Universidade Federal de Pernambuco

Centro de Informática
Graduação em Ciência da Computação

Um Sistema de Recomendação de Promoções Baseado em Posts no Twitter

Luiz Antônio Correia

lac2@cin.ufpe.br

Trabalho de Graduação

Recife, Dezembro de 2011

Universidade Federal de Pernambuco

Centro de Informática
Graduação em Ciência da Computação

Um Sistema de Recomendação de Promoções Baseado em Posts no Twitter

Luiz Antônio Correia

lac2@cin.ufpe.br

Trabalho de Graduação apresentado ao Centro de Informática da Universidade Federal de Pernambuco como requisito para obtenção do Grau de Bacharel em Ciência da Computação.

Orientadora: Flávia de Almeida Barros

fab@cin.ufpe.br

Recife, Dezembro de 2011

Agradecimentos

A minha família por ter acompanhado de perto todo o meu trajeto, desde o princípio deste ciclo até o encerramento.

A minha amiga Clara por me ajudar a escolher o curso.

Aos amigos que fiz durante o curso. Em particular, a Jonathan (js2) e a Osman (otxj) por me ajudarem em momentos decisivos, sempre muito prestativos.

A minha orientadora Flávia pela paciência e atenção com que me ajudou a realizar este trabalho.

Enfim, agradeço a todos que, de alguma forma, contribuíram para que esse trabalho fosse realizado. A todos vocês, a minha sincera gratidão.

Aos meus dois queridos irmãos:

A Isaac por toda confiança que depositou em mim e

A Daniel por estar comigo em todos os momentos que precisei.

Resumo

Atualmente o Twitter é uma das maiores e mais importantes redes sociais. Estima-se que a quantidade de usuários cadastrados no Twitter seja superior a 301 milhões, aproximadamente o tamanho da população dos Estados Unidos [12].

Explorando tais características, muitas empresas utilizam esse sistema para divulgarem seus produtos e serviços, oferecendo promoções. Esse tipo informação tem atraído uma grande quantidade de usuários, pois são significativos os benefícios que podem ser obtidos através dessas promoções.

Contudo, os posts de promoções não são característicos e, nem tão pouco, o sistema oferece uma interface para o usuário filtrar itens de promoção em que ele tem interesse. Assim, para o usuário verificar se a promoção postada o interessa de fato, é necessário verificar "manualmente" os detalhes da promoção. Tal verificação pode ser muito cansativa, pois é muito grande a quantidade de postagens (muitas vezes irrelevantes para este usuário) que é gerada diariamente neste sistema.

Entretanto, o tempo que um usuário gasta executando manualmente esta atividade pode ser substancialmente reduzido se ele utilizar um sistema capaz de extrair do texto as informações referentes aos produtos, classificá-las e exibir para ele apenas os produtos de seu interesse.

Este trabalho tem por objetivo apresentar o *twitterRecommender*, um sistema de recomendação de promoções baseado em posts do Twitter, que classifica os posts e os indica de acordo com os interesses dos usuários finais, de maneira automática e inovadora. Um protótipo do *twitterRecommender* foi desenvolvido e foram realizados alguns experimentos para validar a sua precisão.

Palavras-chave: Sistemas de Recomendação, Filtragem de Informação, Classificação Automática, *twitterRecommender*

Abstract

Today the Twitter is one of the largest and most important social networks. It is estimated that the number of registered Twitter users exceeds 301 million, roughly the size of the U.S. population [12].

Because of that the twitter is becoming a good platform for promoting and marketing of businesses and services especially through promotions. Such information has attracted a large number of users, as are the significant benefits that can be obtained through these promotions.

However the promotion posts are not specific and neither the system offers an interface that allows the user to filter promotional items of its interest. Thereby the user has to check “manually” the promotion details, which can be very tiring because of the largest number of posts (often irrelevant to this user), generated daily in this system.

Nevertheless the time a user spends manually performing this activity can be substantially reduced if it was used a system capable of extracting information regarding the products from text, sort them and display only the products of its interest.

This paper aims to present the *twitterRecommender*, a Recommendation System of promotions based in Twitter posts. It classifies them and indicates them automatically and by an innovative way according to the interests of end users. A prototype has been developed and some experiments were performed to validate its accuracy.

Key-words: Recommendation Systems, Informational Filtering, Automatic Classification, *twitterRecommender*

Sumário

1. Introdução

Desde o início da Web 2.0, seus usuários podem não só acessar informações, como também expressar e compartilhar suas opiniões e conhecimentos, o que tornou a Internet um ambiente mais livre e democrático [3]. Alguns exemplos de aplicações que estão inclusas nesta abordagem são: Twitter[1], Orkut[2], Wikipédia[3], Flickr[4] e Websites sucessores do Napster[5].

Assim, a quantidade de informação na Web tem crescido substancialmente, e a combinação desse "mar" de informações com a natureza pouco estruturada da Web tem dificultado muito a atividade de encontrar informações precisas sobre um item ou tema desejado. Estima-se que o número de informações disponíveis na Web é tão grande que um ser humano não conseguiria analisá-la em toda a sua existência [1].

Os usuários comuns se sentem perdidos diante de tanta informação, e desestimulados diante de tanta dificuldade. Encontrar o que interessa então ficou a mercê de esforço próprio, da sorte, ou de uma indicação de outro usuário [8].
Para minimizar esses entraves, os usuários costumam utilizar algumas recomendações disponíveis na própria Internet. As maneiras mais comuns e simples de se encontrar essas recomendações são: (1) em listas de recomendação de alguns sites, tais como lojas online e blogs (como listas dos 10 livros mais vendidos, por exemplo); (2) em fóruns (como o Yahoo! Groups[6], por exemplo); (3) ou por meio de sistemas de recomendações (como o CineDica[7], por exemplo). Para tanto, geralmente são utilizados Sistemas de Recuperação de Informação e Sistemas que se utilizam de técnicas de personalização, baseados em Filtragem e Recomendação de itens.

1.1. Motivação e Relevância

Atualmente o Twitter é uma das maiores e mais importantes redes sociais. Estima-se que a quantidade de usuários cadastrados no Twitter seja superior a 301 milhões, aproximadamente o tamanho da população dos Estados Unidos [12].

Explorando tais características, muitas empresas utilizam esse sistema para divulgarem seus produtos e serviços, oferecendo promoções. Esse tipo informação tem atraído uma grande quantidade de usuários, pois são significativos os benefícios que podem ser obtidos através dessas promoções.

---

[1] www.twitter.com
[2] www.orkut.com
[3] www.wikipedia.org
[4] www.flickr.com
[5] www.napster.com
[6] groups.yahoo.com
[7] www.cinedica.com.br

Contudo, os posts de promoções não são característicos e, nem tão pouco, o sistema oferece uma interface para o usuário filtrar itens de promoção em que ele tem interesse. Assim, para o usuário verificar se a promoção postada o interessa de fato, é necessário verificar "manualmente" os detalhes da promoção. Tal verificação pode ser muito cansativa, pois é muito grande a quantidade de postagens (muitas vezes irrelevantes para este usuário) que é gerada diariamente neste sistema.

Entretanto, o tempo que um usuário gasta executando manualmente esta atividade pode ser substancialmente reduzido se ele utilizar um sistema automático capaz de extrair do texto as informações referentes aos produtos, classificá-las e exibir para ele apenas os produtos de seu interesse.

1.2. Objetivos e Solução Proposta

Este trabalho teve como objetivo principal desenvolver um método para construir um Sistema de Recomendação, baseado em técnicas de Filtragem de Informação e Classificação Automática.

O método aqui proposto teve como ponto de partida um classificador de posts de promoção do Twitter desenvolvido na disciplina de Mineração Web do Centro de Informática - UFPE. Tal método estende este classificador, sendo capaz de extrair informações sobre quais são os itens que estão em promoção nos posts e recomendá-los para os usuários que têm interesse em receber notificações para o item em questão.

A solução proposta foi desenvolvida, resultando no protótipo *twitterRecommender*. Ressaltamos que não foi encontrado na literatura nenhum Sistema de Recomendação que faça indicações de maneira semelhante ao *twitterRecommender*.

1.3. Organização do Trabalho

Este trabalho está dividido em 5 capítulos:

O capítulo 1 é a parte introdutória do trabalho onde são apresentados o contexto de Sistemas de Recomendação e a inserção da solução proposta neste contexto.

O capítulo 2 apresenta brevemente os principais conceitos de Sistemas de Recomendação, bem como as técnicas e estratégias mais utilizadas.

O capítulo 3 descreve a solução proposta, o *twitterRecommender*. Neste capítulo, são apresentados módulos que compõem a solução, descrevendo as técnicas e procedimentos utilizados para cada um desses módulos.

O capítulo 4 apresenta os experimentos realizados para aviar o t*witterRecommnender*. Também são apresentados como os resultados foram obtidos e a análise dos resultados finais.

O capítulo 5 contém a conclusão deste trabalho, ressaltando as contribuições e apontando as dificuldades encontradas e alguns possíveis trabalhos futuros.

2. Técnicas e Estratégias de Recomendação

Este capítulo tem como objetivo apresentar os fundamentos e alguns exemplos de sistemas utilizados para auxiliar usuários Web na busca e recuperação de informações relevantes para sua necessidade de informação. Como foco principal, serão vistas técnicas de personalização usando Filtragem de Informação e Sistemas de Recomendação de itens que utilizam a filtragem como parte da sua estratégia de recomendação.

A seção 2.1 apresenta conceitos básicos sobre sistemas de RI; a seção 2.2 traz uma descrição sobre a área de Filtragem de Informação, com exemplos de aplicações; e por fim a seção 2.3 apresenta estratégias e exemplos de sistemas para Recomendação de Itens.

2.1. Recuperação de Informação

Sistemas de Recuperação de Informação (SRI) são responsáveis por criar uma representação dos itens de informação em questão, bem como pelo armazenamento, a organização e o acesso a esses itens de informação [9].

