

Personalização de Conteúdo na TVDI através de um Sistema de Recomendação Personalizada de Programas de TV (SRPTV)

Fábio Santos da Silva

Genius Instituto de Tecnologia – Universidade Federal do Amazonas (UFAM)
Manaus – AM – Brasil

fsantos.mail@gmail.com

Abstract. *Abstract. This work presents the Customized-Recommendation System for TV Programs (SRPTV), which is in a tool for recommending programs that are of interest for the user, through an analysis of the data available in his profile. This tool can be used as an application in the Interactive Digital Television environment, being a possible solution for the problem of information overload. During the development of this tool, filtering techniques available in the literature were evaluated, as well as the Rocchio and kNN algorithms, in order to verify which is the most convenient for the recommendation of TV programs.*

Resumo. *Este trabalho apresenta o Sistema de Recomendação Personalizada de Programas de TV (SRPTV), que consiste numa ferramenta para a recomendação de programas que são de interesse do usuário, através de informações disponíveis no seu perfil. Esta ferramenta pode ser utilizada como uma aplicação na Televisão Digital Interativa, sendo uma possível solução para o problema de sobrecarga de informações. Durante o desenvolvimento desta ferramenta, foram estudadas técnicas de filtragem de informação disponíveis na literatura, juntamente com os algoritmos Rocchio e kNN, com o objetivo de verificar qual destes é o mais adequado para a recomendação de programas de TV.*

1. Introdução

O avanço da tecnologia digital contribuiu para o surgimento de novas plataformas de transmissão de programas de televisão, como a Televisão Digital Interativa (TVDI) muito popular na Europa, EUA e Japão. Atualmente estudos estão sendo realizados com o objetivo de definir qual será o padrão de TVDI que será brevemente implantado no Brasil.

Estas plataformas apresentam novos recursos como a flexibilidade para acomodar uma numerosa diversidade de serviços, alta qualidade de vídeo, áudio, acesso à Internet e a execução de aplicativos interativos. Além disso permitem a transmissão simultânea de um número elevado de canais e a tendência é aumentar substancialmente a quantidade de canais cada vez mais.

Nos países onde a TVDI já é uma realidade, o telespectador tem encontrando dificuldade para escolher seus programas favoritos dentre os canais e os milhares de programas disponíveis. As emissoras de televisão destes países procuram ajudar ao telespectador a encontrar informações sobre seus programas favoritos, através da

utilização de guias de programação. Estes guias são oferecidos em diferentes mecanismos de publicação, tais como guias baseados na Web, revistas especializadas na divulgação da programação das emissoras de televisão e os avançados guias de programação eletrônica comumente conhecido como EPG's, ferramenta esta que possui interfaces interativas contendo a lista de programas de canais de televisão. O *set-top-box* é o dispositivo eletrônico de entretenimento interativo que viabiliza o acesso aos canais da TVDI, e também aos aplicativos interativos como o próprio EPG.

Os meios citados anteriormente são os mais utilizados pelas emissoras de televisão. Estes mecanismos não estão atendendo as necessidades de informações dos telespectadores de forma eficiente. Os guias simplesmente apresentam grandes listas de programas, e as preferências dos usuários raramente são consideradas, o que acabam desperdiçando demasiado tempo procurando as informações sobre seus programas de interesse. Muitas vezes não consegue encontrá-las em tempo hábil.

Diante desta grande quantidade de informação, torna-se uma tarefa árdua para o telespectador procurar pelos seus programas favoritos no EPG utilizando um controle remoto. Um outro problema a ser destacado é que as informações apresentadas o freqüentemente são estáticas, não fazendo distinção dos usuários que buscam informações nos guias de programação.

Segundo Ehrmantraut (1996) a obtenção de informações sobre os programas de televisão pelos telespectadores transformou-se em um desafio. A navegação pelo EPG não será rápida devido ao grande número de programas disponíveis nos mecanismos eletrônicos de publicação. A quantidade de programas logo excederá os limites do que pode ser razoavelmente observado nos guias de programação dos canais. Como conseqüência os telespectadores encontraram uma grande quantidade de informação. A este problema que os usuários da TVDI estão enfrentando denominamos sobrecarga de informação (grande quantidade de serviços e programas de televisão).

Diante dos problemas supracitados, em sua maioria relacionada à grande quantidade de informações disponíveis, conclui-se que justifican-se novos estudos para o desenvolvimento de ferramentas que visam auxiliar a filtragem de informações de forma eficaz e fácil uso. Estas ferramentas poderão explorar os recursos e vantagens oferecidas pelas novas plataformas dos dispositivos eletrônicos e de transmissão de programas de televisão, o que conseqüentemente deverá melhorar a interação entre os usuários e as emissoras de televisão.

O dispositivo *set-top-box*, por exemplo, poderia de ser equipado com um sistema de recomendação avançado que reconhece as preferências do usuário e lhe ajuda a selecionar seus programas favoritos. Além disso, pelo fato do sistema ser utilizado por uma grande diversidade de usuários com preferências diferentes em momentos distintos, é pertinente que o sistema forneça personalização em nível de conteúdo baseando-se no perfil do usuário.

