

Implementação e Análise do Algoritmo K-means com o Dataset Human Activity Recognition

Brenda Trindade e Kevin Borges

03 de dezembro de 2024

Resumo

O objetivo deste projeto foi implementar e avaliar o algoritmo K-means para o agrupamento de atividades humanas com base em dados coletados por sensores de smartphones. Utilizando o dataset "Human Activity Recognition Using Smartphones" da UCI Machine Learning Repository, que contém dados de acelerômetro e giroscópio, o projeto seguiu as seguintes etapas: análise exploratória dos dados, aplicação do algoritmo K-means, escolha do número de clusters, e avaliação do desempenho do modelo. A escolha do número de clusters foi feita através do método do cotovelo e do silhouette score. O algoritmo demonstrou boa capacidade de segmentação das atividades, com a escolha de $K=6$ como o número ideal de clusters, refletindo as atividades do dataset.

Introdução

O *Reconhecimento de Atividades Humanas* visa identificar as atividades realizadas por um indivíduo com base em dados coletados de sensores, como acelerômetros e giroscópios. Este problema pode ter diversas aplicações práticas em monitoramento de saúde, assistência a idosos e interação com dispositivos móveis.

Neste projeto, utilizamos o dataset "Human Activity Recognition Using Smartphones" da UCI Machine Learning Repository que contém dados de 30 participantes voluntários que realizaram seis atividades diárias: caminhar, subir escadas, descer escadas, sentar, deitar e ficar em pé. Capturadas através do uso de smartphones com sensores incorporados, acelerômetro e giroscópio.

Metodologia

Uso do K-means

O K-means é um algoritmo de agrupamento não supervisionado amplamente utilizado em tarefas de segmentação e clusterização de dados. Escolhido devido à sua simplicidade, eficiência e boa aplicabilidade na análise de grandes volumes de dados, como é o caso dos dados de sensores. O K-means permite agrupar os dados sem a necessidade de rótulos prévios, sendo adequado para detectar padrões e agrupar atividades. Neste projeto, o

K-means foi aplicado para agrupar os dados coletados por sensores em clusters que representassem diferentes atividades.

Análise Exploratória

A análise exploratória foi realizada visando compreender a estrutura dos dados e identificar padrões, distribuições e correlações. Para isso, utilizamos técnicas como:

- **Visualização das distribuições** das variáveis e identificação de características importantes para o agrupamento.
- **Redução de dimensionalidade** utilizando **PCA (Principal Component Analysis)**, o que permitiu visualizar os dados em um espaço 2D e facilitar o entendimento das relações entre as amostras.

Implementação do K-means

O algoritmo K-means foi aplicado para segmentar as atividades humanas em clusters. A implementação seguiu os seguintes passos:

- **Escolha do número de clusters (K):** A escolha de K foi realizada utilizando dois métodos:
 - **Método do Cotovelo (Elbow Method):** Avaliou a inércia para diferentes valores de K.
 - **Silhouette Score:** Mediu a coesão e separação dos clusters.
- **Processamento dos Dados:** Os dados foram normalizados com o uso da classe StandardScaler do Scikit-learn para garantir que todas as variáveis tivessem a mesma escala.

Avaliação dos Clusters

A avaliação dos clusters gerados foi feita com as métricas:

- **Silhouette Score:** Para medir a coesão e separação entre os clusters.
- **Inércia:** Calculada para cada valor de K, ajudando a visualizar a qualidade do agrupamento para diferentes números de clusters.

Resultados

Métricas de Avaliação

- **Silhouette Score:** O **silhouette score** foi calculado para determinar a qualidade da separação dos clusters. O valor obtido foi **0.46646077254643364**, indicando uma boa separação entre os clusters.
- **Inércia:** A inércia foi **205699.18093589589**, indicando a coesão dos clusters formados.

