

Personalização para Televisão Digital utilizando a estratégia de Sistema de Recomendação para ambientes multiusuário

Adriano S. Lucas, Sergio D. Zorzo

Departamento de Computação - Universidade Federal de São Carlos (UFSCar)
Caixa Postal 676 – 13565-905 – São Carlos – SP – Brazil

{adriano_lucas, zorzo}@dc.ufscar.br

Abstract. *The increasing content offer and the possibility of new services are two characteristics offered by the digital TV technology (TVD). The personalization is one of these services and it can be a possible solution to the information overload problem caused by the increasing content offer. This amount of information makes difficult the research and localization of interesting content. This paper describes a personalization model besides a recommendation system for multi-user environments originated from this model. The data mining algorithms from a case study were analyzed in the DTV domain and later a recommendation system was implemented using an API Java TV as a concept proof of the proposed model.*

Resumo. *O aumento da oferta de conteúdo e a possibilidade de novos serviços são duas características que a tecnologia de televisão digital (TVD) oferece. A personalização é um desses serviços e pode ser uma possível solução para sobrecarga de informação causada pelo aumento da oferta de conteúdo que dificulta a pesquisa e localização de conteúdo de interesse. Este artigo descreve um modelo de personalização, além de um sistema de recomendação para ambientes multiusuário originado deste modelo. A partir de um estudo de caso foram analisados algoritmos de mineração de dados no domínio de TVD, e posteriormente, como prova de conceito do modelo proposto, foi implementado um sistema de recomendação, o qual utilizou a API Java TV.*

1. Introdução

Com a tecnologia de televisão digital (TVD) é possível obter melhorias na qualidade de áudio e vídeo, possibilidades de serviços e aplicações e aumento da oferta de conteúdos televisivos. Este aumento faz com que os telespectadores, ao utilizarem esta tecnologia, encontrem uma grande variedade de conteúdos e recebam uma sobrecarga de informação. Os países pioneiros na implantação da TVD enfrentam este problema, o que torna difícil e demorada a procura pelos programas favoritos.

Para minimizar este problema, revistas impressas e o Guia Eletrônico de Programação (EPG – *Electronic Programm Guide*) são disponibilizados para a procura de programas preferidos. Entretanto, ambos os métodos fazem com que se perca tempo à procura do que assistir [Zhiwen e Xingshe 2004]. Nesse cenário, com objetivo de atenuar a sobrecarga de informação, é relevante empregar técnicas que ofereçam personalização, tornando-a necessária para a emergente era da televisão digital interativa (TVDI). A personalização tende a ser uma área de exploração, pois os telespectadores desejam que a TV seja capaz de identificá-los e fornecer recursos

personalizados e aplicações úteis [Kim e Mcgrath 2004]. Atualmente, as pesquisas dedicam-se ao desenvolvimento de sistema de recomendação de conteúdo, o qual é capaz de propor os conteúdos televisivos que correspondam aos interesses dos telespectadores [Chorianopoulos 2008]. Além do serviço de personalização, o sistema de recomendação também pode diminuir o problema de sobrecarga de informação ao refinar o conteúdo de forma inteligente. Assim, ajuda os telespectadores ou grupo destes a encontrar informações relevantes referentes a conteúdos televisivos que coincidam com as suas preferências [Goren-Bar e Glinansky 2004].

Para desenvolver um sistema de recomendação para TVD é necessário priorizar métodos de modelagem para ambientes multiusuário, já que a TV é essencialmente um meio de informação coletivo e diversos sistemas de recomendação existentes para TVD não o fazem. Chorianopoulos (2008) evidencia a importância de aplicar métodos de recomendação dentro de grupos, como de amigos e família. Deste modo, as recomendações não devem focar na visualização de cada telespectador, mas sim no que tem sido visto pelo grupo. Essas características foram observadas para a concepção da estratégia de personalização utilizada neste trabalho que consiste em um sistema de recomendação de conteúdo para ambientes multiusuário.

A maioria dos sistemas de recomendação para TVD utiliza informações explícitas fornecidas pelos telespectadores para que as preferências sejam utilizadas para uma recomendação mais próxima dos interesses. É possível também obter um *feedback* explícito, ou seja, o telespectador pontua e classifica as recomendações, porém torna o sistema intrusivo. Assim, esses sistemas dependem da ação dos telespectadores para configurá-los caso queiram receber recomendações de conteúdo de seus interesses. Deve-se ressaltar que alguns telespectadores podem não ter afinidade com sistemas computacionais ou não querer gastar tempo para configurá-lo com suas preferências. O desafio desta proposta é utilizar técnicas e algoritmos para refletir as preferências de um grupo de telespectadores de forma totalmente implícita, sendo esta a motivação principal deste trabalho.

Assim, é apresentado um modelo de personalização para TVD que utiliza a estratégia de sistema de recomendação e origina um novo sistema, o RePTVD (Recomendação Personalizada para Televisão Digital). Este oferece recomendações personalizadas de conteúdo para grupos de telespectadores de um mesmo local e considera as preferências observadas anteriormente. Estas são coletadas implicitamente e, a partir de tais informações, o sistema identifica diferentes padrões de comportamento dentro de um grupo e determina as melhores recomendações. Para descobrir os padrões de comportamento no histórico de utilização do grupo, utilizou-se a técnica de mineração de dados que é uma das possibilidades para gerar conhecimento adicional e foi escolhida para demonstrar a viabilidade da estratégia de recomendação.

O objetivo do sistema RePTVD não é identificar os membros de um grupo, mas sim os padrões de comportamento e tendências deste. Sendo assim, ao descobrir os padrões, é possível oferecer conteúdo adequado em determinados dias e horários, independente de quem esteja assistindo. Deste modo, o sistema se adapta e recomenda conteúdos seguindo o padrão de comportamento mesmo com grupos heterogêneos.

