

Matéria: Tecnologia da Informação

Assunto: Aprendizado de Máquina e Normalização de Banco de Dados

Resumo Teórico do Assunto

Para resolver as questões apresentadas, é fundamental dominar dois tópicos distintos, mas igualmente importantes em Tecnologia da Informação: a avaliação de modelos de **Aprendizado de Máquina** e a **Normalização de Banco de Dados**.

1. Aprendizado de Máquina: Avaliação de Modelos de Classificação

No contexto de problemas de **classificação binária**, onde um modelo tenta prever se um evento pertence a uma de duas classes (ex: "sim" ou "não", "ocorre" ou "não ocorre"), a avaliação do desempenho do modelo é crucial. A **Matriz de Confusão** é uma ferramenta fundamental para isso.

1.1. Matriz de Confusão

A **Matriz de Confusão** é uma tabela que resume o desempenho de um algoritmo de classificação, comparando as previsões do modelo com os valores reais. Ela é composta por quatro métricas principais:

- **Verdadeiro Positivo (VP)**: O modelo previu a classe positiva, e o valor real era positivo. (Acerto)
- **Verdadeiro Negativo (VN)**: O modelo previu a classe negativa, e o valor real era negativo. (Acerto)
- **Falso Positivo (FP)**: O modelo previu a classe positiva, mas o valor real era negativo. (Erro Tipo I - "Falso Alarme")
- **Falso Negativo (FN)**: O modelo previu a classe negativa, mas o valor real era positivo. (Erro Tipo II - "Perda")

1.2. Acurácia

A **Acurácia** é uma das métricas mais comuns para avaliar modelos de classificação. Ela mede a proporção de previsões corretas (tanto positivas quanto negativas) em relação ao total de previsões.

Fórmula da Acurácia:

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}}$$

Onde:

- $(\text{VP} + \text{VN})$ representa o número total de previsões corretas.
- $(\text{VP} + \text{VN} + \text{FP} + \text{FN})$ representa o número total de amostras (ou observações).

Conhecimento Essencial: Para resolver problemas envolvendo a Matriz de Confusão e Acurácia, você precisa ser capaz de:

- Identificar e calcular cada um dos quatro componentes (VP, VN, FP, FN).
- Aplicar a fórmula da Acurácia para encontrar o valor percentual ou um dos componentes, caso os outros sejam dados.
- Entender que o denominador da fórmula da Acurácia é sempre o **total de amostras**.

2. Normalização de Banco de Dados Relacionais

A **Normalização** é um processo sistemático de organização das tabelas (relações) e colunas (atributos) em um **banco de dados relacional**. Seu principal objetivo é minimizar a **redundância de dados** e melhorar a **integridade dos dados**, evitando **anomalias** de inserção, atualização e exclusão.

2.1. Conceitos Chave

- **Chave Primária (CP):** Um atributo ou conjunto de atributos que identifica unicamente cada registro (tupla) em uma tabela. É fundamental para garantir a unicidade dos dados.
- **Dependência Funcional (DF):** Uma restrição entre dois conjuntos de atributos em uma relação. Se A e B são conjuntos de atributos, dizemos que A determina funcionalmente B ($A \rightarrow B$) se, para cada valor de A, existe exatamente um valor de B.
 - * **Determinante:** O conjunto de atributos A na dependência $A \rightarrow B$.
 - * **Dependente:** O conjunto de atributos B na dependência $A \rightarrow B$.
- **Dependência Funcional Parcial:** Ocorre quando um atributo não-chave depende funcionalmente de apenas uma parte da chave primária composta.
- **Dependência Funcional Transitiva:** Ocorre quando um atributo não-chave depende funcionalmente de outro atributo não-chave, que por sua vez depende da chave primária.

2.2. Formas Normais (FN)

As **Formas Normais** são um conjunto de regras que um esquema de banco de dados deve seguir para estar "normalizado". As mais comuns são 1FN, 2FN e 3FN. A questão foca na **Terceira Forma Normal (3FN)**.

