

FOIL Propositionnel

Apprendre-concept (Pos, Neg) (C le concept est donné)

Début

Règles \leftarrow vide

Tant que Pos $\neq \emptyset$

 Conditions_Règle \leftarrow vide

 Neg2 \leftarrow Neg

 Pos2 \leftarrow Pos

 Tant que Neg2 $\neq \emptyset$

 Choisir le littéral L qui maximise Gain(L, Pos2, Neg2)

 Ajouter L à Conditions_Règle

 Retirer de Neg2 tous les exemples qui ne satisfont pas L

 Retirer de Pos2 tous les exemples qui ne satisfont pas L

 Fintantque

 Ajouter à Règles la règle (C \leftarrow Conditions_Règle)

 Retirer de Pos tous les exemples qui satisfont Conditions_Règle

Fintantque

Retourner l'ensemble Règles

Fin

La fonction Gain(L, Pos, Neg) est définie de la manière suivante :

Soit P le nombre d'exemples dans Pos et N le nombre d'exemples dans Neg

Soit p le nombre d'exemples dans Pos qui satisfont L et n le nombre d'exemples dans Neg qui satisfont L

Gain retourne $(p \cdot (\log(p/(p+n)) - \log(P/(P+N))))$ (log est la fonction \log_2)

Cas	Prévisions	Température	Humidité	Vent	Sport
E1	Soleil	Chaude	Elevée	Faible	NON
E2	Soleil	Chaude	Elevée	Fort	NON
E3	Couvert	Chaude	Elevée	Faible	OUI
E4	Pluie	Douce	Elevée	Faible	OUI
E5	Pluie	Froide	Normale	Faible	OUI
E6	Pluie	Froide	Normale	Fort	NON
E7	Couvert	Froide	Normale	Fort	OUI
E8	Soleil	Douce	Elevée	Faible	NON
E9	Soleil	Froide	Normale	Faible	OUI
E10	Pluie	Douce	Normale	Faible	OUI
E11	Soleil	Douce	Normale	Fort	OUI
E12	Couvert	Douce	Elevée	Fort	OUI
E13	Couvert	Chaude	Normale	Faible	OUI
E14	Pluie	Douce	Elevée	Fort	NON

Jeu d'exemples.

Exécution de l'algorithme sur ce jeu d'exemples

Recherche de la 1ère règle **R1 : Sport=OUI** ← *Conditions_Règle*

P = 9 N = 5

10 littéraux candidats

Littéral	p	n	Gain
Prévisions = Soleil	2	3	<0
Prévisions = Couvert	4	0	2,56
Prévisions = Pluie	3	2	<0
Température = Chaude	2	2	<0
Température = Douce	4	2	0,24
Température = Froide	3	1	0,66
Humidité = Elevée	3	4	<0
Humidité = Normale	6	1	2,52
Vent = Fort	3	3	<0
Vent = Faible	6	2	1,32

Neg2 = ∅ donc on a trouvé une règle

R1 : Sport=OUI ← Prévisions = Couvert (4 exemples couverts : E3, E7, E12, E13)

On met à jour Pos = {E1, E2, E4, E5, E6, E8, E9, E10, E11, E14}

Cas	Prévisions	Température	Humidité	Vent	Sport
E1	Soleil	Chaude	Elevée	Faible	NON
E2	Soleil	Chaude	Elevée	Fort	NON
E4	Pluie	Douce	Elevée	Faible	OUI
E5	Pluie	Froide	Normale	Faible	OUI
E6	Pluie	Froide	Normale	Fort	NON
E8	Soleil	Douce	Elevée	Faible	NON
E9	Soleil	Froide	Normale	Faible	OUI
E10	Pluie	Douce	Normale	Faible	OUI
E11	Soleil	Douce	Normale	Fort	OUI
E14	Pluie	Douce	Elevée	Fort	NON

Recherche de la 2e règle **R2 : Sport=OUI ← Conditions_Règle**

P = 5 N = 5

10 littéraux candidats

Littéral	p	n	Gain
Prévisions = Soleil	2	3	< Gain de Prévisions=Pluie
Prévisions = Couvert	0	0	-infini
Prévisions = Pluie	3	2	0,78
Température = Chaude	0	2	-infini
Température = Douce	3	2	0,78
Température = Froide	2	1	0,84
Humidité = Elevée	1	4	< Gain de Vent = Fort
Humidité = Normale	4	1	2,72
Vent = Fort	1	3	< Gain de Prévisions=Soleil
Vent = Faible	4	2	1,68

