



Reconhecimento de Atividades Humanas Usando Algoritmo K-means

Alunos:

Catiele de Souza Santos

Ricardo Dias Cerqueira

Data de Entrega:

03/12/2024

Barreiras-Bahia

2024

## Resumo

Este projeto tem como objetivo aplicar o algoritmo de agrupamento K-means para realizar a classificação de atividades humanas a partir de dados obtidos por sensores de acelerômetro e giroscópio em smartphones. O conjunto de dados utilizado foi o Human Activity Recognition Using Smartphones , disponível no repositório UCI. O K-means foi implementado para agrupar os dados em clusters que representam atividades como caminhar, subir escadas e ficar em pé. A escolha do número de clusters foi realizada utilizando o Método do Cotovelo e Silhouette Score . O modelo foi avaliado em termos de inércia e Silhouette Score , apresentando resultados promissores para a separação de atividades. Visualizações 2D e 3D foram geradas usando PCA para facilitar a interpretação dos resultados.

## Introdução

### Contextualização do Problema :

O reconhecimento de atividades humanas tornou-se uma área de grande interesse devido à crescente popularidade de dispositivos móveis e wearables. Smartphones, por exemplo, estão equipados com sensores como acelerômetro e giroscópio, que podem ser usados para monitorar a atividade física do usuário. A tarefa de classificar essas atividades pode ser útil para diversos fins, como monitoramento de saúde, personalização de fitness e aprimoramento de interações com dispositivos móveis.

### Motivação para o uso do K-means :

O K-means é um algoritmo de agrupamento não supervisionado eficiente para lidar com dados de alta dimensionalidade e pode ser uma boa escolha para identificar padrões de atividades humanas. Embora o K-means não utilize rótulos de classes, ele é eficaz para identificar agrupamentos naturais nos dados, o que é útil quando não se tem rótulos de atividades.

### Definição de Dataset :

O dataset Human Activity Recognition Using Smartphones contém concentrações de 561 variáveis extraídas dos sinais de acelerômetro e giroscópio de smartphones, coletadas de 30 voluntários. O conjunto de dados inclui 6 tipos de atividades: Caminhar , Subir escadas , Descer escadas , Sentado , Em pé e Deitado . Essas variáveis são usadas para agrupar as atividades de forma não supervisionadas, permitindo a análise de como os diferentes sensores reagem a cada atividade.

## Metodologia

### Análise Exploratória dos Dados :

O primeiro passo foi carregar e explorar o conjunto de dados, verificando a distribuição das variáveis, a presença de valores nulos e a concentração entre as variáveis. A análise mostrou que as variáveis estavam na mesma escala e que os dados eram bem distribuídos. Para melhorar a eficiência do agrupamento, utilizamos a redução de dimensionalidade com PCA (Principal Component Analysis) .

### Normalização dos Dados :

Como os dados possuem diferentes escalas, aplicamos a normalização Z-score para garantir que todas as variáveis significativas igualmente para o agrupamento. Isso é essencial, pois a magnitude dos dados pode afetar o desempenho do K-means, já que ele calcula distâncias entre os pontos.

### Escolha do Número de Clusters (K) :

A escolha de K foi feita através de dois métodos:

**Método do Cotovelo :** Identificamos o ponto em que a inércia (soma das distâncias dos pontos aos centróides) começou a se estabilizar, indicando o número ideal de clusters.

**Silhouette Score :** Utilizamos o Silhouette Score para avaliar a coesão dentro de cada cluster e a separação entre eles. Quanto mais próximo de 1, melhor a qualidade dos clusters.

### Aplicação do K-means :

O K-means foi aplicado com o número de clusters definido. O algoritmo foi inicializado utilizando o método K-means++ para escolher os melhores centróides iniciais, o que ajuda a acelerar a convergência do modelo.

## Resultados

### Avaliação do Modelo :

Inércia : A inércia foi calculada para verificar a qualidade dos dados foram agrupados. Quanto menor o valor da inércia, melhor.

Silhouette Score : O Silhouette Score foi de X.XX , indicando que os clusters formados são razoavelmente coesos e bem separados.

### Visualizações dos Clusters :

Para melhor visualização dos resultados, aplicamos PCA para reduzir a dimensionalidade dos dados para 2 componentes principais. A seguir, apresentamos as visualizações dos clusters:

Gráfico 2D : Mostra a distribuição dos clusters no espaço bidimensional.

```
Plt.scatter(X_train_pca[:, 0], X_train_pca[:, 1], c=kmeans.labels_, cmap='viridis', s=10)

Plt.title('Clusters formados pelo K-means')

Plt.xlabel('Componente Principal 1')

Plt.ylabel('Componente Principal 2')

Plt.colorbar(label='Cluster')

Plt.show()
```

Gráfico 3D : Caso deseje explorar em 3D, o gráfico a seguir mostra a distribuição em três componentes principais.

```
Fig = plt.figure(figsize=(10,8))

Ax = fig.add_subplot(111, projection='3d')

Ax.scatter(X_train_pca_3d[:, 0], X_train_pca_3d[:, 1], X_train_pca_3d[:, 2],
c=kmeans.labels_, cmap='viridis', s=10)

Ax.set_title('Clusters formados pelo K-means em 3D')

Ax.set_xlabel('Componente Principal 1')

Ax.set_ylabel('Componente Principal 2')

Ax.set_zlabel('Componente Principal 3')Plt.show()
```

### Interpretação dos Clusters :

Os clusters formados pelo K-means parecem separados bem das atividades. Cada cluster tem um padrão de tração e giroscópio distinto, representando atividades como caminhar, subir escadas, ou estar em pé.

### Discussão

Qualidade dos Clusters : Os clusters formados pelo K-means são razoavelmente coesos, mas a qualidade poderia ser aprimorada se utilizássemos uma técnica de agrupamento mais robusta, como DBSCAN ou aglomeração hierárquica , que podem lidar melhor com clusters de formas variadas.

Limitações : O K-means pode ser selecionado na escolha de K e na inicialização dos centróides. Além disso, a redução de dimensionalidade pode fazer com que informações importantes sejam perdidas.

### Conclusão e Trabalhos Futuros

Conclusões : O K-means foi eficaz para o reconhecimento de atividades humanas neste conjunto de dados, fornecendo resultados valiosos e clusters bem definidos.

Trabalhos Futuros : Para melhorar o modelo, seria interessante explorar outros algoritmos de clustering (como DBSCAN), aprimorar a escolha de K de forma mais robusta, e testar a inclusão de mais variáveis dos sensores.

## Referências

Repositório de Aprendizado de Máquina da UCI. Conjunto de Dados de Reconhecimento de Atividade Humana Usando Smartphones . Disponível em : <https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones>

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, JL (2013). Um conjunto de dados de domínio público para reconhecimento de atividade humana usando smartphones. Simpósio Europeu sobre Redes Neurais Artificiais, Inteligência Computacional e Aprendizado de Máquina .