Conforme Chorianopoulos (2008), os sistemas de recomendação combinam técnicas e algoritmos para melhorar as recomendações e inferir as preferências dos telespectadores. Assim, foi realizado um estudo de caso com tarefas de mineração de

dados com objetivo de analisá-los no domínio de TVD. Após a análise, uma aplicação foi implementada com uso da API Java TV [Sun Microsystem 2006] e utilizado o emulador Xletview [Xletview 2008] como ambiente de execução. Esta aplicação é compatível com os outros sistemas de TVD, o que evidencia a portabilidade do modelo.

Este trabalho está organizado da seguinte forma: a seção 2 descreve os trabalhos relacionados; na seção 3 é apresentado o modelo de personalização proposto; na seção 4 é descrita a implementação de um sistema de recomendação; a seção 5 descreve os estudos de caso e a seção 6 apresenta as considerações finais.

2. Trabalhos Relacionados

Vários sistemas de recomendação para TVD foram construídos para oferecer um serviço diferenciado de personalização e auxiliar os telespectadores a lidar com a abundância de programas de TV. A seguir, são apresentados sistemas que possuem alguma intersecção com o trabalho apresentado neste artigo.

Em [O'Sullivan et. al 2004], o aumento do volume de conteúdo nos dispositivos de Gravador de Vídeo Pessoal (PVR) é abordado como um novo desafio para os telespectadores que devem procurar e navegar por coleções de conteúdos de vídeos interessantes e obtê-los de maneira intuitiva e eficiente. Assim, o Guia de Programação Eletrônico Personalizado (pEPG – *personalized Eletronic Program Guide*) é visto como uma possível solução para o problema de sobrecarga de informação. Os autores concentram-se na mineração de dados que pode ser usada para suavizar este problema e na tecnologia de recomendação dos pEPGs. A tarefa de mineração de dados utilizada para extrair novos metadados de programas a partir dos perfis dos telespectadores foi regras de associação. Os metadados podem aumentar significativamente o conhecimento sobre a similaridade e relevância do programa. As avaliações dessa abordagem mostram que essa técnica fornece uma precisão superior de recomendação comparada com abordagens tradicionais. Os autores acreditam que esta oferece uma fundação sólida para o modelo de utilizador individual que foi usado no estudo. Também foi comparado o uso de perfil explícito e implícito e provou-se que os indicadores de interesse implícito são semelhantes aos de interesse explícitos.

Bozios et. al (2001) apresentam uma abordagem que permite a identificação de grandes grupos de telespectadores pelas empresas de publicidade e possibilita o desenvolvimento de técnicas de publicidade e oferta de marketing personalizado. Os dados demográficos, as preferências e a análise da interação do consumidor com a TVD são coletados automaticamente e assim, o sistema define os perfis dos consumidores e seus agrupamentos. Os dados são armazenados a priori no *set-top-box*¹ (STB) e posteriormente enviados para um servidor que aplicará a técnica de mineração de dados. Diferentemente da abordagem adotada por esses autores, o modelo proposto neste trabalho coleta dados de comportamento de visualização do telespectador de forma implícita. Além disso, os dados permanecem no STB para descoberta de padrões de comportamento pela aplicação de mineração de dados.

O trabalho de [Hsu et. al 2007] propõe um mecanismo de recomendação que considera características como atividades, interesses, humores, histórico do uso da TV e

¹ *Set-top-box* - dispositivo responsável pela recepção e decodificação de sinais digitais de áudio e vídeo, além de oferecer serviços interativos.

informações demográficas, as quais alimentam um modelo de rede neural que infere as preferências de programas dos telespectadores. Utiliza-se uma abordagem híbrida baseada nos métodos de filtragem de conteúdo e filtragem colaborativa. Assim, os telespectadores respondem questionários para configurar todo o sistema conforme suas preferências. Esta abordagem diferencia-se do trabalho proposto neste artigo que visa realizar toda coleta implicitamente sem nenhuma intervenção dos telespectadores.

Em [Vildjiounaite et al 2008] é apresentado um método para descobrir modelos de ambiente multiusuário a partir de interações implícitas de escolhas feitas por diferentes subgrupos de um grupo. Este método é capaz de reconhecer pessoas e suas ações, pois utiliza as informações armazenadas nos *logs* de casas inteligentes que podem ser usados por um sistema de recomendação para reduzir o esforço e adaptar o conteúdo para cada telespectador ou para situações multiusuário. Foi focado o desenvolvimento de um sistema baseado somente em dados de cada casa e que adapta as práticas familiares ao invés de forçar uma prática a priori. Uma desvantagem do sistema é o tempo que requer para o aprendizado, além dos telespectadores terem que se identificar pelo controle remoto. Na avaliação com históricos de telespectadores de 20 famílias, a precisão do método foi semelhante ao de um sistema explícito. A maioria dos sistemas de recomendação desenvolvidos requer alguma interação explícita ou a classificação dos conteúdos por parte dos telespectadores para obter o *feedback* explícito. Diferentemente destes, a proposta apresentada neste artigo é desenvolver um sistema de recomendação totalmente implícito para ambiente multiusuário.

O sistema de recomendação desenvolvido por Goren-Bar e Glinansky (2004), denominado FIT, efetua recomendações de programas para membros de uma família por modelos baseados em estereótipos e histórico de visualização para inferir qual subconjunto de membros da família está assistindo TV. O sistema foi avaliado em três grupos de família e em cada um foi aplicado um método. Deste modo, para o primeiro grupo foram criadas predições aleatoriamente, para o segundo, geradas predições pelo FIT e para o terceiro foi usado o conhecimento de quem estava assistindo a TV. A precisão do FIT foi mais baixa que no terceiro grupo e mais alta que no primeiro grupo.