Gráficos e Visualizações

- **Projeção PCA:** A projeção dos dados em duas dimensões (PCA) mostrou a separação clara entre as atividades.
 1. Caminhar
 2. Subir escadas
 3. Descer escadas
 4. Sentado
 5. Deitado
 6. Ficar em pé

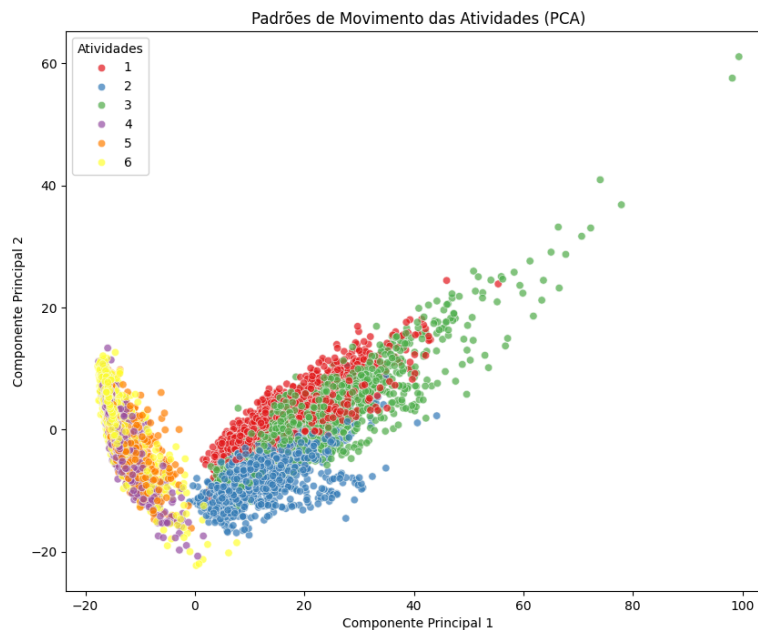


Figura 1: Visualização PCA das atividades com 2 componentes principais.

- **Método do Cotovelo e Silhouette:** Utilizado para determinar o número ideal de clusters K. O gráfico gerado mostrou que a inércia diminui de forma acentuada até K=6, após a diminuição estabiliza indicando 6 clusters como suficientes para representar as atividades.

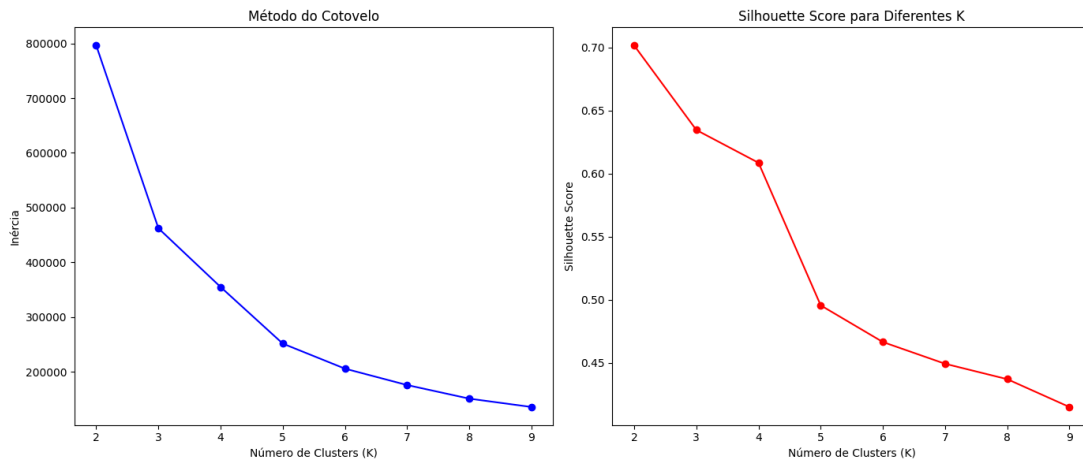


Figura 2: Método do Cotovelo e Silhouette

- **Atividades por Cluster:** As atividades em cada cluster foram analisadas e os clusters foram interpretados com base nas atividades predominantes. A tabela cruzada mostrava os clusters formados por K-means e as atividades associadas.