- **Primeira Forma Normal (1FN):**

- * Todos os atributos devem ser **atômicos** (não multi-valorados, não compostos).

- * Não deve haver grupos repetitivos de colunas.
- * Cada coluna deve conter um único valor.

• Segunda Forma Normal (2FN):

- * A tabela deve estar na **1FN**.
- * Todos os atributos não-chave devem depender funcionalmente da **chave primária inteira**. Isso significa que não pode haver **dependências funcionais parciais** de atributos não-chave em relação à chave primária.

• Terceira Forma Normal (3FN):

- * A tabela deve estar na **2FN**.
- * Não deve haver **dependências funcionais transitivas** de atributos não-chave em relação à chave primária. Em outras palavras, nenhum atributo não-chave pode depender de outro atributo não-chave.

2.3. Mapeamento de Relacionamentos

A normalização frequentemente envolve a decomposição de tabelas grandes em tabelas menores e mais gerenciáveis. Isso é feito com base nos relacionamentos entre as entidades:

- **Relacionamento Um para Muitos (1:N)**: A chave primária da tabela do lado "um" migra como uma **chave estrangeira** para a tabela do lado "muitos".
- **Relacionamento Muitos para Muitos (N:M)**: É necessário criar uma nova tabela, chamada de **tabela de ligação** (ou associativa), que conterá as chaves primárias de ambas as tabelas originais como chaves estrangeiras. A chave primária da tabela de ligação geralmente é a combinação dessas chaves estrangeiras.

2.4. Processo de Normalização para 3FN

Para normalizar um esquema para a 3FN, siga estes passos:

1. **Identifique a Chave Primária** de cada tabela.
2. **Identifique todas as Dependências Funcionais** existentes entre os atributos.
3. **Verifique a 1FN**: Garanta que não há atributos multi-valorados ou grupos repetitivos. Se houver, crie novas tabelas para eles.
4. **Verifique a 2FN**: Para tabelas com chaves primárias compostas, verifique se algum atributo não-chave depende apenas de uma parte da chave primária. Se sim, mova esses atributos para uma nova tabela com a parte da chave primária como sua própria chave.
5. **Verifique a 3FN**: Verifique se algum atributo não-chave depende de outro atributo não-chave (dependência transitiva). Se sim, mova o atributo dependente e seu determinante para uma nova tabela, deixando o determinante como chave primária na nova tabela e como chave estrangeira na tabela original.

Conhecimento Essencial: Para resolver problemas de normalização, você precisa ser

capaz de:

- Identificar chaves primárias e dependências funcionais a partir de uma descrição ou esquema.
 - Reconhecer violações de 1FN, 2FN e 3FN.
 - Decompor tabelas de forma a eliminar redundâncias e anomalias, criando novas tabelas e estabelecendo chaves estrangeiras para manter a integridade referencial, resultando no número mínimo de tabelas que satisfaçam a 3FN.
 - Mapear corretamente relacionamentos 1:N e N:M em esquemas normalizados.
-

Questões de Provas Anteriores

Fonte: [escriturario_agente_de_tecnologia \(1\).pdf](#), Página: 12

pcimarkpci MjgwNDowMTRkOjE0YTU6OTI1ODozOGQ2OjNhMGM6NTM0MzplZmI1:U3VuLCAyNyBKdWwgMjAyNSAyMzo0Nzo0MCAtMDMwMA==

www.pciconcursos.com.br

12

BANCO DO BRASIL

AGENTE DE TECNOLOGIA - Microrregião 16 DF-TI GABARITO 1

TECNOLOGIA DA INFORMAÇÃO

36

Ao tentar resolver um problema de aprendizado de máquina que separava um evento entre duas classes, um desenvolvedor encontrou uma acurácia de exatamente 90%.

