

Núcleo de Avaliação: Núcleo II

Área temática: Ciências Exatas e da Terra

Área do Conhecimento: Matemática

Estudo de decomposição em fatores paralelos PARAFAC para análise de dados

Lara Carla Freitas Rodrigues, Luiza Helena Felix de Andrade, Silvio Fernandes.

Estudos recentes, como o relatório publicado pela Internacional Data Corporation (IDC) em 2017 para o ano de 2025, demonstram que a sociedade contemporânea gera, retém e utiliza grandes quantidades de dados, e a tendência é que nos próximos anos essa demanda só aumente. Dentro desse contexto, surge a necessidade de analisar conjuntos de informações com muitas variáveis correlacionadas, por isso nas últimas décadas, técnicas analíticas multivariadas têm sido utilizadas para uma melhor interpretação de grandes bancos de dados, dentre elas está a Análise de Fatores Paralelos (PARAFAC). O PARAFAC permite a identificação de padrões por meio da decomposição do banco de dados originais, aqui chamado de tensor, em matrizes de carregamento, onde cada elemento do tensor original é representado como uma combinação de produtos das matrizes de carregamento. Essas matrizes refletem a contribuição de diferentes variáveis para cada fator, possibilitando uma visão mais simplificada sobre quais variáveis estão mais correlacionadas com todo o banco de dados. Dentre as características mais importantes do PARAFAC, a unicidade é um destaque. Devido à utilização de baixos graus de liberdade, o método fornece uma decomposição única e um resultado único, sendo uma grande vantagem para o estudo em que se é aplicado. Além disso, para assegurar a precisão do modelo, a decomposição busca minimizar a soma dos erros ao quadrado (SSE), que mede a diferença entre os valores observados e os estimados pelo modelo, sendo um valor mais baixo indicativo de um melhor ajuste. A escolha do número de fatores, porém, é uma das partes mais difíceis e cruciais, ela pode ser feita por meio de métodos estatísticos, como a análise da SSE, que indica a qualidade do ajuste, o Fator Match Score (FMS), responsável por avaliar a estabilidade das componentes extraídas, e o Diagnóstico de Consistência de Núcleo (CORCONDIA), que ajuda a evitar a superestimação do número de fatores necessários para descrever adequadamente os dados. Esta pesquisa tem como objetivo identificar, através do uso do PARAFAC, dentre os microdados do ENEM 2022 fornecidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), qual o principal fator socioeconômico relacionado com o desempenho positivo ou negativo dos alunos. Com esse propósito, primeiro foi necessário calcular a média geral das notas de 2022. O passo seguinte foi a filtragem dos dados e a formação de dois grandes grupos: o grupo A, com maior desempenho, e o segundo, o grupo B, de menor desempenho. Depois um tensor de ordem três foi formado a partir do agrupamento dos dois grupos. Finalmente, o algoritmo do PARAFAC foi implementado no banco de dados, posteriormente o SSE, FMS e CORCONDIA foram utilizados e em todos os casos o modelo com 3 componentes foi o que melhor

representou os dados. Por fim, a análise dos componentes indicou que a renda foi fator mais impactante para o desempenho dos estudantes do ENEM 2022, deixando evidente a necessidade do contínuo investimento em políticas públicas que visam diminuir a disparidade econômica no Brasil.

Palavras-chave: PARAFAC, ENEM, Desigualdade.

Agência financiadora: PIBIC/CNPq.

Campus: Mossoró.
