

ANÁLISE COMPARATIVA ENTRE K-MEANS E DBSCAN PARA CLUSTERIZAÇÃO DE TRAJETÓRIAS

XXXVIII Encontro de Iniciação Científica

Nicolas de Freitas Soares, Nicksson Ckayo, Ticiania Linhares Coelho da Silva

Clusterização é uma técnica de mineração de dados multivariados, que tem por objetivo agrupar os N casos de uma base de dados em K grupos disjuntos denominados clusters. O conhecimento de uma amostra de pontos com referências espaço-temporais de um objeto móvel possibilita a dedução de uma sucessão lógica de observações espaciais denominada trajetória. Algoritmos de clusterização são utilizados para identificação de padrões em dados espaciais, entretanto, sua aplicação em larga escala requer um estudo mais detalhado dos métodos de abordagem. A prática da clusterização de trajetórias de objetos móveis está muito presente no cotidiano, pois facilita análises de tráfegos, migrações, mudanças climáticas, etc. Entre os métodos abordados, há o K-Means, que se baseia em parâmetros espaciais, como centroides e vizinhança de pontos, útil para análises genéricas sobre grande quantidade de dados, e o método DBScan, mais arbitrário, utilizado para análises detalhadas, que se baseia na densidade e coerência dos dados amostrais. O presente trabalho tem por objetivo comparar os referidos dois métodos na clusterização de trajetórias com o intuito de questionar a aplicação mais adequada. Para isso, realizaram-se análises sobre trajetórias de um indivíduo monitorado na cidade de Fortaleza utilizando ambos os métodos. O cálculo utilizado na validação dos clusters considera três critérios principais: (i) Separação inter-clusters, (ii) Densidade intra-clusters e (iii) Coerência entre clusters. Portanto, baseado nos critérios determinados, a validação pelo método DBScan apresentou-se mais apropriada para tal aplicação. No entanto, por ser mais complexo, limita-se a menor quantidade de dados processados. Como forma de contribuição da presente análise, espera-se aprimorar as técnicas de clusterização de trajetórias de objetos móveis a partir de pontos espaço-temporais para extrair informações úteis de uma análise. Por fim, os autores agradecem à Fundação ASTEF pelas bolsas concedidas.

Palavras-chave: Base de dados. Cluster. Trajetórias. pontos espaço-temporais.