Os Sistemas de Recuperação de Informação mais utilizados pelos usuários são os engenhos de busca disponíveis na Web, tais como o Google[8] e Yahoo[9] [1].

Engenhos de Busca na Web

Para utilizar um engenho de busca, o usuário deve digitar sua consulta na interface disponível, e então o software retorna um conjunto de dados em forma de links. Vale lembrar que o usuário precisa transformar a sua necessidade de informação em um conjunto de palavras-chave, para que sua consulta possa ser processada pelo motor de busca [9].

Engenhos de busca geralmente utilizam uma arquitetura de rastreamento-indexação centralizada. Essa arquitetura consiste em dois módulos principais: um que lida com os usuários e o mecanismo de consulta; e outro que consiste nos sub-módulos rastreador e indexador. Nessas arquiteturas, o módulo de rastreamento, também conhecido como *Web Crawler*, é responsável por coletar as páginas da Web que devem fazer parte da base de dados do sistema. Feito isso, as páginas coletadas são enviadas para o servidor principal, onde são realizadas algumas operações sobre o texto (análise léxica, eliminação de *stopwords*, operações de *stemming* e identificação de grupos nominais). A seguir, as informações já tratadas são indexadas, de forma que possam ser recuperadas em uma consulta feita pelo usuário, caso tal informação seja relevante para a consulta [9].

Apesar dos engenhos de buscas ajudarem substancialmente os usuários, esses sistemas apresentam duas grandes desvantagens. A primeira delas é o fato do usuário ser obrigado a "traduzir" sua necessidade de informação em termos de palavras-chave, pois nem

[8] www.google.com e www.scholar.google.com
[9] www.yahoo.com

sempre usuários leigos sabem fazer isso corretamente. A segunda desvantagem acontece devido ao vasto número de informações (em forma de *links*) recuperadas em uma determinada consulta, pois, apesar do bom ranqueamento que os engenhos de busca têm se empenhado em fazer, geralmente apenas os primeiros links (as informações mais bem ranqueadas) apresentam dados relevantes para o usuário [1].

Além dessas desvantagens, esses sistemas não costumam levar em consideração o perfil do usuário para ranquear o resultado da consulta, realizando uma recuperação *ad hoc*. A fim de minimizar esse problema, alguns pesquisadores e desenvolvedores lançam mão de técnicas de personalização, via Filtragem de Informação.

2.2. Filtragem de Informação

Como o que foi discutido na seção 2.1., os sistemas convencionais de RI, em geral, realizam recuperação *ad hoc*, tratando consultas de usuários distintos da mesma maneira. Já a Filtragem de informação trata de forma distinta consultas iguais de usuários diferentes, pois leva em conta o perfil de cada usuário [9]. A tarefa de filtragem é, portanto, indicar os itens que são relevantes para um determinado usuário final considerando seu perfil.

Pelo que foi exposto no parágrafo anterior, percebemos que, para que seja possível indicar itens em Filtragem, é necessário que o indivíduo seja identificado através de seu perfil, no qual são informados seus interesses [2].

As duas maneiras mais usadas de identificação de usuários de um sistema são [4]:

1) **Identificação no servidor:** o sistema geralmente solicita, por meio de uma interface de cadastro, que o usuário informe seus dados pessoais, tais como: email, sexo, nome, CPF e data de nascimento. Além disso, essa abordagem torna obrigatório o uso de login e senha, para que possa identificar precisamente o indivíduo. Essas informações ficam armazenadas em um servidor e o usuário pode acessar seu perfil por meio do login e da senha criados no cadastro [4]. Ver Figura 2.1 a seguir.

Figura 2.1: exemplo de interface para cadastramento de usuários [www.youtube.com].

2) **Identificação no cliente:** Essa abordagem costuma utilizar *cookies* [4] - mecanismo utilizado para informar que dados foram trocados entre uma máquina e um servidor de páginas, criando um arquivo de texto no computador do usuário [6]. Esse método costuma não ser muito eficiente, pois assume que o computador é utilizado sempre pela mesma pessoa [4]. Ver Figura 2.2 a seguir.

Figura 2.2: exemplo de sistema que usa *cookies* para identificar usuários [www.youtube.com]

Após ter identificado usuário, é possível coletar seus interesses, a fim de montar seu perfil. Essa coleta pode ser feita forma *implícita* ou *explícita*.

1) **Coleta explícita:** Na coleta explícita, o usuário informa sobre qual produto, serviço ou pessoa ele tem interesse [4]. A figura 2.3 abaixo exibe um exemplo de uma aplicação que realiza esse tipo de coleta.

Figura 2.3: Exemplo de interface para coleta explícita de perfil. [http://www.peixeurbano.com.br/conta/Criar]

2) **Coleta implícita:** Na coleta implícita de perfil, os dados são coletados a partir de ações executadas pelo o usuário no sistema, como a compra de um livro, por exemplo [4]. Perceba que na figura 2.2, os vídeos estão sendo recomendados mesmo sem que o usuário tenha efetuado login no sistema. Isso é possível porque o sistema realiza a identificação no cliente, por meio de *cookies*, e também faz coleta implícita dos vídeos que foram acessados pelo usuário.

Nas seções subseqüentes, faremos uma breve descrição de duas técnicas Filtragem de Informação: Filtragem Baseada em Conteúdo e Filtragem Colaborativa.

2.2.1. Filtragem Baseada em Conteúdo

Em filtragem baseada em conteúdo, os itens do sistema geralmente possuem atributos com diferentes valorações e estão associados a classes pré-definidas. Para fazer a indicação, o sistema compara os interesses constantes no perfil do usuário com os valores dos atributos de cada item cadastrado no sistema, e, em seguida, indica os itens que "casam" com o seu perfil [6].

Antes de prosseguir, vamos apresentar uma breve explicação, através de um exemplo simples, dos conceitos de **classe**, **atributo** e **valores de atributos**, pois, a partir de agora, essas terminologias serão aqui muito utilizadas.

Vamos supor que em um sistema de uma loja on-line sejam vendidos CDs, livros e revistas, entre outros tipos de produto. Os livros possuem informações sobre o ano de publicação, autor e categoria.

Nesse exemplo, (1) as classes do sistema seriam: CDs, livros e revistar. (2) Os atributos seriam ano de publicação, autor e gênero. (3) Por fim, os valores dos atributos seriam para ano: 2000 ou 2009 ou 2011...; para categoria: “ficção científica”, “infanto-juvenil”, “literatura policial”...; para autor: “Machado de Assis”, “João Cabral de Melo Neto”, “Carlos Drummond de Andrade”... .

Prosseguindo com Filtragem Baseada em Conteúdo, essa técnica se caracteriza por fazer indicação de maneira indistinta para todos os usuários (isto é, todos os usuários recebem as mesmas indicações para um determinado valor de um atributo de um item do sistema no qual possuem interesse).

Devido ao modo que essa técnica utiliza para coletar as informações dos usuários, as recomendações realizadas não costumam surpreendê-los. Na maioria das vezes, o indivíduo poderia tê-la inferido sozinho [7].

Pelo o que foi exposto nesta subseção, podemos perceber que, para que os itens sejam indicados por Filtragem Baseada em Conteúdo, é necessário que tais itens já tenham sido previamente associados a classes que os valores de seus atributos já tenham sido informados.

Essa tarefa é chamada de classificação e pode ser feita de duas maneiras: manualmente ou automaticamente. A seguir, apresentaremos uma descrição para cada uma delas:

2.2.1.1. Classificação Manual

Em classificação manual, os itens são classificados manualmente por um especialista do domínio do sistema. Esse tipo de classificação costuma ter uma ótima precisão. Contudo, uma vez que a classificação é realizada por humanos, quando a base de itens do sistema é muito grande, o trabalho mantê-la atualizada torna-se muito difícil.

2.2.1.2. Classificação Automática

Sistemas de Recomendação também podem utilizar um classificador automático para descobrir a quais classes um item está associado. Uma vez que esses itens são classificados, o sistema pode agrupá-los, filtrá-los e indicá-los para os usuários de acordo os interesses de cada um.

É possível construir um classificador automático de duas formas: manualmente e automaticamente.

Construção Manual do Classificador

A construção manual de classificadores é baseada em conhecimento explícito usando-se regras de classificação (Sistemas Baseados em Conhecimento). Os componentes básicos desses

sistemas são: (1) Uma base de conhecimento, que contém as regras de classificação; e (2) Uma máquina de inferência. A base de conhecimento é construída manualmente por um especialista no domínio da aplicação.

Comparada com a construção automática de classificadores, o classificador geralmente é executado de maneira mais rápida com a construção manual. Essa técnica, entretanto, tem a desvantagem de necessitar de um especialista no domínio da aplicação, pois o trabalho de manter a base de regras atualizada por humanos é uma tarefa muito difícil de ser feita.

Construção Automática do Classificador

Dentro dessa abordagem, o classificador é construído com base em técnicas de aprendizagem de máquina. Utiliza-se algum algoritmo de aprendizagem para induzir regras que serão capazes, no futuro, de classificar corretamente os itens. Existem dois tipos de algoritmos de aprendizagem de máquina: (1) supervisionado e o (2) não-supervisionado.

Algoritmo de Aprendizagem Supervisionada

Utilizamos algoritmos de aprendizagem de máquina supervisionada quando temos conhecimento prévio das classes do sistema. A construção de um classificador com essa abordagem possui basicamente as seguintes fases [10]:

- Primeiramente, é necessário que um especialista escolha um conjunto de exemplo, contendo um subconjunto que possui determinada classe e outro que não possui.

- Feito isso, segue-se para fase de treinamento, na qual é utilizado um algoritmo/técnica de aprendizagem de máquina. Essa fase tem como objetivo gerar uma coleção de regras de classificação baseadas na análise do conjunto de treinamento fornecido.