O sistema deve monitorar o comportamento dos telespectadores conforme forem interagindo e recomendar novos programas baseado nos seus julgamentos que refletem suas necessidades de informações. O telespectador não precisará investir muito tempo pesquisando nos mecanismos responsáveis pela publicação da programação dos canais, pois deve receber a cada dia uma nova lista personalizada de programas de seu provável interesse.

Desta forma podemos resumir que a proposta do trabalho é implementação do Sistema de Recomendação Personalizada de Programas de TV (SRPTV). Este sistema poderá ser uma possível solução para o problema de sobrecarga de informações que poderá futuramente atingir o Brasil como tem acontecido nos países que utilizam a TVDI.

O SRPTV será baseado na implementação de um algoritmo de filtragem de informações que utiliza um modelo matemático conhecido como Modelo Vetorial descrito em Baeza-Yates e Ribeiro-Neto (1998). Como objetivo específico, pretendemos também realizar um estudo e uma análise comparativa entre os algoritmos *Rocchio* descrito em Rocchio (1971) e *KNN* descrito em Cover (1967) para atualização dinâmica do perfil de usuário.

1.1 Trabalhos Relacionados

Serão apresentados a seguir alguns trabalhos que foram desenvolvidos sobre personalização na TV. Os artigos descritos possuem similaridade em vários aspectos com o SRPTV.

Baudisch e Brueckner (2002) apresentam o TV Scout que é um sistema de recomendação desenvolvido para a plataforma Web. O sistema fornece aos usuários as programações personalizadas dos canais de televisão. Neste artigo é discutida a arquitetura desenvolvida para resolver o problema de filtragem de informação utilizando a Abordagem Colaborativa. O TV Scout efetua atualização dinâmica dos perfis dos usuários que é uma funcionalidade implementada também no SRPTV.

P. Cotter e B. Smyth (2000) apresentam PTV que é um sistema que fornece ao usuário recomendação personalizada de programas de TV baseada na abordagem de filtragem colaborativa. O perfil inicial do usuário é registrado de forma explícita onde é especificado seu interesse. A nossa ferramenta é similar com PTV neste aspecto pois também utilizamos o modo explícito para gerar o perfil inicial do usuário.

Kurapati K., Gutta e S. Schaffer (2001) apresentam um Sistema Multi-Agente para Recomendação de TV, ele combina os métodos explícito e implícito para registrar as preferências dos usuários. Este trabalho possui uma similaridade com o SRPTV neste aspecto, pois também optamos pela utilização de uma abordagem híbrida.

Pazzine M. e Bilsus D. (1996) este artigo descreve o funcionamento de um sistema de recomendação que utiliza a filtragem de informações baseada no conteúdo, sendo abordagem utilizada no SRPTV.

2. Conceitos

Nesta seção apresentamos os conceitos que foram utilizados desenvolvimento deste trabalho, tais como Sistemas de Recomendação, Abordagem Colaborativa e não Colaborativa, Perfil do Usuário, Modelo Vetorial, Filtragem de informações, *Relevance Feedback* e Avaliação de Desempenho.

2.1 Sistema de Recomendação Personalizada

Estes sistemas filtram documentos relevantes e descartam documentos irrelevantes de acordo com os interesses de um determinado usuário ou um grupo de usuários levando em consideração os seus perfis. O funcionamento do sistema de recomendação consiste

em analisar o fluxo de documentos que entram no sistema. Em seguida compara estes documentos com os assuntos de interesses dos usuários especificados nos perfis que ficam armazenados em uma base de informação sobre usuários. A figura 1 abaixo apresenta a arquitetura de um sistema de recomendação.

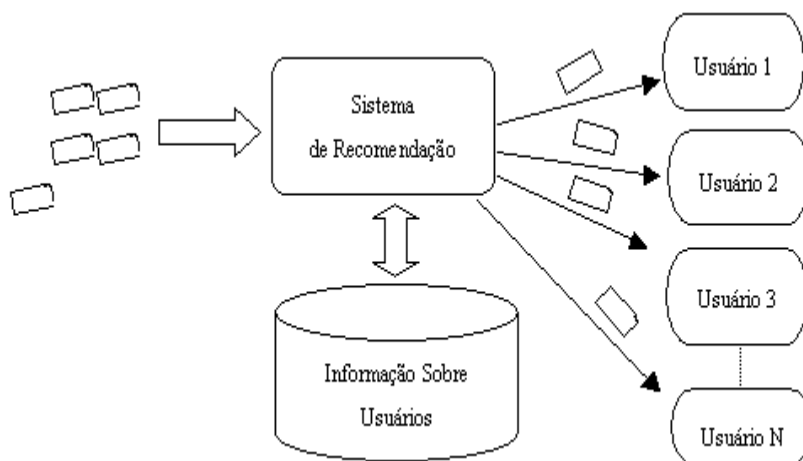


Figura 1. Sistema de Recomendação

O perfil do usuário pode ser visto como um conjunto de palavras que descrevem os interesses do usuário, por exemplo, ficção, política, etc. Para se desenvolver sistemas de recomendação é necessário implementar determinadas funções, tais como a coleta do perfil do usuário, e a recomendação personalizada de forma dinâmica. A coleta do perfil segundo Rich (1979) o perfil pode ser coletado de duas formas:

- *Explícita:* na forma explícita o perfil é construído diretamente pelo usuário através de uma interface com um formulário de cadastro. O usuário deve especificar um conjunto de termos, por exemplo, aventura, comédia, ficção, etc.
- *Implícita:* na forma implícita o sistema infere e deduz o perfil do usuário através da sua interação com o sistema. Nesse modelo cada vez que um usuário julga um documento como relevante ou irrelevante, o sistema utiliza esta informação para atualizar dinamicamente o perfil do usuário.

