

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Personalização de Programas de TV no contexto da TV  
Digital Portátil Interativa

ALUNA: Elaine Cecília Gatto

ORIENTADOR: Prof. Dr. Sergio Donizetti Zorzo

São Carlos-SP

Outubro/2010

Elaine Cecília Gatto

Personalização de Programas de TV no contexto da TV  
Digital Portátil Interativa

*DISSERTAÇÃO DE MESTRADO  
APRESENTADA AO PROGRAMA DE PÓS-  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
DO DEPARTAMENTO DE COMPUTAÇÃO DA  
UNIVERSIDADE FEDERAL DE SÃO CARLOS  
COMO PARTE DOS REQUISITOS PARA  
OBTENÇÃO DO TÍTULO DE MESTRE EM  
CIÊNCIA DA COMPUTAÇÃO*

São Carlos-SP

Outubro/2010

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

G263pp

Gatto, Elaine Cecília.

Personalização de programas de TV no contexto da TV digital portátil interativa / Elaine Cecília Gatto. -- São Carlos : UFSCar, 2011.

87 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2010.

1. Sistemas de recomendação. 2. Algoritmos. 3. Middleware. 4. Sistemas multimídia. I. Título.

CDD: 003.7 (20<sup>a</sup>)

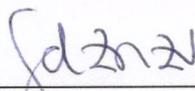
**Universidade Federal de São Carlos**  
**Centro de Ciências Exatas e de Tecnologia**  
**Programa de Pós-Graduação em Ciência da Computação**

**“Personalização de Programas de TV no Contexto  
da TV Digital Portátil Interativa”**

**ELAINE CECÍLIA GATTO**

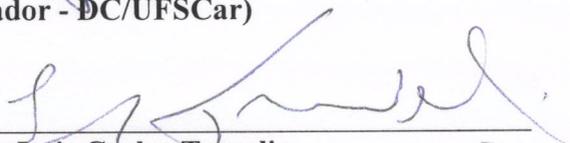
Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação

**Membros da Banca:**



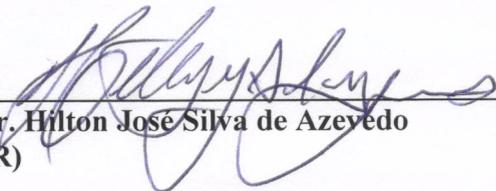
---

**Prof. Dr. Sérgio Donizetti Zorzo**  
(Orientador - DC/UFSCar)



---

**Prof. Dr. Luis Carlos Trevelin**  
(DC/UFSCar)



---

**Prof. Dr. Hilton José Silva de Azevedo**  
(UTFPR)

**São Carlos**  
**Novembro/2010**

*Dedico este trabalho primeiramente a Deus, pois sem Ele e a minha fé n'Ele nada seria possível. Dedico particularmente e de um modo carinhoso todo o resultado dos meus estudos à minha família, que foi a minha fonte de persistência nos momentos mais dolorosos destes quase três anos de estudos e dedicação.*

# AGRADECIMENTOS

Agradeço primeiramente aos meus pais Rosalina e José Valter, aos meus irmãos Edson e Ana e também à minha Avó Emiliana, por ter me auxiliado financeiramente, psicologicamente e espiritualmente em todo o longo e difícil caminho do mestrado. Obrigada por todas as orações, pela fé e confiança depositada em minha capacidade e habilidade.

Agradeço ao meu orientador Sergio Donizetti Zorzo, por ter me dado a oportunidade de conhecer um mundo novo, o mestrado acadêmico. Como consequência, devo agradecê-lo pelas experiências positivas e negativas e por todo o conhecimento adquirido e trabalhado durante todo o período de pesquisa.

Cursando o mestrado eu pude aprender mais sobre o mundo e sobre as pessoas, pude também aumentar minha rede de contatos conhecendo pessoas extraordinárias. Por mais difícil que tenha sido, eu realmente só tenho a agradecer pela oportunidade e confiança.

Agradeço também a uma pessoa em particular, Glauber Mattar Delgobo, que me auxiliou, apoiou e animou durante quase 2 anos. Durante todo o tempo que permanecemos juntos, ele sofreu com as minhas dores e sorriu com as minhas alegrias. Eu também lhe sou grata por tudo isso.

Aos meus “irmãos” de orientação Paulo Ávila, Paulo Cereda, Adriano, Filipe, Thiago, Rodrigo, Gabriel, Lizbeth e também aos meus companheiros do laboratório GSDR, Ricardo, Anderson, Jesus, Danilo, Bruno, Diego, Luciano, João Víctor, Samuel, Douglas, Fabiano e Maurício, que compartilharam e dividiram as minhas alegrias, tristezas, dificuldades, loucuras, entusiasmos, desânimos, desesperos e também o sucesso.

Como cristã que sou, agradeço de todo o coração, de toda a minha alma e de toda a minha mente, a Deus, ao meu Anjo da Guarda, ao Espírito Santo e à Nossa Senhora, que me protegeram e suscitaram as idéias e soluções para todas as situações adversas encontradas no caminho.

Não posso deixar de citar algumas pessoas em particular que, com sua amizade, me ajudaram a transformar os momentos deprimentes em momentos felizes. Meu muito obrigada à Mayra, Maísa, Johana, Juciara, Eliane, Samara, Bárbara, Luz, Juliana, Alexandre Chohfi, Bruno Cagnin e Lucas Ponce.

Não posso esquecer de agradecer a todos os mestrandos da turma 2007, 2008, 2009 e 2010 que também me apoiaram de alguma forma nesse longo caminho. Agradeço também a todas as moças que moraram comigo no pensionato junto com a Irmã Luíza em 2009 e 2010 e torceram para que tudo desse certo no meu mestrado.

Por fim, agradeço a todos que, de forma direta ou indireta, participaram para a conclusão deste trabalho.

*"Todo ser humano é dotado de capacidade infinita". (Masaharu Taniguchi)*

# RESUMO

A Televisão Digital Interativa permite que diversos serviços sejam oferecidos aos usuários, possibilitando entretenimento, educação à distância e novos mecanismos para a inclusão social. Emissoras e programas de TV podem ser criados, o que pode aumentar gradativamente a quantidade de informação disponível a ser visualizada nas telas. Como consequência, os usuários podem ter dificuldades em encontrar as informações que realmente interessam. No ambiente portátil o usuário deseja aproveitar ao máximo o seu tempo de visualização de TV, ou seja, investir tempo demais para procurar programas de TV do seu interesse é algo indesejado. Os sistemas de recomendação permitem minimizar tais dificuldades, auxiliando os usuários na sua busca por conteúdos que sejam do seu interesse e também reduzindo o tempo gasto durante a busca. Este trabalho tem como foco o desenvolvimento de um sistema de recomendação híbrido, denominado BIPODiTVR, que é capaz de recomendar conteúdo adequado a partir da observação do comportamento do usuário durante o uso da televisão no seu dispositivo portátil. O sistema recomenda programas de TV aos usuários utilizando as técnicas de Filtragem Colaborativa e Filtragem Baseada em Conteúdo e foi avaliado a partir de métricas de aceitabilidade de sistemas de recomendação baseado em dados fornecidos pelo IBOPE de seis domicílios em um período de quinze dias.

**Palavras-Chave:** Ginga, Personalização, TV Digital Portátil.

# ABSTRACT

Interactive Digital Television allows several services to be offered to users, in order to provide entertainment, e-learning and new mechanisms for social inclusion. Broadcasters and TV shows may be created, which may gradually increase the amount of information to be displayed on the screen. As a result, users may experience discomfort and difficulties in finding information that really matters. In a portable environment, the user wants to make the most of his time when watching TV. Thus, investing too much time in searching for TV shows of interest is something undesirable. Recommender systems are used to minimize such shortcomings, help users with their searches for contents of interest and also to reduce the time spent on searches. This study focused on the development of a hybrid recommender system called BIPODiTVR, which is able to make recommendations from the observation of users' behavior while watching television on the portable device. This system recommends TV shows through a collaborative and content-based filtering. The proposed system was evaluated by using metrics of acceptability of recommender systems applied to data provided by IBOPE, from six households for the period of fifteen days.

**Keywords:** Ginga, Personalization, Portable Digital TV.

# LISTA DE FIGURAS

Figura 1 - Componentes básicos de um Sistema de Recomendação .....	16
Figura 2 – Etapas básicas de um Sistema de Recomendação.....	16
Figura 3 – Filtragem baseada em conteúdo .....	18
Figura 4 – Funcionamento do BiPODiTVR .....	27
Figura 5 – Arquitetura Cliente .....	28
Figura 6 – Arquitetura Servidor.....	30
Figura 7 – Fluxo entre os módulos e as partes do BiPODiTVR .....	32
Figura 8 – Primeira vez solicitando recomendações.....	33
Figura 9 – Passos da identificação do usuário após a primeira solicitação de recomendação e cadastramento.....	34
Figura 10 – Identificação do usuário e armazenando dos dados.....	35
Figura 11 – Passos para o envio das recomendações ao cliente após a solicitação	37
Figura 12 – Passos da solicitação dos Top 10 Programas e Categorias.....	38
Figura 13 – Passos da solicitação do EPG.....	38
Figura 14 – Passos da solicitação da recomendação.....	39
Figura 15 – Tela inicial e tela de identificação .....	40
Figura 16 – Top 10 Programas e Categorias.....	41
Figura 17 – Tela “Sugestões”: tela retornada quando o usuário seleciona o botão “Temos sugestões para você” .....	41
Figura 18 – Passo a passo da simulação .....	46
Figura 19 – Gráfico FC_NP porcentagem de acertos por dia .....	50
Figura 20 – Gráfico FC_NP porcentagem de acertos por usuário .....	50
Figura 21 – Gráfico FC_C porcentagem de acertos por dia .....	52
Figura 22 – Gráfico FC_C porcentagem de acertos por usuário .....	52
Figura 23 – Gráfico FBC_NP porcentagem de acertos por dia.....	54
Figura 24 – Gráfico FBC_NP porcentagem de acertos por usuário.....	54
Figura 25 – Gráfico FBC_C porcentagem de acertos por dia .....	56

Figura 26 – Gráfico FBC_C porcentagem de acertos por usuário .....	56
Figura 27 – Gráfico FH_NP porcentagem de acertos por dia .....	58
Figura 28 – Gráfico FH_NP porcentagem de acertos por usuário .....	58
Figura 29 – Gráfico FH_C porcentagem de acertos por dia .....	60
Figura 30 – Gráfico FH_C porcentagem de acertos por usuário .....	60
Figura 31 – Gráfico: Média da porcentagem de acertos por dia .....	62
Figura 32 – Gráfico: Média da porcentagem de acertos por usuário .....	63
Figura 33 – Gráfico: comparação do nome do programa por dia .....	64
Figura 34 – Gráfico: comparação da categoria por dia .....	65
Figura 35 – Gráfico: comparação do nome do programa por usuário .....	65
Figura 36 – Gráfico: comparação da categoria por usuário .....	66

# LISTA DE TABELAS

Tabela 1 – Matriz de avaliações usuários x itens .....	20
Tabela 2 – Classificação de gênero do descritor .....	27
Tabela 3 – Exemplo do histórico de visualização .....	34
Tabela 4 – Fragmento do EPG original.....	36
Tabela 5 – Fragmento do EPG convertido em uma matriz .....	36
Tabela 6 – Características dos domicílios .....	44
Tabela 7 – Características socioeconômicas dos domicílios .....	44
Tabela 8 – Exemplo de comparação para a FC .....	47
Tabela 9 – Exemplo de comparação para a FBC .....	47
Tabela 10 – Exemplo de comparação para a FH .....	48
Tabela 11 – Porcentagem de acertos FC_NP .....	49
Tabela 12 – Porcentagem de acertos FC_C.....	51
Tabela 13 – Porcentagem de acertos FBC_NP .....	53
Tabela 14 – Porcentagem de acertos FBC_C .....	55
Tabela 15 – Porcentagem de acertos FH_NP .....	57
Tabela 16 – Porcentagem de acertos FH_C.....	59
Tabela 17 – Média da porcentagem de acertos por dia.....	61
Tabela 18 – Média da porcentagem de acertos por usuário.....	62
Tabela 19 – Média da porcentagem de acertos por nome do programa .....	63
Tabela 20 – Média da porcentagem de acertos por categoria .....	64

# LISTA DE EQUAÇÕES

Equação 1 – Normalização .....	21
Equação 2 – Cosseno.....	23
Equação 3 – Cálculo da porcentagem de acerto .....	45
Equação 4 – Cálculo da média de acertos .....	48

# LISTA DE ABREVIATURAS E SIGLAS

<b>2G</b>	Segunda geração de redes celulares
<b>3G</b>	Terceira geração de redes celulares
<b>3GPP</b>	3rd Generation Partnership Project
<b>ABNT</b>	Associação Brasileira de Normas Técnicas
<b>API</b>	Application Programming Interface
<b>ARFF</b>	Attribute-Relation File Format
<b>BiPODiTVR</b>	Brazilian Portable Digital Television Recommender
<b>BMCO</b>	Broadcast Mobile Convergence
<b>CRC</b>	Cyclic redundancy check
<b>CSV</b>	Comma-separated values
<b>DAB</b>	Digital Audio Broadcast
<b>DCBD</b>	Descoberta de Conhecimento em Base de Dados
<b>DMB</b>	Digital Multimedia Broadcast
<b>DVB</b>	Digital Video Broadcast
<b>DVB-H</b>	Digital Video Broadcast – Handeld
<b>DVB-T</b>	Digital Video Broadcast – Terrestrial
<b>DVR</b>	Digital Video Recorder
<b>EIT</b>	Tabela de informação de eventos
<b>EPG</b>	Electronic Program Guide
<b>ETSI</b>	European Telecommunications Standards Institute
<b>EUA</b>	Estados Unidos da América
<b>FBC</b>	Filtragem Baseada em Conteúdo
<b>FC</b>	Filtragem Colaborativa
<b>FH</b>	Filtragem Híbrida
<b>FIFO</b>	First In First Out

<b>FLUTE</b>	File Delivery Over Unidirectional Transport
<b>GEM</b>	Globally Executable MHP
<b>GMIT</b>	Galway-Mayo Institute of Technology
<b>IDE</b>	Integrated Development Environment
<b>ISDB</b>	Integrated Services Digital Broadcasting
<b>ISDB-T</b>	Integrated Services Digital Broadcasting – Terrestrial
<b>IBOPE</b>	Instituto Brasileiro de Opinião Pública e Estatística
<b>IMDB</b>	The Internet Movie Database
<b>IPTV</b>	Internet Protocol Television
<b>JPA</b>	Java Persistence API
<b>KDD</b>	Knowledge Discovery in Databases
<b>KNN</b>	K Nearest Neighbors
<b>LASER</b>	Light Weight Application Scene Representation
<b>MBMS</b>	Mobile Broadcast/Multicast Service
<b>MEDIAFLO</b>	Media Forward Link Only
<b>MP3</b>	Moving Pictures Expert Group part 1/2 Audio Layer 3
<b>NCL</b>	Nested Context Language
<b>NCM</b>	Nested Context Model
<b>OFDM</b>	Orthogonal Frequency Division Multiplexing
<b>OMA-BCAST</b>	Open Mobile Alliance – Mobile Broadcast Services
<b>ONE-SEG</b>	One Segment
<b>PDA</b>	Personal Digital Assistants
<b>PHP</b>	Hypertext Preprocessor
<b>SAF</b>	Simple Aggregation Format
<b>SDT</b>	Tabela de descrição de serviço
<b>SED</b>	EPG DataSet
<b>SDMB</b>	Satellite Digital Multimedia Broadcast
<b>STB</b>	Set-top-box
<b>SVG</b>	Scalable Vectorial Graphics
<b>TF-IDF</b>	Term Frequency Inverse Document Frequency