Os algoritmos/técnicas mais utilizados nesta etapa são:

- KNN
- Classificador Linear (Rocchio, por exemplo)
- Naive Bayes
- Árvores de Decisão
- Redes Neurais
- Sopport Vector Machine (SVM)

Ao término da etapa de treinamento, são feitas validações para verificar a capacidade de generalização das regras de classificação que foram geradas e, se os resultados não forem satisfatórios, os parâmetros do algoritmo são ajustados e então se retorna à etapa anterior; quando os resultados obtidos são satisfatórios, obtém-se, enfim, um classificador.

Algoritmo de Aprendizagem não-Supervisionada

De maneira oposta à aprendizagem supervisionada, utiliza-se o algoritmo de aprendizagem não-supervisionada quando não se tem conhecimento prévio das classes do sistema. O algoritmo aprende a generalizar sem a intervenção do usuário, baseando-se em observações e descobertas de padrões dos dados, de modo a agrupar elementos similares (formação de *clustering*), encontrando, dessa forma, as classes do sistema às quais os itens estão associados [11].

2.2.2. Filtragem Colaborativa

Essa técnica consiste em filtrar itens para um usuário baseando-se em experiências de outros usuários com gostos similares [7]. A essência dessa técnica está na troca de experiências entre os seus usuários [1], assumindo a idéia de que pessoas com mesmos gostos possuem também os mesmos interesses.

Alguns sistemas conseguem descobrir automaticamente as relações entre os usuários por meio de seus padrões de comportamento (indivíduos com gostos mais parecidos são colocados mais "próximos"). Dessa forma, é possível criar comunidades de usuários, o que permite fazer indicações mais eficientes, pois o sistema consegue indicar itens que interessam aos indivíduos com gostos semelhantes, mas que não acessaram ainda [2].

O quadro 2.1 abaixo mostra um exemplo simples de como essa filtragem pode ser usada [2]. Nesse exemplo, queremos indicar um livro para Mauro. Como queremos utilizar a Filtragem Colaborativa, devemos primeiramente identificar outras pessoas que tenham comportamento parecido ao dele no sistema. Examinado a tabela, podemos perceber que Paulo e João compraram um livro que Mauro também comprou. Em seguida, indicamos a Mauro livros que Paulo e João compraram, mas que ele ainda não comprou. Assim, indicaremos o item Livro1 (comprado por João) e o item Livro5 (comprado por Paulo).

Tabela 2.1: Exemplo de recomendação baseada em filtragem colaborativa

Usuário	**Livro1**	**Livro2**	**Livro3**	**Livro4**	**Livro5**	**Livro6**
Paulo		X			x	
João	x	X				
Márcia			X	X	x	
Carlos			X			
Ana	x			X		
Mauro		X				

A Filtragem Colaborativa se diferencia da Filtragem Baseada em Conteúdo pelo fato de não necessitar do conhecimento automático do conteúdo do item em questão [4]. Essa técnica possui a grande vantagem de oferecer mais novidades para seus usuários, por fazer boas indicações de itens não óbvios [7]. Além disso, é possível criar comunidades de usuários que possuem preferências e gostos parecidos [2].

A figura 2.4 a seguir ilustra uma recomendação de um site que utilizando essa abordagem.

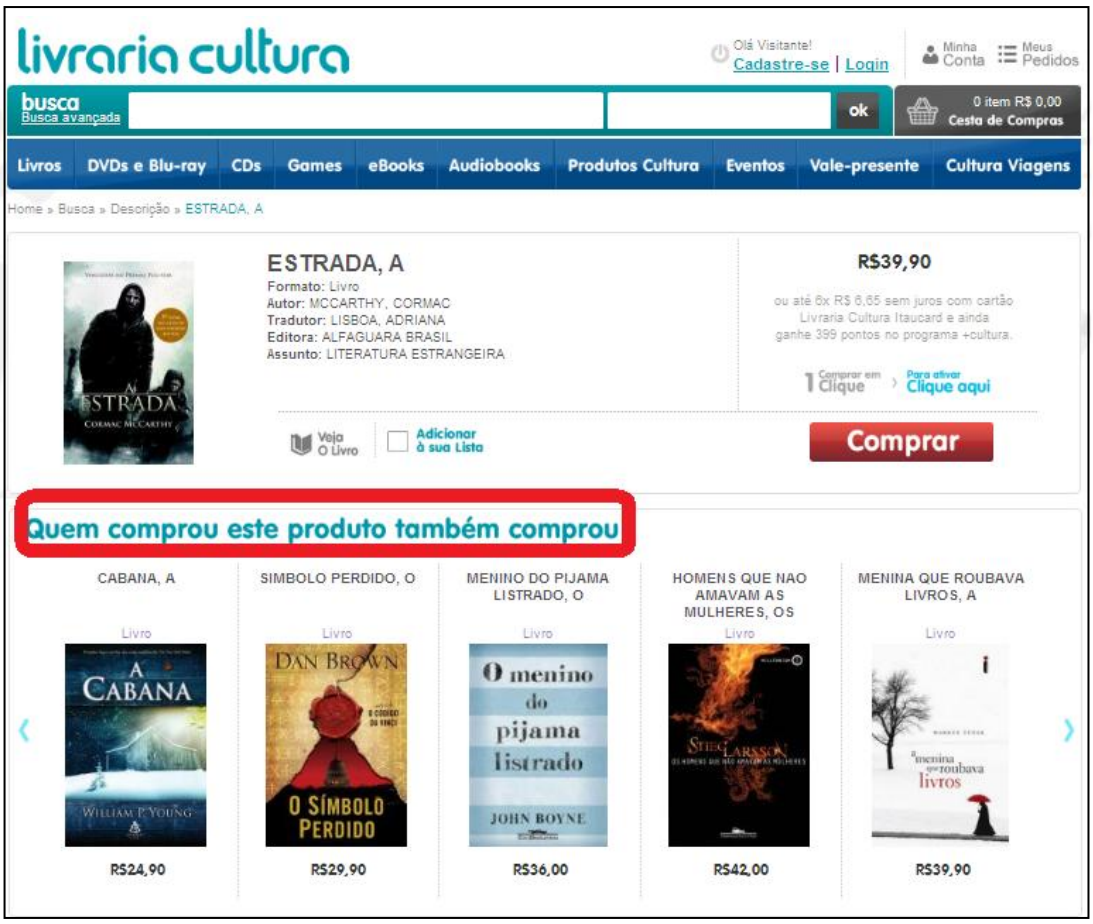


Figura 2.4: exemplo de sistema que faz filtragem colaborativa
[http://www.livrariacultura.com.br]

Segundo [2], apesar dos benefícios, essa abordagem também possui algumas limitações:

1. **O problema do primeiro avaliador**: uma vez que essa abordagem se baseia em indicações através de experiências de outros usuários, um item novo no sistema pode ficar um longo tempo sem que seja recomendado a ninguém até que o primeiro usuário o avalie.
2. **O problema do usuário novo**: pela mesma razão do problema anterior, o sistema encontrará dificuldades em indicar itens para um usuário que ainda não realizou nenhuma ação no sistema.

3. **O problema de pontuações muito esparsas**: quando a quantidade de itens no sistema é muito maior que a quantidade de usuários, as avaliações dos itens podem ficar muito esparsas, o que pode dificultar as indicações.
4. **O problema singularidade**: usuários que tenham gostos singulares podem ter dificuldades para receber boas recomendações.

2.2.3. Filtragem Híbrida

A Filtragem Híbrida combina os pontos positivos da Filtragem Colaborativa e da Filtragem Baseada em Conteúdo a fim de corrigir as defasagens que essas técnicas causam quando são utilizadas isoladamente.

Como mostrado nas duas subseções anteriores, a Filtragem Baseada em Conteúdo tem a desvantagem de indicar itens para cada usuário de maneira incomum; enquanto a Filtragem Colaborativa tem dificuldades tanto em indicar itens novos no sistema, como em filtrar informações para usuários que não executaram nenhuma ação no sistema. Ao fundir essas duas técnicas, a Filtragem Híbrida elimina esses problemas da Filtragem Colaborativa, utilizando as técnicas da Filtragem Baseada em Conteúdo (1) de inserir as informações dos conteúdos nos itens e (2) permitir que usuários informem sobre quais valores de atributos eles possuem interesse – o que torna possível tanto indicar itens a usuários novos como indicar itens novos a usuários no sistema. O problema de avaliações muito esparsas e o problema de usuários com perfis singulares também são resolvidos de maneira similar.

2.3. Sistemas de Recomendação

Como visto acima, a tarefa de filtragem se resume a indicar (ou filtrar, selecionar) os itens que "casam" com o perfil de um determinado usuário, não ficando ao seu encargo fazer o ranqueamento dessas informações, o que ainda pode sobrecarregar o usuário [9].

Para entender melhor, vamos supor que um usuário se inscreveu em um site de compartilhamento de vídeos para receber notificações sobre novos vídeos relacionados a jazz. Vamos supor agora que centenas de vídeos foram adicionados ao sistema em um só dia. Apenas usando filtragem, este usuário receberia notificações sobre todos esses novos vídeos na ordem em que foram inseridos na base, i.e., sem um ranqueamento personalizado. Assim, possivelmente o usuário não teria tempo de verificar todos os vídeos e descobrir quais eram os mais relevantes para ele.

O Sistema de Recomendação, portanto, é responsável por utilizar critérios para ranquear os itens que passaram por um filtro e, então, indicar aqueles que são mais relevantes para o usuário, de modo a reduzir a sobrecarrega de informação. Além disso, Sistemas de Recomendação permitem que os usuários recebam notificações de maneira pró-ativa, sem que seja necessário que o usuário acesse o sistema para poder visualizar as recomendações.

2.3.1. Estratégias de Recomendação

Existem diversas estratégias para se fazer esse ranqueamento dos itens filtrados. Veremos aqui conceitos básicos sobre essa tarefa, bem como algumas das estratégias mais usadas no ranqueamento de itens.

