)	79.61(3.77)
HM	150	75.26(2.51)	67.55(4.53)	150	66.18(3.54)	73.63(2.96)	76.53(3.65)
SEG	140	97.85(0.80)	95.54(1.06)	140	96.54(0.77)	97.19(0.60)	97.19(0.40)
SMD	100	61.63(0.82)	55.29(2.42)	75	56.36(1.48)	69.50(1.91)	69.54(1.97)
THY	150	99.67(0.10)	96.13(0.53)	150	96.26(0.62)	98.25(0.38)	99.17(0.22)
VEH	200	76.27(2.73)	67.96(3.84)	200	70.06(3.04)	74.41(3.04)	75.18(3.01)
VIN	150	98.03(1.93)	97.19(2.35)	150	97.20(1.45)	97.06(1.36)	97.48(1.23)

Resultados 4

Table: Comparison accuracies of FRF and other classifiers

Dataset	Technique	GRA	CIGRA	MLP	C4.5	RBF	Bayes	Cart	GBLM	Fuzzy D.Tree
	FRF									
APE	91.04 _(0,12) MWLF1/MWLF1	86.00	88.70	85.80	84.90	80.20	83.00	84.90	-	-
BCW	97.14 _(1,78) MWLF1/MWLF2	96.20	96.80	96.50	94.70	96.60	96.40	94.40	96.70	96.80
CER	76.41 _(3,96) MWLF2	73.00	74.20	71.60	73.50	75.70	70.40	73.90	-	-
CLA	76.82 _(7,85) MWLF2	57.40	63.20	68.70	65.80	46.70	71.80	63.60	65.40	66.00
ION	96.13 _(2,99) MWLF1	88.50	92.60	92.00	90.90	94.60	85.50	89.50	-	86.50
IRP	97.33 _(4,58) **	95.70	96.10	96.00	94.00	98.00	94.70	92.00	94.70	96.10
PIM	76.70 _(4,34) MWLF1	74.90	76.20	75.80	72.70	72.20	74.70	75.80	73.10	73.10
WIN	97.64 _(3,19) MWLF1	93.30	96.20	98.30	93.30	94.90	94.40	87.60	95.10	91.20

Table: Comparison with tree-based ensembles

	Dataset								
	BCW	BLD	CMC	HEA	PID	SEG	SMO	THY	VEH
Best classification error [18]	2.64	25.80	47.39	17.78	22.93	1.60	34.05	0.28	23.40
Best classification error method [18]	RF	RF	RF	RF	RF	RF	BO(100)WLC-E	BO(100)WLC-E	BO(250)WLC-G
Best classification error - FRF	2.49	24.67	46.57	14.44	19.35	2.55	30.47	0.78	23.17
Best classification error combination method	MWLT1/MWLT2/ MWLFUS1/ MWLFUS2	MWLFUS1	MWLF1	MWLF1	MWLFUS2	MWF2	MWLT2 MWLFUS2	MWLF1/ MWLF1	MWF2

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.