

Desenvolvimento de Aplicações

Germano C. Vasconcelos
Centro de Informática - UFPE

Análise e Solução do Problema

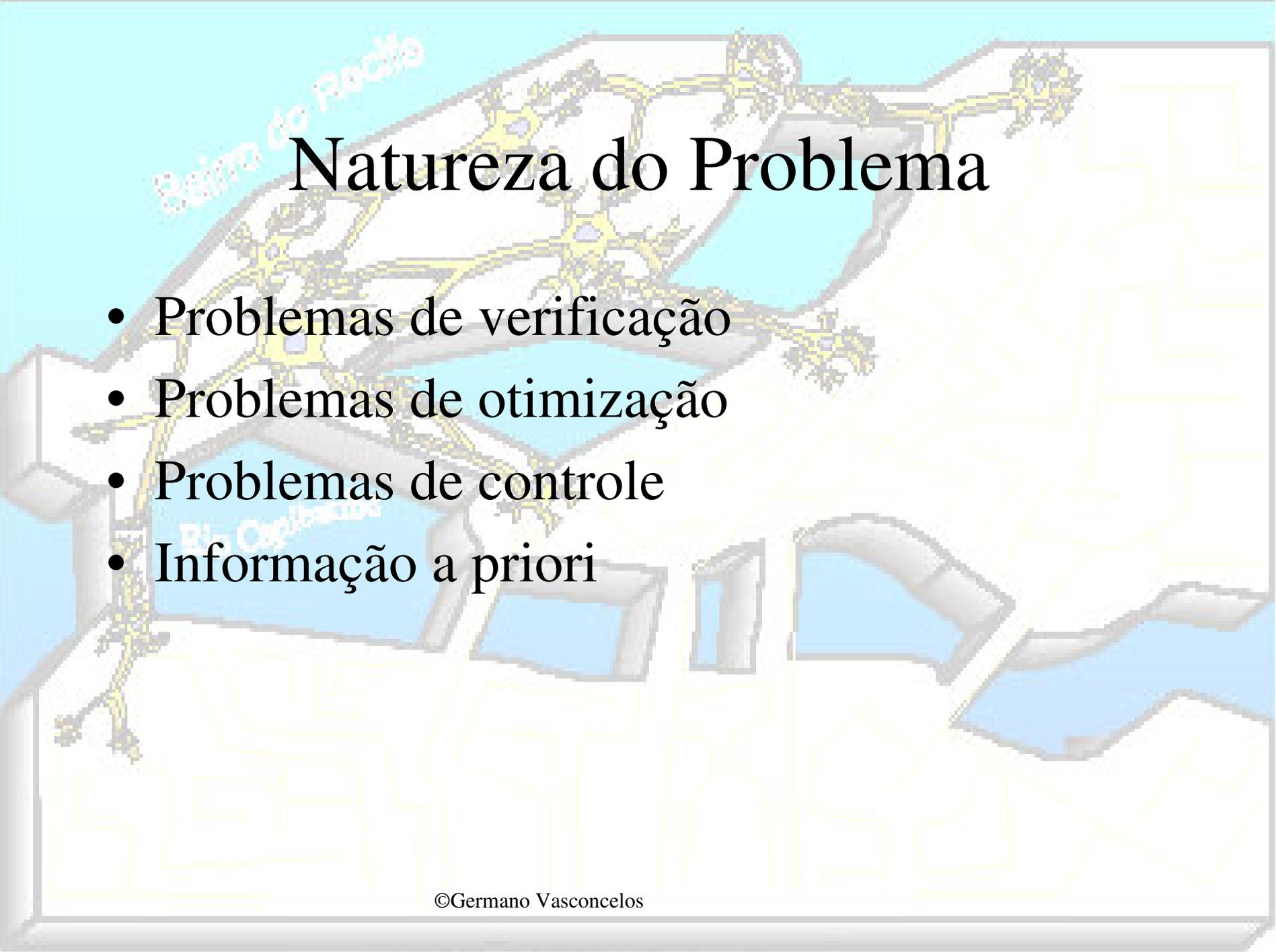
- Qual a natureza do problema ? Ou problemas ?
- Quais as informações de entrada e de saída?
- Pré-processamento dos dados
 - consistência
 - representatividade
 - codificação
 - divisão em lotes

Análise e Solução do Problema

- Normalização
- Escolha dos Modelos
 - feedforward
 - recorrente
 - auto-organizáveis
 - Booleanos
 - construtivas