<b>TVD</b>	Televisão Digital
<b>TVDI</b>	Televisão Digital Interativa
<b>TVDP</b>	Televisão Digital Portátil
<b>TVDPI</b>	Televisão Digital Portátil Interativa
<b>TXT</b>	Text File
<b>UDP</b>	User Datagram Protocol
<b>XML</b>	eXtensible Markup Language
<b>WiFi</b>	Wireless Fidelity

# SUMÁRIO

<b>CAPÍTULO 1 - INTRODUÇÃO .....</b>	<b>1</b>
1.1 Motivações e Justificativas .....	3
1.2 Objetivos .....	4
1.3 Organização do Trabalho .....	5
<b>CAPÍTULO 2 - TVD PORTÁTIL INTERATIVA .....</b>	<b>6</b>
2.1 Breve Histórico da TVDPI .....	6
2.2 Sistemas de TVD para Dispositivos Portáteis.....	7
DVB/DVB-H .....	8
DMB.....	9
ISDB/One-Seg.....	9
MediaFLO .....	10
Outros Sistemas .....	11
2.3 Middleware Ginga.....	11
2.4 Observações.....	13
<b>CAPÍTULO 3 - SISTEMAS DE RECOMENDAÇÃO .....</b>	<b>15</b>
3.1 Filtragem Baseada em Conteúdo .....	17
3.2 Filtragem Colaborativa.....	19
3.3 Filtragem Híbrida .....	20
3.4 Técnica dos vizinhos mais próximos.....	20
3.5 Técnica do Cosseno .....	22
3.6 Mineração de Dados.....	23
3.7 Observações.....	24
<b>CAPÍTULO 4 - SISTEMA DE RECOMENDAÇÃO PARA TVDPI.....</b>	<b>26</b>
4.1 Arquitetura do BiPODiTVR .....	26
4.2 Implementação do BiPODiTVR.....	32
4.3 Aplicação Protótipo .....	40
4.4 Observações.....	41

<b>CAPÍTULO 5 - EXPERIMENTOS E RESULTADOS .....</b>	<b>43</b>
5.1 Dados do IBOPE.....	43
5.2 Testes e Resultados .....	45
5.3 Observações.....	66
<b>CAPÍTULO 6 - TRABALHOS RELACIONADOS.....</b>	<b>69</b>
6.1 Recommender TV: Suporte ao Desenvolvimento de Aplicações de Recomendação para o Sistema Brasileiro de TV Digital .....	69
6.2 Personalização para Televisão Digital utilizando a estratégia de Sistema de Recomendação para ambientes multiusuário.....	71
6.3 Content Personalization for Mobile TV Combining Content-Based and Collaborative Filtering .....	73
6.4 Mobile TV Targeted Advertisement and Content Personalization.....	74
6.5 ZapTV: Personalized User-Generated Content for Handheld Devices in DVB-H Networks.....	76
6.6 A Lightweight Mobile TV Recommender: Towards a One-Click-To-Watch Experience.....	77
6.7 Observações Finais .....	78
<b>CAPÍTULO 7 - CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>80</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>82</b>

A Televisão Digital (TVD) pode ser considerada uma evolução natural da televisão analógica. Sistemas de TVD podem ser definidos como um conjunto de padrões e especificações que definem seu comportamento e funcionamento e podem oferecer, além de áudio e vídeo digital de alta qualidade, interatividade.

Com interatividade o telespectador deixa de ser um usuário passivo e passa a ser um usuário que interage com o conteúdo, usufruindo da TV de forma muito parecida a de um computador. Devido a isso, o usuário pode ser chamado de telenauta (telespectador + internauta), ao invés de telespectador.

Com a digitalização dos sinais de televisão, o espectro de radiofrequências é otimizado, isto é libera mais de dois terços do espectro e amplia o número de canais de comunicação e, conseqüentemente, o conteúdo da TV. (LOSS, 2008)

O termo TVD é utilizado para definir o sistema de TV Digital que não tem suporte algum à interatividade, mas apenas a recepção do sinal digital. TVDI (Televisão Digital Interativa) é definida como o sistema de TV Digital que fornece interatividade. TVDP (Televisão Digital Portátil) é definida como o sistema de TV Digital que fornece apenas recepção de sinal digital em dispositivos portáteis, como celulares, PDAs (*Personal Digital Assistants*), etc., sem interatividade. Por fim, TVDPI (Televisão Digital Portátil Interativa) é definida como o sistema de TV Digital que fornece interatividade a dispositivos portáteis.

Três ajustes ocorrem quando da mudança de um sistema de TV Analógico para um sistema de TV Digital Interativo. A primeira delas ocorre nas redes de TV em que os equipamentos de captura, edição e transmissão interna de áudio e vídeo analógicos, devem ser substituídos por digitais. A segunda ocorre na sociedade pela adoção de um padrão uniforme de codificação, transmissão e modulação, difusão e recepção digital. A terceira define novos modelos de negócios que permitem à

população investir em novos equipamentos e também às redes obter retorno sobre os investimentos.

Em um sistema de TVDI/TVDPI passam a ser fundamentais os provedores de conteúdo, que devem desenvolver os programas televisivos interativos e distribuí-los digitalmente, e os telespectadores, que receberão esse conteúdo, o que resulta em novos nichos mercadológicos e de pesquisa para exploração. Com a implantação da TVDI/TVDPI no Brasil, novos serviços são oferecidos, novos modelos de negócios são desenvolvidos e surgem novas oportunidades para empresas, mercado de trabalho e pesquisadores a cada dia.

A TVDPI surge, nesse contexto, como uma classe de serviços multimídia que oferece uma experiência de uso de TV pessoal, dinâmica e personalizada, diferente daquela utilizada em residências, que normalmente é coletiva e social. Os serviços de personalização podem, então, ser explorados e fornecidos.

Muito utilizados na Internet, os sistemas de recomendação estão sendo empregados no contexto de TVDI/TVDPI, visando à recomendação de programas de TV, publicidade/propaganda e comércio eletrônico. Com base em informações sobre as preferências dos usuários, os sistemas de recomendação fornecem sugestões utilizando técnicas de recomendação. Os sites *Youtube* e *LastFM* são sistemas Web multimídia que utilizam sistemas de recomendação para classificar, avaliar e recomendar vídeos e músicas, conforme as preferências do usuário.

Sistemas de recomendação também são muito utilizados em sistemas de *e-commerce*. Em sistemas de *e-commerce*, o usuário se cadastra no site, define suas preferências e avalia produtos. São recomendados para o usuário produtos de sua preferência. Com frequência esse tipo de sistema de recomendação usa os dados do histórico de visualização de produtos do usuário para fazer a recomendação. Como exemplo, têm-se as lojas virtuais *Submarino* e *Amazon*. (TORRES, 2004) Esse mesmo conceito, denominado de *t-commerce*, pode ser aplicado na TVDI/TVDPI, embora não existam sistemas em operação no Brasil devido ao fato de ainda estar em fase de implantação.

Redes sociais também podem fazer uso de sistemas de recomendação, com o intuito de sugerir amigos e recomendar pessoas a outras pessoas. Os sites de

relacionamento *Facebook*, *Orkut* e *Twitter* já sugerem possíveis amigos aos seus usuários.

O celular, como um dispositivo pessoal e um dispositivo portátil com suporte à TVDPI, permite que a experiência de assistir TV em qualquer lugar e a qualquer hora tenha um nível de personalização diferente do de residências. Com um público diferenciado e características particulares do ambiente portátil, tratar os usuários desses dispositivos de forma pessoal representa um fator que deve ser levado em conta.

Portanto, um sistema de recomendação para TVDI/TVDPI pode contribuir para o aumento da satisfação do usuário em sua experiência com programas de TV direcionados aos seus interesses.

## **1.1 Motivações e Justificativas**

Na TV analógica, os telespectadores ficam cientes dos programas a serem exibidos através de propagandas da própria emissora durante a programação diária ou pela consulta à grade de programação disponível nos sites das emissoras. Entretanto, a grade de programação disponível nos sites é normalmente a grade de programação nacional e, muitas vezes, não corresponde à grade das afiliadas. Por exemplo, em São Paulo capital, é exibida uma grade de programação, enquanto uma filial na região noroeste paulista exibe outra grade com a inclusão de programas regionais.

Sistemas de TVDI/TVDPI possuem um Guia Eletrônico de Programação (EPG – Electronic Program Guide) para auxiliar os usuários na busca dos programas de TV que estão sendo exibidos, com acesso pelo controle remoto. A oferta do EPG na própria televisão possibilita que o usuário consulte localmente a grade de programação de todas as emissoras sem a necessidade de acesso à internet, como é usual nos sistemas analógicos.

A TVDI permite a criação de novas emissoras e programas de TV, o que pode aumentar gradativamente a quantidade de informação a ser visualizada no EPG e

gerar dificuldades para os usuários encontrarem as informações que realmente lhes interessam. A TVDPI também passará pelas mesmas situações e sofrerá esse impacto de forma mais acentuada, devido às características peculiares dos dispositivos portáteis como, por exemplo, a tela reduzida para exibir grande quantidade de informação.

Investir tempo demasiado a fim de procurar programas de TV de interesse do usuário também é algo indesejado, principalmente na TVDPI, ambiente em que o usuário é objetivo e deseja aproveitar ao máximo o seu curto ou médio tempo de visualização. Quanto menos tempo utilizar na busca pela informação desejada, mais tempo o usuário passará visualizando o programa escolhido. A atenção às necessidades especiais dos usuários de TVDPI, o mercado de trabalho emergente, a área de pesquisa em crescimento e a proposta de trabalho multidisciplinar to são motivações deste trabalho.

As justificativas do trabalho englobam o aumento na carga de informações disponíveis sobre programas televisivos, as dificuldades do usuário em procurar o programa desejado, a demasiada carga de informação apresentada em uma tela reduzida, as dificuldades na escolha de programas e o longo tempo gasto para realizar a escolha. Essas dificuldades podem ser minimizadas através do emprego de sistemas de recomendação.

## **1.2 Objetivos**

Na maior parte das vezes, a TVDPI é usada em tempos de espera ou deslocamentos curtos/médios e a recomendação de programas de televisão compatíveis com a preferência do usuário tende a ser uma solução rápida e eficaz no aproveitamento do tempo de utilização da TVDPI.

O objetivo principal deste trabalho é apresentar e validar um modelo de recomendação de conteúdos de TV de acordo com as preferências dos usuários, durante o uso da televisão no seu dispositivo, contribuindo com a área de personalização no domínio da TVDPI. Informações como a categoria de programas

mais assistidos e o horário de utilização da televisão são armazenadas para a descoberta implícita das preferências de programas do usuário.

Outro objetivo deste trabalho foi a avaliação de técnicas que possam ser utilizadas em sistemas de recomendação para a personalização de conteúdo e verificar como as mesmas podem ser aplicadas no contexto da TVDPI.

### **1.3 Organização do Trabalho**

Esta dissertação está dividida em sete capítulos, organizados como descrito a seguir.

No Capítulo 1, consta a introdução da dissertação, abordando as motivações, justificativas e objetivos do trabalho.

O Capítulo 2 é dedicado à TVDPI. Descreve-se um pouco da história e da evolução dessa tecnologia e os sistemas mais difundidos mundialmente.

O capítulo 3 trata da personalização e dos sistemas de recomendação. Os principais métodos de personalização são comentados, assim como os problemas que surgem quando de sua utilização.

O capítulo 4 é dedicado ao sistema desenvolvido como prova de conceito desta dissertação. Sua arquitetura, funcionamento e implementação são descritos detalhadamente.

O capítulo 5 apresenta os resultados obtidos a partir dos testes realizados.

No capítulo 6, seis trabalhos relacionados são brevemente descritos e, então, comparados.

O capítulo 7 trata dos trabalhos futuros que podem ser desenvolvidos a partir desta dissertação e também apresenta as conclusões obtidas ao final deste trabalho. Por fim, encontram-se as referências bibliográficas.

# 2

## TVD PORTÁTIL INTERATIVA

---

Este capítulo tem o objetivo de apresentar a revisão bibliográfica sobre a TVDPI e está organizado da seguinte forma. A seção 2.1 contém o histórico da TVDPI. A seção 2.2 trata dos padrões existentes de TVDPI. As seções 2.3, 2.4, 2.5, 2.6 e 2.7 tratam dos sistemas de TVDPI mais utilizados atualmente. E, finalmente, a seção 2.8 encerra o capítulo com as observações finais sobre os estudos realizados.

### 2.1 Breve Histórico da TVDPI

Tentativas de tornar a TVDPI uma realidade iniciaram-se há alguns anos. A Sony, por exemplo, na década de 80, lançou no mercado TVs portáteis com recepção de sinal analógico. Entretanto, não era possível compensar o efeito Doppler Spread<sup>1</sup> e, por essa razão, as TVs não tiveram aceitação pelos usuários. Nos sistemas de TVDPI esse efeito pode ser minimizado pela aplicação de algoritmos de processamento de sinal e métodos de diversidade.

Outra tentativa partiu de uma emissora alemã na década de 90, a qual realizou testes de um serviço DVB-T (*Digital Video Broadcasting-Terrestrial*) em um bonde em movimento no centro de *Cologne*. Os japoneses também realizaram pesquisas para a TVDPI e em 1998 aprovaram um sistema que suporta a transmissão para dispositivos portáteis.

Os europeus, em 2004, pelo ETSI (*European Telecommunications Standards Institute*), aprovaram o sistema para a TVDPI, denominado DVB-H (*Digital Video Broadcasting-Handheld*). A Coreia também disponibilizou telefones celulares com capacidade para receber sinais de TV em 2002. Os sinais eram transmitidos por uma rede de telefonia celular padrão, o que significava tarifas para assistir TV

---

<sup>1</sup> Desvio da frequência da portadora devido ao deslocamento relativo entre a fonte transmissora e a fonte receptora.

cobradas por minuto e contas telefônicas com alto valor. Os serviços são oferecidos agora na Coreia através do padrão SDMB (*Satellite Digital Multimedia Broadcast*) que fornece um *link* direto para o celular.

Nos EUA (Estados Unidos da América) empresas de TVDPI pagas, como *MobiTV*, *SprintTV* ou *SmartVideo*, fornecem assinaturas de planos de serviços. No Brasil, já existe a transmissão gratuita de TVDPI em grande parte do país.

A próxima seção abordará sistemas de TVDPI. Esses sistemas são importantes para que dispositivos portáteis, como celulares, possam receber o sinal digital de TV e também utilizar a interatividade. (YEUN, 2007)

## **2.2 Sistemas de TVD para Dispositivos Portáteis**

Um dispositivo de TVDPI, como o celular, tem a capacidade de receber sinais de rádio nas bandas de frequência alocadas para os sinais de TV e também nas bandas alocadas para os dados e voz do telefone celular.

A TVDPI combina dois produtos de consumo considerados os mais vendidos na história: televisão e telefone celular. Os aparelhos celulares atuais oferecem diversas funcionalidades, tais como: e-mails, mensagens multimídia, MP3 (MPEG-1/2 Audio Layer 3), rádio, *internet*, *download* de imagens, vídeos e agenda. Também oferecem diferentes modos de conectividade como, por exemplo, Infravermelho, *Bluetooth* e *WiFi*. Como uma evolução natural, já é possível assistir TV em aparelhos celulares e, em alguns países, usufruir de alguns serviços interativos. (SANTOS ET AL, 2005; SEBASTIAN ET AL, 2007)

Vários modos de transmissão podem ser usados entre eles, satélite, *broadcast* e *internet*. Em grande parte do mundo é possível assistir TV por um dispositivo portátil com transmissão digital, seja de forma paga ou gratuita.

