

MINERAÇÃO DE DADOS UTILIZANDO K-MEANS

ANDRESSA STÉFANY SILVA DE OLIVEIRA*, VITOR RAMOS GOMES DA SILVA†

Emails: `astefanysoliveira@gmail.com`, `ramos.vitor89@gmail.com`

Abstract— O processo de revisão do CBA 2014 será DOUBLE BLIND, portanto NAO inclua autores na versão que será submetida para revisão. Leve isso em consideração ao citar seus próprios trabalhos.

Keywords— Template, Examples.

Resumo— O processo de revisão do CBA 2014 será DOUBLE BLIND, portanto NAO inclua autores na versão que será submetida para revisão. Leve isso em consideração ao citar seus próprios trabalhos.

Palavras-chave— Dados, Mineração, Clusters, k-Means.

1 Introdução

Com o advento da tecnologia, a geração de dados foi bastante impulsionada e assim existe uma vasta quantidade de informações, porém, esse dado é encontrado em sua forma bruta, ou seja, não passou por um processamento o qual proporcione um significado. Atualmente, é fácil encontrar na internet Data Science disponíveis para a sociedade, a qual pode utilizá-lo para inúmeros fins, por exemplo: *European Soccer Database* (Kaggle) o qual possui milhares de dados das temporadas europeias de 2008 a 2016, outro é o Gapminder que disponibiliza dados da Organização Mundial de Saúde, do Banco Mundial.

A Mineração de Dados é o "desenvolvimento e a aplicação de técnicas que permitem analisar e obter conhecimentos novos e úteis a partir de grandes bases de dados", pois tem o objetivo de descobrir padrões nos dados através de um processo automático ou semiautomático e assim, obter significado importante de forma que proporcione alguma vantagem, a qual poderá ser: prever tendências e comportamentos de uma determinada empresa, dessa forma fazendo com que seja possível a tomada de decisões proativas; detecção de fraude; campanhas de marketing; estudos de perfil de consumo; e medicina, na indicação de diagnósticos mais precisos. Esses são apenas alguns exemplos, pois a Mineração de Dados está presente em muitas áreas, as três consideradas com maior expressão são Estatística, Aprendizado de Máquina e Banco de Dados.

Há métodos para que haja essa análise, alguns deles são: Agrupamento (*Clustering*), Árvores de Decisão (*Decision Trees*), Classificação Bayesiana (*Bayesian Classification*), Classificação Baseada em Regras (*Rule-Based Classification*), Redes Neurais (*Neural Networks*), SVM (*Support Vector Machines*), Algoritmo Genético (*Genetic Algorithm*), Métodos Hierárquicos (*Hierarchical Methods*) e Métodos de Particionamento (*Partitioning Methods*), onde este último pode ser subdividido em *k-Means* e *k-Medoids*.

Neste trabalho foi utilizado o Banco de Dados

Abertos da Universidade Federal do Rio Grande do Norte - UFRN, especificamente a tabela que fornece dados da avaliação dos docentes, onde essa avaliação é com relação a vários anos levando em consideração todas as disciplinas lecionadas pelo professor. Para a mineração desses dados foi utilizado o método *k-Means*, o qual divide em *Clusters* os professores de acordo com os valores das colunas: quantidade de discentes, média da postura profissional, postura profissional no departamento, média da atuação profissional, atuação profissional no departamento, média da autoavaliação do aluno e autoavaliação do aluno no departamento.

1.1 Título, autores e afiliações

O título do artigo é declarado por meio do comando `\title`. Por exemplo, o título deste artigo foi gerado com o comando:

```
\title{Artigo Exemplo}
```

O título não deve conter agradecimentos, os quais devem aparecer em uma seção separada, no final do artigo.

