

KDD E MINERAÇÃO DE DADOS

Métodos de Mineração de Dados Baseados em Algoritmos Genéticos

Prof. Ronaldo R. Goldschmidt

ronaldo@de9.ime.eb.br

rribeiro@univercidade.br

geocities.yahoo.com.br/ronaldo_goldschmidt

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Inspiração e Motivação

Identificar padrões de empresas em um BD

Exemplo: Qual o padrão das empresas do Cluster 1?

Se receita_serviço 1 (Voz) = 5000<R\$<7000 &
receita_serviço 2 (Telex) = 7000<R\$<8000 &
código_atividade = 13 (Ind. Mat.Elétrico Eletrônico e
Comunicação) &
10 < #_Filiais < 50 &
.....
#_Empregados > 100

Então Empresa pertence ao Cluster 1(In-Transaction)

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Cromossoma $\equiv$ Regra

Genes $\equiv$ atributos do banco de dados

cruzamento

P_1	Receita Serviço 1 1000<R\$<2000	Receita Serviço 2 4000<R\$<9000	COD_ATIV = 13	10<#_Filiais<50	Empregados>100
P_2	Receita Serviço 1 5000<R\$<7000	Receita Serviço 2 7000<R\$<8000	COD_ATIV = 14	30<#_Filiais<60	Empregados>300
F_1	Receita Serviço 1 1000<R\$<2000	Receita Serviço 2 4000<R\$<9000	COD_ATIV = 14	30<#_Filiais<60	Empregados>300
F_2	Receita Serviço 1 5000<R\$<7000	Receita Serviço 2 7000<R\$<8000	COD_ATIV = 13	10<#_Filiais<50	Empregados>100

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Regras – Forma Geral

SE C_1 E C_2 E ... E C_n ENTÃO CONCLUSÃO

Onde

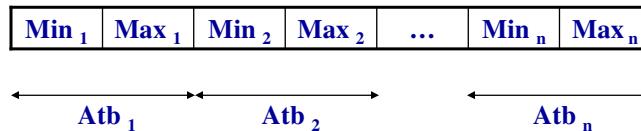
C_i , $i=1,...,n$ são condições envolvendo **atributos preditivos**

Conclusão – envolve um **atributo objetivo** com valor fixo, não manipulado pelo AG

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Representação do Cromossoma



Onde Atb_i pode ser:

- Quantitativo – Faixa de valores: [Min_i,Max_i]
- Categórico – Contém um único símbolo codificado em Min_i

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operadores de Crossover Disponíveis

- Um Ponto
- Dois Pontos
- Uniforme
- Média
- Aritmético
- Lógico
- Híbrido

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Op. Lógico OU p/ Atributos Categóricos

- Representação Real
- Interpretação do valor de um atributo categórico (alfabeto de cardinalidade k):
 - Real → Binário com k bits
 - Cada posição indica ausência (0) ou presença (1) do símbolo correspondente

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Op. Lógico OU p/ Atributos Categóricos

Ex: Tipo Residência: Dom. = {própria, alugada, parente, funcional}

Alelo	Decodificação	Tipo Res
1	0001	própria
2	0010	alugada
3	0011	própria ou alugada
...	...	...
15	1111	própria ou alugada ou parente ou funcional (don't care)
0	0000	Não informada (Null)

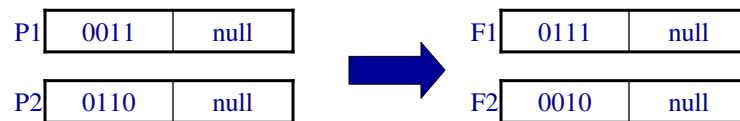
MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operador de Crossover Lógico (E/OU)

$$F_1 = P_1 \text{ OU } P_2$$

$$F_2 = P_1 \text{ E } P_2$$



MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operador de Crossover Híbrido

- O Crossover Aritmético para Atributos Quantitativos
- O Crossover Lógico proposto para Atributos Categóricos

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operador de Mutação Disponíveis

- Mutação Simples
- Mutação Don't Care
- Mutação Creep
- Mutação Lógica
- Mutação Híbrida
- Mutação de Extensão Lógica

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operador de Mutação Lógica (Negação)

$I = \text{NOT } I$



MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operador de Mutação Híbrida

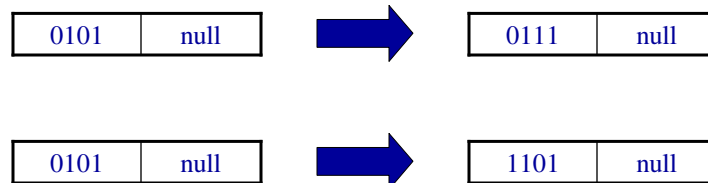
- A Mutação CREEP para Atributos Quantitativos
- A Mutação Lógica para Atributos Categóricos

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Operador de Mutação de Extensão Lógica