Os cinco sistemas mundiais de televisão digital terrestres mais utilizados são o Americano, europeu, japonês, chinês e nipo-brasileiro. Cada um desses sistemas

possui características que possibilitam, por exemplo, a interatividade, capacidade de recepção em dispositivos portáteis, inclusão social, entre outros.

O sistema americano, por exemplo, privilegia a televisão de alta definição, enquanto o europeu a multiprogramação, a interatividade e novos serviços. O japonês, por sua vez, favorece a televisão de alta definição e a recepção portátil e o chinês oferece programas de alto nível, de acordo com os interesses dos usuários, tais como, entretenimento, investimentos de seguros, entre outros. Por fim, o nipo-brasileiro, que possui características do japonês acrescentado de tecnologia brasileira na camada de *middleware*, tem como foco a inclusão digital. (ALENCAR, 2007)

Além dos sistemas terrestres citados, existem também outros padrões desenvolvidos especificamente para TVDPI, os quais são descritos nas próximas seções.

## **DVB/DVB-H**

O sistema europeu, denominado DVB (*Digital Video Broadcasting*), teve início em 1993 auxiliado por 35 países e por um consórcio formado por 300 empresas. A DVB-H é a norma europeia para TVDPI que tem como base a especificação do DVB-T acrescentada de um conjunto de características que levam em conta as especificações dos dispositivos portáteis e utiliza satélites e torres para a transmissão do sinal de TV.

Uma vantagem da Europa no uso do DVB é a reutilização da estrutura existente, permitindo que sinais DVB-T e DVB-H sejam multiplexados e transmitidos ao mesmo tempo. Uma característica do DVB-H é o *Time Slicing*; uma técnica que permite reduzir o consumo de energia nos terminais em até 90% - o *stream* de áudio e vídeo é agrupado e transmitido em rajadas dentro de um intervalo de tempo. (DVB, 2010)

## **DMB**

O DMB (*Digital Multimedia Broadcasting*) foi inicialmente desenvolvido por alguns países como Coréia do Sul, Alemanha, Inglaterra e Espanha e tem como base o DAB (*Digital Audio Broadcast*), existindo ainda outra extensão DMB baseada no ISDB-T (*Integrated Services Digital Broadcasting Terrestrial*). O DMB oferece uma ampla gama de serviços inovadores, tais como TVDPI, tráfego e segurança de informação, programas interativos, informações de dados e muitas outras aplicações. Países da Europa e da Ásia têm implementado serviços comerciais com o uso deste sistema.

Como o DMB é baseado em DAB, dispositivos DMB são sempre compatíveis e recebem não apenas os serviços de multimídia DMB, mas também serviços de áudio DAB. Do lado da empresa de radiodifusão, isso significa que o sistema de transmissão convencional DAB pode ser usado para a transmissão DMB simplesmente através de um codificador de vídeo DMB para o sistema DAB existente. (DMB, 2010; DAB, 2010)

## **ISDB/One-Seg**

O ISDB já foi especificado de modo a suportar serviços de TV para dispositivos portáteis que têm a designação One-Seg. O nome é uma referência ao fato de que o sistema japonês divide o sinal em 13 segmentos: os 12 primeiros de alta definição e o 13º de baixa, para dispositivos portáteis.

No ISDB-T os sinais de recepção de serviços fixos e portáteis podem ser combinados em transmissão, através do uso de camadas hierárquicas. Essa ocorre em fluxo contínuo, minimizando os atrasos na aquisição de sinal quando o usuário muda de um canal para outro. Isso também significa que o receptor deve ser alimentado continuamente enquanto os programas estão sendo vistos.

O serviço *One-Seg* é baseado na tecnologia de transmissão OFDM (*Orthogonal Frequency Division Multiplexing*) segmentada. Além disso, adota uma tecnologia exclusiva, chamada de acolhimento parcial, utilizada para reduzir o consumo de energia do receptor, cujo fator mais importante é a diminuição da velocidade de processamento de sinal no receptor.

*One-Seg* é um serviço de TV Digital para telefones celulares, PDA's, miniTVs, etc. O serviço usa um segmento de 6MHz e não precisa de outro canal, o que leva à economia da frequência de recursos e custos de infraestrutura de radiodifusão. A interatividade é baseada em um conceito de tela separada, denominada *Split Screen*; o vídeo é visualizado em uma parte da tela e a interatividade ocorre em uma segunda tela do dispositivo. (DIBEG, 2010)

## **MediaFLO**

O MediaFlo (*Media Forward Link Only*) é uma plataforma desenvolvida especificamente para trazer qualidade à transmissão de vídeo portátil, além de um padrão global, o qual é dedicado ao sucesso da cadeia de valor da televisão móvel e foi desenvolvido pela empresa *Qualcomm* dos EUA. Além disso, utiliza algoritmos de código turbo que permitem a correção de erros mais agressivos, o CRC (*Cyclic Redundancy Check*), que verifica se o pacote é recebido perfeitamente. Seu receptor não tem a necessidade de calcular o código *Reed Solomon*<sup>2</sup> exterior, o que permite ao receptor a economia de energia e intercalação de dados no tempo que reduzem o impacto do ruído de impulso, mas que pode também aumentar o tempo de aquisição do canal.

A interatividade ocorre recorrendo-se a uma rede 3G (redes celulares de terceira geração) e a transmissão em *broadcast* ponto a ponto. MediaFlo também utiliza a técnica de *Time-Slicing* do DVB-H. (MEDI AFLO, 2010)

---

<sup>2</sup> Os códigos Reed-Solomon foram criados em 1960 por Irving S. Reed e Gustave Solomon. São códigos cíclicos que podem ser construídos de forma a corrigir erros múltiplos e são usados em telecomunicações, tanto em meios de armazenamento, como em CDs e em transmissões distantes. Esse tipo de código não é recomendado para a transmissão de grande quantidade de dados, entretanto, apresenta a vantagem de possibilitar a correção de um elevado número de erros dentro da mesma palavra. (BORGES, 2008)

## Outros Sistemas

Existem ainda outros sistemas menos conhecidos como: GMIT/HisTV (*Galway-Mayo Institute of Technology*), LAsER (*Lightweight Application Scene Representation*), 3GPP (*3rd Generation Partnership Project*) e OMA (*Open Mobile Alliance*), que dão suporte à transmissão de TV para dispositivos portáteis.

No sistema GMIT/HisTV a transmissão ocorre pelo FLUTE<sup>3</sup> (*File Delivery Over Unidirectional Transport*), um canal de *broadcast* definido pelo DVB-H. O HisTV é um *framework* de aplicação usado como *middleware* para TVDPI. (GMIT, 2010)

LAsER é um grupo que tem como objetivo o desenvolvimento de um padrão de descrição de cenas e um formato de agregação para TVDPI, a fim de estender o uso de conteúdos multimídia nos dispositivos portáteis, nomeados respectivamente de LAsER e SAF (*Simple Aggregation Format*). (LASER, 2010)

3GPP desenvolveu o padrão MBMS (*Mobile Broadcast/Multicast Service*), específico de transmissão para TVDPI. Seu *middleware* permite interações locais e remotas usando como linguagem uma extensão própria do SVG (*Scalable Vectorial Graphics*) Tiny. (3GPP, 2010)

OMA oferece soluções e padrões para sistemas de TVDPI. BCAST (*Mobile Broadcast Services*) é um subprojeto do OMA, que teve início em 2003 e já possui alguns dispositivos da Nokia disponíveis com suporte ao sistema, entretanto, ainda não foi especificado um *middleware* próprio para o OMA-BCAST. (OMA, 2010)

### 2.3 Middleware Ginga

Ginga é o nome do *middleware* aberto do sistema brasileiro de TV Digital e é constituído por um conjunto de tecnologias padronizadas e inovações brasileiras que

---

<sup>3</sup> FLUTE é um protocolo para a entrega unidirecional de arquivos pela Internet, o que é particularmente adequado para redes *multicast*. (FLUTE, 2010)

o tornam a especificação de *middleware* mais avançada e a melhor solução para os requisitos do Brasil.

O *middleware* Ginga é subdividido em Ginga-J e Ginga-NCL. O Ginga-J é a parte procedural do *middleware* e é utilizado para o desenvolvimento de aplicações procedurais Java. Já o Ginga-NCL é a parte declarativa do *middleware* e é utilizado para o desenvolvimento de aplicações declarativas em NCL (*Nested Context Language*). Dependendo das funcionalidades requeridas no projeto de cada aplicação, um paradigma é mais adequado que o outro.

O *middleware* Ginga-NCL é obrigatório para dispositivos portáteis, enquanto o Ginga-J é opcional. O Ginga-NCL utiliza as linguagens NCL, declarativa e Lua, imperativa. A linguagem NCL é uma linguagem declarativa para autoria de documentos hipermídia baseados no modelo conceitual NCM (*Nested Context Model*) e foi desenvolvida utilizando uma estrutura modular. Seus módulos para a especificação de conectores e de *templates* de composição podem ser incorporados a outras linguagens existentes. Além disso, a linguagem separa claramente os conteúdos de mídia e a estrutura de uma aplicação, além de definir como os objetos de mídia são estruturados e relacionados no tempo e no espaço.

Lua é uma linguagem rápida e leve, projetada para estender aplicações para programação procedural e foi escolhida como uma linguagem de *script* para compor o *middleware* Ginga-NCL. Possui poderosas construções para a descrição de dados, baseadas em tabelas associativas e semânticas extensíveis. Lua é tipada dinamicamente e interpretada a partir de bytewcodes para uma máquina virtual baseada em registradores. Tem gerenciamento automático de memória com coleta de lixo incremental e é uma linguagem ideal para configuração, automação e prototipagem rápida. (GINGA-NCL, 2010; LUA, 2010; NCL, 2010; SOARES, 2005; SOARES, 2010)

O objetivo do GINGA-J é oferecer uma implementação completa, não emulada e com código aberto, de um *middleware* de TV Digital compatível com o padrão brasileiro e executável em ambiente de PC, de forma a oferecer as APIs (Interfaces de Programação de Aplicativos) Java definidas na especificação Ginga-J.

Ginga-J é composto por APIs projetadas para suprir todas as funcionalidades necessárias para a implementação de aplicativos para TVDI, desde a manipulação de dados multimídia até protocolos de acesso. As APIs Ginga-J são divididas em três partes. As vermelhas são inovações que dão suporte às aplicações brasileiras, em especial as de inclusão social. As amarelas, também inovações brasileiras, podem ser exportadas para outros sistemas. E, por fim as verdes, que seguem o núcleo comum do padrão GEM (*Globally Executable Multimedia Home Platform*).

Os ambientes de apresentação e execução do *middleware* Ginga se complementam, unidos por uma ponte em uma implementação, o que confere ao sistema uma ótima eficiência, tanto em termos de uso de processador quanto de ocupação de memória. O Ginga foi desenvolvido com foco nos dois ambientes de programação desde o seu projeto inicial. (GINGA-J, 2010; SOFTWARE PÚBLICO, 2010)

## **NCLua SOAP**

Para fazer a comunicação com o servidor remoto, foi utilizada a biblioteca NCLua SOAP, desenvolvida pelo Manoel Campos. A biblioteca feita em Lua permite que uma aplicação NCLua se comunique com um Webservice, Java ou não, permitindo enviar e receber informações, assim como ocorre com aplicações Web e Celulares. A biblioteca deve ser incluída no projeto NCL para que possa ser utilizada corretamente e permitir que a aplicação NCLua comunique-se com o servidor. Atualmente a biblioteca está na versão 0.5.6.6 beta e pode ser utilizada gratuitamente pelos desenvolvedores de TVD. (CAMPOS, 2011)

## **2.4 Observações**

Neste capítulo foram conceituadas a TVDPI e os padrões mais conhecidos e utilizados. Mobilidade e flexibilidade são consideradas as principais vantagens da TVDPI, pois permitem obter serviços de TV durante uma viagem, no transporte público ou em momentos de espera.

Segundo os estudos realizados na BMCO (*Broadcast Mobile Convergence*), os usuários de TVDPI gastam cerca de 15 minutos em uma visualização de TV e gostam de conteúdos como notícias, esportes e música. A TVDPI pode ter um impacto significativo no modelo de negócios vigente na indústria de TV e poderia mudar o entendimento de horário nobre, devido à sintonia de mais espectadores assistindo seus programas favoritos no trem, no metro, no ônibus, etc., em horário diferente do considerado nobre. (BMCO, 2010)

A TVDPI provavelmente evoluirá além de uma simples TV com tela pequena, devido justamente às suas características peculiares, como portabilidade e conectividade. Conteúdos novos, adaptados para celulares e a interatividade são algumas das características para sua evolução. Enquanto utiliza a TVDPI, o usuário tem uma atenção limitada aos programas que assiste, mas com disponibilidade crescente de interação com os mesmos.

A TVDPI não oferece o mesmo conforto que a TV em residências, a qual se pode assistir tranquilamente, por longos períodos, entretanto, permitirá experiências mais interativas e pessoais. (SEBASTIAN ET AL, 2007; YEUN, 2007)

Devido ao fato da TVDPI proporcionar uma experiência mais particular para cada usuário, a personalização pode, então, ser utilizada para melhorar e aperfeiçoar essa experiência, proporcionando serviços diversos como *ranking* de programas, recomendações de programas de TV, favoritos, etc.

No próximo capítulo apresenta-se a revisão bibliográfica dos sistemas de recomendação.

# 3

## SISTEMAS DE RECOMENDAÇÃO

---

Hoje em dia, existe uma busca constante e uma necessidade de atualização das informações pertinentes às necessidades e interesses de cada indivíduo. De modo geral, essas informações são as mais diversificadas possíveis, desde sugestões de itens e objetos de consumo, até mesmo coordenadas de locais ou eventos. Uma parcela das pessoas realiza essa busca através de consultas baseadas em confiança, tais como, conselhos ou sugestões oriundos de outras pessoas que estejam inseridas no contexto das buscas.

A experiência que uma pessoa tem sobre um assunto também influencia no momento de efetuar uma escolha dentre várias alternativas disponíveis. Em algumas situações pessoas recorrem ao conhecimento de outras, com mais experiência que elas, sobre um determinado assunto. Os sistemas de recomendação seguem basicamente esse conceito intuitivo de interação, trabalha com as opiniões e experiências das pessoas sobre determinados assuntos, lidando com a sobrecarga de informações, aumenta a capacidade e eficácia e também automatiza o processo de indicação. (ADOMAVICIUS, 2005; RESNICK, 1997)

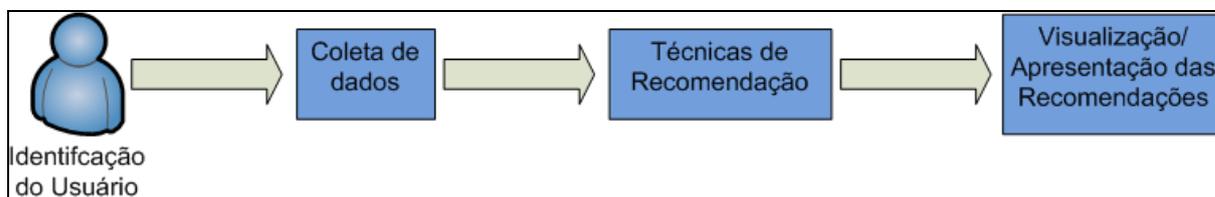
Um sistema de recomendação pode usar avaliações para indicar o nível de preferência de um usuário. Avaliações indicam as notas que os usuários podem dar a um item. Em um sistema de recomendação um item é o objeto que está sendo avaliado. Por exemplo, dar uma nota de zero a cinco para um livro que tenha comprado em uma loja virtual pode indicar o nível de sua preferência.

A Figura 1 apresenta os componentes básicos de um sistema de recomendação. O componente “Usuários” corresponde aos usuários que vão interagir e usar o sistema de recomendação. O componente “Itens” corresponde às informações sobre os itens do sistema e o componente “Sistemas de Recomendação” corresponde à maneira com que os dados dos itens e usuários serão trabalhados para gerar as recomendações.



