

MUSIC4U : Sistema de recomendação de músicas

Obede Carvalho¹, Otávio Lara¹

¹ Departamento de Ciência da Computação - Universidade Federal de Lavras (UFLA)
Caixa Postal 3037 - 37200-000 - Lavras(MG) - Brasil

{obedejc,otaviolara}@computacao.ufla.br

Resumo. *Nesse artigo é descrita a proposta da ferramenta Music4U, um sistema de recomendação que usa técnicas de mineração de texto, através de bibliotecas poderosas, visando facilitar a descobertas de novas músicas pelos usuários.*

Abstract. *In this paper, it is described the proposal of the tool Music4U, a recommendation system that uses text mining techniques*

1. Introdução

Em um mundo cada vez mais globalizado, temos acesso a muitas músicas, vindas de diversas origens, com diferentes temas e diversos artistas. Cada música tem suas características que a irá diferenciar de outras músicas, mas mesmo assim muitas vezes as músicas que nos chamam mais atenção são aquelas que tratam de um tema específico, mas há uma infinidade de música, não fácil encontrar aquelas que mais se adequam ao que gostamos.

Com base nisso foi proposto nesse artigo a ferramenta Music4U, que com base em músicas que um usuário já tenha ouvido e gostado, serão sugeridas aquelas que podem ser do seu agrado. Para isso foram empregadas diversas técnicas de mineração de texto que tornaram tal processo mais preciso.

Com essa ferramenta espera-se que a descoberta de novas músicas sejam mais precisa, assim como músicas que não o agradaria já serem descartas. Esse artigo descreve tal ferramenta, na seção 2 serão mostrados conceitos básicos para o entendimento desse artigo, na seção 3 será apresentado a implementação e funcionamento de tal ferramenta, e encerraremos na seção 4.

2. Background

Nessa seção serão apresentados um pouco sobre conceitos úteis para o melhor entendimento desse artigo. Serão apresentados o que é *Web Scraping*, Sistemas de recomendação, a técnica para mineração de texto Naive Bayes, e a aplicação WEKA.

2.1. Web Scraping

Web Scraping mais conhecida como mineração de web sites, geralmente utilizada para recolher informações de forma automatizada de uma ou várias páginas web, um exemplo são os *bot* ou *spiders* da *Google* que ficam navegando pelas páginas para indexar e atualizar seu enorme banco de dados de informações de sites do mundo todo.

No caso desse artigo, utilizamos uma biblioteca de *Python* de alto nível com capacidade de navegar de maneira simples utilizando o protocolo *HTTP*(*Hypertext Transfer*

Protocol), padrão da web para requisições de serviços. *Beautiful Soup* é uma biblioteca completa para mineração de *HTML(HyperText Markup Language)* que com hierarquia de *tags* e com buscas nos elementos das páginas estáticas tem a capacidade de retirar os dados de sites prontos para serem transformados em informações depois com algum algoritmo de mineração de dados.

O site escolhido para ser minerado foi um dos maiores sites de músicas do Brasil: *Letras.mus[let]* e dentro desse ainda selecionamos as 1000 (mil) melhores músicas de MPB e a partir daí foi rodado o algoritmo de Web Scraping para baixar todas as letras e coloca-las em uma estrutura de dados para ser trabalhada.

2.2. Sistemas de recomendação

Um Sistema de recomendação é a combinação de vários algoritmos de DM (Data Mining), IR (Information Retrieval), AI (Artificial Intelligence), cada um com seu papel fundamental na escolha da melhor recomendação possível para usuários. Geralmente o sistema de recomendação é utilizado em web sites de e-commerce, músicas, links patrocinados, vídeos, anúncios e muitas outras aplicações onde a recomendação pode gerar algum lucro ou mais informações úteis para um usuário.

Neste artigo foi utilizado mineração de texto para coletar letras de músicas e criar um sistema capaz de reconhecer e recomendar uma música de acordo com o histórico das últimas ouvidas por um usuário.

A técnica de recomendação utilizada é a baseada em conteúdo que é feita com base na comparação entre o conteúdo/descrição de um conjunto de itens e o perfil de um dado usuário.

2.3. WEKA

A biblioteca WEKA é uma poderosa fonte de algoritmos uteis para diversos propósitos dentro dos campos de mineração de dados, inteligência artificial e qualquer outra área que atue sobre grandes quantidades de dados. O WEKA é uma ferramenta livre muito utilizada no meio acadêmico, foi desenvolvida dentro do ambiente acadêmico na universidade de Waikato, na Nova Zelândia.