Em seguida, definem-se os autores com seus endereços eletrônicos e afiliações por meio dos comandos `\author` e `\address`. Por exemplo, o nome e endereço do primeiro autor deste artigo, João da Silva, é introduzido por meio dos comandos:

```
\author{João da Silva}
      {jsilva@silva.com}
\address{Endereço do João e da Maria\\
        Em algum lugar\\
        Cidade, Estado, País}
```

Note-se que o comando `\author` requer dois argumentos: o nome e o endereço eletrônico. Quebras de linha podem ser obtidas na afiliação por meio do comando `\\`. Segue-se, neste exemplo, a declaração do nome e afiliação do segundo autor:

```
\author{Joaquim Pereira}
      {jpereira@pereira.org}
\address{Endereço do Joaquim\\
        Em algum lugar\\
        Cidade, Estado, País}
```

Em geral, para cada comando `\author` deve

haver um comando `\address` correspondente, salvo o caso de compartilhamento de endereços. Neste caso, omite-se o comando `\address` para o segundo autor e indica-se o número do endereço que se deseja compartilhar como um parâmetro adicional do comando `\author`. Por exemplo, o comando:

```
\author[1]{Maria da Silva}
        {maria@pereira.org}
```

indica que a Maria da Silva utilizará o endereço de número 1, que corresponde ao endereço de João da Silva.

Os autores seguintes possuem endereços independentes e são introduzidos da maneira convencional:

```
\author{Pedro Manoel}
        {pedro@pedro.com}
\address{Endereço do Pedro e do Rafael\\
        Em algum lugar\\
        Cidade, Estado, País}
```

```
\author{José Rodrigues}
        {jose@rodrigues}
\address{Endereço do José\\
        Em algum lugar\\
        Cidade, Estado, País}
```

Finalmente, o autor Rafael é definido como

```
\author[3]{Rafael Pires}
        {rafael@pires}
```

o que indica que Rafael compartilha o endereço de número 3 com Pedro Manoel. Lembre-se de que, apesar de Pedro Manoel ser o autor de número 4, o seu endereço é o de número 3, uma vez que a Maria já compartilha – sem segundas intenções – do mesmo endereço com o João.

O comando

```
\maketitle
```

coloca o título e os autores no formato adequado.

1.2 Resumo, abstract e palavras-chave

Os artigos devem conter um resumo em português (ou espanhol) e em inglês. O ambiente para definição dos resumos, em qualquer idioma, é o `abstract`. Para diferenciar os idiomas utiliza-se o comando `\selectlanguage`, do pacote `babel`. Por exemplo, a seqüência de comandos produz um *Abstract* seguido de um *Resumo*:

```
\selectlanguage{english}
\begin{abstract}
    Yossarian says, ...
\end{abstract}

\selectlanguage{brazil}
\begin{abstract}
    O Largo da Sé ...
\end{abstract}
```

As *palavras-chave* devem ser definidas por meio do comando:

```
\keywords{Exemplo, Ilustração}
```

Deste ponto em diante selecione o idioma a ser utilizado e inicie o texto da sua contribuição.

IMPORTANTE: Caso esteja se produzindo um texto com a opção `conference`, colocam-se os comandos que vão de `\maketitle` a `\keywords` dentro de um ambiente `twocolumn`. Isto é necessário para que os resumos sejam formatados em coluna simples¹. Veja seção 1.5.

1.3 Teoremas, lemas, corolários e provas

Quatro ambientes já se encontram pré-definidos no estilo SBAT_{EX}. São eles

- `theorem`: ambiente para teoremas.
- `corollary`: ambiente para corolários.
- `lemma`: ambiente para lemas.
- `proof`: ambiente para provas.

Segue-se um exemplo de utilização destes ambientes:

Lema 1 (Desigualdade Subtrativa) *Em um espaço vetorial linear normado $\|x\| - \|y\| \leq \|x - y\|$ para quaisquer vetores x, y .*

Prova:

$$\begin{aligned} \|x\| - \|y\| &= \|x - y + y\| - \|y\| \\ &\leq \|x - y\| + \|y\| - \|y\| \\ &= \|x - y\|. \end{aligned}$$

□

Este lema é produzido pela seguinte seqüência de comandos:

```
\begin{lemma}[Desigualdade Subtrativa]
    Em um espaço vetorial linear ...
\end{lemma}
\begin{proof}
    ...
\end{proof}
```

Estes ambientes, com exceção do ambiente `proof`, aceitam um parâmetro opcional que define um nome. No exemplo acima, o nome *Desigualdade Subtrativa* foi produzido por meio deste argumento.

1.4 Bibliografia

As referências são reunidas ao fim do manuscrito, arranjadas alfabeticamente pelo primeiro autor e cronologicamente para cada autor.

IMPORTANTE: Todas referências citadas devem aparecer em algum outro ponto do texto.