**Figura 1 - Componentes básicos de um Sistema de Recomendação**

Sistemas de recomendação concentram-se em encontrar informações que podem ser úteis ao usuário. A recomendação é obtida a partir da identificação das preferências, perfil do usuário e da interação com o próprio sistema. A Figura 2 apresenta as etapas básicas de um sistema de recomendação.



**Figura 2 – Etapas básicas de um Sistema de Recomendação**

A primeira etapa refere-se à forma de identificação do usuário e à coleta dos dados. Nessa etapa realiza-se a coleta de informações do comportamento do usuário e o armazenamento das mesmas para a composição do seu perfil, ou seja, suas preferências como, por exemplo, em um site de *e-commerce*, um usuário pode gostar muito de livros de ficção científica. As preferências podem ser identificadas ou informadas e são armazenadas como parte do perfil e utilizadas para indicar àquele usuário outros produtos do mesmo gênero.

A coleta dos dados pode ser feita de forma explícita, implícita ou de ambas as formas. A forma explícita ocorre quando o usuário fornece as suas preferências usualmente pelo preenchimento de um formulário e a implícita quando o sistema obtém as informações do usuário pelo registro de sua navegação.

A partir dos dados coletados, técnicas são aplicadas para descobrir as similaridades entre as preferências dos usuários e os itens a serem recomendados e, em seguida, gerar as recomendações. No caso de uma loja virtual, os itens são os produtos e, portanto, o sistema de recomendação deverá encontrar os que poderão ser indicados para aquele usuário a partir do conhecimento que o sistema de recomendação detém do mesmo.

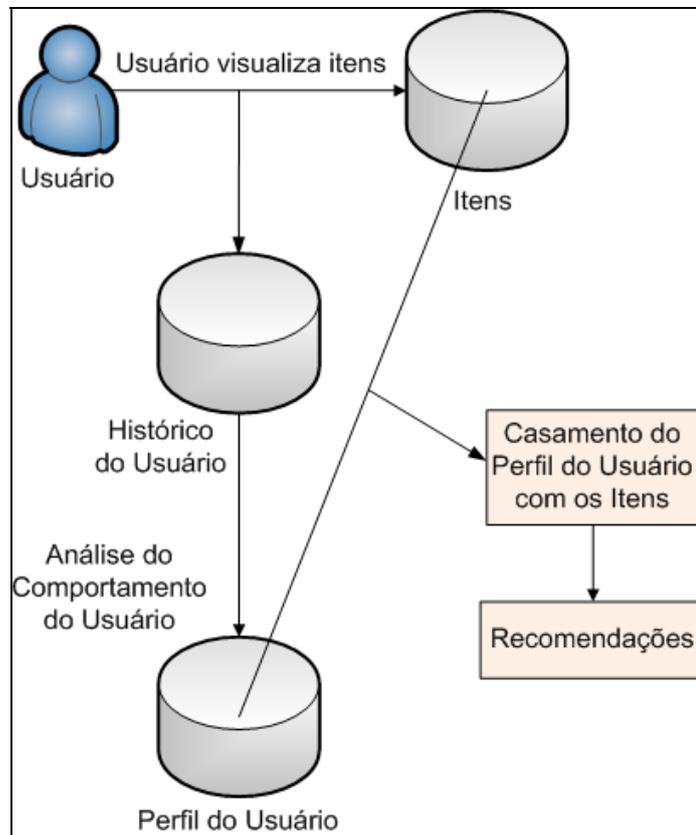
A segunda etapa corresponde às técnicas de recomendação, que constituem o núcleo do sistema de recomendação. Essas manipulam os dados fornecidos e geram as recomendações, podendo utilizar algoritmos de filtragem de informação, recuperação de informação, mineração de dados ou, até mesmo, uma mistura de todos para chegar a um resultado que satisfaça o usuário.

A terceira etapa consiste da entrega da recomendação ao usuário. Duas formas- *push* e *pull*- são utilizadas com frequência para entregar a recomendação, ilustrada na Figura 2 como apresentação/visualização das recomendações. Na forma *push*, o usuário não precisa pedir a recomendação; o sistema se encarrega de apresentá-las. Na forma *pull*, é necessário que o usuário solicite a recomendação para que a mesma seja apresentada. As próximas seções tratam de algumas das técnicas que são utilizadas com frequência em sistemas de recomendação.

### **3.1 Filtragem Baseada em Conteúdo**

Sistemas de recomendação que aplicam Filtragem Baseada em Conteúdo (FBC), realizam a filtragem de informação baseada na análise do conteúdo do item e no perfil do usuário, para recomendar itens relevantes e descartar os itens não relevantes. Itens semelhantes aos que um usuário tenha gostado no passado são recomendados, baseando-se no princípio de que os usuários tendem a se interessar por itens que já se interessaram antes. A similaridade é medida entre itens ao invés de usuários. A técnica do Cosseno, descrita detalhadamente na seção 3.5, pode ser utilizada para medir a similaridade entre itens e usuários.

Em um sistema de recomendação FBC, o usuário tem preferência por itens, os quais são descritos por suas características, normalmente chamadas de atributos. Se o item é uma música, por exemplo, seus atributos são: nome da música, compositor, cantor, banda, álbum, ano, gênero, etc. O perfil do usuário é descrito pelos atributos dos itens que ele visualiza. O sistema de recomendação baseado em conteúdo recomenda itens que estão de acordo com o perfil do usuário, o que normalmente é chamado de “casamento”. A Figura 3 ilustra o que ocorre em um sistema FBC.



**Figura 3 – Filtragem baseada em conteúdo**

Supondo-se um sistema FBC para músicas, um usuário poderia ouvir diversas de vários gêneros, as quais seriam armazenadas, tornando-se o seu histórico. A partir dele, descobre-se que o usuário gosta mais de músicas do gênero Rock. Para sugerir ao usuário outras músicas desse gênero, o sistema procura na base de dados as que tenham o atributo gênero igual a Rock.

## 3.2 Filtragem Colaborativa

Na Filtragem Colaborativa (FC), ao invés de recomendar itens que são semelhantes aos que um usuário tenha preferido no passado, recomenda-se itens semelhantes às preferências de outros usuários que tenham interesses em comum. A FC baseia-se na troca de experiências entre as pessoas.

Um sistema de recomendação desenvolvido com FC que permite ao usuário avaliar cada item, armazena as informações de avaliações para gerar as recomendações. Com base nisso, a FC encontra usuários com perfis semelhantes, normalmente chamados de “vizinhos mais próximos”. Esses são agrupados e os itens que foram avaliados pelos vizinhos com notas altas são recomendados. (BALABANOVIC, 1997)

Normalmente, um sistema FC utiliza uma matriz em que os usuários são representados por linhas e os itens por colunas. Os dados na matriz correspondem à avaliação dada por cada usuário a um item. Isso é exemplificado na Tabela 1, na qual os itens são títulos de livros, as avaliações vão de 0 a 5 e o hífen indica itens não avaliados. No exemplo, existe uma interrogação no usuário Valter, indicando que o sistema fará a previsão de qual avaliação o usuário Valter dará para o livro 3.

Ana e Valter avaliaram os mesmos livros com pontuações semelhantes, o que os tornam vizinhos próximos e, portanto, a avaliação de Ana para o livro 3 influenciará na previsão de Valter para o mesmo livro. Rosalina, Edson e Paula avaliaram os mesmos livros, entretanto, pelas pontuações atribuídas, eles não compartilham os mesmos gostos de Ana e Valter, portanto, terão menor influência na previsão da avaliação de Valter, comparados a Ana.

A previsão normalmente é feita por um cálculo que utiliza medidas de similaridade como o Cosseno ou a Correlação de Pearson<sup>4</sup>.

---

<sup>4</sup> O coeficiente de correlação de Pearson mede o grau de relação linear entre duas variáveis quantitativas e varia entre os valores -1 e 1. O valor 0 indica que não há relação linear, o valor 1 indica relação linear perfeita. O valor -1 indica relação linear perfeita, porém inversa, quando uma das variáveis aumenta, a outra diminui. Quanto mais próximo estiver de 1 ou de -1, mais forte é a associação linear entre as duas variáveis. (TORRES, 2004)

**Tabela 1 – Matriz de avaliações usuários x itens**

Usuários	Itens		
	Sherlock Holmes - Edição Definitiva - Comentada e Ilustrada - Vol. 1	A Batalha do Apocalipse - Da Queda dos Anjos ao Crepúsculo do Mundo	O Vendedor de Sonhos
Ana	3	3	4
Edson	-	2	1
Paula	5	-	2
Rosalina	-	5	-
Valter	3	3	?

A técnica dos vizinhos mais próximos e do Cosseno podem ser utilizadas na FC e são detalhadas nas seções 3.4 e 3.5.

### **3.3 Filtragem Híbrida**

Sistemas híbridos são aqueles que combinam mais de uma técnica de recomendação. Um sistema de recomendação que utiliza Filtragem Híbrida (FH) pode ser implementado de várias formas. Uma delas é implementar o sistema com FBC e FC separadamente, gerando listas de recomendação separadas para cada uma das técnicas. Após a geração, as listas são mescladas através de algum critério utilizado para selecionar as recomendações que serão apresentadas ao usuário. Outra forma é implementar o sistema com FBC adicionando características da FC ou, então, com FC adicionado de características FBC.

### **3.4 Técnica dos vizinhos mais próximos**

Um usuário pode ter preferências similares às de outros usuários e são, então, denominados os vizinhos mais próximos. A técnica é implementada como um algoritmo de classificação em que, para cada item/usuário não conhecido, verifica-se a similaridade entre ele e todos os outros itens/usuários da base, através de uma medida de distância, como: o Cosseno, a distância euclidiana, a manhatan, etc. A

técnica consegue garantir um número suficiente de vizinhos para a geração das recomendações.

A técnica do vizinho mais próximo é um classificador baseado em aprendizagem por analogia, ou seja, compara-se uma tupla de teste com tuplas de treinamento semelhantes a ela. As tuplas de treinamento são descritas por “ $n$ ” atributos. Cada tupla representa um ponto em um espaço  $n$ -dimensional. Dessa forma, todas as tuplas de treinamento são armazenadas em um espaço de padrões  $n$ -dimensional.

Dada uma tupla não classificada ou desconhecida, o classificador procura o espaço de padrões para as  $k$  tuplas de treinamento mais próximas da tupla desconhecida. Essas  $k$  tuplas de treinamento são os  $k$  "vizinhos mais próximos" da tupla desconhecida. Já para a classificação do  $k$  vizinho mais próximo, a tupla desconhecida é atribuída à classe mais comum entre os seus  $k$  vizinhos mais próximos. Quando  $k = 1$ , a tupla desconhecida é atribuída à classe da tupla de formação mais próxima a ela no espaço de padrões.

A técnica dos  $k$  vizinhos mais próximos pode também ser usada para previsão, ou seja, para retornar uma previsão de valor real para uma determinada tupla desconhecida. Nesse caso, o classificador retorna o valor médio real dos rótulos associados aos  $k$  vizinhos mais próximos da tupla desconhecida. Como já mencionado, uma medida de distância é utilizada pra calcular a similaridade entre duas tuplas. É preciso normalizar os valores de cada atributo da tupla antes da aplicação do cálculo de similaridade, a fim de que todos permaneçam em um mesmo intervalo de variação. Para normalizar os valores, utiliza-se a Equação 1. (WITTEN, 2005)

$$w^f = \frac{w - \min At}{\max At - \min At}$$

#### Equação 1 – Normalização

$w$  é o valor do atributo  $At$ ,  $\min At$  é o valor mínimo que o atributo  $At$  pode assumir,  $\max At$  é o valor máximo que o atributo  $At$  pode assumir e  $w^f$  é o valor

normalizado. Exemplificando: seja o atributo tempo de visualização de um programa de TV e supondo que  $\text{minAt}$  seja igual a 5 minutos,  $\text{maxAt}$  a 20 minutos e  $w$  a 7 minutos, então, o valor normalizado do atributo  $w'$  será

$$w' = \frac{7 - 5}{20 - 5} = \frac{2}{15} = 0,13333$$

### 3.5 Técnica do Cosseno

O Cosseno é uma medida de similaridade, um método de previsão, uma métrica que calcula a similaridade entre itens e usuários ou entre usuários e usuários. No cálculo itens-usuários, o Cosseno consulta itens similares a um dado item e faz o casamento entre conteúdo de item e perfil de usuário e é aplicado para descobrir se um item possui ou não correlação com o perfil do usuário. No cálculo usuários-usuários, o Cosseno calcula quão similares os usuários são.

Técnicas de recomendação aplicadas em sistemas de recomendação para Web podem utilizar a avaliação dos itens feita pelos usuários. Por exemplo, em um sistema de recomendação que não exige a avaliação dos usuários para os seus produtos, considera-se “produto consumido” e “produto não consumido” como uma forma de avaliação. Portanto, o Cosseno pode ser utilizado em ambas as situações, para calcular a similaridade tanto em sistemas com FBC quanto com FC.

Um vetor binário é formado por dois elementos,  $x$  e  $y$ , os quais contêm informações. Em um espaço  $n$ -dimensional, onde  $n$  é o número de itens do vetor, é possível, portanto, calcular o Cosseno entre os vetores, medindo, assim, a similaridade entre o perfil do usuário e o seu histórico ou, então, entre os usuários. A similaridade é considerada alta quando o valor do Cosseno é alto, ou seja, quanto mais próximo do valor 1, mais similar será e quanto mais próximo de 0, menos similar. Matematicamente, o Cosseno mede o ângulo existente entre dois vetores que representam usuários e itens ou usuários e usuários e retorna valores entre 1 –

máxima correlação – e 0 – mínima correlação. A Equação 2 apresenta o Cosseno, em que x e y são os vetores itens-usuários ou usuários-usuários. (CRISTO, 2009)

$$\cos(x, y) = \frac{\sum(xy)}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$$

### Equação 2 – Cosseno

Aplicando o Cosseno para o cálculo da similaridade entre usuários, para um perfil de um usuário chamado João com os valores de  $x = [0, 1, 5, 9, 7]$  e um perfil de outro usuário chamado Douglas com os valores de  $y = [9, 5, 1, 3, 7]$ . Efetuando-se o cálculo, nota-se que João e Douglas não são tão similares.

$$\cos(\text{João}, \text{Douglas}) = \frac{(0 * 9) + (1 * 5) + (5 * 1) + (9 * 3) + (7 * 7)}{\sqrt{(0^2 + 1^2 + 5^2 + 9^2 + 7^2)} \sqrt{(9^2 + 5^2 + 1^2 + 3^2 + 7^2)}} = 0,54$$

## 3.6 Mineração de Dados

Descoberta de Conhecimento em Base de Dados (DCBD), também conhecida como Busca de Conhecimento em Banco de Dados ou *Knowledge Discovery in Databases* (KDD), é um processo para obter informações e conhecimento úteis a partir de grandes bancos de dados.

Mineração de Dados se refere a um passo particular do processo de KDD; o processo de extração de conhecimento através da aplicação de algoritmos a fim de extrair padrões de um grande volume de dados.

Técnicas de mineração de dados podem ser aplicadas separadamente ou em conjunto com outras técnicas a sistemas de recomendação, com o intuito de melhorar as recomendações. Muitos desses sistemas desenvolvidos para a Web utilizam mineração de dados particularmente em lojas virtuais, justamente devido a grande quantidade de informação disponível nos bancos de dados.