Em [Zhiwen et. al 2006], é proposto uma estratégia de recomendação de programa para múltiplos telespectadores baseada na junção do perfil do telespectador. O estudo analisou três estratégias para efetuar as recomendações de conteúdo e a análise proporcionou a escolha da estratégia baseada na junção do perfil. Os resultados comprovaram que a união do perfil dos telespectadores pode refletir de forma adequada nas preferências da maioria dos membros de um grupo. Observou-se que o sistema é aceitável para grupos homogêneos, porém membros de grupos heterogêneos não ficaram satisfeitos com as recomendações.

3. Modelo de personalização para TVD

O modelo proposto estabelece as fases para a oferta da recomendação de conteúdo e como estas interagem entre si. A Figura 1 apresenta o modelo de personalização para TVD proposto. A seqüência de passos e as características que compõem o modelo são descritas ao admitir um possível cenário de uso.

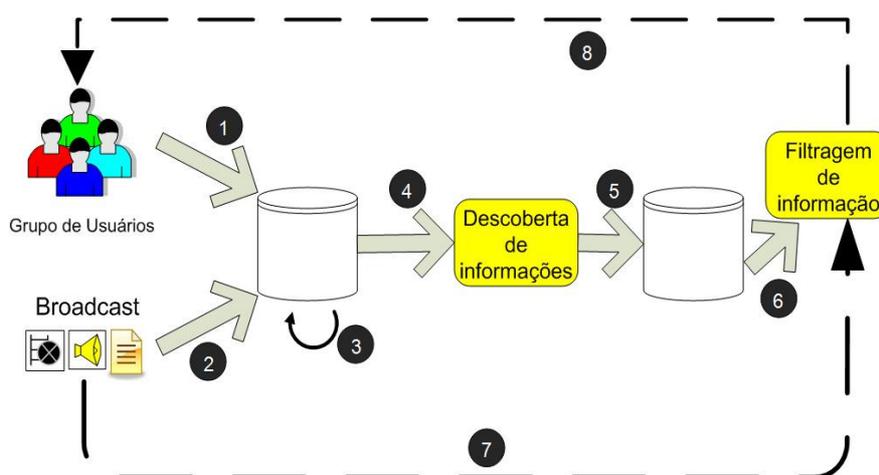


Figura 1 – Esquema dos passos do modelo proposto

- (1) Enquanto o grupo de telespectadores interage com a programação da TV, algumas informações referentes à visualização são coletadas implicitamente e armazenadas e assim é obtido o histórico de visualização.
- (2) As emissoras enviam o vídeo, áudio e metadados referentes à programação exibida. Os metadados são extremamente importantes para o modelo e, conseqüentemente, para efetuar recomendações. Alguns dados referentes aos programas que os telespectadores estão assistindo são coletados e armazenados.
- (3) Os dados coletados referentes ao comportamento de visualização e ao conteúdo proveniente das emissoras são armazenados em um arquivo que fica localizado no STB. Este arquivo é atualizado dinamicamente e alimentado por essas informações durante o uso da TV.
- (4) Com a utilização de técnicas e algoritmos, informações úteis são descobertas no arquivo armazenado no STB, as quais refletem as preferências dos telespectadores. Esse modelo permite que diversos algoritmos sejam explorados para implementar o sistema de recomendação.
- (5) As informações relevantes encontradas são armazenadas em um novo arquivo que contém as preferências do grupo de telespectadores. Com isso, é possível oferecer recomendações que reflitam as preferências do grupo.
- (6) e (7) É realizada uma intersecção entre as preferências descobertas e os metadados referentes à programação para efetuar a filtragem de informação com o objetivo de encontrar os conteúdos adequados para serem recomendados.
- (8) Por fim, as recomendações são apresentadas aos telespectadores quando solicitadas.

Este modelo é compatível com qualquer sistema de TVD existente. O principal objetivo é descobrir diferentes padrões de comportamento e regras para efetuar recomendações direcionadas a um grupo de telespectadores. Este modelo foi implementado como prova de conceito e originou o sistema de recomendação RePTVD. Para isso, foram consideradas características inerentes a TVD como o baixo envolvimento com o ambiente, dispositivos de entrada, distância de visualização da TV, grupo de visualização, entre outros. Sendo assim, o RePTVD não é simplesmente uma adaptação de um sistema de recomendação da WEB que pode ser utilizada para a TVD. É um sistema totalmente baseado nas características da TVD que não está atrelado a nenhum vídeo e sua existência é independente dos programas audiovisuais em

transmissão. O RePTVD contempla todos os passos descritos no modelo e a seguir, sua arquitetura e seu funcionamento são detalhados.

4 Sistema RePTVD

Para que o sistema RePTVD possa recomendar itens, produtos ou conteúdos são necessárias técnicas e estratégias para coleta e manipulação de informações, além da apresentação das sugestões aos telespectadores.

Primeiramente, a coleta de informações para a adaptação ou personalização de conteúdo pode ser realizada de maneira explícita ou implícita. O método explícito exige tempo e esforço por parte dos telespectadores e não se adapta a evolução de suas preferências [Chorianopoulos 2008] [Vildjiounaite et. al 2008]. Já a coleta implícita pode ser realizada por monitoramento e proporciona diminuição do esforço dos telespectadores, além de deduzir seus interesses mediante o histórico de visualização [Goren-Bar e Glinansky 2004]. Pesquisa realizada por O'Sullivan et. al (2003), mostrou que os perfis implícitos são tão precisos, se não mais, quanto suas contrapartes explícitas. O RePTVD coleta as informações de forma implícita mediante o monitoramento do comportamento de visualização dos telespectadores.