Para melhor entendimento das estratégias de recomendação, é importante que sejam analisados os graus de personalização que uma recomendação pode ter. Quanto ao grau de personalização, a recomendação pode ser [8]:

1. **Não personalizada:** Quando a mesma recomendação é feita para todos os usuários igualmente;

2. **Efêmera:** Quando o sistema realiza indicações com base apenas nas ações/informações do usuário relativas ao instante em que o indivíduo está acessando o sistema (a cada acesso individual ao sistema, sem levar em conta seu perfil e o histórico de ações desse usuário); ou

3. **Persistente**: Quando o sistema utiliza informações armazenadas no perfil do usuário para fazer a indicação, levando em conta seu histórico de acessos.

Como dito acima, existem muitas estratégias de recomendação, as quais podem variar de acordo com o domínio e os objetivos do sistema. A seguir serão mostradas algumas delas.

2.3.1.1. Reputação do Produto

Essa estratégia utiliza as experiências de acessos[10] dos usuários a determinado item para determinar a reputação do produto em questão. Para isso, geralmente o software solicita (ou apenas possibilita) que o indivíduo avalie o produto, depois que ele o tenha acessado [2]. Essas avaliações costumam ser muito úteis, principalmente quando o usuário nunca teve experiência com o tipo do item em questão.

Essa técnica, entretanto, tem algumas desvantagens, pois comumente os usuários sentem maior impulso para avaliar um item quando a experiência não foi boa, o que pode prejudicar a reputação real do item. Para minimizar esse problema, alguns sites (os quais não são muitos) costumam dar brindes e descontos para estimular que seus usuários façam avaliações.

É muito comum encontrar esse tipo de recomendações em redes sociais e em Websites de lojas online, como ilustram as figuras 2.5 e 2.6 a seguir. Observe que essa estratégia de recomendação é não personalizada.

[10] Experiências de acesso: Algo que o usuário tenha comprado, lido, assistido, enfim uma ação específica do domínio do sistema.

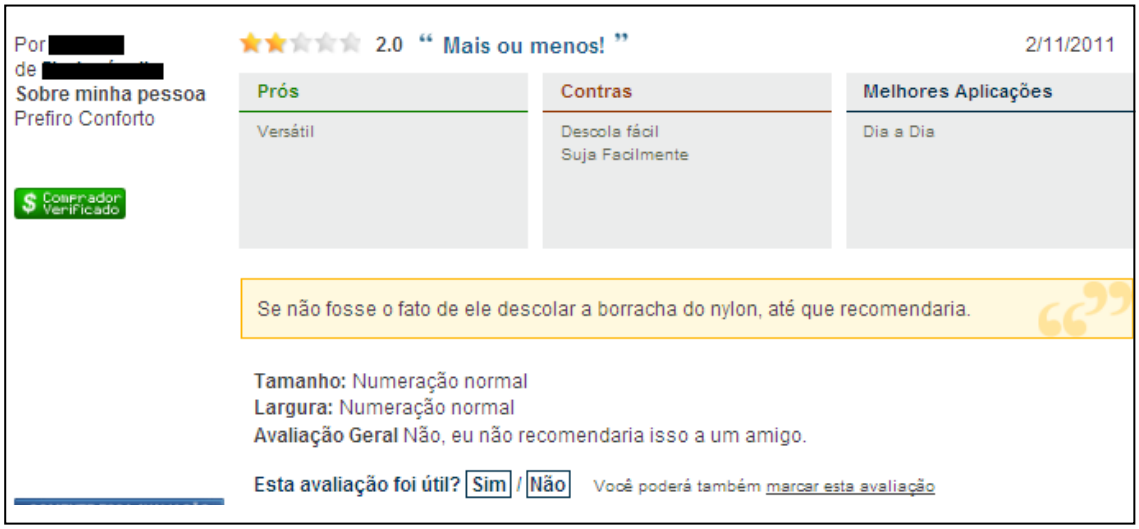


Figura 2.5: exemplo de avaliação do usuário em sistema que faz recomendação por reputação do produto. [www.netshoes.com.br]

Figura 2.6: exemplo de sistema que faz recomendação por reputação do produto [www.netshoes.com.br]

2.3.1.2. Recomendação por Associação

Essa técnica busca associar um usuário a outros usuários do sistema que realizaram ações similares às suas, criando, assim, clusters de usuários. A seguir, o sistema recomenda a esse usuário os produtos que foram escolhidos pelos usuários no mesmo cluster. Esse tipo de recomendação é um dos mais complexos, pois o sistema pode exigir uma análise profunda das ações do usuário no sistema, para "aproximar" usuários com gostos mais parecidos [2].

Dependendo do sistema, essa estratégia é muito utilizada por Websites de lojas online e ela pode ser persistente ou efêmera (quando os grupos são formados a cada nova consulta ao sistema).

2.3.1.3. Associação por Conteúdo

Com essa estratégia, o sistema faz indicações de itens que possuam atributos com valorações similares aos do item ao qual o usuário teve acesso. Esse tipo de recomendação é

mais aconselhável em sistemas em que existe conhecimento automático sobre os valores dos atributos de cada item, tais como bibliotecas e livrarias, por exemplo. Dependendo do sistema, essa estratégia pode ser efêmera ou persistente.

Em um sistema de uma livraria, por exemplo, o software poderia indicar para um usuário itens que fossem do mesmo autor, área ou editora de um livro que foi acessado por ele no sistema [2]. Na figura 2.8, o sistema está indicando alguns livros que são do mesmo autor dos livros que foram visitados pelo usuário no sistema.

Figura 2.7: exemplo de sistema realizando recomendação por associação de conteúdo [http://www.livrariasaraiva.com.br/produto/3091757/se-houver-amanha-nada-dura-para-sempre-vira-vira-saraiva/?ID=BB4EC9287DB0C051616291151]

2.3.1.4. Listas de Recomendação

Também conhecida como *Recomendação Top-N,* essa estratégia consiste em manter listas de itens organizadas por assuntos, tais como literatura, cinema e música [8]. Com essa

estratégia, todos os usuários têm acesso às mesmas recomendações, logo ela é não personalizada.

Esse tipo de recomendação é muito comum em Websites de lojas online, em sites de compartilhamento de arquivos e em blogs, sob a forma de, respectivamente, “Os mais vendidos”, “Os mais assistidos/baixados” e “Os melhores segundo...”.

Essa estratégia costuma ser muito útil quando o indivíduo deseja informações de um assunto sobre o qual não se manteve atualizado. Para exemplificar, vamos supor que uma pessoa não tem muita idéia de quais foram os bons filmes ou os mais assistidos no ano de 2010. Ele poderia encontrar essa informação facilmente em um blog de cinema sob a forma de “Os melhores filmes de 2010” ou “Os filmes mais vistos em 2010”. As figuras 2.9 e 2.10 a seguir mostram exemplos de listas de recomendação.

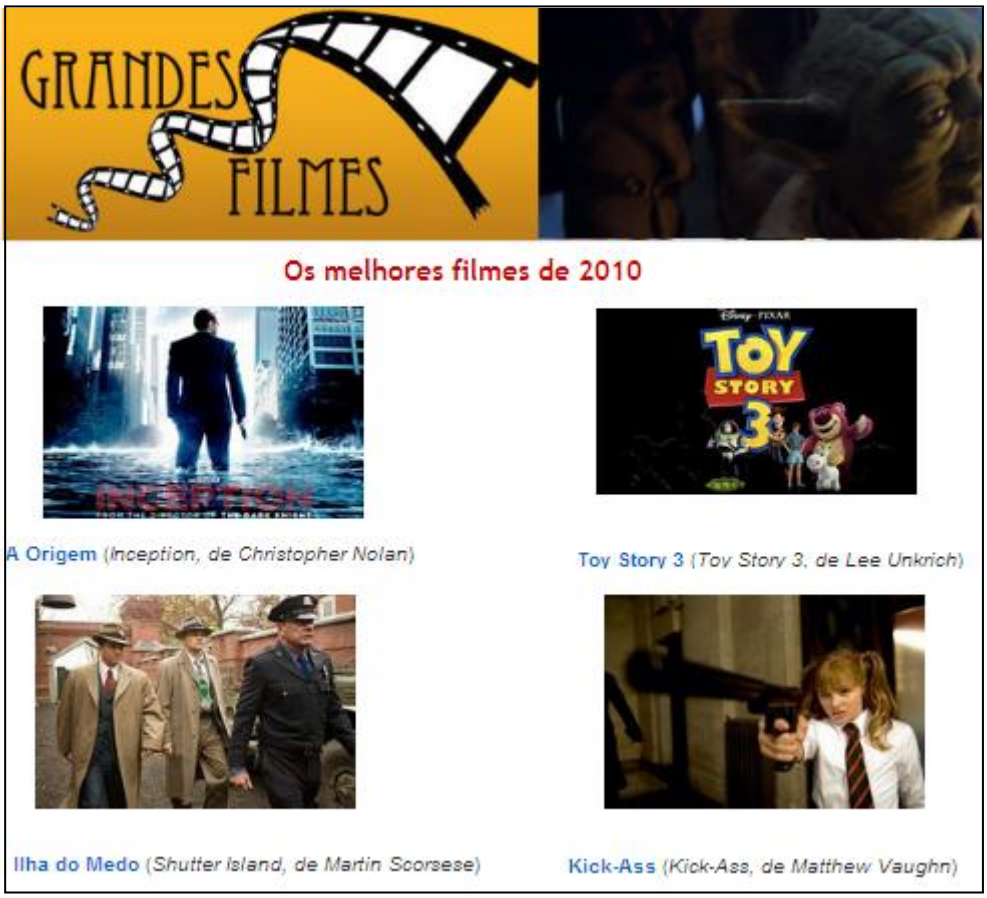


Figura 2.8: exemplo de Lista de Recomendação do tipo “Os melhores de ...” [http://www.grandesfilmes.com.br/2010/09/os-melhores-filmes-de-2010-ate-agora.html]

Figura 2.9: exemplo de Listas de Recomendação do tipo "Os mais acessados em ..." [http://www.adorocinema.com/cinenews/confira-os-filmes-mais-vistos-no-brasil-em-2010-6067/]

3. twitterRecommender

Nos capítulos precedentes, nós mostramos como Sistemas de Recomendação podem ajudar a reduzir a sobrecarga de informações nos usuários da internet, ajudando os usuários a encontrar informações sobre um dado assunto em que se tem interesse de maneira prática e rápida.