Um sistema de recomendação pode usufruir de técnicas baseadas em heurísticas assim como de técnicas baseadas em modelos. Por exemplo, em um sistema totalmente FBC, podem ser aplicadas TF-IDF (*Term Frequency–Inverse Document Frequency*), árvores de decisão, classificadores bayesianos, redes neurais artificiais, etc. Em um sistema totalmente FC, podem ser aplicadas técnicas como *clustering*, regressão linear, modelos probabilísticos, etc. Em um sistema híbrido, as técnicas podem ser mescladas para a obtenção de melhores resultados e de um bom desempenho do sistema (ADOMAVICIUS, 2005; CÔRTES ET AL, 2002; MAXIMINO, 2007; SCHNEIDER, 2002)

### **3.7 Observações**

Um sistema de recomendação pode utilizar técnicas de FC e FC e/ou mineração de dados, para fornecer sugestões de itens a um determinado usuário e essas técnicas podem ser combinadas em um sistema híbrido.

No contexto da TVDPI, sistemas de recomendação podem ser utilizados para recomendar programas, propagandas, em *t-commerce*, entre outras possibilidades. Nesse sentido, diversos trabalhos vêm sendo desenvolvidos e alguns deles encontram-se descritos na seção de trabalhos correlatos, no capítulo 6 desta dissertação.

Um sistema de recomendação para TVDPI pode aumentar a satisfação do usuário em sua experiência com programas de televisão direcionados aos seus interesses. Os interesses dos usuários podem ser descobertos tanto explicitamente - quando o usuário fornece as preferências conscientemente, implicitamente - por meio da coleta de dados de visualização de TV, ou de ambas as formas.

Além disso, caso haja um grande aumento na carga de informação na TV, a partir da utilização da multiprogramação, por exemplo, os sistemas de recomendação podem reduzir tal sobrecarga, o que facilita a busca e o retorno da pesquisa dos programas interessantes ao usuário.

Portanto, no contexto da TVDPI as pessoas desejam ver programas interessantes, do seu gosto e, particularmente, sem perder tempo na busca desses programas.

O próximo capítulo apresenta o sistema desenvolvido neste trabalho.

# 4

## SISTEMA DE RECOMENDAÇÃO PARA TVDPI

---

Este capítulo tem o objetivo de apresentar o sistema desenvolvido neste trabalho, denominado BiPODiTVR (*Brazilian Portable Digital Television Recommender System* – Sistema de Recomendação para Televisão Digital Portátil Interativa Brasileira). O capítulo está organizado da seguinte forma: a seção 4.1 apresenta detalhes da arquitetura e do funcionamento do BiPODiTVR; a seção 4.2 de sua implementação e a seção 4.3 encerra este capítulo com as observações finais.

### 4.1 Arquitetura do BiPODiTVR

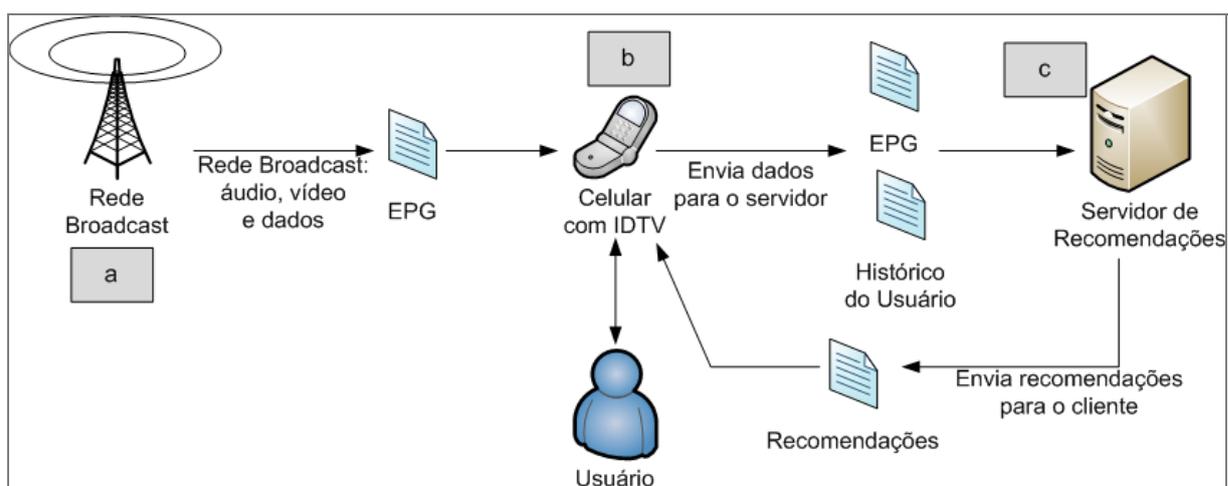
O sistema BiPODiTVR foi desenvolvido com o intuito de fornecer personalização na TVDPI, mais especificamente, para os celulares. O sistema visa à recomendação de programas de acordo com as preferências dos usuários e esses são recomendados pela utilização de técnicas de FBC, FC e FH.

Dois atributos são usados como parâmetros para a geração da recomendação, o nome do programa propriamente dito e a categoria do programa. A Categoria pode ser definida como o gênero a que um programa de TV pertence, por exemplo, filme, jornalismo, debate, humorístico, etc. Os gêneros são organizados na TVDPI brasileira de acordo com a norma NBR 15603-2 da ABNT. A Tabela 2 apresenta a classificação do gênero do descritor contida na norma, utilizada no desenvolvimento de BiPODiTVR.

**Tabela 2 – Classificação de gênero do descritor**

Classificação de gênero	Descritor de conteúdo
0x0	Jornalismo
0x1	Esporte
0x2	Educativo
0x3	Novela
0x4	Minissérie
0x5	Série/seriado
0x6	Variedade
0x7	Reality show
0x8	Informação
0x9	Humorístico
0xA	Infantil
0xB	Erótico
0xC	Filme
0xD	Sorteio, televidas, premiação
0xE	Debate/entrevista
0xF	Outros

A Figura 4 apresenta o fluxo do funcionamento básico do BiPODiTVR. A rede de TV (a) envia áudio, vídeo e dados para o dispositivo receptor de TVDPI, exemplificado na Figura por um aparelho celular (b). O celular com recepção de TVD recebe o EPG, coleta e armazena os dados do usuário e envia (c) essas informações para o servidor de recomendações.



**Figura 4 – Funcionamento do BiPODiTVR**

O servidor recebe e armazena o EPG e o histórico do usuário, que são enviados pelo celular, e geram o perfil e as recomendações, as quais são enviadas de volta ao celular e, apresentadas em sua tela.

A grande vantagem de se utilizar uma arquitetura cliente-servidor é que o processamento pesado fica por conta do servidor e não do dispositivo portátil. A Figura 5 apresenta a arquitetura projetada para a parte Cliente do BiPODiTVR, a qual é composta por 5 módulos e 3 bases de dados.

A parte Cliente executa em um dispositivo portátil, por exemplo, um aparelho celular. Foi desenvolvida nas linguagens NCL e LUA e necessita do *middleware* Ginga-NCL para executar e se comunicar remotamente com o servidor pelo canal de retorno. A seguir, são descritas as funcionalidades de cada módulo da parte Cliente.

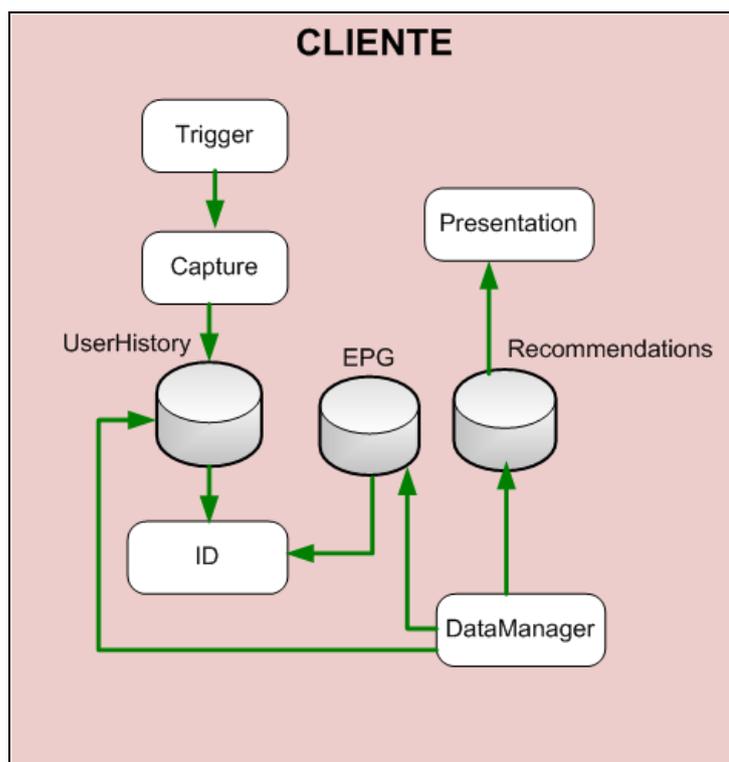


Figura 5 – Arquitetura Cliente

O módulo *Capture* insere informações na base de dados *UserHistory*. O módulo *ID* obtém as informações armazenadas na base de dados *EPG* e *UserHistory*. *Presentation* obtém informações armazenadas na base de dados

*Recommendations* e *DataManager* tem acesso a todas as bases de dados para o gerenciamento dos dados armazenados.

O módulo *Trigger* dá início a todo o processamento do sistema. Quando o usuário liga a TV Digital no seu celular, *Trigger* notifica o módulo *Capture*, o qual captura a interação do usuário com a TV. Esse módulo é ativado para realizar o monitoramento do comportamento do usuário. O módulo *ID* é responsável por fazer e tratar a identificação do usuário que está utilizando o aparelho e enviar os dados do histórico de visualização e do EPG para o servidor. O módulo *Presentation* é responsável por organizar e apresentar as recomendações na tela, quando solicitadas pelo usuário. *User History* é a base de dados que armazena todas as informações que são coletadas pelo módulo *Capture*.

*EPG* é a base de dados que contém as informações do guia eletrônico de programação, mais popularmente conhecido como grade de programação da TV. Usualmente, um EPG contém informações como: nome e código numérico da emissora, nome, data da exibição, hora de início e de término do programa seu código numérico e tempo de duração.

Já a parte Servidor é composta por 4 módulos e 5 bases de dados, como se observa na arquitetura apresentada na Figura 6. *Generates Profile* obtém dados da base *UserHistory* e os insere na base *Profile*. *Generates Recommendations* obtém dados da base *Profile* e *EPG* e os insere na base *Recommendations*. *ID* obtém dados da base *Users* e *DataManager* gerencia os dados armazenados em todas as bases de dados.

A parte Servidor consiste de um *WebService*, implantado em um container *Tomcat*, foi desenvolvida em Java e emprega JPA (*Java Persistence API*) para realizar as transações com o banco de dados *MySQL*. Além de fornecer uma lista de recomendações, o *WebService* do BiPODiTVR oferece serviços adicionais, como: uma lista com os *top 10* programas, uma lista com as *top 10* categorias e o EPG. A seguir, são descritas as funcionalidades de cada módulo da parte Servidor.

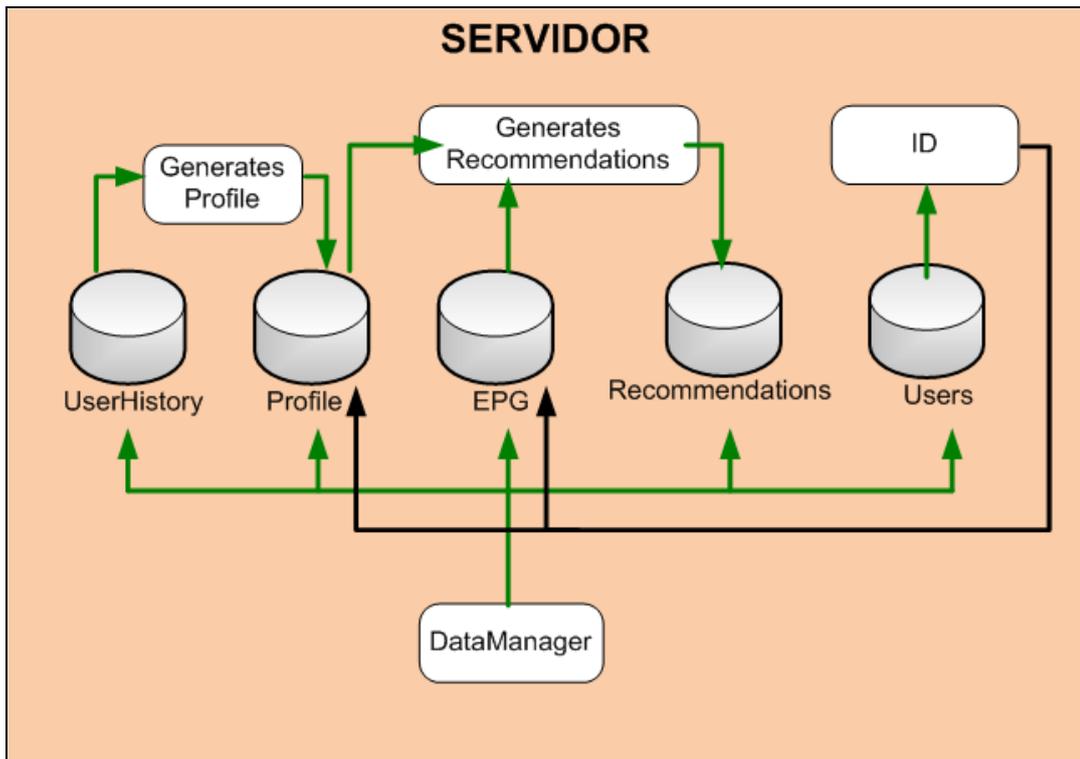


Figura 6 – Arquitetura Servidor

O módulo *DataManager* é o responsável por gerenciar os dados e excluí-los assim que se tornam obsoletos. O módulo *ID* é responsável por fazer e tratar a identificação do usuário que está utilizando o aparelho. O módulo *Generates Profile* é responsável por gerar o perfil do usuário, a partir do histórico de visualização, independente da técnica de recomendação escolhida. O módulo aplica uma técnica que permite descobrir as categorias e também os programas que o usuário prefere assistir. O perfil do usuário é atualizado com a coleta dos dados de visualização e com as recomendações aceitas.

O módulo *Generates Recommendations* encontra, a partir do perfil gerado e do EPG, as recomendações, ou seja, os programas que podem ser sugeridos. Para isso, é aplicada também uma técnica, que pode ser igual ou diferente da utilizada no módulo de geração de perfil. Os programas candidatos a recomendações podem ser descobertos tanto com base na categoria, quanto no nome do programa. O BiPODiTVR faz uma mescla, obtendo tanto sugestões baseadas no nome do programa, quanto na categoria. Esse módulo seleciona, dentre todos os programas candidatos, no máximo 5 para serem exibidos pelo módulo *Presentation*.

*User History* é a base de dados que armazena as informações coletadas pelo módulo *Capture* que, por sua vez as envia ao servidor para serem utilizadas na geração do perfil. *Profile* é a base de dados que armazena todas as informações dos perfis gerados. Aqui são armazenadas informações como: nome da emissora, nome, data da exibição e hora de início do programa, tempo de visualização do usuário, data e hora da visualização do programa pelo usuário.

*Recommendations* é a base de dados que armazena todas as recomendações geradas. Essas contêm informações como: nome da emissora, nome, categoria, data da exibição e hora de início do programa. Dentre essas informações, foram somente utilizadas: nome da emissora, nome, horário de início e de término, data de exibição e categoria do programa.

O fluxo entre os módulos e as partes do Cliente e do Servidor é descrito na Figura 7, a seguir.

Primeiramente, o usuário liga a TV em seu celular. O dispositivo recebe o sinal digital da rede de TV e os dados. O módulo *Trigger* notifica o módulo *Capture*. Em um dado momento, o usuário solicita uma recomendação. Nesse momento o módulo *ID* identifica o usuário e todas as informações relativas a ele são armazenadas com um número de identificação. Então, a parte *Cliente* se comunica com a parte Servidor, já que a identificação é feita com o auxílio da base de dados *Users*, localizada no Servidor.