Natureza do Problema

- Disponibilidade de respostas desejadas
 - feedforward X auto-organizáveis
- Domínio do problema
 - contínuo ou discreto
- Informação estática
- Informação temporal
- Problemas de classificação



Natura do Problema

- Problemas de verificação
- Problemas de otimização
- Problemas de controle
- Informação a priori

Aspectos a Considerar na Aplicação

- Pré-processamento e tratamento dos dados
 - Normalização
 - Centralização
 - Invariâncias
 - Dados ausentes (“missing data”)
 - Inconsistências
 - Codificação

Aspectos a Considerar na Aplicação

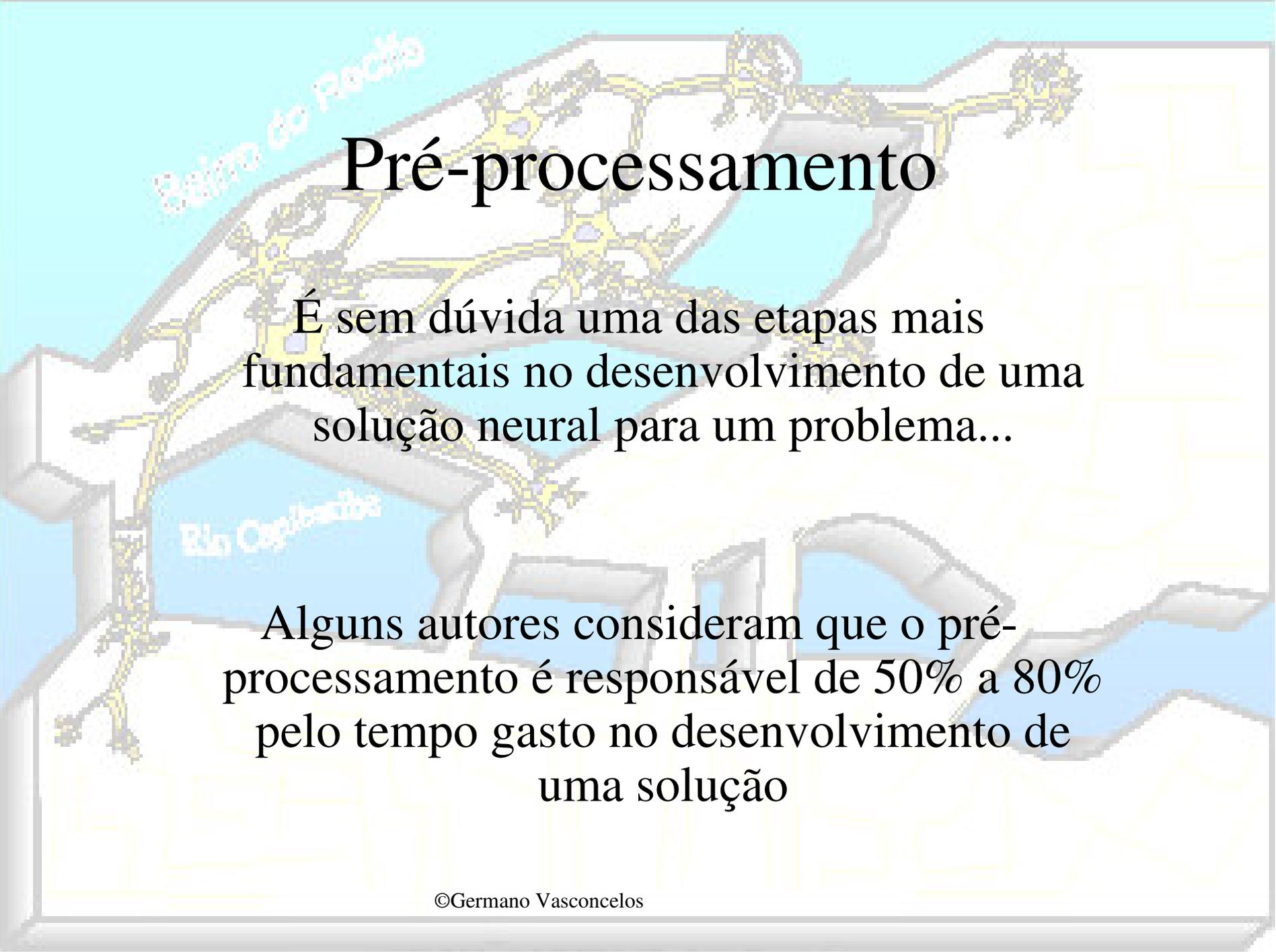
- Extração de características
 - Técnicas simples
 - Extração dos componentes principais (PCA)
 - Outras transformadas
 - Momento
 - Morfologia matemática

