

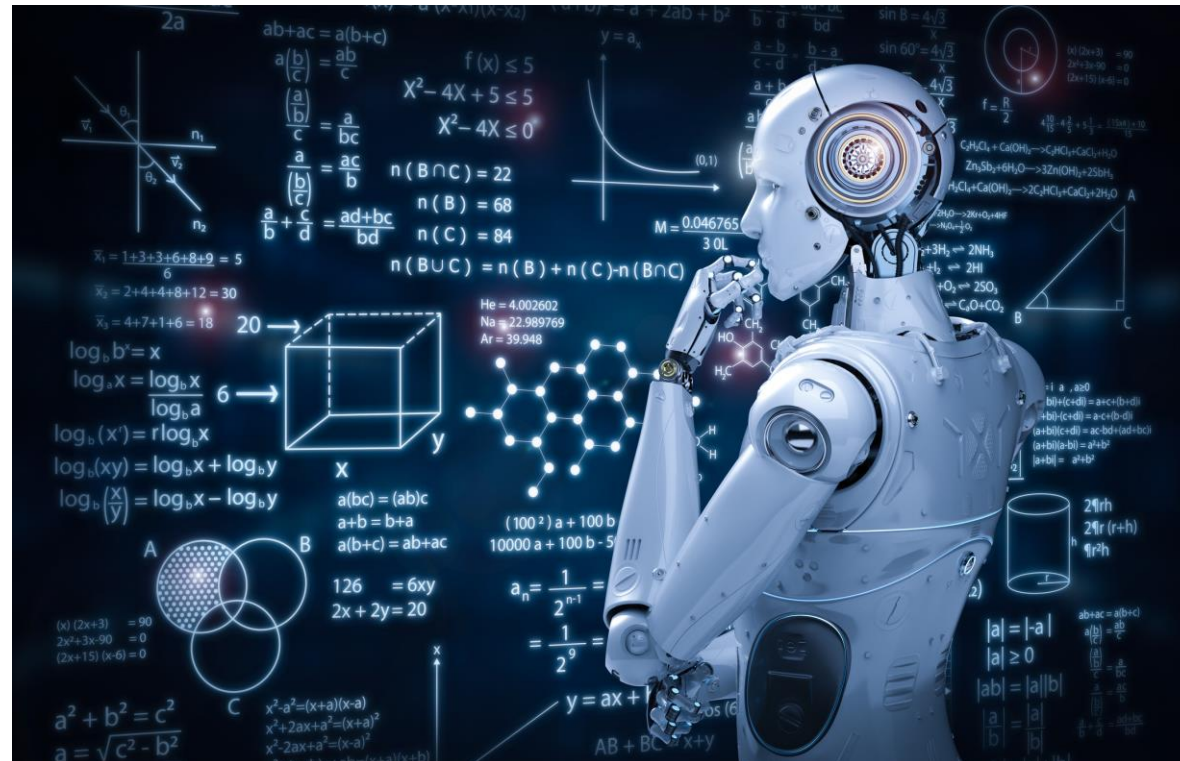


APRENDIZADO DE MÁQUINA: FORMULAÇÃO DO PROBLEMA

André Gustavo Adami
Daniel Luis Notari

INTRODUÇÃO

“Obter uma compreensão mais clara dos objetivos da investigação e planejar as demais etapas” (Webb e Copsey, 2011)



DEFINIÇÃO DO PROBLEMA

A definição do problema nem sempre é tão óbvia e simples quanto parece ser

“Desenvolver um sistema que relacione as músicas de interesse de um usuário”

A subjetividade de “músicas de interesse” pode produzir diferentes abordagens para a tarefa. É possível extrair informações das músicas que permitam descrever o “interesse” do usuário?

DEFINIÇÃO DO PROBLEMA

A definição do problema deve ser o mais clara e simples possível a fim de permitir a definição da metodologia a ser aplicada na sua solução

“Desenvolver um sistema que identifique se um paciente possui predisposição ou não para uma forma mais grave do CoVid-19”

Pode-se ir além, utilizando o formalismo de aprendizagem proposto por Tom Mitchell



“

Um programa de computador é dito aprender da experiência **E** com respeito a alguma classe de tarefas **T** e medida de desempenho **P**, se o seu desempenho nas tarefas em **T**, como medido por **P**, melhora com a experiência **E**

”