A segunda função citada anteriormente tem sido muito estudada nos últimos anos. As soluções propostas são diferenciadas com relação à abordagem de filtragem utilizada.

2.2 Abordagem Colaborativa

A Abordagem Colaborativa, de acordo com Herlocker (1999) também conhecida como Filtragem Colaborativa, baseia-se na correlação entre os perfis dos usuários para recomendar informações que possam ser de interesse dos mesmos. Para recomendar informação para o usuário, o seu perfil é comparado a uma base de dados contendo perfis de outros usuários com o objetivo de encontrar os usuários que mostraram ter preferências semelhantes a ele em um determinado momento.

2.3 Abordagem Não Colaborativa

A segunda abordagem largamente utilizada nos sistemas de recomendação analisa a similaridade do conteúdo do perfil do usuário com os termos contidos nos documentos

com o objetivo de recomendar os documentos relevantes e ignorar os documentos insignificantes.

2.4 Modelo Vetorial

No modelo Vetorial descrito em Baeza-Yates e Ribeiro-Neto (1998) os elementos a serem modelados são representados como vetores de termos localizados dentro de um espaço vetorial. Os perfis de usuários e os documentos da coleção são vistos como vetores cujas coordenadas são determinadas pelos valores associados aos termos que o descrevem. O número de termos distintos da coleção determina a dimensão do espaço onde os documentos e os perfis serão representados. Conforme mostra a figura abaixo:

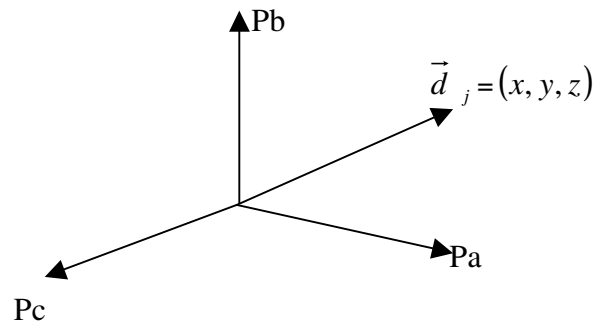


Figura 2. A Dimensão do espaço é igual à quantidade de termos

2.4.1 Filtragem de Informações com o Modelo Vetorial

Na filtragem com o Modelo Vetorial a base de dados deve conter os perfis dos usuários no lugar dos documentos. Os documentos que entram no sistema são tratados como consultas.

As distâncias entre os vetores do documento e o perfil indicam o valor de similaridade. Desta forma a similaridade de cada documento com o perfil fornecido é encontrado através do cálculo do cosseno entre os vetores correspondentes de cada elemento. A idéia é verificar o ângulo entre os vetores para descobrir a similaridade.

$$\vec{d}_j = (2,2) \quad \vec{p} = (2,1)$$

$$Sim(d_j, p) = Cos(d_j, p)$$

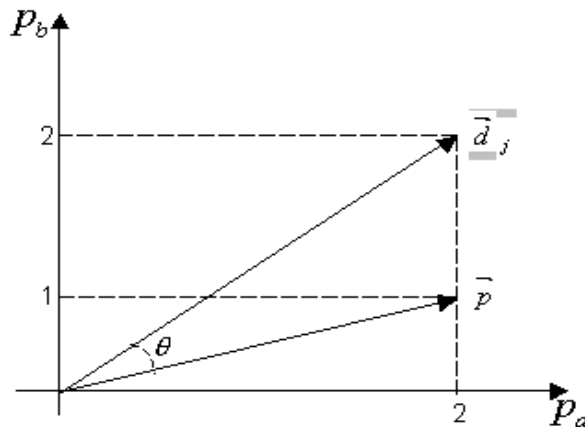


Figura 3: Ângulo entre dois vetores.

O valor da similaridade é calculado pela fórmula abaixo:

$$sim(d, p) = \cos \theta = \frac{\sum_{i=1}^t w(i, d) \times w(i, p)}{\sqrt{\sum_{i=1}^t (w(i, d))^2} \times \sqrt{\sum_{i=1}^t (w(i, p))^2}} \quad (3)$$

Norma de d
Norma de p

Onde,

d é o vetor de termos do documento;

p representa o vetor de termos do perfil

$w(i, d)$ são as coordenadas dos termos do documento; e

$w(i, p)$ são as coordenadas dos termos do perfil.

Os documentos que forem mais similares para um determinado perfil são considerados relevantes para o proprietário do respectivo perfil, e são enviados para o mesmo. Quando os cálculos dos graus de similaridades forem concluídos, é possível gerar uma lista ordenada (*ranking*) baseada nos graus de relevância dos documentos em relação ao perfil, do maior documento para o de menor relevância.