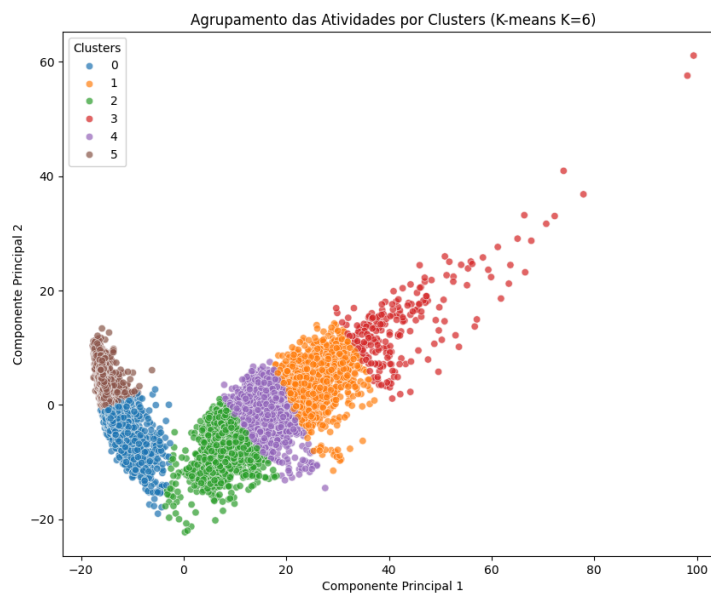


Figura 3: Atividades por Clusters

- **Centroides dos Clusters:** Representando o centro de cada grupo de atividades, foram plotados e comparados com as distribuições das amostras.

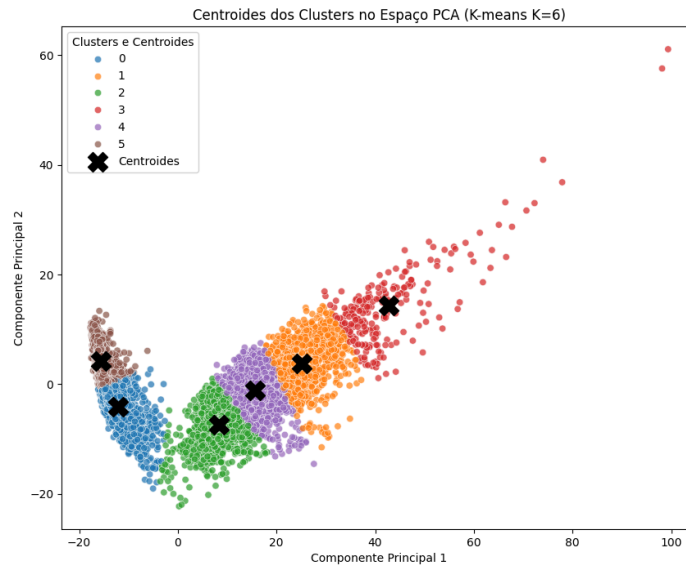


Figura 4: Centroides dos clusters.

Discussão

O algoritmo K-means demonstrou boa capacidade de segmentar as atividades humanas com o número ideal de clusters $k=6$. No entanto, houve uma certa sobreposição entre atividades como "sentar" e "ficar em pé", sugerindo que essas atividades possuem características semelhantes nos dados do acelerômetro e giroscópio, o que dificulta a separação entre elas.

Limitações

- O K-means é sensível à escolha inicial dos centroides e à presença de outliers. Isso pode afetar a qualidade dos resultados. A utilização do método K-means++ ajudou a melhorar a inicialização dos centroides.
- A redução de dimensionalidade para 2D pode perder informações relevantes para o agrupamento.

Conclusão

O projeto demonstrou que o algoritmo K-means é eficaz para agrupar atividades humanas com base em dados de sensores de smartphones. A escolha do número de clusters foi fundamental para garantir uma boa separação das atividades e qualidade dos resultados. As métricas de avaliação adotadas como o silhouette score e a inércia, indicaram uma boa separação entre os clusters

Referências

1. UCI Machine Learning Repository. Human Activity Recognition Using Smartphones. Disponível em: [UCI Machine Learning Repository](#)
2. Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). Human Activity Recognition Using Smartphones. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning ESANN. Disponível em: [es2013-84.pdf](#)