Em segundo lugar, o componente central dos sistemas de recomendação é a técnica de recomendação a ser empregada para manipular as informações. Para realizar a filtragem de informação, os sistemas de recomendação tradicionais usam as técnicas de filtragem colaborativa, filtragem baseada em conteúdo e a junção das duas anteriores, o método híbrido. A filtragem colaborativa utiliza a relação entre perfis de telespectadores distintos com preferências semelhantes. A filtragem baseada em conteúdo relaciona os conteúdos semelhantes com as preferências dos telespectadores. Porém, para aprimorar a qualidade das recomendações, diversos autores adotam combinações de técnicas juntamente com filtragem de informação para gerar as recomendações. A rapidez para gerar a recomendação também é um fator relevante em alguns domínios. O RePTVD descobre as informações pelo uso de algoritmos de mineração de dados e depois utiliza filtragem de conteúdo para ofertar a recomendação.

Por fim, há diversas possibilidades para representar as recomendações e uma destas é uma lista de recomendação com os conteúdos mais similares às preferências dos telespectadores. Preferencialmente, os conteúdos que são assistidos frequentemente não devem ser recomendados. Porém, há relevância em recomendar um programa já assistido caso não tenha sido visualizado por algum tempo, além de outros programas com conteúdos semelhantes que podem ser desconhecidos pelo telespectador.

O desafio de um sistema de recomendação para TVD, incluindo o RePTVD, é representar e oferecer recomendações de acordo com as preferências individuais dos integrantes de um grupo, identificando-os entre todos em um ambiente multiusuário de forma implícita. Chorianopoulos (2008) explorou o estado da arte da personalização na TVD relacionado ao processo de recomendação de conteúdo e evidenciou que os desafios da área concentram-se nos métodos para recomendar conteúdo e nas técnicas de coleta de dados.

4.1 Arquitetura do RePTVD

A arquitetura do RePTVD foi desenvolvida com a opção de alocação das decisões no cliente, ou seja, a aplicação permanecerá residente no STB. Um requisito fundamental

para o RePTVD é o STB possuir memória suficiente para a persistência de informações referentes ao histórico de visualização dos telespectadores. Por ser uma aplicação residente e fazer uso de memória, é necessário que a implementação seja otimizada. O sistema de recomendação RePTVD é composto de três módulos como demonstrado na Figura 2.

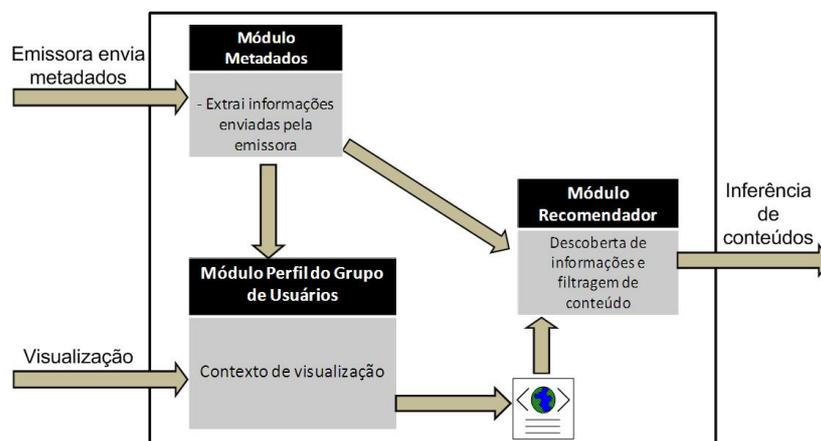


Figura 2 – Módulos do sistema RePTVD

O RePTVD possui interação sem canal de retorno, pois considera a transmissão unidirecional, na qual o STB funciona como servidor de aplicações. Assim, diferentemente de outros tipos de serviços como comércio e banco eletrônico e e-mail, o RePTVD não considera a infra-estrutura de canal de retorno. Novas possibilidades poderão ser oferecidas pelo canal de retorno, porém há questões ainda não equacionadas quanto a sua eficiência em desempenho e aplicabilidade.

4.1.1 Módulo Metadados

O módulo de metadados analisa os dados transmitidos pela tabela SI (*Service Information*) e extrai as informações referentes aos conteúdos televisivos importantes para o RePTVD. Com os dados transmitidos por *broadcasting* pelas tabelas SI, torna-se possível selecionar canais e eventos existentes através de um receptor digital de TV. Dentre as várias tabelas que compõem o SI, as mais importantes para este trabalho são a EIT (*Event Information Table*) - tabela de informação de eventos, a TDT (*Time and Date Table*) - tabela de data e horário e a TOT (*Time Offset Table*) - tabela de referência de data e horário. A tabela EIT oferece informações em ordem cronológica sobre os eventos existentes por canal. A TDT deve ser utilizada para informar data e hora do sistema. Já a TOT é responsável por informar ao receptor a hora, data, fuso horário e a existência de horário de verão, sendo transmitida obrigatoriamente pelo radiodifusor [ABNT 2007].

A tabela EIT contém os dados da grade de programação por um período de até 14 dias, tanto da programação futura como da programação atual. Assim, algumas informações relevantes para o sistema RePTVD são coletadas à medida que os telespectadores assistem a programação e armazenadas para uso posterior. Na Figura 3 é possível observar todas as métricas adotadas, as quais são relevantes para o sistema RePTVD, sendo que as métricas número do canal, nome do programa e categoria do programa são destacadas por serem coletadas por esse módulo. O RePTVD armazena os dados coletados em um arquivo ARFF (*Attribute Relation File Format*), que é um

formato próprio do pacote Weka (*Waikato Environment for Knowledge Analysis*) e este pacote contém os algoritmos de mineração de dados utilizados neste trabalho [Witten e Frank 2005].