Neste capítulo, será descrito em detalhes a solução proposta para realizar a recomendação de posts de promoção de Twitter, o *twitterRecommender.*

3.1. Caracterização do Problema

Os posts de promoções não são característicos e, nem tão pouco, o sistema oferece para o usuário uma interface de filtragem para um item de promoção em que se tem interesse. Assim, para o usuário verificar se a promoção postada o interessa de fato, é necessário verificar "manualmente" os detalhes da promoção. Tal verificação pode ser muito cansativa, pois é muito grande a quantidade de postagens (muitas vezes irrelevantes para este usuário) que é gerada diariamente neste sistema.

O *twitterRecommender* resolve esses problemas oferecendo uma interface para que os usuários informem especificamente sobre quais itens possuem interesses, reduzindo, dessa forma, a sobrecarga de informação.

3.2. Abordagem Adotada

Abordagem adotada para o *twitterRecommender* consiste nas abordagens de Classificação (com um Classificador construído manualmente) e Filtragem Baseada em Conteúdo, as quais foram descritas nos capítulos anteriores.

Devido à efemeridade dos prazos das promoções e devido à dinamicidade das atualizações das informações do Twitter, concluímos que a utilização de Filtragem Baseada em Conteúdo seria mais vantajosa que a utilização da Filtragem Colaborativa.

No entanto, para filtrar itens para um usuário por conteúdo, é necessário que saibamos, previamente, a quais classes os itens estão associados e, para isso, construímos um classificador automático. Como descrevemos nos capítulos anteriores, há duas abordagens com que podemos construir um classificador automático: manualmente e automaticamente. Devido às restrições do tamanho dos posts publicados no Twitter (de tamanho máximo de 140 caracteres), percebemos que a abordagem de construção automática do classificador se tornaria pouco efetiva, pois os textos dos posts, além de pequenos, são muito parecidos - o que poderia limitar muito a capacidade de generalização dos algoritmos de aprendizagem de

máquina utilizados por essa abordagem. Escolhemos, portanto, a abordagem de Construção Manual do Classificador.

3.3. Arquitetura Geral

A figura a seguir ilustra um esboço da arquitetura geral do *twitterRecommender*. O sistema foi projetado de forma modular, priorizando manutembilidade e a escalabilidade. Os módulos do sistema são Módulo Coletor, Módulo Classificador, Interfaces do usuário e Filtragem.

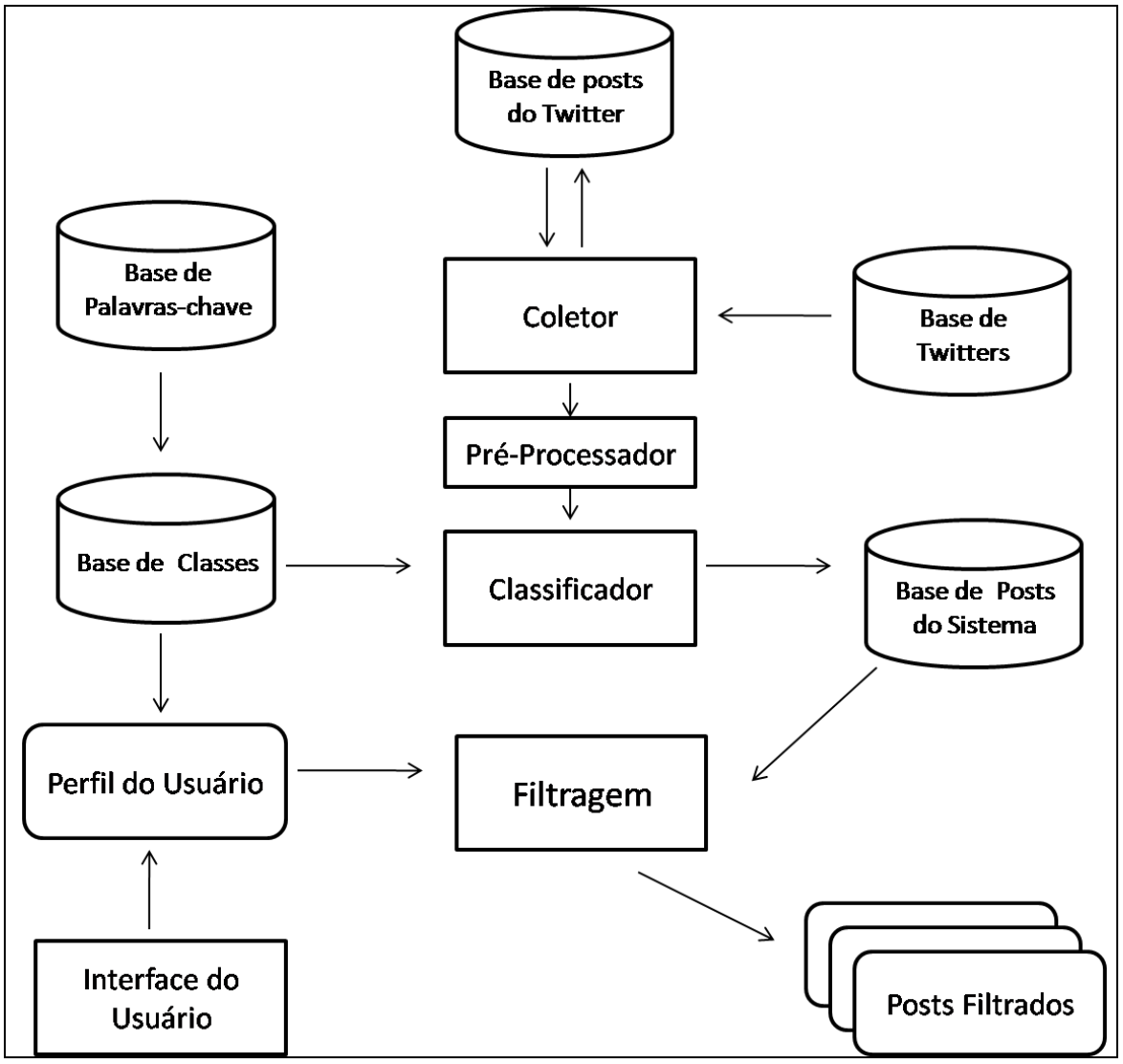


Figura 3.1: Arquitetura geral do *twitterRecommender*
Fonte: elaboração própria

O **Módulo Coletor** é responsável por coletar os posts do Twitter a partir de uma coleção de twitters pré-estabelecida. Feito isso, as informações coletadas passam pelo Módulo de Classificador.

O **Módulo Pré-processador**. Antes que o texto de um post seja submetido ao processo de classificação, o texto passa por um pré-processador, onde são feitas algumas operações sobre o texto, a fim de facilitar as atividades executadas no classificador.

O **Módulo Classificador** é responsável por classificar os posts a partir de um conjunto de classes pré-definidas, para que o sistema possa filtrá-los e indicá-los, segundo as preferências de cada usuário.

As **Interfaces do Usuário** são responsáveis por fazer a identificação do usuário no sistema e por fazer a coleta das informações que ele tem interesse, para, por fim, construir o seu perfil, o qual será armazenado no sistema e usado para fazer suas indicações.

Após identificar o usuário e coletar suas preferências, o sistema pode fazer a **Filtragem** dos posts de promoção que foram classificados e armazenados e então recomendá-los para o usuário em questão.

3.4. Base de Twitters

A base de twitters é formada por usuários do Twitter costumam gerar posts de promoções. Esses posts são inseridos manualmente na base do sistema pelo administrador. Para inserir um twitter na base, é necessário que se saiba para que local (cidade, estado, todo o Brasil) as promoções publicadas por ele são direcionadas. Isso, porque, futuramente será necessário filtrar tais informações para os usuários por cidade.

A seguir, há exemplo de twitters que publicam para promoções para a cidade de João Pessoa (PB).

cidade	twitter
PB:João Pessoa	aofertas_jp
PB:João Pessoa	ClickOnJPessoa
PB:João Pessoa	degusteAqui_jp
PB:João Pessoa	regateiojampa
PB:João Pessoa	ricardopeixejp

Figura 3.1: exemplo de twitters cadastrados na base do sistema
Fonte: elaboração própria

3.5. O Coletor

O primeiro módulo do sistema é Módulo Coletor. Ele é responsável por coletar os posts do Twitter que foram publicados pelos twitters da base do *twitterRecommender*. Esse procedimento é automático e acontece da seguinte forma: (1) Primeiro, buscam-se todos os twitters da base do sistema. (2) Para cada twitter retornado, faz-se uma busca no Twitter pelos últimos posts que foram publicados por ele. (3) O Twitter envia um arquivo *XML* contendo

essas informações. (4) Os posts coletados são enviados para o pré-processador para que, futuramente, possam ser classificados e armazenados.

