

Estudos Azure Speech Studio e Language Studio

Nome: Giovanna Pereira

Curso: Introdução a Serviços de IA na Cloud com a Azure – DIO

Instrutora: Valéria Baptista

A inteligência artificial serve para:

- Prever resultados e reconhecer padrões com base em dados históricos;
- Extrair informações de fontes para obter conhecimento
- Compreender linguagens e participar de conversas
- Reconhecer eventos anormais e tomar decisões
- Interpretar informações visuais

Os princípios de uma IA são:

- Imparcialidade
- Inclusão
- Transparência
- Responsabilidade

Quando se fala em aprendizado de máquina, tem que se levar em consideração 5 passos:

1. Dados de treinamento (Observações anteriores)
2. Algoritmo (Generaliza a relação entre x e y como uma função)
3. Modelo (Encapsula a função)
4. Inferência de dados (Recursos não rotulados) / Previsões
5. Avaliação do modelo (Comparar rótulos previstos com os rótulos reais)

Aprendizado profundo com a rede neural humana:

- Neurônios disparam em resposta a estímulos eletroquímicos
- Quando disparado, o sinal é passado para neurônios conectados
- Cada neurônio é uma função que opera com um valor de entrada (x) e um peso (w)
- A função é envolvida em uma função de ativação que determina se a saída deve ser transmitida

Sobre o Azure Machine Learning: É uma plataforma baseada em nuvem para aprendizado de máquina. A nuvem do Azure da Microsoft oferece escalabilidade e confiabilidade em armazenamento de dados, computação e serviços.

Serviços de IA no Microsoft Azure:

- Aprendizado de Máquina Azure: Uma plataforma para treinar, implantar e gerenciar modelos de aprendizado de máquina
- Serviços de IA no Azure: Um conjunto de serviços que abrange visão, fala, linguagem, decisão e IA generativa

- Pesquisa Cognitiva do Azure: Extração, enriquecimento e indexação de dados para pesquisa inteligente e mineração de conhecimento

Esses serviços são consumidos por aplicativos via endpoint REST (<https://endereco>) e com chave de autenticação.

Processamento de linguagem natural e IA conversacional no Azure:

- Tradução de texto
- Tradução de documentos
- Tradução personalizada

Simulação de uma análise de textos:

“Passei dias terríveis em São Paulo”

Idioma predominante: Português

Sentimentos: 0,88 (negativo)

Frases-chave: “dias terríveis”

Entidade: São Paulo

No serviço de bot do Azure, você encontra:

- Uma plataforma baseada em nuvem para desenvolvimento e gerenciamento de bots
- Integração com AI Language e outros serviços
- Conectividade através de vários canais

Você também acha recursos de conversão de texto em fala do serviço de fala para gerar fala audível a partir de texto.

IA Generativa: Cria conteúdo original, como IA gerativa que foi incorporada a aplicativos de chat. Os aplicativos de IA gerativa usam entrada em linguagem natural e retornam respostas apropriadas em uma variedade de formatos:

- Geração de linguagem natural
- Geração de código
- Geração de imagem

Modelos de linguagem grandes: Os aplicativos de IA gerativa são alimentados por LLM's (modelos de linguagem grandes), que são um tipo especializado de modelo de machine learning que você pode usar para executar tarefas de PLN (processamento de linguagem natural), incluindo:

- Determinar sentimento ou classificar de outra forma o texto em idioma natural
- Resumir um texto
- Comparar várias fontes de texto quanto à similaridade semântica
- Geração de novo linguagem artificial

Modelos de linguagem grandes - Transformador:

- Bloco codificador: cria representações semânticas do vocabulário de treinamento
- Bloco decodificador: gera novas sequências de linguagem

Modelos de linguagem grandes - Tokenização:

- Etapa 01: Decompor o texto de treinamento em tokens
- Etapa 02: As relações entre tokens são capturadas como vetores, conhecidos como inserções
- Etapa 03: Captura a força das relações entre tokens usando a técnica de atenção

Copilotos: Integrados a outros aplicativos e fornecem uma maneira para os usuários obterem ajuda em tarefas comuns a partir de um modelo generativo de IA. Os desenvolvedores podem criar copilotos que enviam prompts para grandes modelos de linguagem e geram conteúdo para uso em aplicativos.

Para aprimorar as respostas de IA generativa com a engenharia de prompts, é necessário melhorar a qualidade da IA gerativa usando linguagem direta, mensagens do sistema, exemplos e/ou dados de fundamentação.