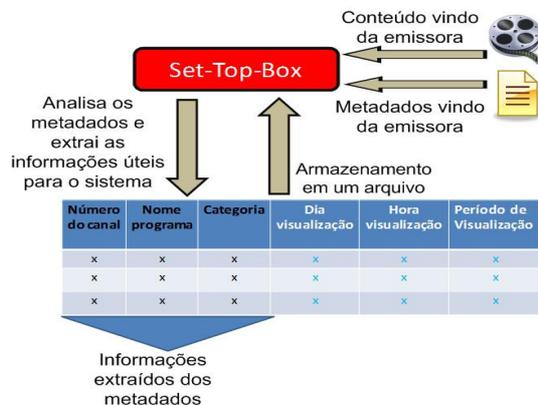


Figura 3 – Extração dos metadados

O módulo metadados atua em conjunto com o módulo perfil do grupo de telespectadores e ambos armazenam as informações no mesmo arquivo ARFF. Este módulo de metadados foca nos dados referentes aos programas, enquanto o módulo perfil do grupo de telespectadores, nos dados referentes ao comportamento de visualização dos telespectadores. Assim que a TV é ligada, os dois módulos tornam-se ativos e em cada troca de canal o arquivo ARFF é atualizado. Com o objetivo de minimizar as informações insignificantes devido ao “zapping²”, o registro é armazenado no arquivo ARFF somente após 60 segundos em um mesmo canal.

4.1.2 Módulo Perfil do Grupo de Telespectadores

O módulo perfil do grupo de telespectadores coleta, de forma implícita, informações de visualização do grupo. A coleta implícita de dados é realizada pelo monitoramento da troca de canais e tempo que o telespectador assiste cada canal. Para realizar a coleta implícita, é necessário que o sistema residente a um STB seja habilitado pelo grupo de telespectadores. Dessa forma, contempla-se a privacidade e opção do grupo para autorizar o sistema a coletar dados e efetuar as recomendações. A Figura 4 ilustra o esquema de coleta das métricas do comportamento de visualização dos telespectadores.

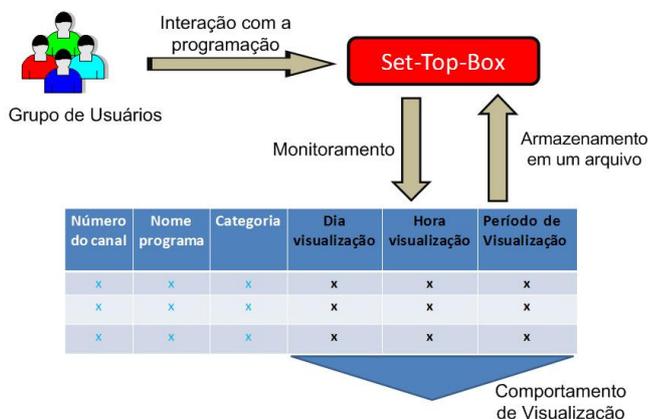


Figura 4 – Informações dos telespectadores

² Constante mudança de canal com uso do controle remoto

Esse módulo utiliza a experiência do contexto de visualização como previsor de recomendação. Sendo assim, para obter essa experiência e com o objetivo de aumentar a precisão da recomendação, o arquivo ARFF armazena os dados de visualização. No STB o arquivo é armazenado e atualizado de forma contínua e autônoma conforme a interação dos telespectadores com a programação da TV. As métricas desse módulo são dia, hora e período de visualização em cada canal e são coletadas simultaneamente com as métricas do módulo de metadados. Deste modo, é possível descobrir características inerentes ao grupo de telespectadores de um STB.

4.1.3 Módulo Recomendador

Após a coleta e armazenamento das informações no arquivo ARFF, este se torna o histórico de visualização do grupo. Após determinado tempo de uso da TV, esse conjunto de informações certamente conterá muitos dados. Parte dessas informações pode não ser relevante e, conseqüentemente, não ser aproveitada pelo sistema de recomendação. Dessa forma, o módulo recomendador possui duas funções: descobrir informações e filtrá-las para ofertar a recomendação.

A primeira função é responsável por utilizar o arquivo armazenado no STB para realizar a captura de informações específicas e descobrir padrões ou regras que permitam uma melhor compreensão dos dados. Para isso, é utilizada a mineração de dados, mais especificamente a tarefa de regras de associação. Com isso, as informações obtidas tornam-se relevantes para o contexto em que o telespectador está inserido e as recomendações de conteúdos refletem as preferências do grupo de telespectadores. A mineração permite que o sistema de recomendação de conteúdo efetue recomendações baseadas principalmente no conhecimento acumulado e armazenado no STB. A Figura 5 ilustra a primeira função do módulo recomendador.



Figura 5 – Descoberta de informações úteis

A técnica de mineração de dados foi escolhida por evidenciar tendências e comportamentos implícitos que são previamente desconhecidos e potencialmente úteis, além de demonstrar a viabilidade da estratégia do sistema de recomendação proposto. Essa técnica depende basicamente do tipo de tarefa a ser realizada sendo que cada tarefa tem um algoritmo que melhor se adapta e obtém resultados. Assim, é relevante o estudo relacionado à utilização de diferentes algoritmos para identificar aquele que apresenta melhores resultados para determinado domínio. Os conhecimentos obtidos pelas técnicas de mineração de dados são geralmente expressos na forma de regras e padrões [Han e Kamber 2006].