Após a identificação do usuário, os dados capturados até aquele momento são enviados para a parte Servidor. Os dados são recebidos e armazenados nas bases correspondentes do servidor. O módulo *Generates Profile* descobre um perfil com base no histórico de visualização e o armazena.

Em seguida, o módulo *Generates Recommendations*, baseado nas informações do EPG e do Histórico de Visualização, gera, armazena e envia os programas escolhidos para o celular, na forma de uma lista de recomendações. Por fim, o módulo *Presentation* recebe os dados e exibe as recomendações na tela do dispositivo.

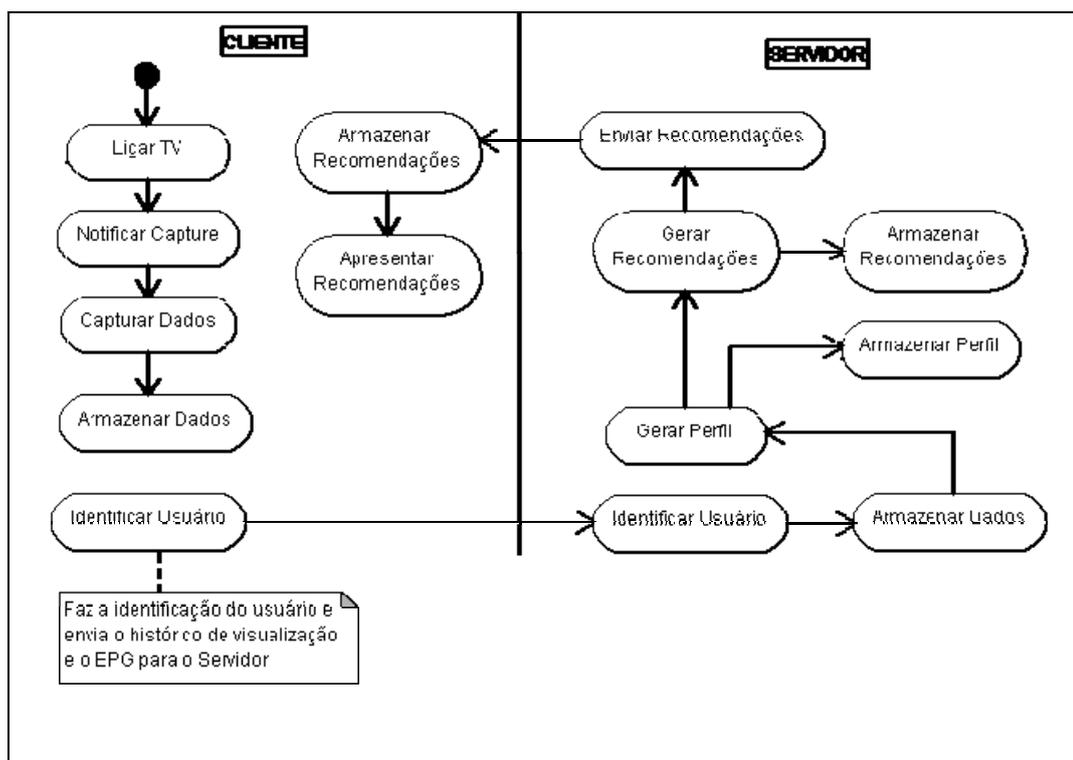


Figura 7 – Fluxo entre os módulos e as partes do BiPODiTVR

## 4.2 Implementação do BiPODiTVR

O BiPODiTVR foi implementado em uma arquitetura cliente-servidor, estruturado em módulos e desenvolvido em quatro fases.

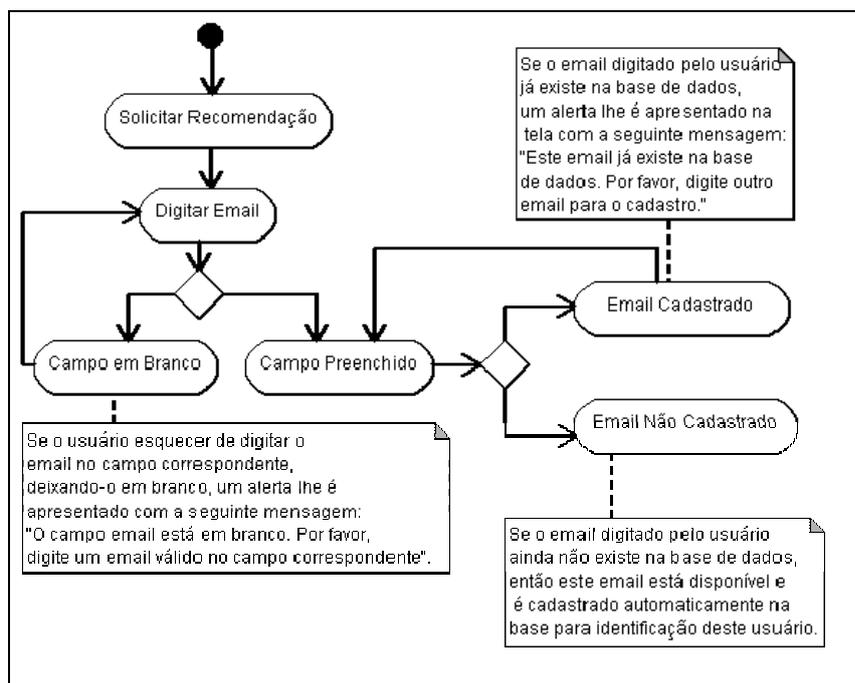
A primeira delas consistiu em definir as tecnologias e as linguagens que seriam utilizadas para o desenvolvimento do sistema. A segunda teve como foco o desenvolvimento da aplicação no lado servidor, que é um Webservice que fornece serviços de recomendação. As recomendações são geradas para um usuário em particular e não para um grupo de pessoas, como ocorre em sistemas de recomendação para TV digital em residências. Para atender a esse requisito, foi necessário definir uma forma de identificação do usuário, a fim de garantir a entrega correta da recomendação.

Mesmo que um dispositivo, por exemplo, um aparelho celular, pertença a um determinado usuário, seu uso pode ser compartilhado com outros. Portanto, o usuário que utilizar o BiPODiTVR deverá se identificar através de um *e-mail*. A

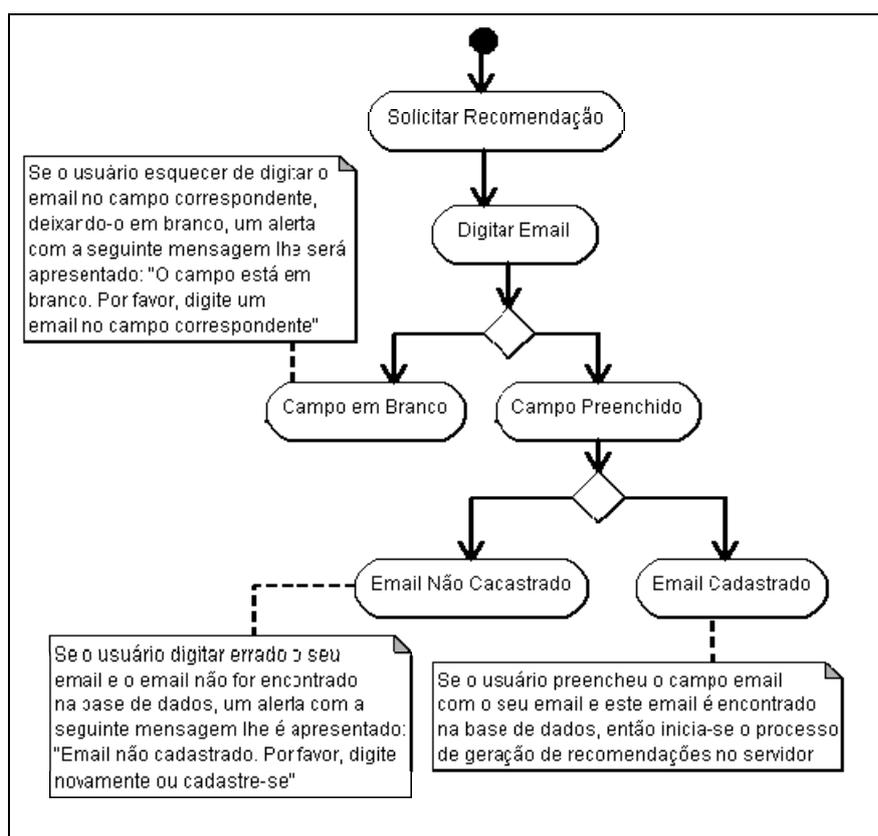
utilização do email garante que aquele usuário em particular esteja utilizando o dispositivo. Vale ressaltar que essa é a única identificação explícita exigida pelo sistema. As informações armazenadas em um servidor e atualizadas através da rede garantem que o perfil e o histórico do usuário sejam mantidos, caso o usuário perca o dispositivo ou esse venha a ser roubado.

O estágio atual de desenvolvimento dos receptores portáteis de TVDPI ainda não permite que soluções mais avançadas sejam propostas e implementadas.

Quando o usuário solicitar uma recomendação pela primeira vez, ele é cadastrado no sistema, o qual passa a armazenar e gerar o histórico do usuário e as recomendações de forma identificada. A partir dessa etapa, sempre que o usuário solicitar uma recomendação, o seu endereço de *e-mail* será requisitado. A Figura 8 apresenta os passos relacionados à primeira utilização do BiPODiTVR, em que é necessário cadastrar o usuário e a Figura 9 ilustra os passos após o primeiro acesso.



**Figura 8 – Primeira vez solicitando recomendações**



**Figura 9 – Passos da identificação do usuário após a primeira solicitação de recomendação e cadastramento**

Quando o usuário solicita uma recomendação, além da identificação pelo e-mail, arquivos que contêm o histórico de visualização e o EPG, que estão no dispositivo portátil, são enviados para o servidor. A Tabela 3 apresenta um fragmento do histórico de visualização.

**Tabela 3 – Exemplo do histórico de visualização**

<i>Id</i>	000001	000002
<i>Dia</i>	08/09/2010	06/09/2010
<i>Horário</i>	13:15:00	10:10:00
<i>Duração</i>	00:29:00	01:00:00
<i>Dia da semana</i>	Quarta-feira	Domingo
<i>Período do dia</i>	Tarde	Manhã
<i>Nome do Programa</i>	As visões da raven	Bom Dia Mulher
<i>Nome da Emissora</i>	SBT	Rede TV
<i>Categoria</i>	Série/Seriado	Variedade

Nessa tabela constam: a identificação do usuário, no formato 00000X; o dia de exibição do programa, no formato dd/mm/aaaa; o horário em que o programa

foi assistido; a duração – quanto tempo o usuário assistiu aquele programa, no formato hh:mm:ss; o dia da semana, de domingo à sábado; o período do dia, manhã, tarde, noite ou madrugada; o nome do programa, o nome da emissora e a categoria do programa.

O histórico de visualização é armazenado em uma tabela específica do banco de dados no servidor, juntamente com os dados coletados anteriormente. O EPG, por sua vez, é armazenado em uma tabela e, em seguida, convertido em uma matriz que também é armazenada. Essa conversão é necessária devido à utilização da técnica do Cosseno no BiPODiTVR. Então, uma nova tabela é criada, com as informações constantes da tabela do EPG, além das categorias dos programas. Esses campos são preenchidos com 0 ou 1, conforme a categoria do programa da tabela original se encaixa ou não na categoria da nova tabela e são utilizados para calcular o Cosseno. A Figura 10 ilustra os passos para a identificação do usuário, a recepção e o armazenamento dos dados.

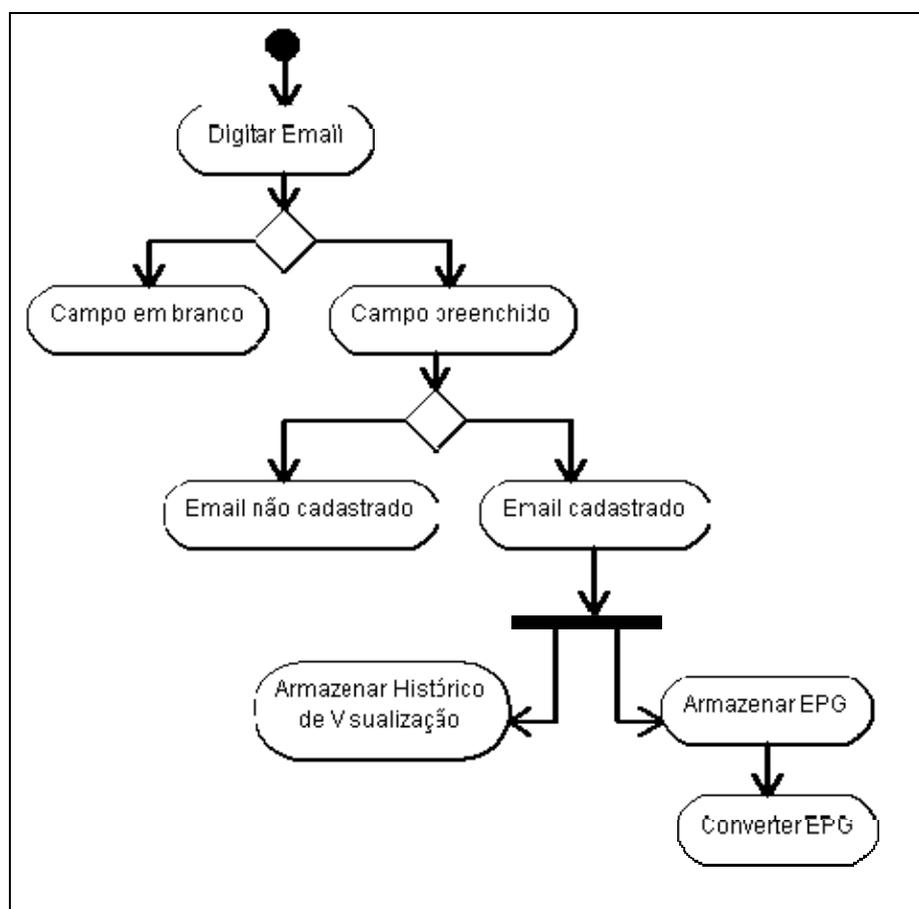


Figura 10 – Identificação do usuário e armazenando dos dados

A Tabela 4 apresenta um fragmento do EPG original. Nela constam os atributos do EPG: dia, no formato dd/mm/aaaa; código da emissora, no formato 00Y; nome da emissora, hora de início e término do programa; tempo de duração, no formato hh:mm:ss; nome e gênero, no formato 0xY, e categoria do programa.

**Tabela 4 – Fragmento do EPG original**

<i>Dia</i>	07/09/2010	07/09/2010	08/09/2010
<i>Código Emissora</i>	005	007	013
<i>Nome Emissora</i>	RedeTV!	SBT	MTV
<i>Hora Início Programa</i>	14:05:00	20:59:00	11:59:00
<i>Hora Término Programa</i>	16:53:00	21:40:00	12:28:00
<i>Duração Programa</i>	02:48:00	00:41:00	00:29:00
<i>Nome Programa</i>	A Tarde é Sua	SBT Brasil	My Own
<i>Gênero</i>	0x6	0x0	0x5
<i>Categoria</i>	Variedade	Jornalismo	Série/Seriado

A Tabela 5 apresenta um fragmento do EPG convertido em uma matriz. Nela constam os atributos do EPG adicionados das 10 categorias de programas. No dia 07/10/2009, às 14:05:00, tem início o programa “A Tarde é Sua” na emissora RedeTV! O programa se encaixa na categoria Variedade e, portanto, o campo desse atributo é preenchido com o valor 1, enquanto os outros campos são preenchidos com 0.

