

Inteligência Artificial

Guilherme Henrique de Souza Nakahata

Universidade Estadual do Paraná - Unespar

09 de Maio de 2024

- Algoritmo C4.5;
- Melhoras ao ID3:
 - Tratamento de atributos numéricos;
 - Razão de ganho;
 - Valores ausentes;
 - Estimativa de probabilidade;
 - Poda de árvore (Minimizar o superajuste).

- Restrito a criar nós com dois filhos;
 - Atributos $\leq x$;
 - Atributos $> x$.
- Criação de partições de forma supervisionada;
- Exemplo:

Temp.:	64	65	68	69	70	71	72	75	80	81	83	85
Classe:	yes	no	yes	yes	yes	no	no yes	yes yes	no	yes	yes	no

Atributos numéricos

Outlook	Temperature	Humidity	Windy	Play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

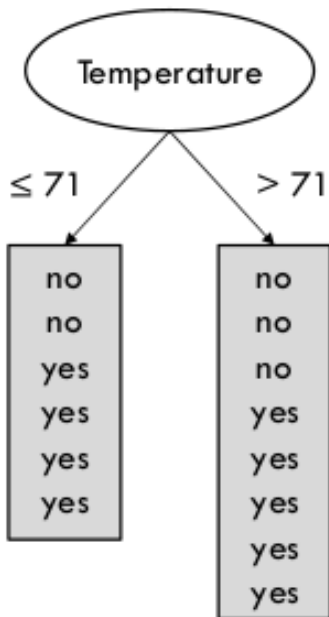
Atributos numéricos

Temp.:	64	65	68	69	70	71	72	75	80	81	83	85
Classe:	yes	no	yes	yes	yes	no	no yes	yes yes	no	yes	yes	no

$$S_{\leq 71} = [4_{yes}, 2_{no}] \rightarrow Entropia(S_{\leq 71}) = 0,918$$

$$S_{> 71} = [5_{yes}, 3_{no}] \rightarrow Entropia(S_{> 71}) = 0,954$$

$$Entropia(\text{temperature} > 71) = \frac{6}{14} \times Entropia(S_{\leq 71}) + \frac{8}{14} \times Entropia(S_{> 71}) = 0,939$$



- ID3 \rightarrow ganho de informação;
- Favorece atributos com muitos valores possíveis;
- Alta ramificação;
- Atributos com mais valores possíveis;
- Maior probabilidade de gerar filhos puros;
- Exemplo "id";
- **Razão de ganho (GR).**

$$Entropia(S, A) = \sum_{v \in \{a_1 \dots a_n\}} -\frac{|S_v|}{|S|} \times \log_2 \frac{|S_v|}{|S|}$$

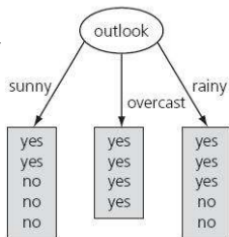
$$GR(S, A) = \frac{IG(S, A)}{Entropia(S, A)}$$

Razão de ganho

Para o atributo outlook:

$$Entropia(S, outlook) = -\frac{5}{14} \log_2 \frac{5}{14} - \frac{4}{14} \log_2 \frac{4}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 1,577$$

$$GR(S, outlook) = \frac{IG(S, outlook)}{Entropia(S, outlook)} = \frac{0,247}{1,577} = 0,157$$



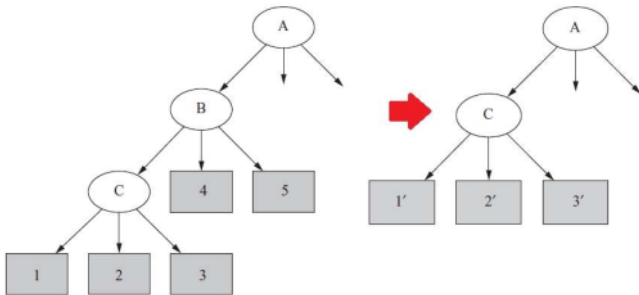
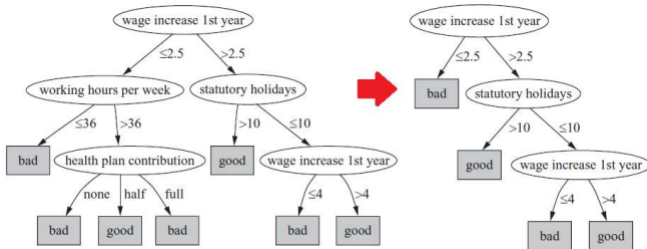
Razão de ganho

Outlook		Temperature		Humidity		Windy	
Entropia(A)	0,693	Entropia(A)	0,911	Entropia(A)	0,788	Entropia(A)	0,892
IG(S, A)	0,247	IG(S, A)	0,029	IG(S, A)	0,152	IG(S, A)	0,048
Entropia(S, A)	1,577	Entropia(S, A)	1,557	Entropia(S, A)	1,000	Entropia(S, A)	0,985
GR(S, A)	0,157	GR(S, A)	0,019	GR(S, A)	0,152	GR(S, A)	0,049

- Razão de ganho;
- Baixo nível de ramificação;
- Maximizar a razão de ganho;
- Médias dos ganhos de informação.

- Assume-se que os resultados de testes desconhecidos são distribuídos probabilisticamente;
- Frequência relativa;
- Valores conhecidos;
- Fórmulas da razão de ganho e ganho de informação são adaptadas.

- Construir a árvore completa;
- Realizar a poda bottom-up;
- Subtree replacement;
 - Substitui um nó interno por uma folha;
- Subtree raising;
 - Substitui um nó interno por uma subárvore



- J48;
- Weka;
- Plataforma de código aberto;
- Análise de dados;
- Mineração;
- Gráficamente;
- Waikato;
- Algoritmos de aprendizado de máquina;
- Pesquisa acadêmica.



- Implemente um algoritmo que calcule a razão de ganho para um atributo;
- Crie uma base de dados contendo informações e atributos que possam ser utilizados no algoritmo C4.5 ou ID3.

Table: Base de Dados das Aulas de Ciência da Computação

Dia	Professor	Dia da Semana	Laboratório	Matéria	Aula Legal
Dia 1	Guilherme	Segunda	Sim	Algoritmos	Não
Dia 2	James	Terça	Não	Estrutura de Dados	Não
Dia 3	William	Quarta	Sim	Inteligência Artificial	Sim
Dia 4	Jones	Quinta	Não	Computação Gráfica	Sim
Dia 5	Marcia	Sexta	Não	Redes de Computadores	Não
Dia 6	David	Segunda	Não	Banco de Dados	Sim
Dia 7	Guilherme	Terça	Sim	Segurança da Informação	Não
Dia 8	William	Quarta	Não	Sistemas Operacionais	Sim
Dia 9	David	Quinta	Sim	Engenharia de Software	Sim
Dia 10	James	Sexta	Não	Algoritmos	Sim

Obrigado! Dúvidas?

Guilherme Henrique de Souza Nakahata

guilhermenakahata@gmail.com

<https://github.com/GuilhermeNakahata/UNESPAR-2024>