Foi realizada, como estudo de caso, a análise de algoritmos de mineração que realizam a tarefa de classificação, associação e agrupamento. O objetivo foi verificar o

comportamento de cada uma dessas tarefas no domínio de TVD. Assim, este módulo realiza a mineração pela utilização do arquivo de comportamento de visualização do grupo de telespectadores e depois de descobertos as regras e o conhecimento, um novo arquivo é armazenado no STB. Por motivos de desempenho, a mineração é programada para ser realizada uma vez a cada 24 horas e o arquivo com as preferências é constantemente atualizado com os resultados descobertos.

A segunda função utiliza a filtragem de conteúdo para comparar as preferências descobertas com o conteúdo enviado pelas emissoras e efetuar a criação de uma lista de recomendação dinâmica. Este processo é contínuo e ocorre sempre que o telespectador solicita uma recomendação. Ao solicitá-la, o RePTVD verifica o relógio interno do STB e ao obter a data e horário, realiza uma pesquisa no arquivo que contém as preferências para encontrar as recomendações adequadas e mostrar ao telespectador. Além disso, uma lista de recomendação para o grupo é disponibilizada e atualizada a cada 24 horas. A Figura 6 apresenta como é realizada a filtragem de informação.

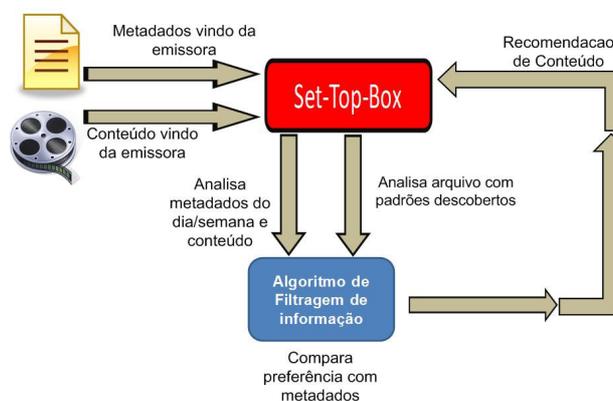


Figura 6 – Filtragem de informação

5. Estudo de Caso

Para validar a implementação do sistema RePTVD, foi desenvolvida uma aplicação *Xlet* que contempla os três módulos do sistema. Antes da implementação do sistema foi realizado um estudo de caso que consistiu em analisar as tarefas de mineração para serem aplicadas nos dados coletados implicitamente. Sendo assim, o estudo de caso foi dividido em duas etapas: na primeira foi realizada uma análise das tarefas de mineração e na segunda, a modelagem computacional do sistema.

5.1. Experimento de análise de tarefas de mineração

Este estudo analisou três tarefas de mineração de dados no domínio de TVD que são regras de associação, classificação e agrupamento, implementadas pelos algoritmos *Apriori*, *C4.5* e *K-means*, respectivamente, disponíveis no software Weka, o qual foi adotado para essa fase do trabalho [Witten e Frank 2005]. A associação é aplicada para descobrir padrões escondidos em conjunto de dados que contém itens relacionados à ocorrência de outros itens. A classificação consiste em construir um modelo que possa ser aplicado a dados não classificados com o objetivo de categorizá-los em classes. O agrupamento é a tarefa que procura dividir populações heterogêneas em subgrupos homogêneos [Han e Kamber 2006]. Os testes dos algoritmos foram realizados com a utilização de conjuntos de dados correspondentes ao comportamento de visualização de uma amostra de telespectadores.

O modo de teste utilizado no experimento foi o *Cross-validation*, um dos mais comuns e úteis quando a disponibilidade de dados é pequena. Esta opção realiza a validação cruzada com os subconjuntos da base de dados e para isso é escolhido um número de conjuntos n , no qual $n-1$ serão utilizados para treinamento e o n -ésimo é utilizado para teste. O modo de teste *Cross-validation* foi repetido com o objetivo de se obter maior consistência na análise dos resultados. Deste modo, três configurações de experimentos foram analisadas: uma repetição com cinco subconjuntos, duas repetições com cinco subconjuntos e dez repetições com cinco subconjuntos para cada algoritmo.

Após a análise dos algoritmos, foi observado que o *Apriori* foi o que melhor se adaptou ao domínio de TVD devido ao número de regras encontradas e a diversidade destas, as quais tiveram relevância para este contexto. O número de regras obtidas pelo C4.5, a partir do histórico de visualização, foi insuficiente nas primeiras semanas para inferir conteúdo conforme o que tinha sido visto na TV. Isto acontece pois o algoritmo utiliza o histórico de visualização para construir um modelo e a partir deste, inferir conteúdo. O *K-means* agrupou registros baseados na similaridade e descobriu padrões, porém, no domínio de TVD os agrupamentos similares causaram uma especialização das regras geradas, ou seja, não houve diversidade das regras conforme o aumento do histórico de visualização. O agrupamento seria pertinente se o sistema identificasse os usuários, pois ajudaria na descoberta de grupos distintos de telespectadores.

Dentre os algoritmos utilizados e considerando a arquitetura proposta, observou-se que as regras de associação resultaram no melhor comportamento por serem aplicáveis a dados que não tenham uma classe determinada para efetuar a mineração, como a categoria e o tempo de visualização do programa. Deste modo, reflete exatamente o que o sistema necessita, ou seja, obter informações escondidas sem treinar ou adotar uma classe específica para minerar a partir de um conjunto de dados brutos, os quais correspondem ao histórico de visualização da TV.