A seguir, há um exemplo de consulta feita pelo *twitterRecommender* para twitter ***regateio*** referente à cidade de Recife, bem como um trecho do XML retornado pelo Twitter.

https://twitter.com/statuses/user_timeline/regateio.xml

Quadro 3.1: exemplo de url de consulta de posts de um twitter
Fonte: elaboração própria

```
− <status>
    <created_at>Thu Dec 15 18:00:02 +0000 2011</created_at>
    <id>147375377206099969</id>
  − <text>
      52% em 2 diárias com café da manhã para 2 pessoas (e uma criança) no Viver Hotel Fazenda (de R$605 por R$290) http://t.co/FDdV7qJm
    </text>
```

Figura 3.2: exemplo de resposta de uma consulta realizada para se obter posts de um twitter.
Fonte: elaboração própria

3.6. O Pré-Processador

O Módulo Pré-Processador é o segundo módulo do sistema. Nesse módulo, os documentos são submetidos a algumas operações que auxiliam na extração de informação dos mesmos, antes de seguirem para serem avaliados pelo classificador do sistema. No processo de preparação de documentos do *twitterRecommender*, são feitas as seguintes operações:

1. Análise Léxica
    a. Remoção de acentuação
    b. Remoção de pontuação
    c. O texto é convertido para minúsculo
    d. Os dígitos são removidos
    e. Os preços são adaptados

2. Eliminação de *stopwords*

    São removidas palavras e expressões ambíguas, pois, quando aparecem, podem prejudicar a precisão do processo de classificação do sistema.

3. *Stemming*

    Os textos não passam por uma operação de *stemming* bem definida. O que o sistema faz, na verdade, é preservar a menor parte dos radicais das palavras-chave escolhidas que permita identificar uma palavra-chave inteira. A seguir, há algumas palavras-chave do sistema e algumas situações em que o sistema consideraria o documento como possuindo esta palavra.

Palavra-chave do sistema	Encontra no documento
Pizz	(pizza, pizzas, pizzarias)
Hospedag	(hospedagem, hospedagens)
Churrasc	(churrasco, churrascaria)

Quadro 3.2: exemplo de operação simplificada de *stemming* do sistema

3.7. O Classificador

O classificador do *twitterRecommender* foi construído manualmente. Ele é composto por duas etapas:

1. Na primeira etapa, o classificador verifica se o post é referente a alguma promoção. Essa verificação é feita de maneira independente em relação às classes às quais o post pode estar associado.

2. Na segunda etapa, o classificador verifica se o post está associado a alguma classe do sistema. Isso é feito por meio de um conjunto de palavras-chave cadastradas no sistema que estão associadas a determinadas classes. Se o post estiver associado alguma classe do sistema, ele é indexado e armazenado para que possa ser filtrado futuramente, de acordo com o perfil do usuário.

3.8. As Classes as Palavras-Chave do Sistema

O processo para se escolher uma classe e o processo para se escolher as palavras-chave e grupos nominais que permitam que tal classe seja identificada acontece simultaneamente. A seguir, há uma descrição detalhada de como esse processo é realizado.

Ao término da execução do classificador, é gerado um arquivo que contém os posts que passaram pela primeira etapa do classificador (promoções foram identificadas), mas que não passaram no segundo módulo (não foi associado a nenhuma classe do sistema). O arquivo é analisado manualmente e os posts são agrupados de acordo com o conteúdo das promoções.

Para cada grupo, verifica-se se o conjunto de posts pode ser subdividido em grupos menores de classes de promoções e, em seguida, os posts são reagrupados. Feito isso, a partir do texto dos posts, é escolhido, para cada classe de promoção, um grupo de palavras-chave e grupos nominais que permitam identificar a classe.

Em seguida, as palavras-chave e grupos nominais são reduzidos e simplificados, preservando-se a menor parte de suas estruturas, de modo que seja possível tanto identificar a expressão original bem como algumas variações. Por fim, o grupo de promoções é

armazenado no banco de dados do sistema, juntamente com suas classes e as respectivas palavras-chave de cada classe.

Para melhor entendimento, a seguir há um exemplo simplificado de como esse procedimento é realizado. Assuma que, neste exemplo, os posts já foram agrupados e que estamos organizando o grupo de posts relativos à “estética”.

Passo 1: posts agrupados por tipo (neste exemplo, “estética”).

64% de desconto em Escova Progressiva no New´s Fashion Hair Estética e Beleza
72% OFF em Depilação a Cera Feminina OU Masculina na NovaDepil
77% OFF Limpeza de pele+Massagem Facial+Hidratação Facial+Máscara Facial de Argila+Massagem Relaxante
90% OFF em Escova Selante Queratinização Manicure no Le Liss Cabelo e Corpo (de até R$290,00 por R$29,00)
79% OFF em Ultrassom + Drenagem Linfática + Corrente Russa no London by Tânia Espaço de Beleza
Banho de Lua + Gomagem Corporal + Depilação de Buço + Design de Sobrancelhas: de R$128,00 por APENAS R$29,90!!!
84% OFF em Limpeza de Pele com Extração + Revitalização no CENTRO DE ESTÉTICA SOLANGE DE OLIVEIRA de...

Quadro 3.3: exemplo de promoções agrupadas por tipo
Fonte: elaboração própria

Passo 2: subdividir o grupo em classes

Cabelos -64% de desconto em Escova Progressiva no New´s Fashion Hair Estética e Beleza -90% OFF em Escova Selante Queratinização Manicure no Le Liss Cabelo e Corpo (de até R$290,00 por R$29,00)
Depilação -72% OFF em Depilação a Cera Feminina OU Masculina na NovaDepil -Banho de Lua + Gomagem Corporal + Depilação de Buço + Design de Sobrancelhas: de R$128,00 por APENAS R$29,90!!!
Cuidados com a pele -77% OFF Limpeza de pele+Massagem Facial+Hidratação Facial+Máscara Facial de Argila+Massagem Relaxante -84% OFF em Limpeza de Pele com Extração + Revitalização no CENTRO DE ESTÉTICA SOLANGE DE OLIVEIRA de...

Quadro 3.4: exemplo de grupo de promoções subdividido em classes

Fonte: elaboração própria

Passo 3: escolher um conjunto de palavras-chave/grupos nominais que permitam identificar a classe

Cabelos -Escova Progressiva -Escova Selante -Queratinização **Depilação** -Depilação a Cera -Depilação de Buço **Cuidados com a pele** - Limpeza de pele -Massagem Facial -Hidratação Facial -Máscara Facial de Argila -Limpeza de Pele com Extração

Quadro 3.5: exemplo de palavras-chave e grupos nominais escolhidas por classe
Fonte: elaboração própria

Passo 4: reduzir, se possível, o conjunto de palavras-chave/grupos nominais

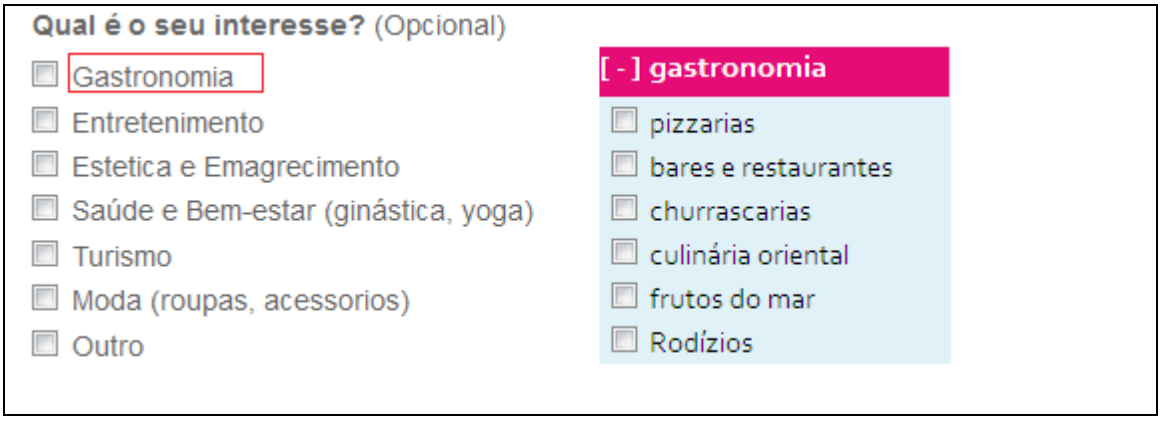
Cabelos -escova progressiva -escova selante -queratinizac **Depilação** -depila **Cuidados com a pele** -limpeza de pele -massagem facial -hidratacao facial

Quadro 3.6: exemplo de redução de palavras-chave e grupos nominais
Fonte: elaboração própria

Vale ressaltar que essa tarefa é um dos pontos de originalidade desse trabalho, pois Sistemas de Recomendação similares não subdividem os grupos de promoção em classes de promoções mais específicas, o que pode sobrecarregar o usuário com informações que ele não tem interesse.

A imagem a seguir ilustra um exemplo em que é possível perceber a variedade de promoções que o *twitterRecommender* oferece sobre um determinado assunto, enquanto que os outros sistemas oferecem apenas uma opção.

Peixe Urbano / *twitterRecommender*



Quadro 3.3: parte de interface do cadastro Peixe Urbano e do *twitterRecommender*
Fontes: [elaboração própria] e [http://www.peixeurbano.com.br/conta/Criar]

3.9. Interfaces do Usuário e Recomendação

Para que o *twitterRecommender* possa fazer recomendações para um usuário, é necessário que este informe, no seu perfil do sistema, sobre quais promoções ele possui interesse e em que cidade. Para editar o seu perfil, o usuário deve estar cadastrado e logado no sistema.

As recomendações são feitas baseadas no conteúdo dos posts armazenados que casam som os interesses informados pelo usuário. O usuário pode visualizar promoções indicadas pelo sistema a partir da interface *"Minhas Promoções"* do *twitterRecommender*. Entretanto, mesmo que o usuário não efetue login no sistema, o sistema envia as indicações proativamente.

Se o usuário não informar de qual cidade ele deseja receber promoções, o sistema só fará recomendações de promoções destinadas a todo o território nacional.

A seguir, estão ilustradas as interfaces do *twitterRecommender* de **cadastro**, de **log in** e de **configurações do perfil**, de **indicações** e um exemplo de **email enviado pelo sistema**.