R2 : Sport=OUI ← Humidité = Normale

Pos2 = {E5, E9, E10, E11} Neg2 = {E6}

Littéraux candidats

Je ne vais pas donner le résultat du calcul de gain. En regardant les valeurs de p et n on peut décider quel sera le meilleur gain (ligne grisée)

Littéral	p	n	Gain
Prévisions = Soleil	2	0	
Prévisions = Couvert	0	0	-infini
Prévisions = Pluie	2	1	
Température = Chaude	0	0	-infini
Température = Douce	2	0	
Température = Froide	2	1	0,84
Vent = Fort	1	1	
Vent = Faible	3	0	

Neg2 = \emptyset donc on a trouvé une règle

R2 : Sport=OUI \leftarrow Humidité = Normale ET Vent = Faible (3 exemples couverts : E5, E9, E10)

Cas	Prévisions	Température	Humidité	Vent	Sport
E1	Soleil	Chaude	Elevée	Faible	NON
E2	Soleil	Chaude	Elevée	Fort	NON
E4	Pluie	Douce	Elevée	Faible	OUI
E6	Pluie	Froide	Normale	Fort	NON
E8	Soleil	Douce	Elevée	Faible	NON
E11	Soleil	Douce	Normale	Fort	OUI
E14	Pluie	Douce	Elevée	Fort	NON

On met à jour Pos = {E4, E11}

Recherche de la 3e règle **R3 : Sport=OUI \leftarrow Conditions_Règle**

P=2 N=5

Littéraux candidats

Littéral	p	n	Gain
Prévisions = Soleil	1	3	
Prévisions = Couvert	0	0	-infini
Prévisions = Pluie	1	2	
Température = Chaude	0	2	-infini
Température = Douce	2	2	
Température = Froide	0	1	-infini
Humidité = Elevée	1	4	
Humidité = Normale	1	1	
Vent = Fort	1	3	

Littéral	p	n	Gain
Vent = Faible	1	2	

R3 : Sport=OUI ← *Température = Douce*

Pos2 = {E4,E11} Neg2 = {E8, E14}

Cas	Prévisions	Température	Humidité	Vent	Sport
E4	Pluie	Douce	Elevée	Faible	OUI
E8	Soleil	Douce	Elevée	Faible	NON
E11	Soleil	Douce	Normale	Fort	OUI
E14	Pluie	Douce	Elevée	Fort	NON

Littéral	p	n	Gain
Prévisions = Soleil	1	1	
Prévisions = Couvert	0	0	-infini
Prévisions = Pluie	1	1	
Humidité = Elevée	1	2	
Humidité = Normale	1	0	
Vent = Fort	1	1	
Vent = Faible	1	1	

R3 : Sport=OUI ← *Température = Douce ET Humidité = Normale (1 exemple couvert E11)*

On met à jour Pos = {E4}

P=1 N=5

Cas	Prévisions	Température	Humidité	Vent	Sport
E1	Soleil	Chaude	Elevée	Faible	NON
E2	Soleil	Chaude	Elevée	Fort	NON
E4	Pluie	Douce	Elevée	Faible	OUI
E6	Pluie	Froide	Normale	Fort	NON
E8	Soleil	Douce	Elevée	Faible	NON
E14	Pluie	Douce	Elevée	Fort	NON

Littéral	p	n	Gain
Prévisions = Soleil	0	3	-infini
Prévisions = Couvert	0	0	-infini
Prévisions = Pluie	1	2	
Température = Chaude	0	2	-infini
Température = Douce	1	2	
Température = Froide	0	1	-infini
Humidité = Elevée	1	4	
Humidité = Normale	0	1	
Vent = Fort	0	3	
Vent = Faible	1	2	

On peut choisir n'importe lequel de ces littéraux (donc plusieurs solutions

R4 : Sport=OUI ← Prévisions = Pluie

Pos2 = {E4} Neg2={E6,E14}

et on obtient

R4 : Sport=OUI ← Prévisions = Pluie ET Vent = Faible

d'où les 4 règles

R1 : Sport=OUI ← Prévisions = Couvert (4 exemples couverts : E3, E7, E12, E13}

R2 : Sport=OUI ← Humidité = Normale ET Vent = Faible (3 exemples couverts : E5, E9, E10}

R3 : Sport=OUI ← Température = Douce ET Humidité = Normale (1 exemple couvert E11)

R4 : Sport=OUI ← Prévisions = Pluie ET Vent = Faible (1 ex couvert E4)

On peut pour un jeu de données trouver différentes solutions suivant la stratégie adoptée pour choisir le « meilleur » littéral quand plusieurs candidats ont le même score : choix du premier rencontré (mais dépend de l'ordre dans lequel vous examinez vos candidats), choix aléatoire, ...