5.1.1 Resultados

A oferta de conteúdos desejáveis aos telespectadores realizada pelo RePTVD foi efetuada com o uso de regras de associação. A análise foi realizada pela medida da precisão do sistema para apresentar conteúdos relevantes às expectativas dos telespectadores. Na fórmula $\delta = \beta / \alpha$, δ expressa a eficiência do sistema e varia de 0 a 1, β é o número de recomendações visualizadas pelos telespectadores e α é o número de recomendações efetuadas. Quanto maior o resultado de δ , maior é a precisão que o sistema de recomendação possui em ofertar conteúdo de acordo com as preferências do grupo de telespectadores. Entende-se que o valor no intervalo entre 0 e 1 corresponde ao fator equivalente a um percentual entre 0 e 100 de aceitação das recomendações. Assim, o valor de $\delta = 0,28$ corresponde a 28% de aceitação.

Os dados foram extraídos de uma amostra de treze telespectadores subdivididos em quatro grupos: o primeiro, segundo e terceiro grupos possuíam três pessoas cada, e o quarto, quatro. O comportamento de visualização foi coletado conforme a grade de programação real da TV aberta e o tempo de visualização de cada telespectador variou de acordo com a sua rotina. Durante quatro semanas, o histórico de visualização dos grupos foi submetido ao sistema de recomendação. Assim, no primeiro dia o sistema coletou o comportamento de visualização e realizou a mineração ao final daquele dia. Após isso, foi realizada a filtragem de informação conforme as regras obtidas, as quais

eram comparadas com a grade de programação do dia seguinte, com o objetivo de gerar uma lista de recomendações. Deste modo, a contabilização da eficiência do sistema foi mediante o uso de *feedback* implícito, ou seja, se o telespectador teria optado em assistir as recomendações oferecidas. No segundo dia, foi realizada novamente a coleta do comportamento de visualização, porém a lista de recomendações para o terceiro dia foi gerada com referência aos dois dias anteriores de coleta, e assim sucessivamente. A Figura 7 ilustra os resultados observados.

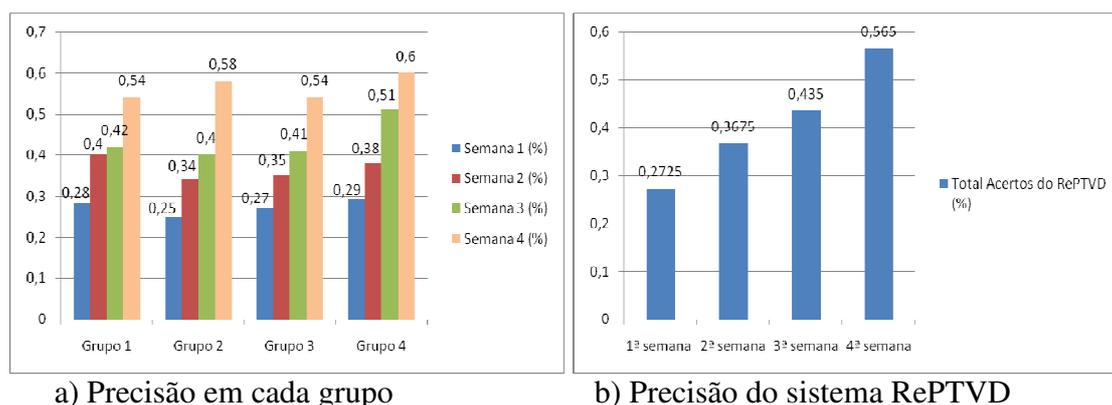


Figura 7 – Gráficos comparativos da precisão do RePTVD

Na primeira avaliação foi constatada a precisão das recomendações em cada grupo. Como pode ser observado na Figura 7a, verificou-se que a precisão do RePTVD na primeira semana variou entre 25% e 29%, já na quarta semana, entre 54 e 60%. Assim, observou-se que após quatro semanas o sistema obteve uma variação de melhora de 29 a 31%. A segunda avaliação foi em relação à média da precisão de todos os grupos durante as quatro semanas e pode ser observado na Figura 7b. Deste modo é visível que o sistema RePTVD obteve média de precisão de 27,25% na primeira semana e de 56,5% na quarta semana, na qual foi constatada melhora de 29,25%. Fica evidente que nas primeiras semanas, como são poucos os dados coletados, o algoritmo *Apriori* não extraiu informações relevantes das preferências do grupo. Com o aumento dos dados no histórico de visualização na terceira e quarta semana, o algoritmo *Apriori* pode obter melhores resultados e o índice de aceitação das recomendações aumentou.

5.2 Implementação do RePTVD

Para realizar a implementação do RePTVD, utilizou-se a linguagem de programação Java, mais especificamente a API Java TV. A maioria dos sistemas de TVD utiliza a especificação Java TV pois permite o desenvolvimento de aplicações *Xlet* e fornece facilidade para o desenvolvimento de aplicações para TVD. Foi utilizada uma emulação em um computador devido à inviabilidade do uso do sistema de televisão digital brasileiro (ISDTV-T - *Internacional Standard for Digital Television Terrestrial*). Este utiliza o *middleware* Ginga, o qual é especificado pela norma ABNT (Associação Brasileira de Normas Técnicas) [Ginga 2006] e possui apenas o interpretador GINGA-NCL disponível, faltando ainda o módulo GINGA-J para suportar a execução de *Xlets*. Em relação ao desenvolvimento de aplicativos, o *middleware* é a parte mais importante de um sistema de TVD. Sendo assim, para efetuar a emulação de um ambiente de TVD em um desktop e executar o RePTVD, foi utilizado o Xletview [Xletview 2008], um emulador baseado no *middleware* MHP que foi escolhido por ser Open Source. A execução do sistema RePTVD é ilustrado pela Figura 8.