**Tabela 5 – Fragmento do EPG convertido em uma matriz**

<i>Dia</i>	07/09/2010	07/09/2010	08/09/2010
<i>Código Emissora</i>	005	007	013
<i>Nome Emissora</i>	RedeTV!	SBT	MTV
<i>Hora Início Programa</i>	14:05:00	20:59:00	11:59:00
<i>Hora Término Programa</i>	16:53:00	21:40:00	12:28:00
<i>Duração Programa</i>	02:48:00	00:41:00	00:29:00
<i>Nome Programa</i>	A Tarde é Sua	SBT Brasil	My Own
<i>Gênero</i>	0x6	0x0	0x5
<i>Categoria</i>	Variedade	Jornalismo	Série/Seriado
<i>Jornalismo</i>	0	1	0
<i>Esporte</i>	0	0	0
<i>Educativo</i>	0	0	0
<i>Novela</i>	0	0	0
<i>Minissérie</i>	0	0	0
<i>Série/seriado</i>	0	0	1
<i>Variedade</i>	1	0	0
<i>Reality show</i>	0	0	0
<i>Informação</i>	0	0	0
<i>Humorístico</i>	0	0	0

Realizada a identificação do usuário no momento da solicitação e após o armazenamento dos novos dados, a FC e a FBC são aplicadas a esses dados,

gerando recomendações dos dois tipos. É importante ressaltar que o BiPODiTVR gera recomendações de programas para o horário atual de visualização, ou seja, toma como base o horário em que a recomendação foi solicitada, em um intervalo de uma hora a mais e uma hora a menos. Se não forem encontrados programas que estejam sendo transmitidos nesse faixa de horário, o BiPODiTVR recomenda um programa que reflita o perfil do usuário, independentemente do horário. A Figura 11 exemplifica os passos relacionados ao envio das recomendações do servidor para o cliente.

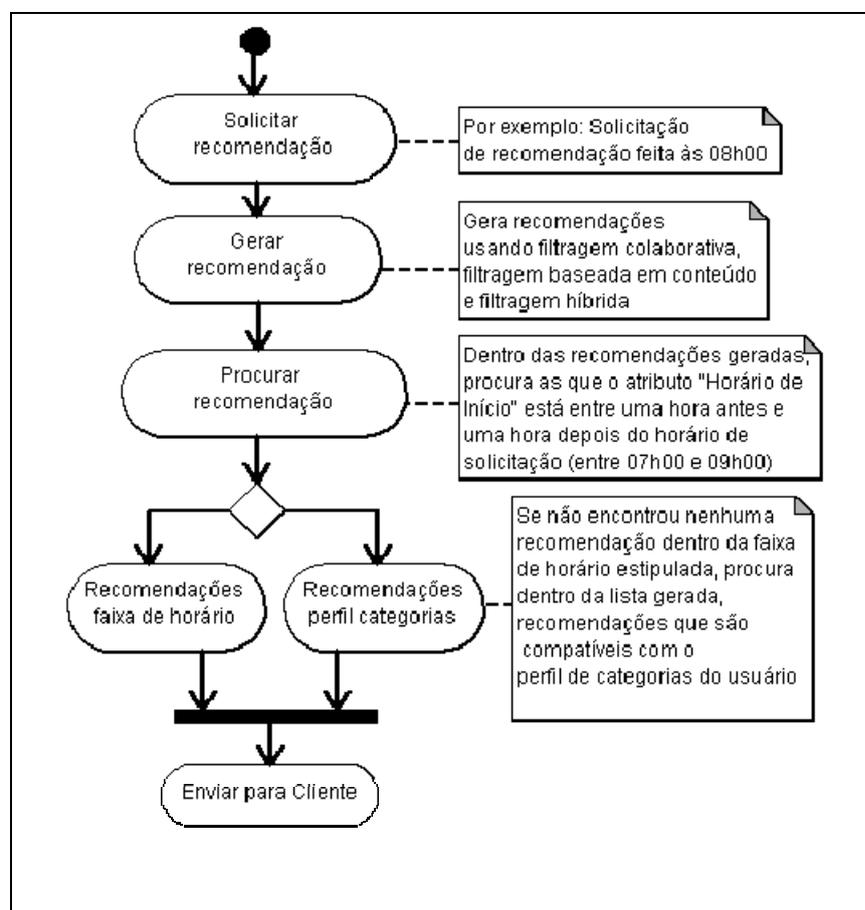
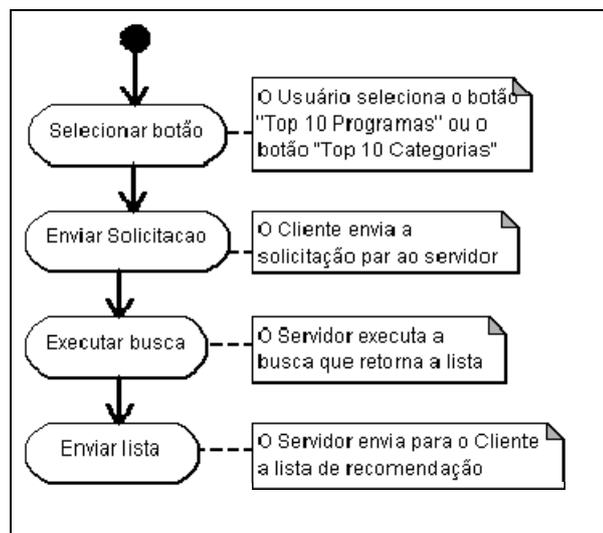


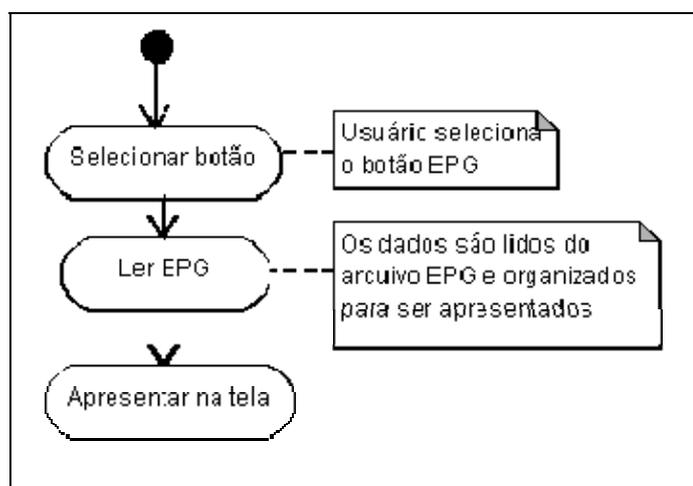
Figura 11 – Passos para o envio das recomendações ao cliente após a solicitação

A terceira fase envolveu o desenvolvimento do cliente. A aplicação foi desenvolvida em NCL/LUA com o auxílio da IDE Eclipse, que tem um *plugin* específico para o desenvolvimento de aplicações desse tipo. A compilação e execução dessas aplicações são realizadas na máquina virtual Ginga-NCL, que é uma implementação de referência do *middleware* Ginga para TVD brasileira.



**Figura 12 – Passos da solicitação dos Top 10 Programas e Categorias.**

O cliente pode solicitar serviços que não necessitam de identificação, como os *top 10* Programas, as *top 10* Categorias e o EPG. Quando o usuário quer verificar quais os 10 melhores programas ou categorias, ele deve selecionar os botões correspondentes na tela do dispositivo, os quais enviam uma requisição ao servidor e esse, por sua vez, processa e retorna a resposta, que é exibida na tela. A Figura 12 ilustra os passos para a solicitação dos top 10 programas e das top 10 categorias.



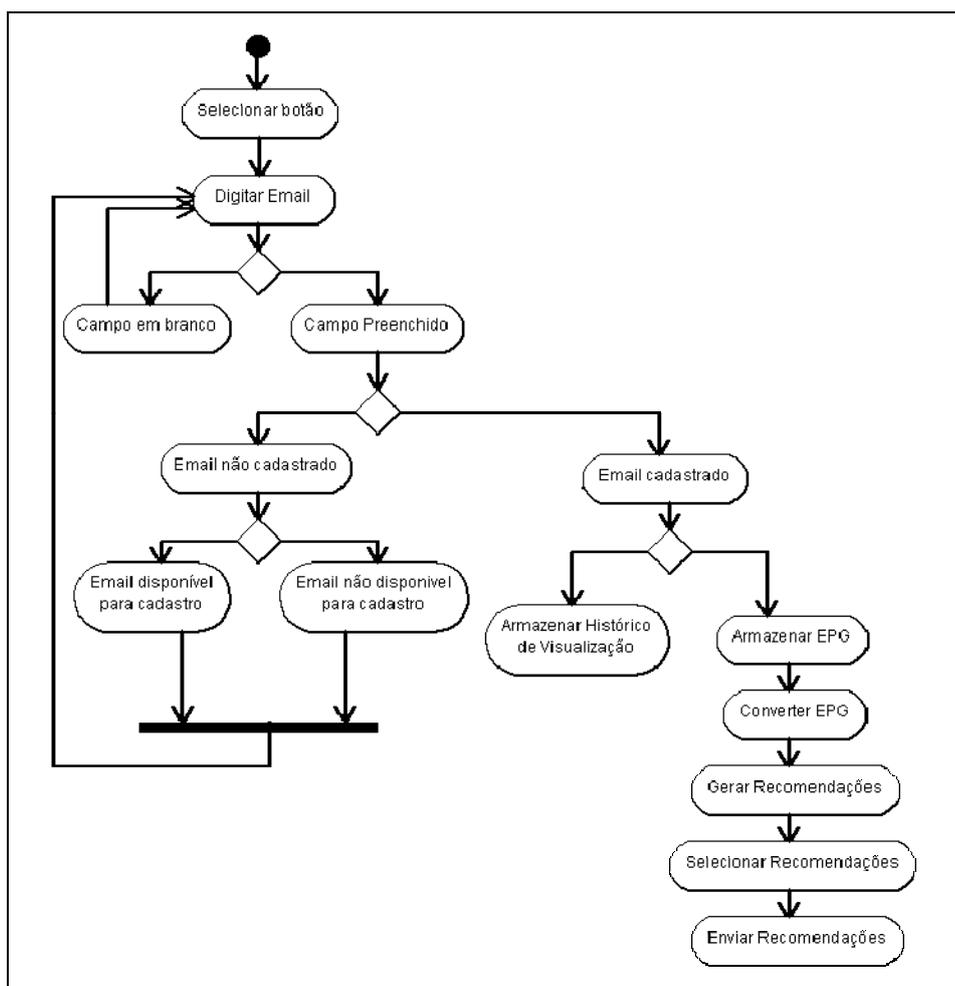
**Figura 13 – Passos da solicitação do EPG.**

O EPG é o único serviço que não necessita de comunicação remota, pois o dispositivo recebe e armazena o EPG dentro dele. Quando o usuário deseja ver o EPG, ele seleciona o botão correspondente e os dados são lidos e apresentados na tela.

A Figura 13 ilustra os passos para a solicitação do EPG. Para solicitar uma recomendação, o usuário deve selecionar o botão “Temos sugestões para você”, o que inicia o processamento descrito nesta seção.

A Figura 14 ilustra o que ocorre quando um usuário solicita a recomendação, desde o momento em que ele clica no botão até o momento em que a recomendação é enviada.

A quarta fase consistiu de testes a partir do sistema já pronto, os quais são detalhados no capítulo 5, juntamente com os resultados.



**Figura 14 – Passos da solicitação da recomendação.**

### 4.3 Aplicação Protótipo

Para validar o BiPODiTVR, o sistema de recomendação e uma aplicação NCL/LUA foram implementados. A Figura 15 apresenta duas capturas de tela da aplicação Cliente. À esquerda, tem-se a captura da tela inicial e à direita, a tela de identificação do usuário. A tela inicial possui 4 botões, o vermelho, para o usuário solicitar os top 10 programas, o verde, para solicitar as top 10 categorias, o amarelo, para solicitar o EPG e o azul, para solicitar uma recomendação. A tela de identificação possui um campo para a digitação do email e dois botões; o verde, para enviar o email e efetuar a identificação e o azul, para cancelar e retornar à tela inicial.



Figura 15 – Tela inicial e tela de identificação

A Figura 16 apresenta duas capturas de tela. O lado direito apresenta o retorno das top 10 categorias, quando o usuário seleciona o botão top 10 categorias na tela inicial. O lado esquerdo apresenta o retorno dos top 10 programas, quando o usuário seleciona o botão top 10 programas na tela inicial. Essas telas apresentam as informações e contêm apenas o botão voltar, que retorna à tela inicial.



Figura 16 – Top 10 Programas e Categorias

A Figura 17 apresenta o retorno das recomendações, quando o usuário seleciona o botão temos sugestões para você na tela inicial. A tela apresenta as recomendações enviadas e um único botão, o voltar, para retornar à tela inicial.



Figura 17 – Tela “Sugestões”: tela retornada quando o usuário seleciona o botão “Temos sugestões para você”

#### 4.4 Observações

Este capítulo teve como objetivo apresentar a arquitetura e a implementação

do sistema de recomendação desenvolvido para a televisão digital portátil interativa brasileira. O sistema tem como objetivo recomendar programas de TV com base no que o usuário já assistiu no passado.

O sistema gera as recomendações utilizando como parâmetros o nome do programa, a categoria e o horário de solicitação. Baseado nisso,, o sistema utiliza o perfil do usuário para conhecer as suas preferências e o EPG para buscar programas que estejam dentro dos parâmetros indicados.

Várias tecnologias foram utilizadas no desenvolvimento do sistema, como: *WebServices*, máquina virtual da implementação de referência do *middleware* Ginga-NCL, *Tomcat*, *MySQL* e as linguagens de programação NCL, LUA e JAVA, o que demonstra a multidisciplinaridade do trabalho realizado.

O funcionamento do sistema, a comunicação entre partes e os módulos da arquitetura, além das solicitações do usuário são explicados e ilustrados pelas Figuras e Tabelas. Os testes realizados após o término de desenvolvimento do sistema são apresentados no próximo capítulo.

# 5

## EXPERIMENTOS E RESULTADOS

---

Este capítulo apresenta os experimentos conduzidos com o BiPODiTVR, que permitiram a avaliação dos resultados obtidos com a aplicação protótipo. O capítulo está organizado da seguinte forma: a seção 5.1 apresenta os dados que foram utilizados nos testes, a seção 5.2 detalha os testes realizados e os resultados obtidos, a seção 5.3 destaca a aplicação protótipo desenvolvida e, por fim, a seção 5.4 encerra este capítulo com as observações finais.

### 5.1 Dados do IBOPE<sup>5</sup>

O IBOPE forneceu os dados utilizados nos testes com o BiPODiTVR. Esses dados são o histórico de visualização e o EPG correspondente a 15 dias de programação e monitoramento de 6 domicílios brasileiros. Os dados foram tratados manualmente quase que em sua totalidade, a fim de adequá-los aos testes.

O EPG é composto por 15 arquivos TXT, denominados arquivos de programação, um para cada dia entre 05/03/2008 e 19/03/2008, com a grade de 10 emissoras de TV Aberta. Todos esses dados foram organizados e armazenados no banco de dados MySQL. O comportamento do usuário é descrito em várias planilhas, denominadas planilhas de sintonia, que apresentam informações dos programas visualizados pelos usuários nesses 15 dias. Detalhes da organização dos dados do IBOPE podem ser encontrados em (GATTO, 2009; GATTO, 2010a; GATTO, 2010b; GATTO, 2010c).

A Tabela 6 mostra a quantidade de indivíduos e TVs por domicílio. Como pode ser observado, os domicílios 1, 2 e 5 possuem apenas uma TV na residência,

---

<sup>5</sup> <http://www.ibope.com.br>

enquanto os domicílios 3, 4 e 6 possuem duas. Os domicílios 1, 4 e 5 possuem dois indivíduos, enquanto os domicílios 2, 3, e 6 possuem três indivíduos. No total há 6 domicílios, 9 televisores e 15 indivíduos.