2.5 Atualização de Perfil do Usuário através do *Relevance Feedback*

Depois de muitos estudos realizados na área de Recuperação de Informações, de acordo com Buckley (1995), os pesquisadores concluíram que o usuário não consegue recuperar documentos que possam ser de seu interesse na primeira vez que efetua uma consulta ao sistema. Tipicamente o que ocorre é o usuário efetuar uma consulta inicial como uma tentativa, e vai refinando-a, melhorando-a, conforme os resultados que ele obtém.

As próximas consultas passam a recuperar cada vez mais documentos relevantes ao usuário, pois ele vai contextualizando mais precisamente o assunto que é de seu interesse, utilizando novas palavras, e eliminando as palavras que desvirtuam as suas consultas, obtendo documentos que estão fora de seu interesse. Desta forma, são produzidas novas consultas que teoricamente serão mais precisas e refletindo seus interesses, recuperando mais documentos relevantes.

Segundo Buckley (1995) afirma que *Relevance Feedback* é o processo automático de refinamento de uma consulta inicial, utilizando informações fornecidas pelo usuário sobre a relevância dos documentos previamente recuperados (em uma consulta anterior).

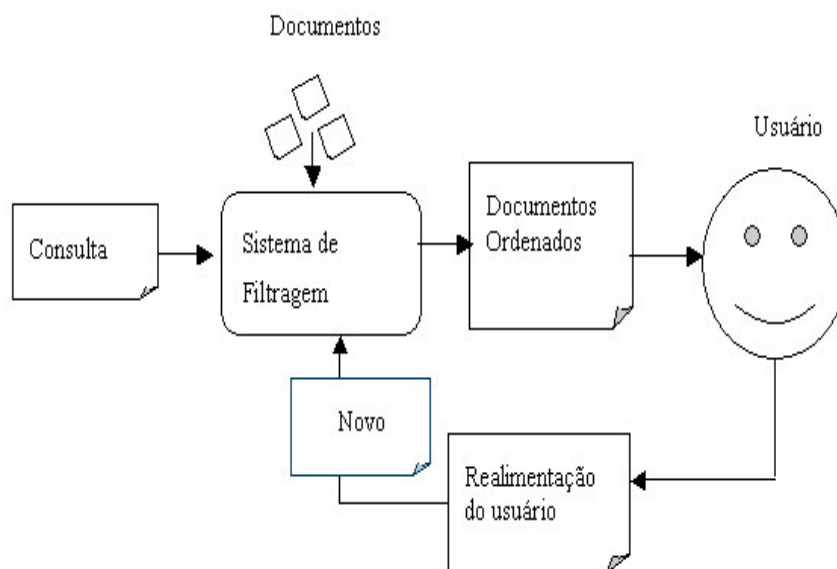


Figura 4: Método Relevance Feedback.

No *Relevance Feedback* mostrado na figura acima, o processo inicia-se com uma consulta inicial. Com esta consulta uma lista de documentos ordenados (*rankiados*) é gerada. Esta lista é apresentada ao usuário com os documentos classificados pelo algoritmo de classificação. O usuário checa a classificação dos documentos e faz um julgamento dos documentos que considerou relevantes e irrelevantes.

Com a realimentação do usuário, o algoritmo gera uma outra lista de documentos ordenados pelo valor de similaridade que satisfaça melhor as necessidades de informações do usuário. Esta nova lista é apresentada ao usuário e processo é repetido até que a necessidade de informação do usuário seja atingida.

Entendemos que o *Relevance Feedback* é um método de refinamento, que visa melhorar a consulta do usuário tornando-a mais precisa, recuperando somente os documentos relevantes e descartando os irrelevantes. Este método dividi-se em dois segmentos: a seleção dos termos, e a alteração dos pesos dos termos nas consultas utilizando fórmulas matemáticas complexas. Os algoritmos de recomendação, tais como *Rocchio* e o *kNN*, no decorrer deste trabalho implementados, estão baseado neste método *Relevance*

Feedback. Neste trabalho o *Relevance Feedback* é referenciado, como *Realimentação de Relevantes*.

Este método foi aplicado no nosso trabalho, com o objetivo de melhorar o perfil do usuário e recomendar os documentos como relevantes ou irrelevantes baseando-se na avaliação dos usuários.

2.6 Métricas para Avaliação de Desempenho dos Algoritmos

A Revocação (Recall) e a Precisão (Precision) são as métricas de qualidade mais utilizadas e conhecidas na atividade de avaliação dos sistemas de Recuperação de Informações.

$$\text{Precisão} = \frac{|N \cap R|}{|R|}$$

Onde,

N representa o conjunto de documentos relevantes que foram identificados pelos especialistas;

R é o conjunto de programas recomendados pelo sistema que foram analisados.

$$\text{Revocação} = \frac{|N \cap R|}{|N|}$$

Onde,

N representa o conjunto de documentos relevantes que foram identificados pelos especialistas;

R é o conjunto de programas recomendados pelo sistema que foram analisados.

Admitindo-se que ambas as medidas possuem o mesmo grau de importância, foi utilizada a medida F (*F-measure*), que combina em uma única fórmula as medidas de precisão e revocação. Esta medida é calculada pela seguinte fórmula:

$$\text{Medida } F = \frac{2 \times Pr \times Rc}{Pr + Rc}$$

Onde,

Pr é a precisão; e

Rc é a revocação.

De acordo com Baeza-Yates e Ribeiro-Neto (1998), um sistema de RI eficiente deve ter o seu valor de medida F o mais perto de um, que ocorre quando a revocação e a precisão se aproximam de um.