Figura 8 – Sistema RePTVD

Os algoritmos de mineração foram obtidos do Weka, um software que efetua mineração de dados e aprendizado de máquina, desenvolvido em Java e distribuído com licença pública GNU (*General Public License*). As características e técnicas implementadas são descritas de forma detalhada em [Witten e Frank 2005]. Os algoritmos podem ser aplicados tanto a um conjunto de dados quanto em chamadas no código Java. Como o Xletview não tinha suporte para a biblioteca Weka, foi necessário acoplar essa nova funcionalidade e para isso, realizou-se uma extensão deste emulador para tornar possível a utilização dos algoritmos de mineração em chamadas no *Xlet*.

6. Considerações Finais

O domínio de TVD contribuiu para o surgimento de novas aplicações na área de personalização de conteúdo e este trabalho apresentou uma dessas possibilidades. Foi desenvolvida uma aplicação *Xlet*, a qual consiste em um sistema de recomendação para ambientes multiusuário. Mediante o estudo de caso das tarefas de mineração no domínio de TVD, observou-se uma melhor aplicabilidade da tarefa de regras de associação com uso do algoritmo *Apriori*, sendo evidenciado pelos resultados apresentados. Deste modo, a aplicação desenvolvida utilizou o *Apriori* juntamente com filtragem de informação para ofertar recomendações personalizadas de programas de TV mediante informações do comportamento no uso da TV. Além de oferecer um serviço, o sistema de recomendação pode ser uma possibilidade para solucionar o problema de sobrecarga de informação, o qual serviu de motivação para a pesquisa demonstrada neste artigo.

Para realização de trabalhos futuros, tem-se expectativa no lançamento do Ginga para acoplar os módulos do RePTVD diretamente neste *middleware*. Além disso, com o advento chamado TV social, no qual cada receptor poderá ser um potencial emissor de conteúdo, os telespectadores poderão compartilhar os dados referentes à visualização e preferências para outro grupo e assim, aplicar a filtragem colaborativa.

7. Agradecimentos

Este trabalho tem o apoio da FINEP e CAPES. FINEP financia o projeto N° 01.07.0110.00, “Avaliação do Middleware Ginga”.

Referências

- ABNT NBR 15603-2:2007 (2007) Televisão digital terrestre - Multiplexação e serviços de informação (SI) - Parte 2: Sintaxes e definições da informação básica de SI.
- Bozios, T., Lekakos, G., Skoularidou, V. e Chorianopoulos, K. (2001). Advanced techniques for personalized advertising in a digital TV environment: The imedia system. *In Proceedings of the eBusiness and eWork Conference*, pages 1025 -1031.
- Chorianopoulos, K. (2008). Personalized and mobile digital TV applications. *In Proceedings of the Multimedia Tools and Applications*, pages 1- 10, Vol.36.
- Goren-Bar, D. e Glinansky, O. (2004). FIT-recommending TV programs to family member. *In Proceedings of the Computers & Graphics*, pages 149-156, Vol. 28.
- Ginga (2006). Portal do GINGA. Available: <http://www.ginga.org.br>.
- Han, J. e Kamber, M. (2006), *Data Mining: Concepts and Techniques*, Morgan Kaufmann.
- Hsu, S. H., Wen, M. H., Lin, H. C., Lee, C. C. e Lee, C. H. (2007). AIMED – A personalized TV Recommendation System. *In Proceedings of the Interactive TV: A Shared Experience*, pages 166-174, Vol 4471, Springer Berlin / Heidelberg.
- Kim, J., Pan, Y. e Mcgrath, B. (2004). Personalization in Digital Television: Adaptation of Pre-Customized UI Design. *In Proceedings of the 2nd European Conference on Interactive Television: Enhancing the Experience*, pages 169-171.
- O’Sullivan, D., Smyth, B., Wilson, D. C., McDonald, K. e Smeaton, A. (2004). Interactive Television Personalization: From Guides to Programs. *Personalized Digital Television: Targeting Programs to Individual Viewers*. L. Ardissono, A. Kobsa and M. Maybury editors, pages 73-91, Kluwer Academic Publishers.
- O’Sullivan, D., Smyth, B., Wilson, D. C., McDonald, K. e Smeaton, A. (2003). Improving the Quality of the Personalised Electronic Program Guide. *In: L. Ardissono, M. Maybury and Alfred Kobsa (eds.): UMUAI Special Issue on User Modeling and Personalization for Television*, In Press.
- Sun Microsystems (2006). Java TV API 1.1 (JSR-927). Available: <http://java.sun.com/javame/reference/apis/jsr927>.
- Vildjiounaite, E., Kyllonen, V., Hannula, T. e Alahuhta, P. (2008). Unobtrusive Dynamic Modelling of TV Program Preferences. *In Proceedings of the Changing Television Environments, 6th European Conference, EuroITV 2008*, pages 82-91.
- Witten, I. H. e Frank, E. (2005), *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2th edition.
- Xletview (2008). Projeto Xletview. Available: <http://xletview.sourceforge.net/index.php>.
- Zhiwen, Y. e Xingshe, Z. (2004). TV3P: an adaptive assistant for personalized TV. *In Proceedings of the IEEE Transactions on Consumer Electronics*, pages 393 – 399.
- Zhiwen, Y., Xingshe, Z., Yanbin, H. e Jianhua, G. (2006). TV program recommendation for multiple viewers based on user profile merging. *In Journal of the User Modeling and User-Adapted Interaction*, pages 63-82. Publishing Springer Netherlands.