Aspectos a Considerar na Aplicação

- Desempenho dos modelos
 - requerimentos de memória
 - tempo de treinamento
 - tempo de classificação
 - capacidade de memória
 - adaptabilidade a mudanças no ambiente
 - capacidade de generalização

Outros Fatores Gerais

- Disponibilidade de conhecimento a priori
- Dimensão da massa de dados disponível
- Disponibilidade de ambiente de dados integrado
- Custo/tempo para a preparação dos dados
- Qualidade da massa de dados
- Tempo de treinamento
- Tempo de resposta exigido
- Adaptatividade a mudanças de comportamento
- Nível de desempenho desejado (e.g. taxa de acerto/erro)
- Explicabilidade / interpretabilidade das decisões
- Tolerância a ruído
- Requisitos de memória e processamento



Pré-processamento

É sem dúvida uma das etapas mais fundamentais no desenvolvimento de uma solução neural para um problema...

Alguns autores consideram que o pré-processamento é responsável de 50% a 80% pelo tempo gasto no desenvolvimento de uma solução

Tratamento dos Dados

- Representatividade dos dados
- Incompletude nos dados (missing data)
- Inconsistência nos dados
- Padronização de atributos
- Codificação de atributos
- Normalização de atributos
- Conjugação de atributos
- Uso de conhecimento a priori
- Atributos pré-computados
- Extração de características

Representatividade dos Dados

- Os dados são representativos ?
- Como saber se os dados são representativos?
 - Conhecimento sobre o domínio ou contexto
 - Análise matemática ou visual dos dados
- Se os dados brutos não são representativos...
 - Coleta de novos dados ou dados adicionais
 - Selecionar sub-conjuntos dos dados
 - Empregar conhecimento a priori
 - Gerar dados adicionais sinteticamente

Incompletude nos Dados

- Dados ausentes (missing data)
 - Ineficiência do processo de coleta
 - Indisponibilidade da informação
- O que fazer ?
 - Buscar informações adicionais
 - Eliminar atributos
 - Definir o valor baseado no restante dos dados
 - Codificar a ausência

Inconsistências nos Dados

- Dados que não representam a realidade
 - Eliminação dos exemplos
- Dados que não representam mais a realidade atual
 - Eliminação dos exemplos
 - Busca de informações adicionais
- Dados contraditórios
 - Identificação do exemplo inconsistente (se houver)
 - Eliminação dos exemplos contraditórios

Padronização de Atributos

- Ocorre principalmente com dados simbólicos
 - Mesma informação preenchida de maneiras diferentes
 - É necessária uma padronização
- Como resolver ?
 - Manualmente (pode ser ineficiente e custoso)
 - Algoritmos de casamento de padrões

Codificação de Atributos

- RN só aceitam como entrada dados numéricos...
- Contínuos
 - inteiros
 - reais
- Discretos
 - binários
 - inteiros
 - reais
 - simbólicos

Como Codificar ?

- Contínuos
 - Já estão adequadamente codificados
- Discretos
 - binários
 - Já estão adequadamente codificados
 - A não ser que uma transformação melhore a qualidade da representação
 - inteiros
 - Já estão adequadamente codificados
 - A não ser que uma transformação melhore a qualidade da representação
 - reais
 - Já estão adequadamente codificados
 - A não ser que uma transformação melhore a qualidade da representação

Como Codificar ? (continuação)

- Discretos
 - Simbólicos
 - necessitam de transformação
 - codificação binária
 - densa
 - esparsa (1 de N)
 - intermediária (M de N)

Normalização de Atributos

- O que é a normalização de atributos
 - É a transformação dos atributos para valores dentro de um intervalo
- Por que é necessário normalizar ?
 - Para que as informações sejam apresentadas de maneira consistente
- Direta
 - Se os dados brutos estiverem uniformemente distribuídos no espaço ou se assim for desejado
- Após transformações
 - Para realizar a distribuição uniforme dos dados
- Aplicada em valores inteiros ou reais

Conjugação de Atributos

- Objetiva uma melhor representação da informação
- Pode-se dar através de somas, multiplicações, subtrações ou funções, mas é muito comum divisões