3. Sistema de Recomendação Personalizada de Programas de TV (SRPTV)

Esta seção descreve o Sistema de Recomendação Personalizada de Programas de TV (SRPTV) e as técnicas utilizadas para recomendação de programas, tais como algoritmos de classificação *Rocchio* e *kNN* (*k Nearest Neighbour*).

3.1 Algoritmos de Recomendação Estudados

No sistema foram experimentados dois algoritmos para implementar a recomendação de programas de televisão; *Rocchio* (1971) e o *kNN* (*K Nearest Neighbour*); Cover (1967). Cada um desses algoritmos implementados apresenta uma diferente abordagem para a recomendação. Todos utilizam um conjunto de documentos pré-classificados em relevantes e irrelevantes para recomendar novos programas baseados em técnicas de classificação. O objetivo dos algoritmos é recomendar automaticamente novos documentos de acordo com o comportamento dinâmico dos usuários.

3.1.1 Algoritmo Rocchio

O algoritmo *Rocchio* descrito em Rocchio (1971) é um algoritmo de realimentação de relevantes usado em sistemas de recuperação de informação. Ele assume que a indexação já está feita e que os documentos, incluindo o perfil, estão mapeados. Cada documento é tratado como um vetor baseado no Modelo Vetorial. A idéia básica é fazer com que o perfil melhore com o tempo.

O objetivo do algoritmo é encontrar um perfil ideal que expresse melhor as necessidades de informação do usuário. O perfil deve produzir um ranking de todos os documentos relevantes, isto é, programas de televisão, e eles são destacados em relação aos não relevantes.

3.1.2 Algoritmo kNN

O algoritmo do *k Vizinhos mais Próximos*, proposto em 1967, é um algoritmo de classificação conforme afirma Cover (1967). É referenciado, neste trabalho, como o *kNN*, que tem sido largamente usado para resolver problemas de reconhecimento de padrão e classificação de documentos. Para cada documento desconhecido, é verificada a similaridade entre ele e todos os documentos da base de treinamento, através de uma medida de distância.

A classificação é efetuada através da atribuição ao documento desconhecido à classe que for predominante, e é verificado a quais classes pertencem os *k* documentos mais próximos. Onde *k* é igual ao número de vizinhos. O *k Vizinhos mais Próximos* também utiliza o modelo convencional do espaço vetorial para representar cada documento como um vetor de termos.

3.2 Sistema Implementado

O SRPTV é um sistema de recomendação que combinou as duas abordagens de coleta do perfil do usuário a Explícita e Implícita. O sistema identifica os interesses do usuário através da interação com o sistema em que o telespectador de forma explícita especifica o seu perfil. O sistema possui uma fase de treinamento no qual são apresentados para o telespectador através de uma lista, exemplos de programas que poderão ser relevantes ou irrelevantes. A coleção de exemplos de treinamento será utilizada para melhorar o

perfil inicial baseado no interesse do telespectador, representado pela realimentação do telespectador.

O perfil é atualizado através do julgamento dos programas que são apresentados na lista. O sistema compara o perfil com os documentos que foram coletados. Em seguida gera uma lista ordenada de programas de acordo com o valor de similaridade. Os programas que possuem os maiores valores são colocados no topo de uma lista gerada dinamicamente em uma página Web.

Um aspecto importante de se ressaltar é que um determinado tipo de programa poderá ser de interesse para o telespectador por um determinado período, pois uma vez que o mesmo recebeu uma quantidade razoável de informação é provável que ele deseje receber outros tipos de programas, esta é a razão para que o sistema recomende programas de acordo com comportamento do mesmo. O sistema gera diariamente uma nova lista de recomendação.

3.3 Arquitetura do Sistema

Nós especificamos e comparamos duas arquiteturas de sistemas, com o objetivo de escolher qual é a mais adequada para o sistema de recomendação. Segue uma breve descrição na próxima seção da arquitetura que foi escolhida e utilizada.

3.3.1 Arquitetura do Sistema com Servidor

A arquitetura de sistema, é apresentada na figura abaixo, que define alguns módulos:

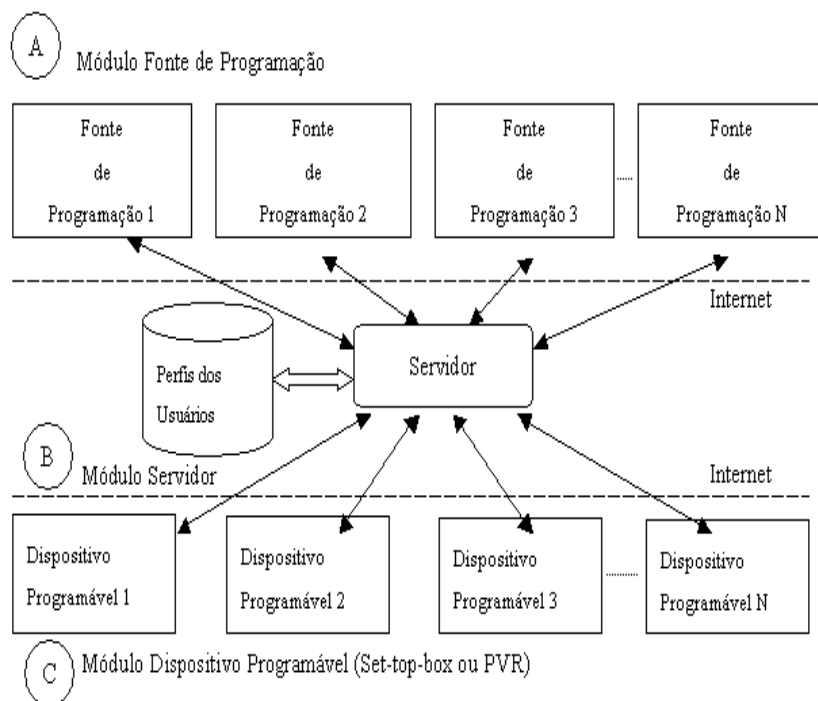


Figura 5. Arquitetura do Sistema

A arquitetura de sistema, além de possuir os módulos Fonte de Programação e Dispositivo Programável, utilizar um novo módulo: o *Servidor*, este módulo central processa os algoritmos responsáveis pela recomendação de programas de televisão,

foram incorporados no módulo o componente de recomendação e a base de perfil dos usuários. O resultado do processamento é enviado para os dispositivos eletrônicos.

3.4 Visão Geral do SRPTV

O sistema SRPTV foi implementado em HTML, PHP e JavaScript. As listas de programas exibidas aos usuários são geradas dinamicamente pelo CGI, executado pelo servidor Web Apache. O usuário interage com o sistema através de páginas Web geradas dinamicamente. O aplicativo cliente recebe as solicitações, interpreta e apresenta os resultados enviados pelo servidor em formato HTML. Para o protótipo do sistema foi utilizado um Web *browser*. Em trabalhos futuros poderá ser utilizado um *browser* embarcado em *set-top-box*. O núcleo funcional do sistema fica totalmente centralizado no servidor, pois os recursos computacionais são maiores comparados com o dispositivo do lado cliente.

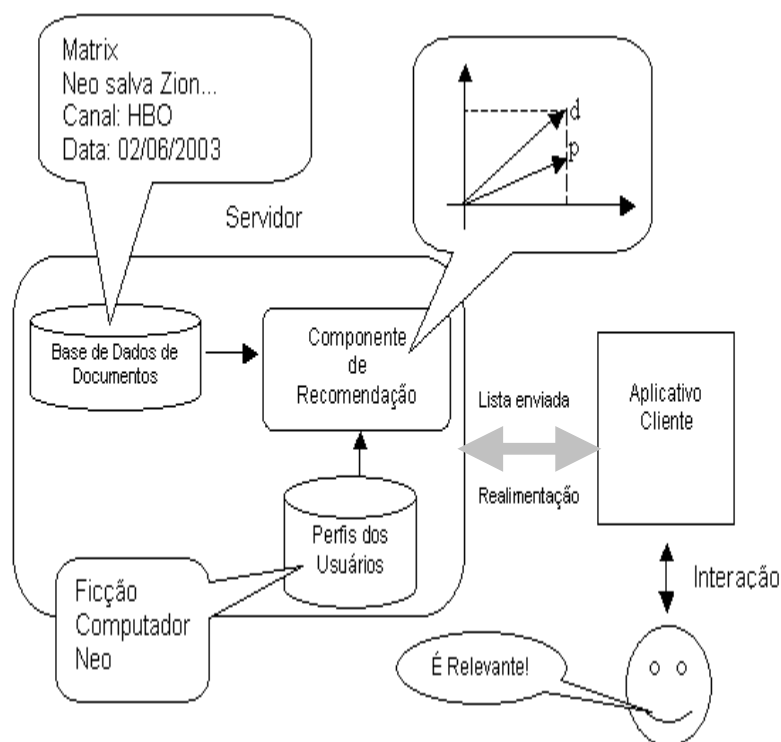


Figura 6. Visão Geral do SRPTV

Base de Dados de Documentos: esta base de dados contém todos os documentos coletados pelos *Crawlers* nas fontes de programação. Estes documentos possuem informações sobre cada programa de televisão, incluem detalhes, tais como gênero, data, canal, a hora inicial, final e algum texto que descreve o programa em questão (sinopse).

Base de Perfis dos Usuários: esta base de dados contém um conjunto de perfis de usuários que utilizam o sistema.

Servidor de Recomendação: componente de recomendação é o núcleo principal do SRPTV. Sua tarefa é examinar a informação do perfil e selecionar novos programas que serão recomendados para usuário.

Cliente: componente responsável pela visualização da lista personalizada gerada pelo componente de recomendação.

4. Avaliação dos Resultados Obtidos

Nesta seção serão apresentadas avaliações dos resultados dos experimentos realizados com os dois algoritmos para recomendação Rocchio e kNN. A seguir, serão explicados o Ambiente de Experimentação, a Descrição dos Experimentos e os Resultados.

4.1 Ambiente de Experimentos

Objetivo dos experimentos é destacar o algoritmo de recomendação que apresentar os melhores resultados. Desejamos verificar qual destes algoritmos é o mais adequado para a recomendação de programas de TV.