As citações seguem um estilo autor/ano (o mesmo usado na revista *Automatica*). O pacote `harvard` (já incluído pelo SBAT_{EX}) provê vários comandos para inclusão de citações. Dois dois mais utilizados são o tradicional `\cite` e o `\citeasnoun`. Veja um exemplo:

¹Como este arquivo está adaptado para utilizar o formato adequado para o *Congresso Brasileiro de Automática* e o *Simpósio Brasileiro de Automação Inteligente*, isto já está feito

O resumo deste artigo é um trecho do livro~\citeasnoun{serafim}. Já o abstract é extraído de~\citeasnoun{catch22}. Informações sobre o estilo bibliográfico \verb+harvard+ encontram-se em~\citeasnoun{harvard}. Os interessados devem consultar referências adicionais sobre \LaTeX~\cite{latex,% latex:guide,latex:companion}.

produz o trecho:

O resumo deste artigo é um trecho do livro de Andrade (1933). Já o abstract é extraído de Heller (1996). Informações sobre o estilo bibliográfico harvard encontram-se em Williams and Schnier (1998). Os interessados devem consultar referências adicionais sobre L^AT_EX (Lamport, 1986; Kopka and Daly, 1993; Gossens and Mittelbach, 1993).

Recomenda-se a utilização o programa BibTeX para gerar as suas referências (Lamport, 1986; Kopka and Daly, 1993; Gossens and Mittelbach, 1993). Note-se que não é necessário definir o estilo bibliográfico com o comando \bibliographystyle, pois o mesmo é definido automaticamente pelo estilo SBAT_EX. Basta incluir o arquivo de bibliografias, neste exemplo o artigo exemplo.bib por meio do comando.

...

\bibliography{exemplo}

\end{document}

Note-se que este comando deve ser colocado logo antes do encerramento do artigo, o que é feito pelo comando \end{document}, de tal forma a garantir que a bibliografia seja o último item do artigo.

1.5 Esqueleto deste arquivo no formato SBAT_EX

Este arquivo, adaptado para gerar o formato adequado para o *Congresso Brasileiro de Automática* e o *Simpósio Brasileiro de Automação Inteligente*, faz uso das seguintes instruções:

```
\documentclass[conference,harvard,brazil,english]{sbatex}
\usepackage[latin1]{inputenc}
\usepackage{ae}

...

\twocolumn[
  \maketitle

  \selectlanguage{english}
  \begin{abstract}
    Yossarian says, ...
  \end{abstract}

  \selectlanguage{brazil}
  \begin{abstract}
    O Largo da Sé ...
  \end{abstract}

  \keywords{Exemplo, Ilustração}
]
```

...

Caso a opção journal seja utilizada, remova o comando twocolumn[antes do comando \maketitle e o] depois do comando \keywords. Nos textos em português ou espanhol substitua o pacote ae pelo pacote fontenc com a opção T1. Os trechos modificados são os seguintes:

```
\documentclass[journal,harvard,brazil,english]{sbatex}
\usepackage[latin1]{inputenc}
\usepackage[T1]{fontenc}

\begin{document}

  \title{Artigo Exemplo}

  \author{João da Silva}
    {jsilva@silva.com}
  \address{Endereço do João e da Maria\\
    Em algum lugar\\
    Cidade, Estado, País}

  \author{Joaquim Pereira}
    {jpereira@pereira.org}
  \address{Endereço do Joaquim\\
    Em algum lugar\\
    Cidade, Estado, País}

  \author[1]{Maria da Silva}
    {maria@pereira.org}

  \author{Pedro Manoel}
    {pedro@pedro.com}
  \address{Endereço do Pedro e do Rafael\\
    Em algum lugar\\
    Cidade, Estado, País}

  \author{José Rodrigues}
    {jose@rodrigues}
  \address{Endereço do José\\
    Em algum lugar\\
    Cidade, Estado, País}

  \author[3]{Rafael Pires}
    {rafael@pires}

  \maketitle

  \selectlanguage{english}
  \begin{abstract}
    Yossarian says, ...
  \end{abstract}

  \selectlanguage{brazil}
  \begin{abstract}
    O Largo da Sé ...
  \end{abstract}

  \keywords{Exemplo, Ilustração}

  ...

  \bibliography{exemplo}

\end{document}
```

2 Como processar este arquivo sem um aplicativo LaTeX

Para processar este arquivo via linha de comandos utiliza-se a sequência de comandos:

```
latex cba
bibtex cba
latex cba
```

Estes comandos geram um arquivo do tipo *Device Independent* (cba.dvi), que pode ser visualizado em programas como o xdvi ou yap.