A biblioteca foi desenvolvida na linguagem de programação JAVA, permitindo que seja distribuída de maneira fácil, e possa ser importada dentro de outros projetos em JAVA.

A ferramenta disponibiliza módulos para lidar com arquivos de dados, diversos arquivos para limpeza de dados, classificação, clusterização, associação e visualização de dados.

3. Music4U

Nessa seção será apresentado a ideia da aplicação *Music4U*, assim como sua implementação e seu funcionamento.

3.1. Ideia

O presente trabalho tem como objetivo permitir que um usuário conheça novas músicas de acordo com as músicas que ele gostou em um outro momento. Para que seja feita

a escolha de uma música que mais se adeque aos gostos dos usuário, é realizado uma avaliação sobre as letras de músicas e aquelas que mais se assemelhem às músicas já ouvidas pelo usuário são então recomendadas.

Para que seja feita a comparação de semelhança entre letras de música é utilizada a técnica descrita na seção ??, chamada Naive Bayes. Tal modelo se mostrou mais adequado para o propósito da aplicação, visto que através dele é calculado a possibilidade de uma música se adequar aos gostos do usuário ou não. O Naive Bayes tenta inferir quais palavras tem maior influência para que o usuário goste de uma música.

Para que o Naive Bayes seja treinado é necessário que sejam inseridos dados históricos do usuário e também se esses dados, no caso letras de músicas, agradaram ou não ao usuário. Com base nisso que o modelo poderá fazer novas previsões.

De acordo com as novas letras recomendadas para o usuário, o ideal é que o modelo se adapte cada vez mais aos gostos do usuário, para que com o tempo as previsões se tornem cada vez mais próximas as expectativas do usuário para uma nova música.

3.2. Implementação

Para que fosse implementado o Music4U foi utilizado como base o WEKA, descrito na seção 2.3, que já possui os algoritmos necessários para a classificação utilizando o Naive Bayes, além de outras funcionalidades importantes que foram também utilizadas na aplicação.

A implementação de Naive Bayes utilizada foi a Naive Bayes Multinomial Text (NaiveBayesMultinomialText) que como o próprio nome já diz atua sobre dados textuais. Para a melhor funcionamento de tal classe foram configurados os parâmetros mínimo de frequência de uma palavra para que apenas palavras significativas fossem analisadas, além disso foram removidos palavras não úteis através do uso de *stopwords*. Também é levado em conta o número de vez que uma palavra aparece em um texto, para que palavras mais repetidas tenham maior peso que as demais.

A aplicação possui duas entradas, a primeira um conjunto de letras que representam o histórico do usuário, assim como a avaliação do usuário de cada uma dessas músicas (gostou ou não). A segunda entrada esperada é um conjunto de letras de músicas ainda não exploradas pelo usuário, tais letras passaram pelo modelo treinado usando a primeira entrada. As músicas que possuem maior probabilidade de se adequarem aos gostos do usuário são então mostradas para ele.

A medida que o usuário avalia as músicas recomendadas para ele como gostei/não gostei, o modelo usado para avaliar as recomendações é então atualizado, com isso espera-se que o modelo evolua com o passar do tempo.

3.3. Funcionamento

Para melhor utilização do usuário e tal ferramenta foi feita um tela onde o usuário visualiza as músicas recomendadas e onde o usuário pode dizer se gostou ou não da sugestão. A medida que é o usuário avalia as sugestões, novas sugestões vão surgindo com base em suas novas avaliações que atualizam o modelo de mineração de texto gerado.

4. Conclusão

Com diversos serviços que disponibilizam músicas para os usuários, e a infinidade de seus catálogos, nem sempre é fácil descobrir novas músicas que nos interessam, o que faz com que percamos tempo descartando músicas que não foram úteis. Diante disso foi proposto nesse trabalho a ferramenta Music4U, que faz recomendações de novas músicas que mais se adequam ao gosto do usuário, com base nas letras de tais músicas.

Para desenvolver essa ferramenta, foi feito o uso de técnicas de mineração web, mineração de texto, e a utilização de renomadas bibliotecas de mineração de textos. Foi utilizada a técnica Naive Bayes que trás um análise estatística da probabilidade de que determinada letra seja agradável ao usuário. Para a utilização da ferramenta, foi feita uma interface gráfica amigável, que permite ao usuário avaliar as sugestões de músicas.

Referências

[let] Letras músicas. <https://www.letras.mus.br/>. Accessed: 2017-08-14.