Para a realização dos experimentos com os algoritmos *Rocchio* e *kNN*, foram desenvolvidos experimentos em um domínio específico: filmes de televisão. Este domínio possui características peculiares. Primeiro, a grande quantidade e diversidade de filmes que são apresentados pelas emissoras de televisão; gêneros podem estar presentes em diferentes filmes ou em um único filme, e não há gêneros exclusivos de um único filme.

Segundo, os telespectadores normalmente possuem interesse por vários gêneros de programas, e o seu interesse é dinâmico, pois pode ser alterado com transcorrer do tempo. Desta forma, o domínio de filmes representa satisfatoriamente este espectro de necessidades dos telespectadores, e por esta razão, o ambiente de experimentação foi focado neste domínio a princípio.

Nos experimentos, foi utilizada uma coleção composta de 217 documentos, com o objetivo de efetuar uma análise comparativa entre os dois algoritmos de recomendação. Cada documento da coleção corresponde a um filme específico e possui um texto com dados estruturados, tais como descrição, nome, data, horas e canal. Estes dados foram coletados de forma não automatizada, através de pesquisas realizadas em vários *sites* e revistas especializadas na publicação da programação das emissoras de televisão.

4.2 Descrição dos Experimentos

Nos experimentos foram simuladas as utilizações diárias do sistema durante o período de uma semana. O objetivo do experimento era verificar qual algoritmo apresentava a lista de recomendação mais próxima das necessidades do usuário. Por esta razão, o procedimento de teste foi efetuado seguindo uma ordem seqüencial que correspondia aos dias da semana de domingo a sábado. Os usuários receberam 31 filmes filtrados pelo sistema em relevantes ou irrelevantes, a partir segundo dia, após a fase de treinamento, que foi executada no primeiro dia.

As filtragens de filmes geradas pelos algoritmos de recomendação e a realimentação de relevantes pelos usuários foram registradas em arquivos de *log* criados pelo sistema. Estes mecanismos armazenaram a quantidade de filmes checados pelos usuários e os filtrados pelo sistema.

A realimentação de relevantes consistia na avaliação em tempo de execução dos filmes classificados apresentados na lista de recomendação. O usuário indicava, através de

componentes de checagem localizados na lista de recomendação, quais filmes o mesmo considerava relevantes ou irrelevantes. Estes dados registrados foram utilizados para avaliar os resultados obtidos.

4.3 Resumo dos Resultados Avaliados

Os resultados seguintes são baseados nos dados registrados pelo sistema durante os procedimentos experimentais. O objetivo foi encontrar medidas de similaridade que indicassem a qualidade das recomendações do sistema SRPTV. A tabela 1 mostra o resultado final da análise comparativa.

Tabela 1. Médias das Medidas F calculadas durante o período de 6 dias

Resultado	2° Dia	3° Dia	4° Dia	5° Dia	6° Dia	7° Dia
Final	SRPTV	SRPTV	SRPTV	SRPTV	SRPTV	SRPTV
Algoritmos	Medida F	Medida F	Medida F	Medida F	Medida F	Medida F
Rocchio	0,50357	0,50176	0,51203	0,44829	0,43564	0,53403
kNN	0,45383	0,48723	0,49875	0,44077	0,41413	0,39565

Os usuários que participaram dos experimentos mostraram-se satisfeitos com os guias personalizados apresentados pelo SRPTV, entretanto, quando examinamos a tabela 1, nota-se que os melhores valores das médias das medidas F estão normalmente a 50 % do valor ideal, que seria *um*. Este dado significa que o módulo servidor de recomendação do nosso sistema SRPTV deverá ser melhorado, apesar dos resultados serem considerados aceitáveis. Um fator crítico foi a revocação no SRPTV que foi maior do que a precisão o que implicou no valor da medida F. De acordo com Baeza-Yates e Ribeiro-Neto (1998), em um sistema de recomendação eficiente deve existir um equilíbrio entre as métricas de avaliação. Existem determinadas técnicas que são sabidas para melhorar os resultados dos algoritmos de recomendação. Nós observamos que os resultados mostrados na tabela 1 indicam que o algoritmo *Rocchio* apresentou melhores resultados em praticamente todos os dias de experimentos. Pode-se notar na tabela 1 que os valores das médias finais do algoritmo *Rocchio* estão mais próximas de *um*, o que indica um desempenho melhor.

5. Conclusão e Trabalhos Futuros

Este trabalho apresentou uma possível solução para o problema da sobrecarga de informações na TVDI. Problema que atinge a maioria dos países onde a TVDI já é uma realidade. O problema surgiu devido ao crescimento do número de canais e conseqüentemente o número de programas de televisão que são disponibilizados. Em um futuro breve o Brasil também poderá ser afetado por este problema quando a TVDI for implantada. Desta forma antevendo este problema no Brasil, nós apresentamos uma proposta de solução que consiste em um protótipo de sistema de recomendação personalizada de programas de televisão (SRPTV). Esta ferramenta tratará o problema de sobrecarga de informação de maneira adaptativa e dinâmica, monitorando o comportamento dos usuários. O sistema utiliza os perfis dos usuários, técnica de

filtragem de informações e método de realimentação de relevantes baseado na avaliação e interação dos usuários com o sistema.

O trabalho também apresenta uma proposta de arquitetura de sistema que poderá ser utilizada para disponibilizar um serviço de recomendação personalizada de programas de televisão, vislumbrando o lado cliente e servidor. Uma aplicação deste tipo poderá ser disponibilizada na televisão digital interativa.

