

Descobrendo segmentos de adolescentes em redes sociais

Pós Graduação em Big Data

Prof. Fabrício Jailson Barth

1 Introdução

Interagir com amigos em redes sociais, tais como Facebook e MySpace, tem se tornado um ritual entre os adolescentes do mundo todo. Estes adolescentes, ao mesmo tempo que interagem com os seus amigos, também estão sujeitos a propagandas de diversos produtos e empresas.

É muito importante para as empresas, e até mesmo para os usuários de redes sociais, que tais propagandas sejam melhor direcionadas, ou seja, que sejam entregues para aqueles que realmente tem interesse naquele tipo de produto.

Dado o texto produzido por adolescentes em Serviços de Redes Sociais, acredita-se que podemos identificar grupos de adolescentes que compartilham interesses em comum, tais como esporte, religião ou música.

Algoritmos de clustering podem automatizar este processo de descoberta de segmentos naturais desta população. No entanto, cabe a vocês decidirem se tais agrupamentos são interessantes ou não para o uso em propaganda.

2 Dados

Para esta análise, nós vamos utilizar um dataset que representa uma amostra aleatória de 30.000 estudantes de *high school* americana que possuem perfil em uma rede social¹ em 2006. Para proteger o anonimato dos usuários da rede, o nome da rede social não é fornecido.

O dataset pode ser encontrado em ². Os atributos deste dataset são:

- **gradyear**: ano de graduação.
- **gender**: sexo, masculino ou feminino.
- **age**: idade representada por um número real.
- **friends**: quantidade de amigos na rede social.
- **demais atributos**: basketball, football, soccer, entre outros indicam classes onde os textos das pessoas foram classificados. Textos de usuários publicados na rede social foram processados para gerar esta tabela. Como esta tabela foi gerada é um problema de *text mining* e deverá ser abstraído, por enquanto.

3 Método de trabalho

Sugere-se seguir as seguintes etapas para a realização desta análise:

- Aquisição, análise exploratório e pré-processamento dos dados;
- Construção do modelo;
- Análise do modelo, e;
- Preparação do relatório e principais artefatos que devem ser entregues.

¹Em inglês, Social Network System - SNS

²<https://raw.githubusercontent.com/fbarth/posGraduacao/master/ExemplosClustering/data/snsdata.csv>

4 Itens que devem ser entregues

Os seguintes itens devem ser entregues:

- Um relatório apresentando os principais resultados da análise. Este relatório pode ser entregue na forma de slides, documento PDF ou documento HTML.
- Independente do formato do relatório, o conteúdo do mesmo deve possuir os seguintes itens:
 - introdução, com descrição do contexto e objetivo do trabalho;
 - descrição do método utilizado para alcançar o objetivo, e;
 - conclusão, ou seja, descrição dos resultados encontrados com a análise realizada.
- O relatório deve possuir duas figuras que ilustram e resumizam bem os achados. Cada uma das figuras deve ter um *caption* com uma descrição autocontida da figura.
- O script que realiza todas as transformações e análises no dataset. Este item geralmente não é entregue para o cliente final. No entanto, será relevante para a avaliação deste trabalho.

5 Equipes e prazos

Este trabalho poderá ser feito por equipes com até quatro (4) integrantes e deverá ser entregue até o dia 22 de setembro de 2016 por email para fabricao.barth@gmail.com.