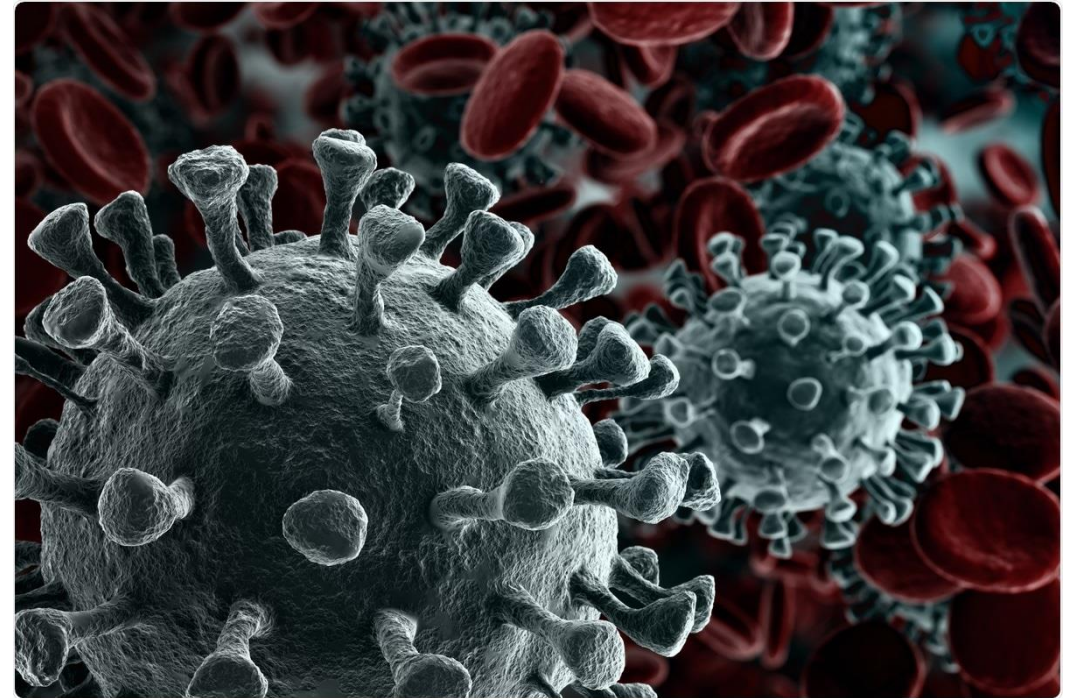
Tom M. Mitchell, 1997
Carnegie Mellon University

DEFINIÇÃO DO PROBLEMA

Tarefa T: classificar se um paciente possa desenvolver uma forma mais grave do CoVid-19

Experiência de Treinamento E: conjunto de dados clínicos relevantes

Medida de Desempenho P: percentual de pacientes que desenvolveram forma mais grave



DEFINIÇÃO DO PROBLEMA



Tarefa T: prever qual é a cotação do dólar para a hora/dia seguinte

Experiência de Treinamento E: série temporal da cotação do dólar

Medida de Desempenho P: erro da cotação estimada

DEFINIÇÃO DO PROBLEMA

Na definição do problema, existem algumas questões básicas que devem ser respondidas a fim de selecionar a melhor abordagem para construir um sistema inteligente

Qual é a tarefa (categorizar, prever, ordenar, aproximar, agir)?

Temos dados de entrada para aprender (quanto) ? Caso negativo, temos como coletar?

Temos os dados de saída? Caso negativo, temos como gerar?

Temos um método para avaliar a qualidade do resultado?

DEFINIÇÃO DO PROBLEMA

A compreensão exata do problema permite definir qual é o objetivo que se pretende alcançar (associação, classificação, previsão, compressão, extração de conhecimento, ranqueamento, ...), quais são os dados necessários e a forma de como o sistema deve ser avaliado

Em certas situações, pode-se concluir que a solução pode ser obtida utilizando-se métodos convencionais!



SUPOSIÇÕES DO PROBLEMA

Listar um conjunto de suposições sobre o problema (heurísticas, senso comum, conhecimento específico do domínio) a fim de tornar o problema mais específico e mais simples de ser resolvido

Para o problema da predisposição do CoVid-19

- A idade é muito importante para o modelo
- A raça e a estação do ano não é importante para o modelo
- Pessoas infectadas podem ter uma pré-disposição maior

SUPOSIÇÕES DO PROBLEMA

A simplificação de um problema tende a facilitar a construção de um sistema de regressão ou de classificação

- Classificação binária x multi-classe
- Regressão unidimensional x multidimensional
- Classificação x regressão

Por exemplo, no caso de um problema de predição de quando uma peça de uma máquina vai quebrar, pode-se trabalhar com duas abordagens:

- Determinar o número de dias até o evento (regressão)
- Determinar se a peça vai quebrar em 1, 2, 3 semanas ou mais (classificação)

SUPOSIÇÕES DO PROBLEMA

Para a tarefa utilize o modelo mais simples possível, pois é mais fácil de entender e implementar

- Classificador linear x não-linear
- Não quer dizer que a solução proposta será utilizada na vida real, mas é um bom começo para entender as características do problema que está sendo enfrentado

Occam's Razor



"When faced with two equally good hypotheses, always choose the simpler."

AVALIAÇÃO DO RESULTADO

A tarefa proposta deve ser avaliada de tal forma que permita medir se o objetivo foi atingido

Como saber se o sistema está
fazendo um bom trabalho?

O tipo de problema define a métrica a ser utilizada (e existem diversas métricas para diferentes problemas e contextos)

PROBLEMAS SIMILARES

Deve-se pesquisar e listar qualquer problema que seja similar ao que está sendo tratado

Isto pode ajudar a se chegar a uma solução, delimitar melhor o problema, identificar dados relevantes e reconhecer problemas potenciais não vistos na definição

- A metodologia utilizada na tarefa de reconhecimento de letras do alfabeto pode ser seguida na tarefa de reconhecimento de dígitos



CONCLUINDO

Todas as etapas seguintes dependem de uma definição clara e precisa do problema a ser tratado

A má definição do problema pode produzir sistemas que não podem alcançar o objetivo estabelecido

Medidas de desempenho quantitativa ou qualitativa devem ser levantadas a fim de avaliar se o objetivo foi atingido

Aprenda das experiências dos outros buscando problemas (e consequentemente soluções) similares