No desenvolvimento do trabalho experimentos foram efetuados com dois algoritmos para recomendação de documentos. O objetivo foi verificar qual algoritmo apresentava os melhores resultados e, por conseguinte ser indicado para o módulo servidor de recomendação do sistema desenvolvido. Os resultados indicaram que o *Rocchio* (1971) foi ligeiramente superior ao *kNN* (1996). O desenvolvimento do protótipo nos permitiu validar a solução proposta, verificar sua viabilidade.

Algumas considerações sobre o sistema de recomendação são:

- A filtragem de informação e recomendação dos documentos localizados no módulo servidor significa que o processamento de maior custo computacional será executado em uma máquina ou *cluster* com mais recursos disponibilizado por um provedor de serviços ou fonte de programação, o que, viabilizará a utilização de plataformas de hardware de menor custo pelos dispositivos eletrônicos;
- O módulo servidor de recomendação localizado conforme a arquitetura do sistema, também permitirá que versões de guias de recomendações possam ser geradas customizadas para dispositivos eletrônicos diferentes de acordo com suas características. Poderá ser utilizado além do perfil do usuário o perfil do dispositivo eletrônico.

Concluimos, que a implantação da TV Digital Interativa no Brasil e a convergência com dispositivos móveis deve contribuir para o surgimento de um espectro de possibilidades de novas aplicações. O avanço dos recursos computacionais dos dispositivos eletrônicos também é um fator que dever incentivar a realização de novos estudos com o objetivo de explorar os recursos avançados que estão sendo disponibilizados. Esperamos que este trabalho possa contribuir com novos estudos relacionados com o tema. Durante o desenvolvimento deste trabalho novos temas que possuem relação com o problema investigado foram identificados, tais como:

- Uma extensão deste trabalho seria o aprimoramento do algoritmo de filtragem na tentativa de melhorar a qualidade das recomendações do sistema;
- No processamento dos documentos técnicas podem ser aplicadas para o tratamento dos termos que agregam pouco significado semântico ao texto, através de técnicas, tais como Análise Léxica, Steaming, etc;
- Um outro trabalho seria é o desenvolvimento, ou customização de um aplicativo cliente tipo *browser* para ser adaptado em dispositivos eletrônicos, tais como *set-top-box*, *Personal Video Record* (PVR) e outros. Este aplicativo é o responsável pela interpretação dos dados enviados e da apresentação do guia personalizado aos usuários;

- Uma outra possibilidade é a utilização dos conceitos apresentados para o desenvolvimento de sistemas similares em outros domínios de informação, por exemplo, sistema de recomendação de informações sobre lazer para terminais móveis;

6. Referência

- Ribeiro-Neto, Berthier; Baeza-Yates, Ricardo (1998): Modern Information Retrieval.
- Rocchio, J. (1971) “Relevance Feedback in information retrieval” In: The SMART Retrieval System: Experiments in Automatic Document Processing, p. 313-23, Prentice-Hall. Inc.
- Yang, Y.; Pedersen, J. O. (1997) “A Comparative Study on Feature Selection in Text Categorization” In: International Conference On Machine Learning, 14. San Francisco: Morgan Kaufmann Publishers.
- Cover, T. M., and P. E. Hart (1967). Nearest Neighbor Classifiers. IEEE Transactions on Computers, 23-11, November, 1974, pp. 1179-1184.
- Rich, E.A. (1979) “User modeling via stereotypes” Cognitive Science, 3:329-354.
- Herlocker, J. L., Konstan, J. A., Borchers, A., Riedl, J. (1999) “An Algorithmic Framework for Performing Collaborative Filtering” In: Proceedings of the ACM SIGIR Conference on Information Retrieval, Berkley, CA, p. 230-37.
- Ehrmantraut, M., Harder, T., Wittig, H., Steinmetz, R. (1996) “The Personal Electronic Program Guide—towards the pre-selection of individual TV Programs”. In Proc. of CIKM'96, Rockville, MD, p. 243-50.
- Buckley, C., Salton, G. (1995) “Optimization of Relevance Feedback Weights” In: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington. USA. 9-13.
- Kurapati, K., Gutta, S., Schaffer, D., Martino, J., Zimmerman, J. (2001) “A Multi-Agent TV Recommender”. Adaptive Systems Department, Philips Research Briarcliff, p.1-2.
- Baudisch, P.; Brueckner, L. (2002) “TV Scout Lowering the entry barrier to personalized TV program recommendation” In: Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems (AH2002), Malaga, Spain, p. 29-31.
- P. Cotter and B. Smyth (2000) “PTV: Intelligent Personalised TV Guides” In: Proceedings of the 12th Innovative Applications of Artificial Intelligence IAAI Conference. AAAI Press.
- Kurapati K., Gutta, S. Schaffer, D. Martino, J. Zimmerman J. (2001) “A Multi-Agent TV Recommender” In: The First Workshop on Personalization in Future TV in conjunction with User Modeling 2001, Sonthofen, Germany, July 13-14.
- Pazzine M., Bilus D., Syskill and Webert (1996) “Identifying Interesting Web Sites” In: Proceedings of 13th National Conference in AL. p. 54-61.