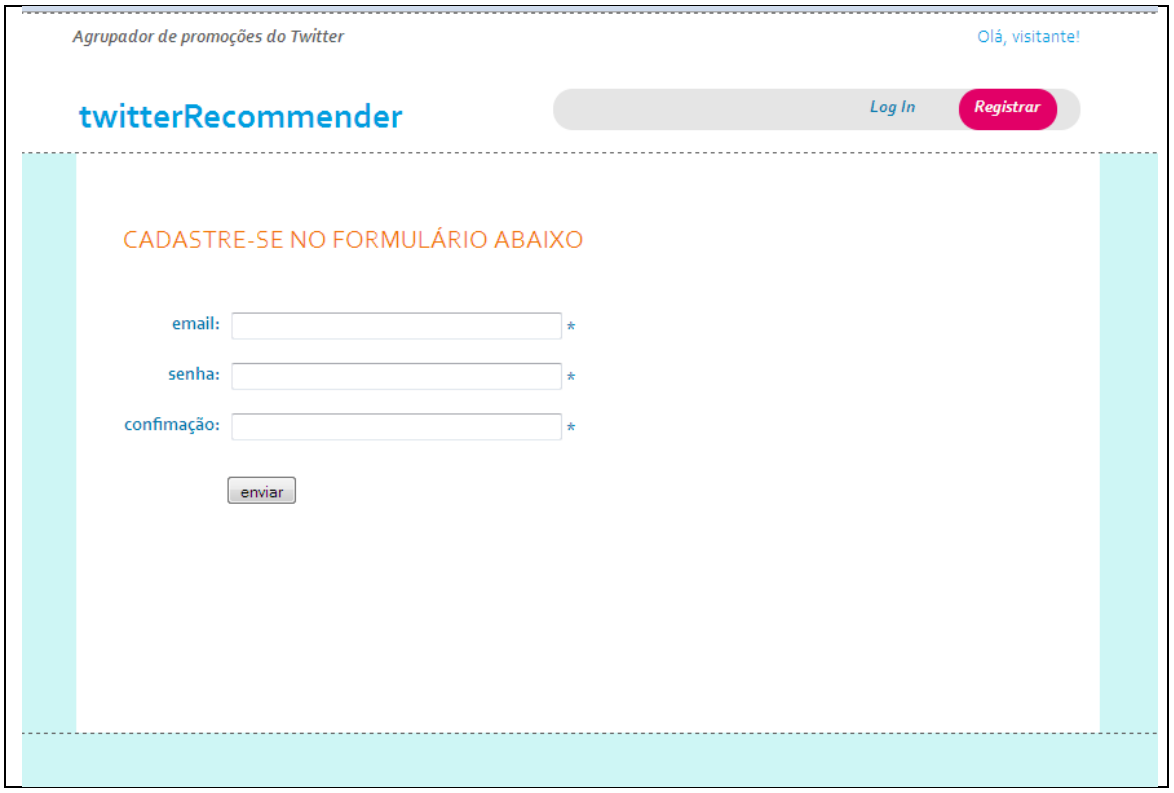


Figura 3.4: Interface de cadastro do *twitterRecommender*
Fonte: elaboração própria

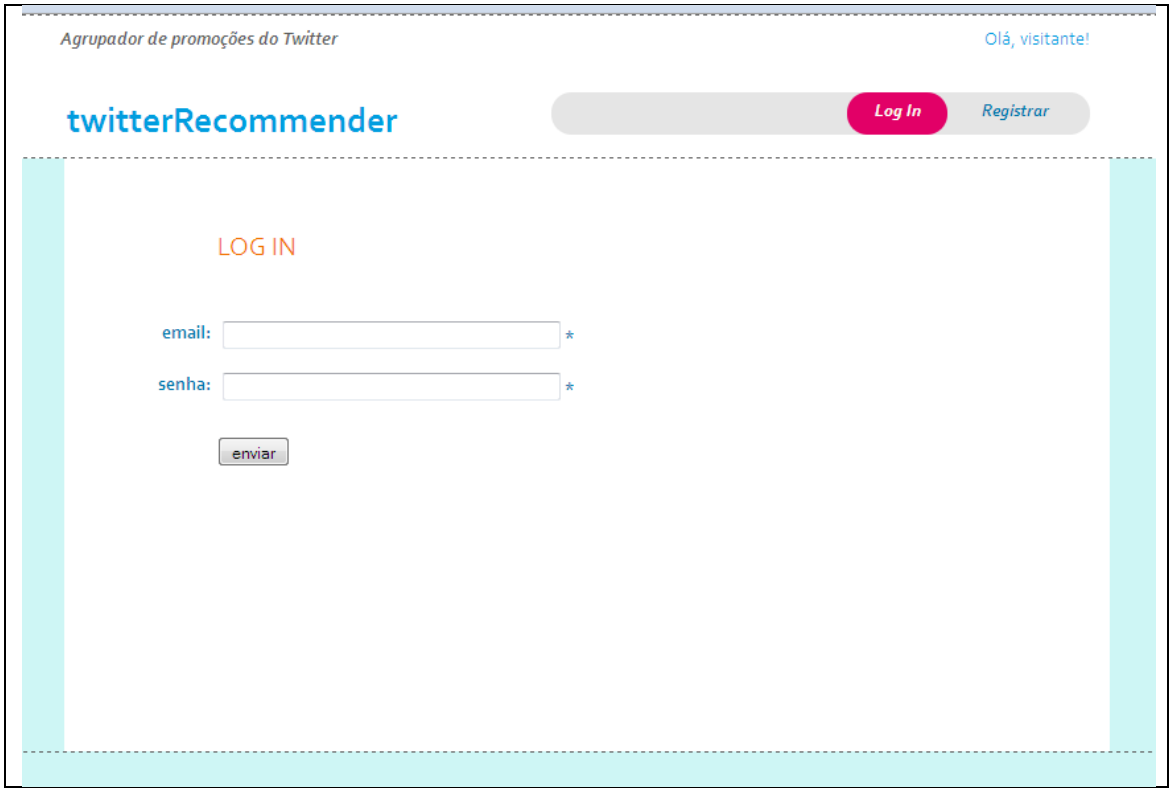


Figura 3.5: Interface de log in do *twitterRecommender*
Fonte: elaboração própria

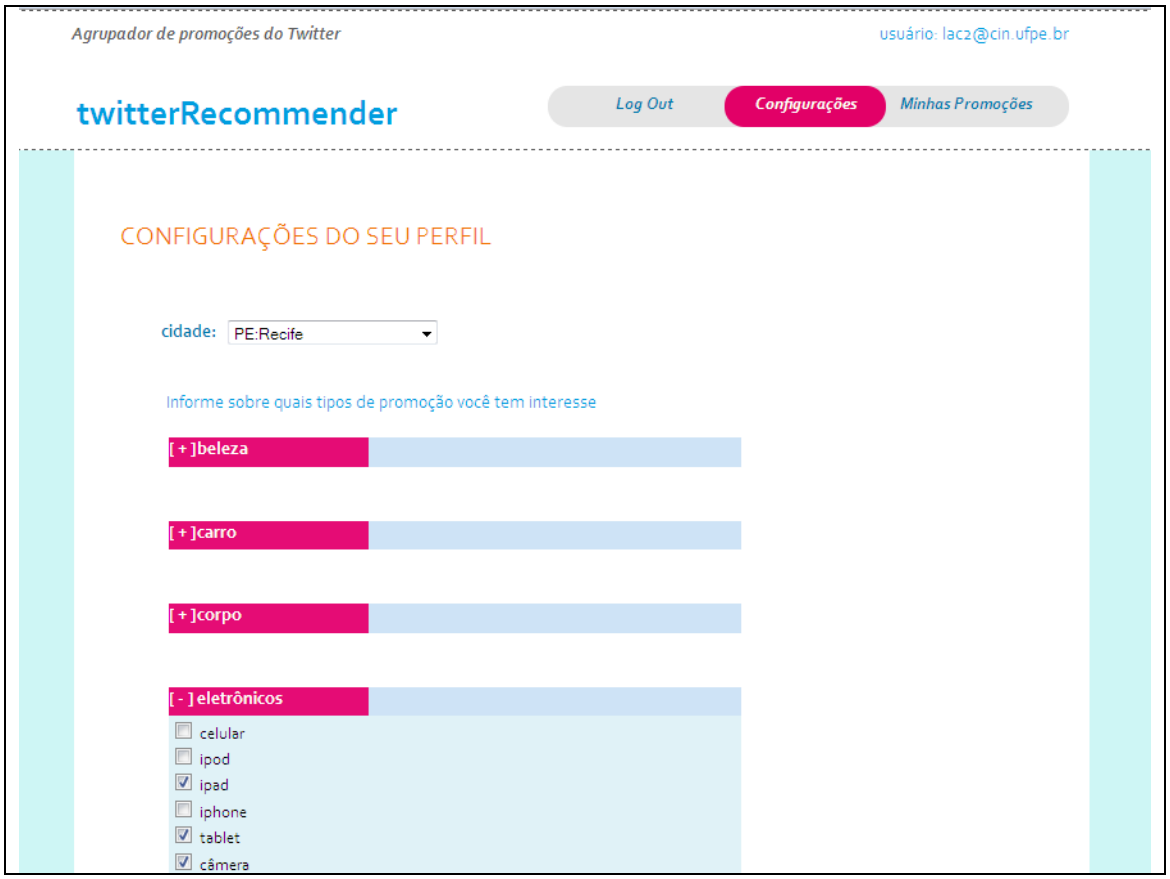


Figura 3.6: Interface de Configurações do Perfil do usuário do *twitterRecommender*
Fonte: elaboração própria

Figura 3.7: Interface de visualização de promoções indicadas pelo *twitterRecommender*
Fonte: elaboração própria