Conhecimento a Priori

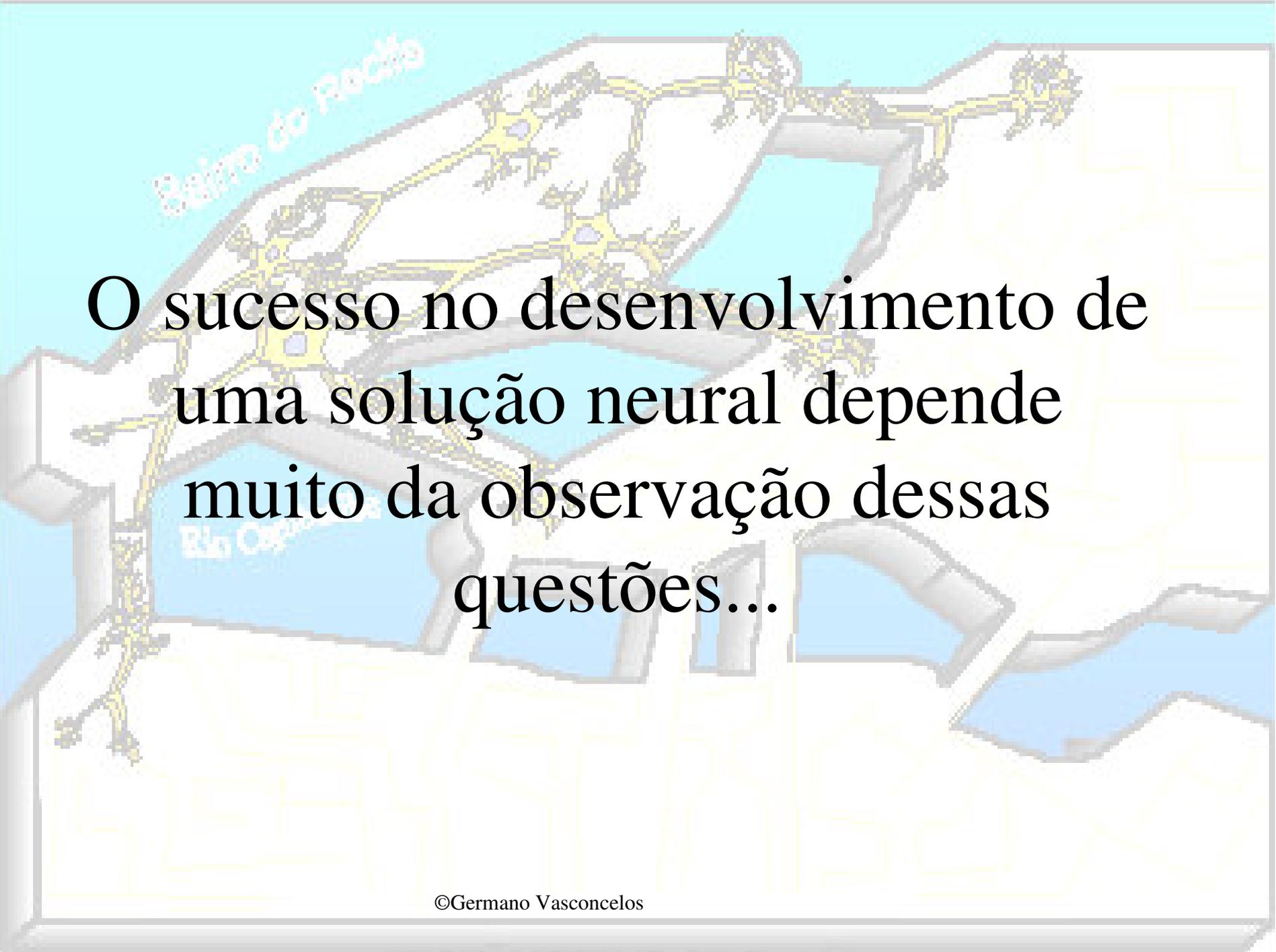
- Emprego de informações que facilitem o desenvolvimento da solução ou o aprimoramento da solução
 - Eliminação, conjugação ou adição de atributos
 - Informações baseadas em observações dos dados
 - Conhecimento de especialistas no domínio do problema
- Ex: Sazonalidade dos dados, tendência de comportamento dos dados no ambiente, distribuição dos dados no espaço...

Atributos Pré-computados

- Emprego de janelas de dados (*telescoping inputs*)
- Retardos
- Médias flutuantes (*moving averages*)

Extração de Características

- Por que extrair características ?
 - ...
- Tipos de características
 - características conceituais
 - Ex: área, volume, número de arestas horizontais cortantes
 - não-conceituais
 - Através de algoritmos de transformação de espaços
 - Transformadas geométricas (ex: moments, Walsh-Hadamard)
 - Transformadas estatísticas (ex: PCA)



O sucesso no desenvolvimento de
uma solução neural depende
muito da observação dessas
questões...