Analizando a matriz de confusão, o desenvolvedor constatou que os verdadeiros positivos eram 14169, que os verdadeiros negativos eram 15360, os falsos positivos eram 1501, e os falsos negativos eram

(A) 1778

(B) 1779

(C) 1780

(D) 1781

(E) 1782

37

Ao construir um protótipo que calculava o risco sanitário de um terreno, um desenvolvedor principiante começou a desenvolver um modelo relacional, o que o levou a propor as seguintes tabelas (onde o sublinhado representa a chave primária):

Terreno(rua,numero,tipoTerreno,CNPJ,nomeEmpresa,codigoRisco,nomeRisco,probabilidadeRisco)

Empresa(CNPJ,nomeEmpresa,CPFs,nomePessoas)

Ao revisar o trabalho, um segundo desenvolvedor, mais experiente, detectou as seguintes

propriedades:

1. $\text{codigoRisco} \rightarrow \text{nomeRisco}$
2. $\text{CNPJ} \rightarrow \text{nomeEmpresa}$
3. $\text{CPF} \rightarrow \text{nomePessoa}$
4. $\{\text{rua}, \text{numero}\} \rightarrow \text{CNPJ}$
5. Uma empresa possui vários donos, cada um com um CPF único, e um dono pode possuir várias empresas
6. Um terreno pode possuir vários riscos

Esse segundo desenvolvedor decidiu, então, colocar as tabelas na terceira forma normal, usando o número mínimo de tabelas.

O esquema que possui as seis propriedades detectadas pelo segundo desenvolvedor e que representa todas as informa-

ções desejadas pelo primeiro desenvolvedor é

(A) $\text{Terreno}(\text{rua}, \text{numero}, \text{tipoTerreno}, \text{CNPJ})$

$\text{Empresa}(\text{CNPJ}, \text{nomeEmpresa})$

$\text{Dono}(\text{CNPJ}, \text{CPF})$

$\text{Risco}(\text{rua}, \text{nome}, \text{codigoRisco}, \text{probabilidadeRisco})$

$\text{TipoRisco}(\text{codigoRisco}, \text{nomeRisco})$

$\text{PessoaFisica}(\text{CPF}, \text{nome})$

(B) $\text{Terreno}(\text{rua}, \text{numero}, \text{tipoTerreno}, \text{CNPJ}, \text{codigoRisco}, \text{probabilidadeRisco})$

$\text{Empresa}(\text{CNPJ}, \text{nomeEmpresa})$

$\text{Dono}(\text{CNPJ}, \text{CPF})$

$\text{TipoRisco}(\text{codigoRisco}, \text{nomeRisco})$

$\text{PessoaFisica}(\text{CPF}, \text{nome})$

(C) $\text{Terreno}(\text{rua}, \text{numero}, \text{tipoTerreno}, \text{CNPJ})$

$\text{Empresa}(\text{CNPJ}, \text{nomeEmpresa})$

$\text{Risco}(\text{rua}, \text{nome}, \text{codigoRisco}, \text{probabilidadeRisco})$

$\text{TipoRisco}(\text{codigoRisco}, \text{nomeRisco})$

$\text{PessoaFisica}(\text{CPF}, \text{nome}, \text{CNPJ})$

(D) $\text{Terreno}(\text{rua}, \text{numero}, \text{tipoTerreno}, \text{CNPJ}, \text{nomeEmpresa}, \text{codigoRisco}, \text{nomeRisco}, \text{probabilidadeRisco})$

$\text{Empresa}(\text{CNPJ}, \text{CPF}, \text{nomeEmpresa}, \text{nomePessoa})$

(E) $\text{Terreno}(\text{rua}, \text{numero}, \text{tipoTerreno}, \text{CNPJ}, \text{nomeEmpresa}, \text{codigoRisco}, \text{probabilidadeRisco})$

$\text{Risco}(\text{codigoRisco}, \text{nomeRisco})$

$\text{Empresa}(\text{CNPJ}, \text{CPF}, \text{nomeEmpresa}, \text{nomePessoa})$