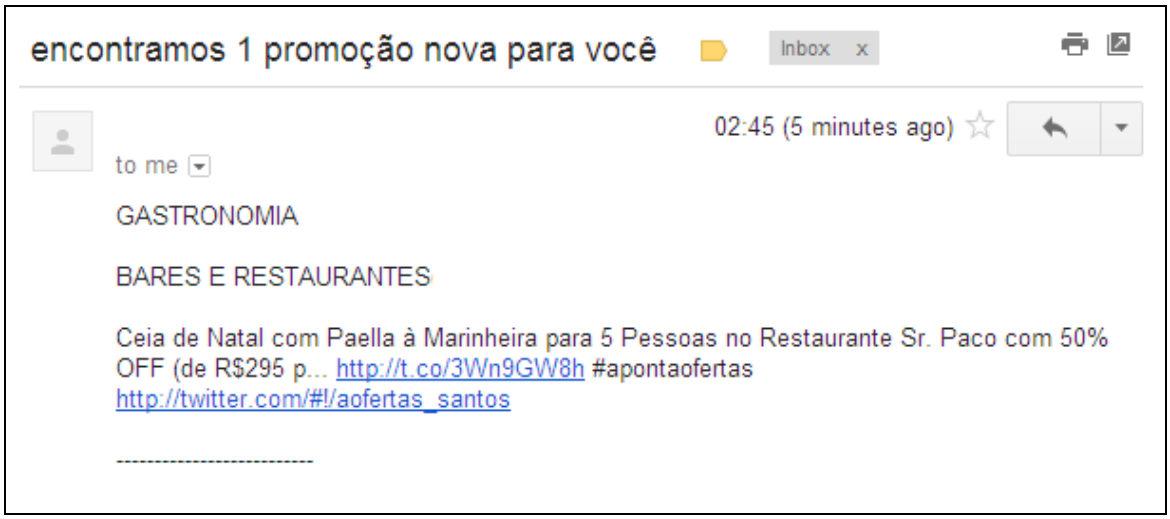


Figura 3.8: exemplo de indicação por email realizada pelo *twitterRecommender*
Fonte: elaboração própria

3.10. Considerações Finais

Foi apresentado neste capítulo um sistema de recomendações baseado em posts do Twitter. Cada um dos módulos foi descrito em detalhes, bem como as técnicas e os procedimentos utilizados, apresentando, quando necessárias, as justificativas para a utilização das mesmas e os pontos de originalidades desse projeto.

4. Avaliação do Sistema

Os testes quanto à eficiência do sistema, sobretudo, quanto à precisão do classificador foram realizados manualmente.

A avaliação do sistema foi feita utilizando técnicas de avaliação de sistemas de recuperação de informação e as medidas utilizadas foram precisão e cobertura. A precisão verifica se as respostas retornadas são relevantes, a cobertura verifica se os documentos relevantes foram retornados.

Como o classificador possui duas etapas, a precisão de cada etapa foi avaliada isoladamente; já a cobertura foi avaliada para o sistema como um todo. Para isso, foram utilizados 1180 posts. Ao término da execução do classificador, foram gerados três arquivos: um arquivo contendo os posts em que nenhuma promoção foi encontrada; outro arquivo contendo os posts de promoção que estão associados a alguma classe do sistema; e um último contendo os posts de promoção que não estão associados a nenhuma classe do sistema. Em seguida os arquivos foram analisados e se obteve os seguintes resultados:

4.1. Precisão

1. Testes realizados para a primeira etapa do classificador: encontrar posts de promoção (associadas ou não a alguma classe do sistema):
    - Acertos: 341
    - Erros: 12
    - Total: 353

    Taxa de acerto: 95%

2. Testes realizados para a segunda etapa do classificador: associar posts de promoção a classes do sistema:
    - Acertos: 154
    - Erros: 5
    - Total: 159

    Taxa de acerto: 97%

4.2. Cobertura

Para medir a cobertura, foi necessário analisar o arquivo com os posts em que nenhuma promoção foi detectada e o arquivo que com posts de promoções que não foram associados a nenhuma classe do sistema. Nessa avaliação, foram contados os posts de promoção associados a alguma classe do sistema que não foram retornados. Feito isso, o resultado foi somado à taxa de acerto da segunda fase do classificador, para se obter a quantidade de posts relevantes total. Os resultados obtidos foram os seguintes:

- Total de posts relevantes: 177
- Posts relevantes não retornados: 23

Taxa de Cobertura do *twitterRecommender*: 87%

4.3. Considerações Finais

Os posts de promoções possuem estruturas relativamente esperadas. Isso ajudou bastante para a baixa taxa de erro que obtivemos na primeira etapa do classificador. A maior complexidade, entretanto, acontece tanto pela grande quantidade de palavras-chave que algumas classes possuem, quanto pela existência de algumas palavras-chave ambíguas - pois o uso destas palavras pode diminuir a precisão do classificador. No classificador do *twitterRecommender,* em alguns casos, optou-se por não utilizar palavras-chave e grupos nominais ambíguos, o que um dos motivos determinantes para que a cobertura do sistema fosse menor em relação à precisão.

5. Conclusão

Este trabalho teve como objetivo apresentar o Sistema de Recomendação *twitterRecommender.* Na seção 5.1, apresentaremos quais foram as principais contribuições deste trabalho; Na seção 5.2, apresentaremos quais foram as principais dificuldades encontradas; Por fim, na seção 5.3, apontaremos alguns possíveis trabalhos futuros a serem realizados no *twitterRecommender*.

5.1. Principais Contribuições

Para começar, não foi encontrado nenhum Sistema de Recomendação capaz de tratar posts de promoções do Twitter. Além disso, dos sistemas que realizam recomendação de promoção atualmente, nenhum deles permite os usuários escolham sobre quais itens de promoção, especificamente, desejam receber notificações – o que acaba sobrecarregando o usuário com promoções ele não tem interesse.

Outro ponto importante é que os esses Sistemas de Recomendação de promoções não oferecem uma interface facilitada para que os usuários possam filtrar por tipo, de maneira rápida, as promoções que estão disponíveis.

5.2. Dificuldades Encontradas

As maiores dificuldade encontradas foram concernentes aos problemas causados pela presença de palavras-chave e grupos nominais de valores ambíguos nos textos dos posts. Como já explicado na seção 5.3, em alguns desses casos, optamos por não utilizar a expressão ambígua.

Apesar da variedade de twitters geradores de promoções, encontramos dificuldades em selecionar uma quantidade relevante desses twitters. Isso, porque alguns deles não informam, explicitamente para qual local as promoções estão sendo direcionadas.

5.3. Trabalhos Futuros

Atualmente as recomendações realizadas pelo *twitterRecommender* são feitas utilizando unicamente Filtragem Baseada de Conteúdo. Entretanto, o sistema permite que outras formas de recomendação também sejam feitas, citaremos algumas delas a seguir.

Como já foi citado na seção 5.2, o trabalho de selecionar twitters geradores de promoções, nem sempre é uma tarefa fácil. Por conta disso, o sistema poderia fornecer uma interface para que o usuário pudesse não só inserir novos twitters como também ignorar certos twitters que

foram pré-estabelecidos pelo sistema. A partir dessas ações, seria possível recomendar twitters a usuários que possuem perfis parecidos.

Além disso, o sistema poderia permitir que os usuários pudessem informar qual o gral de relevância da promoção para fazer Recomendação por Reputação dos Itens. Note que, geralmente os posts de promoção são de curto período de validade, por conta disso, as avaliações seriam usadas principalmente para ranquear o twitter que a gerou e priorizar os futuros posts gerados por ele.

Referências Bibliográficas

[1] Aplicando a Relevância da Opinião de Usuários em Sistema de Recomendação para Pesquisadores
Disponível em:
http://www.lume.ufrgs.br/bitstream/handle/10183/8424/000575704.pdf?sequence=1
Último acesso em 18 de Dezembro de 2011

[2] A Ciência da Opinião: Estado da arte em Sistemas de Recomendação
Disponível em: http://www.dcomp.ufs.br/~gutanunes/hp/publications/JAI4.pdf
Último acesso em 18 de Dezembro de 2011

[3] Tim O'Reilly
Disponível em: http://tim.oreilly.com/
Último acesso em 18 de Dezembro de 2011

[4] Sistemas de Recomendação
Disponível em:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.2811&rep=rep1&type=pdf
Último acesso em 18 de Dezembro de 2011

[5]Cookie
Disponível em: http://pt.wikipedia.org/wiki/Cookie 9/10/2011
Último acesso em 18 de Dezembro de 2011

[6] Sistemas de Recomendação
Disponível em: http://kessia.blogs.unipar.br/files/2008/07/sistemas-de-recomendacao.pdf
Último acesso em 18 de Dezembro de 2011

[7] Sistemas de recomendação de notícias na Internet baseados em filtragem colaborativa
Disponível em: http://www.ime.usp.br/~cef/mac499-07/monografias/rec/allan-renato-sidney-victor/monografia.pdf
Último acesso em 18 de Dezembro de 2011

[8] Sistemas de recomendação
Disponível em: http://www.slideshare.net/berg_pe/sistemas-de-recomendao-9889295
Último acesso em 18 de Dezembro de 2011

[9] BAEZA-YATES, R. A.; RIBEIRO-NETO, B. Modern information retrieval. Addison- Wesley Longman Publishing Co., Inc., 1999

[10] Categorização/Classificação de texto - CIn
Disponível em: www.cin.ufpe.br/~in1152/aulas/classificacao-texto-2010.ppt
Último acesso em 18 de Dezembro de 2011

[11] Data Mining
www.di.ufpe.br/~compint/aulas-IAS/agentes/taci1-981/DataMining.ppt
Último acesso em 18 de Dezembro de 2011

[12] Número de usuários do Twitter se aproxima da população dos EUA
Disponível em: http://tecnologia.terra.com.br/noticias/0,,OI5139396-EI12884,00-Numero+de+usuarios+do+Twitter+se+aproxima+da+populacao+dos+EUA.html
Último acesso em 18 de Dezembro de 2011