- Inclusão de valores de atributos categóricos na regra
- Parâmetro: escopo da extensão
- Sorteio do sentido da aplicação



MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Funções de Avaliação: FAbrangência e FAcurácia

Seja R uma Regra qualquer:

SE P ENTÃO O

Define-se:

- Acurácia de R : $|P \wedge O| / |P|$
- Abrangência de R : $|P \wedge O| / |O|$

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Funções de Avaliação: FAbrangência e FAcurácia - Exemplo:

- BD contém 100 Registros
- Registros estão segmentados em 2 Grupos:
 - 80 regs. do G1 e 20 regs. do G2
- Procura-se regras (C) para G1 (P = pertence a G1)

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver

Funções de Avaliação: FAbangência e Facurácia - Exemplo:

- Uma determinada regra encontrada, resulta em:
 - 60 Registros satisfazem a regra e são do G1 (C&P)
 - 20 Registros do G1 não satisfazem a regra (C&P)
 - 12 Registros do G2 satisfazem a regra (C&P)

$$Ac = \frac{|60|}{(|60| + |12|)} = 0.833$$

$$Ab = \frac{|60|}{(|60| + |20|)} = 0.75$$

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver – Exemplo de Aplicação

- BD ZOO com 101 registros distribuídos em 7 classes

Classe	# Reg	# Reg Treino	# Reg Teste
1	41	20	21
2	20	10	10
3	05	3	2
4	13	7	6
5	04	2	2
6	08	4	4
7	10	5	5

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver – Exemplo de Aplicação

Estrutura do Banco de Dados ZOO

- | | |
|----------------------|------------------|
| • Venomous (y/n) | • Eggs (y/n) |
| • Toothed (y/n) | • Domestic (y/n) |
| • Tail (y/n) | • Catsize (y/n) |
| • Predator (y/n) | • Breathes (y/n) |
| • Milk (y/n) | • Backbone (y/n) |
| • Legs (0,2,4,5,6,8) | • Aquatic (y/n) |
| • Hair (y/n) | • Airborne (y/n) |
| • Fins (y/n) | • Type (1 a 7) |
| • Feathers (y/n) | |

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver – Exemplo de Aplicação

Metodologia de Avaliação

- a) Identificar a melhor função de avaliação para cada classe e obter as regras descritoras de cada classe
- b) Aplicar as regras obtidas sobre o BD de testes

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver – Exemplo de Aplicação

- Parametrização do GA utilizada nos experimentos:
 - a) Inicialização aleatória da população
 - b) Pop Size = 100; Gerações = 50; Rodadas = 5
 - c) Critério de Parada do GA: Evolução das 20 Melhores Regras.
 - d) Critério para Inclusão de Regra no BR (Tolerância da melhor Regra) : 50 gerações sem sofrer alterações.
 - e) Técnica de Elitismo e Normalização Linear

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver – Exemplo de Aplicação

CLASSE	TEXTO DAS REGRAS OBTIDAS DO BD DE TREINO
1	IF (LEGS>= 1.05 AND LEGS<= 5.52) AND HAIR=1 AND FEATHERS=0 AND BACKBONE=1 AND AQUATIC=0 THEN TYPE= 1
2	IF MILK=0 AND (LEGS>= 0.74 AND LEGS<= 2.82) THEN TYPE=2
3	IF PREDATOR= 1 AND (LEGS>= 0 AND LEGS<= 1.79) AND FINS= 0 AND DOMESTIC= 0 AND BACKBONE= 1 THEN TYPE= 3
4	IF (LEGS>= 0 AND LEGS<= 1.74) AND FEATHERS=0 AND EGGS=1 AND AQUATIC=1 THEN TYPE= 4
5	IF TOOTHED=1 AND PREDATOR=1 AND MILK=0 AND (LEGS>= 3.36 AND LEGS<= 5.6) AND EGGS=1 AND CATSIZE=0 AND AQUATIC=1 AND AIRBORNE=0 THEN TYPE= 5
6	IF TAIL=0 AND (LEGS>= 4.73 AND LEGS<= 6.76) AND EGGS=1 AND AQUATIC=0 THEN TYPE= 6
7	IF VENOMOUS=0 AND TOOTHED=0 AND (LEGS>= 2.09 AND LEGS<= 6.2) AND HAIR=0 AND FINS=0 AND FEATHERS=0 AND EGGS=1 AND DOMESTIC=0 AND CATSIZE=0 AND BREATHE=0 AND BACKBONE=0 THEN TYPE= 7

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Rule Evolver – Exercícios

- 1 - Utilize o software Rule Evolver II para evoluir regras a partir do banco de dados de animais (ZOO)
- 2 - Utilize o mesmo software para evoluir regras a partir do banco de dados de doenças cardíacas (Heart Disease)

MÉTODOS BASEADOS EM ALGORITMOS GENÉTICOS

Evolver

- Plug-in do Excel
- Diversos exemplos ilustrativos de aplicações
- Exemplo de Aplicação em Clusterização