Para gerar um arquivo no formato *Portable Document Format* (cba.pdf) utilize os comandos:

```
dvips -Ppdf -t a4 cba
ps2pdf -dMaxSubsetPct=100 -dSubsetFonts=true
       -dEmbedAllFonts=true -dCompatibilityLevel=1.2
       -sPAPERSIZE=a4 cba.ps
```

Observação: Dependendo da sua instalação do dvips o parâmetro `-Ppdf` pode acarretar a troca de algumas seqüências de letras nos arquivos .ps e .pdf. Um típico exemplo é a troca de toda sílaba ‘fi’ pelo símbolo da libra esterlina. Para evitar este indesejável efeito você deve desabilitar (ou habilitar, dependendo da instalação) o remapeamento de caracteres. Faça isto adicionando à linha de comando do dvips o parâmetro `-go` (ou `-G` para habilitar).

3 Contribuição deste artigo

3.1 Os vinte e sete erros mais comuns (de uma lista de cem)

Erros gramaticais e ortográficos devem, por princípio, ser evitados. Alguns, no entanto, como ocorrem com maior freqüência, merecem atenção redobrada. O primeiro capítulo deste manual inclui explicações mais completas a respeito de cada um deles. Veja os cem mais comuns do idioma e use esta relação como um roteiro para fugir deles (Filho, 1992).

1. “Mal cheiro”, “mau-humorado”. Mal opõe-se a bem e mau, a bom. Assim: mau cheiro (bom cheiro), mal-humorado (bem-humorado). Igualmente: mau humor, mal-intencionado, mau jeito, mal-estar.
2. “Fazem” cinco anos. Fazer, quando exprime tempo, é impessoal: Faz cinco anos. / Fazia dois séculos. / Fez 15 dias.
3. “Houveram” muitos acidentes. Haver, como existir, também é invariável: Houve muitos acidentes. / Havia muitas pessoas. / Deve haver muitos casos iguais.
4. “Existe” muitas esperanças. Existir, bastar, faltar, restar e sobrar admitem normalmente o plural: Existem muitas esperanças. / Bastariam dois dias. / Faltavam poucas peças. / Restaram alguns objetos. / Sobravam idéias.
5. Para “mim” fazer. Mim não faz, porque não pode ser sujeito. Assim: Para eu fazer, para eu dizer, para eu trazer.

6. Entre “eu” e você. Depois de preposição, usa-se mim ou ti: Entre mim e você. / Entre eles e ti.
7. “Há” dez anos “atrás”. Há e atrás indicam passado na frase. Use apenas há dez anos ou dez anos atrás.
8. “Entrar dentro”. O certo: entrar em. Veja outras redundâncias: Sair fora ou para fora, elo de ligação, monopólio exclusivo, já não há mais, ganhar grátis, viúva do falecido.
9. “Venda à prazo”. Não existe crase antes de palavra masculina, a menos que esteja subentendida a palavra moda: Salto à (moda de) Luís XV. Nos demais casos: A salvo, a bordo, a pé, a esmo, a cavalo, a caráter.
10. “Porque” você foi? Sempre que estiver clara ou implícita a palavra razão, use por que separado: Por que (razão) você foi? / Não sei por que (razão) ele faltou. / Explique por que razão você se atrasou. Porque é usado nas respostas: Ele se atrasou porque o trânsito estava congestionado.
11. Vai assistir “o” jogo hoje. Assistir como presenciar exige a: Vai assistir ao jogo, à missa, à sessão. Outros verbos com a: A medida não agradou (desagradou) à população. / Eles obedeceram (desobedeceram) aos avisos. / Aspirava ao cargo de diretor. / Pagou ao amigo. / Respondeu à carta. / Sucedeu ao pai. / Visava aos estudantes.
12. Preferia ir “do que” ficar. Prefere-se sempre uma coisa a outra: Preferia ir a ficar. É preferível segue a mesma norma: É preferível lutar a morrer sem glória.
13. O resultado do jogo, não o abateu. Não se separa com vírgula o sujeito do predicado. Assim: O resultado do jogo não o abateu. Outro erro: O prefeito prometeu, novas denúncias. Não existe o sinal entre o predicado e o complemento: O prefeito prometeu novas denúncias.
14. Não há regra sem “excessão”. O certo é exceção. Veja outras grafias erradas e, entre parênteses, a forma correta: “paralizar” (paralisar), “beneficiente” (beneficente), “xuxu” (chuchu), “previlégio” (privilégio), “vultuoso” (vultoso), “cincoenta” (cinquenta), “zuar” (zoar), “frustado” (frustrado), “calcáreo” (calcário), “advinhar” (adivinhar), “benvindo” (bem-vindo), “ascensão” (ascensão), “pixar” (pichar), “impecilho” (empecilho), “envólucro” (invólucro).
15. Quebrou “o” óculos. Concordância no plural: os óculos, meus óculos. Da mesma forma:

Meus parabéns, meus pêsames, seus ciúmes, nossas férias, felizes núpcias.

16. Comprei “ele” para você. Eu, tu, ele, nós, vós e eles não podem ser objeto direto. Assim: Comprei-o para você. Também: Deixe-os sair, mandou-nos entrar, viu-a, mandou-me.
17. Nunca “lhe” vi. Lhe substitui a ele, a eles, a você e a vocês e por isso não pode ser usado com objeto direto: Nunca o vi. / Não o conheci. / A mulher o deixou. / Ela o ama.
18. “Aluga-se” casas. O verbo concorda com o sujeito: Alugam-se casas. / Fazem-se consertos. / É assim que se evitam acidentes. / Compram-se terrenos. / Procuram-se empregados.
19. “Tratam-se” de. O verbo seguido de preposição não varia nesses casos: Trata-se dos melhores profissionais. / Precisa-se de empregados. / Apela-se para todos. / Conta-se com os amigos.
20. Chegou “em” São Paulo. Verbos de movimento exigem a, e não em: Chegou a São Paulo. / Vai amanhã ao cinema. / Levou os filhos ao circo.
21. Atraso implicará “em” punição. Implicar é direto no sentido de acarretar, pressupor: Atraso implicará punição. / Promoção implica responsabilidade.
22. Vive “às custas” do pai. O certo: Vive à custa do pai. Use também em via de, e não “em vias de”: Espécie em via de extinção. / Trabalho em via de conclusão.
23. Todos somos “cidadões”. O plural de cidadão é cidadãos. Veja outros: caracteres (de caráter), juniores, seniores, escrivães, tabeliães, gangsteres.
24. O ingresso é “gratuito”. A pronúncia correta é gratuito, assim como circuito, intuito e fortuito (o acento não existe e só indica a letra tônica). Da mesma forma: flúido, condôr, recôrde, aváro, ibéro, pólipó.
25. A última “seção” de cinema. Seção significa divisão, repartição, e sessão equivale a tempo de uma reunião, função: Seção Eleitoral, Seção de Esportes, seção de brinquedos; sessão de cinema, sessão de pancadas, sessão do Congresso.
26. Vendeu “uma” grama de ouro. Grama, peso, é palavra masculina: um grama de ouro, vitamina C de dois gramas. Femininas, por exemplo, são a agravante, a atenuante, a alface, a cal, etc.

27. “Porisso”. Duas palavras, por isso, como de repente e a partir de.

3.2 Observação importante

Os erros de português porventura presentes neste artigo

- ou foram introduzidos propositalmente para deleite dos leitores atentos,
- ou não constam da lista da seção 3.1.

4 Conclusões

Liste suas conclusões nesta seção, em vez de simplesmente relatar o que foi feito.

Agradecimentos

Mencione aqui seus agradecimentos às agências de fomento e aos colaboradores do trabalho.

Referências

- de Andrade, O. (1933). *Serafim Ponte Grande*, Editora Globo.
- Filho, E. L. M. (1992). *Manual de Redação e Estilo*, Maltese.
- Gossens, M. and Mittelbach, F. (1993). *The L^AT_EX Companion*, Addison-Wesley, Reading, MA.
- Heller, J. L. (1996). *Catch - 22*, reprint edn, Scribner.
- Kopka, H. and Daly, P. W. (1993). *A Guide to L^AT_EX: Document Preparation for Beginners and Advanced Users*, Addison-Wesley.
- Lamport, L. (1986). *L^AT_EX: A Document Preparation System*, Addison-Wesley.
- Williams, P. and Schmier, T. (1998). The harvard family of bibliography styles. Documentação que acompanha o pacote `harvard`.