**Tabela 6 – Características dos domicílios**

<i>Domicílios</i>	<i>Qtde TVs</i>	<i>Qtde Indivíduos</i>
1	1	2
2	1	3
3	2	3
4	2	2
5	1	2
6	2	3
<i>Total</i>	9	15

A Tabela 7 corresponde aos dados exibidos na Tabela 6, porém descreve o perfil sócio econômico da família e os indivíduos por sexo e idade.

**Tabela 7 – Características socioeconômicas dos domicílios**

<i>Domicílio</i>	1	2	3	4	5	6
<i>Classe Social</i>	DE	C	C	AB	C	AB
<i>Idade da dona-de-casa</i>	44	45	39	32	60	36
<i>Grau de instrução do chefe-de-casa</i>	Primário Incompleto	Ginasial Incompleto	Ginasial Incompleto	Colegial Completo	Colegial Incompleto	Colegial Completo
<i>Sexo do indivíduo 1</i>	Feminino	Feminino	Feminino	Feminino	Feminino	Feminino
<i>Idade do indivíduo 1</i>	44	45	39	32	60	36
<i>Sexo do indivíduo 2</i>	Feminino	Masculino	Masculino	Masculino	Masculino	Masculino
<i>Idade do indivíduo 2</i>	8	48	40	30	77	38
<i>Sexo do indivíduo 3</i>	Não informado	Feminino	Feminino	Não informado	Não informado	Masculino
<i>Idade do indivíduo 3</i>	Não informado	17	13	Não informado	Não informado	14

## 5.2 Testes e Resultados

O BiPODiTVR foi testado através de sua simulação por um período de 15 dias, com os 15 usuários da amostra. Os dados do IBOPE foram utilizados na simulação, separando-se os dados de cada usuário e de cada dia em arquivos CSV (*Comma-separated values*), por exemplo, Usuario1Dia1.csv, Usuario1Dia2.csv, Usuario2Dia1.csv, Usuario2Dia2.csv, etc.

A simulação ocorreu da seguinte forma: na simulação do primeiro dia, o histórico de visualização do dia foi inserido no banco de dados de cada um dos usuários. Em seguida, foi simulado o acesso de cada usuário ao sistema, o que resultou na geração das listas de recomendação. Só então, a porcentagem de acertos do sistema foi calculada.

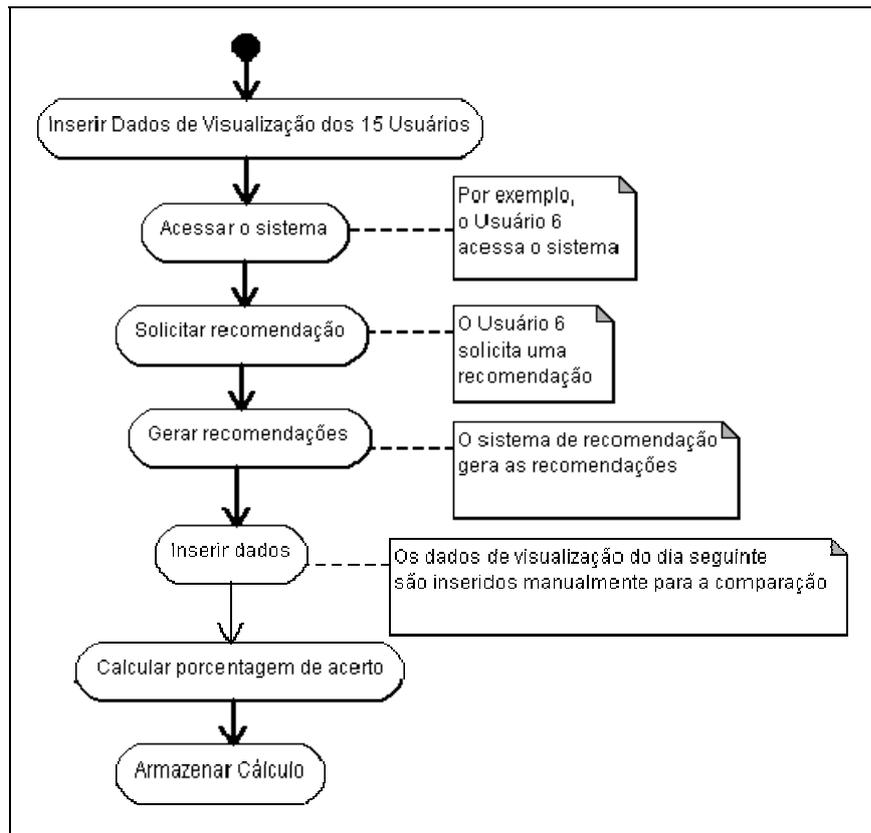
Na simulação do segundo dia, os dados correspondentes a todos os usuários foram inseridos no banco. Para a geração das recomendações, portanto, foram utilizados os dados do primeiro e do segundo dia. O acesso de cada usuário ao sistema foi simulado, gerando as listas e efetuando o cálculo da porcentagem de acertos. Isso se repetiu para todos os outros dias. A Figura 18 ilustra o passo a passo da simulação para um dia e para um único usuário.

A simulação ocorreu de forma iterativa, ou seja, um dia após o outro e as recomendações foram geradas conforme descrito no capítulo 4. Assim como a simulação do uso do BiPODiTVR, o cálculo da porcentagem de acertos também foi feito de forma automática. Esse cálculo ajudou a verificar se o BiPODiTVR estava gerando e oferecendo recomendações correspondentes às preferências do usuário.

A análise dos resultados foi realizada pela medida da precisão em apresentar conteúdos relevantes às expectativas dos usuários sendo utilizada a Equação 3 para isto.

$$\text{porcentagem de acerto} = \frac{\text{número de acertos}}{\text{número de recomendações}}$$

Equação 3 – Cálculo da porcentagem de acerto



**Figura 18 – Passo a passo da simulação**

Como o BiPODiTVR é um sistema híbrido, três listas principais de recomendações são geradas e armazenadas, sendo uma lista de recomendação para FC, outra para FBC e a lista FH. A lista híbrida é formada pelas recomendações encontradas nas listas FC e FBC. Essas se subdividem ainda em atributos que se aplicam à categoria e ao nome do programa, resultando, então, em um total de seis listas. O cálculo da porcentagem de acerto foi aplicado a essas listas e a porcentagem de acerto foi computada separadamente.

Chega-se ao número de acertos da maneira descrita a seguir. Após o BiPODiTVR gerar as listas, os dados correspondentes ao dia seguinte são incluídos no banco de dados e, então, o serviço do Webservice do BiPODiTVR denominado “análise” é executado. O número de acertos é o resultado de uma comparação: se um nome de programa ou categoria gerado na lista de recomendação foi assistido pelo usuário no dia seguinte, o BiPODiTVR acertou e esse acerto é armazenado.

Por exemplo, as listas de recomendação são geradas na simulação do primeiro dia. Os dados do segundo dia são, então, inseridos no banco. Realiza-se a comparação entre os dados resultantes das listas e os dados do segundo dia.

A Tabela 8 apresenta um exemplo de comparação para a FC. A Tabela é constituída pelas colunas principais “FC por categoria” e “FC por nome do programa” e pelas secundárias “recomendações geradas” e “programas visualizados no dia seguinte”. A categoria, ou nome do programa presente em ambas as colunas indica o acerto e são marcados em cinza. Portanto, na “FC por categoria” obteve-se 2 acertos - Novela e Variedade e na “FC por nome do programa”, obteve-se 1 acerto, Chiquititas.

**Tabela 8 – Exemplo de comparação para a FC**

<i>FC por categoria</i>		<i>FC por nome do programa</i>	
<i>Recomendações geradas</i>	<i>Programas visualizados no dia seguinte</i>	<i>Recomendações geradas</i>	<i>Programas visualizados no dia seguinte</i>
Humorístico	Infantil	Castelo Ra Tim Bum	TV Globinho
Novela	Filme	Chiquititas	Sessão da Tarde
Série/Seriado	Novela	As visões da Raven	Chiquititas
Variedade	Variedade	Bom Dia Mulher	Jornal da Globo

A Tabela 9 exemplifica a comparação para a FBC e é constituída pelas colunas principais “FBC por categoria” e “FBC por nome de programa” e pelas “recomendações geradas” e “programas visualizados no dia seguinte”. A categoria ou nome do programa presente em ambas as colunas indica o acerto e são marcados em cinza. Portanto, na “FBC por categoria” obteve-se 1 acerto - Debate/Entrevista e na “FBC por nome de programa”, obteve-se 2 acertos - CSI Miami e Mulheres em Foco.

**Tabela 9 – Exemplo de comparação para a FBC**

<i>FBC por categoria</i>		<i>FBC por nome do programa</i>	
<i>Recomendações geradas</i>	<i>Programas visualizados no dia seguinte</i>	<i>Recomendações geradas</i>	<i>Programas visualizados no dia seguinte</i>
Esporte	Debate/Entrevista	A Praça é nossa Noturno	Mulheres em Foco
Minissérie	Educativo	CSI Miami	Chaves
Informação	Reality Show	Globo Universidade	O melhor do Brasil
Debate/Entrevista	Filme	Mulheres em Foco	CSI Miami

A Tabela 10 apresenta um exemplo de comparação para a FH e é constituída pelas colunas principais “FH por categoria” e “FH por nome de programa” e pelas

secundárias “recomendações geradas” e “programas visualizados no dia seguinte”. A categoria ou nome do programa presente em ambas as colunas, indica o acerto e são marcados em cinza. Portanto, na “FH por categoria” obteve-se 3 acertos - Outros, Jornalismo e Sorteio / Televentas / Premiação e na “FBC por nome do programa”, obteve-se 2 acertos - Pânico Reprise e TV Culinária.

**Tabela 10 – Exemplo de comparação para a FH**

<i>FH por categoria</i>		<i>FH por nome do programa</i>	
<i>Recomendações geradas</i>	<i>Programas visualizados no dia seguinte</i>	<i>Recomendações geradas</i>	<i>Programas visualizados no dia seguinte</i>
Outros	Jornalismo	Pânico Reprise	TV Culinária
Jornalismo	Outros	Programa da Tarde	
Erótico	Sorteio/Televentas/Premiação	Siga bem caminhoneiro	
Sorteio/Televentas/Premiação	Infantil	TV Culinária	Pânico Reprise

Também foi calculada a média da porcentagem de acertos para a FH, a FBC e a FC, como indica a Equação 4.

$$m\acute{e}d\grave{a} = \frac{\sum(\% \text{ acerto})}{n\acute{u}m\acute{e}r\text{o de dias}} = \frac{\sum\left(\frac{n\acute{u}m\acute{e}r\text{o de acertos}}{n\acute{u}m\acute{e}r\text{o de recomenda\c{c}o\~{e}s}\right)}{n\acute{u}m\acute{e}r\text{o de dias}}$$

**Equação 4 – Cálculo da média de acertos**

Os testes foram repetidos para três situações. Uma em que o sistema foi configurado para gerar 20 recomendações, 10 recomendações e 5 recomendações sendo que o percentual de acerto foram comparados. Nas tabelas foi utilizado o termo máxima para 20 recomendações, média para 10 e mínima para 5. Dessa forma, os resultados gerados proporcionaram uma melhor avaliação das listas nas três situações para os 15 dias e os 15 usuários.

### **Filtragem Colaborativa: Atributo Nome do Programa (FC\_NP)**

A Tabela 11 apresenta os resultados obtidos, por dia e por usuário, para a análise da lista de recomendação FC aplicada ao nome do programa. A tabela é

composta por duas colunas principais, uma classificada por Dia e outra por Usuário. As colunas secundárias correspondem ao número de dias, ao número de usuários e à porcentagem de acertos correspondente a cada uma das configurações. Pela tabela observa-se que o sistema acertou em média 2% com a configuração máxima, 31% com a configuração média e 62% com a mínima. As Figuras 19 e 20 apresentam os gráficos gerados a partir dessa tabela.

**Tabela 11 – Porcentagem de acertos FC\_NP**

<i>Dia</i>				<i>Usuário</i>			
<i>Dia</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>	<i>Usuário</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1	18%	16%	32%	1	2%	29%	59%
2	6%	25%	49%	2	2%	26%	53%
3	3%	35%	71%	3	2%	48%	96%
4	1%	36%	72%	4	1%	32%	64%
5	0%	15%	31%	5	4%	61%	100%
6	0%	4%	8%	6	3%	20%	40%
7	0%	49%	97%	7	0%	14%	27%
8	0%	40%	80%	8	1%	6%	13%
9	0%	33%	67%	9	3%	9%	18%
10	0%	47%	95%	10	2%	48%	97%
11	0%	61%	100%	11	0%	22%	44%
12	0%	23%	47%	12	2%	30%	60%
13	0%	4%	8%	13	3%	46%	93%
14	0%	45%	89%	14	5%	15%	30%
				15	3%	60%	100%
<i>Média</i>	2%	31%	62%	<i>Média</i>	2%	31%	62%
<i>Máximo</i>	18%	61%	100%	<i>Máximo</i>	5%	61%	100%
<i>Mínimo</i>	0%	4%	8%	<i>Mínimo</i>	0%	6%	13%

A Figura 19 ilustra o gráfico da porcentagem de acertos por dia da lista de recomendação FC aplicada ao nome do programa. Observa-se que nos dias 7 e 11, tanto para a configuração mínima quanto para a média, os valores são altos - 97% e 100%, enquanto para a configuração máxima, o valor mais alto é de 18% no primeiro dia. Nos dias 6 e 13 os valores são baixos para as três configurações, variando de 0% a 8%.

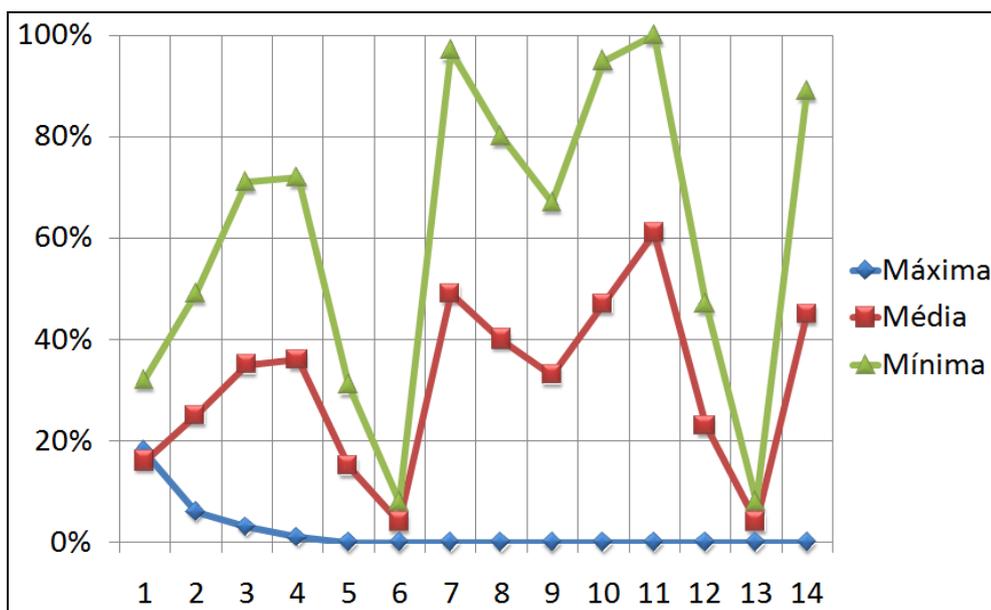


Figura 19 – Gráfico FC\_NP percentagem de acertos por dia

A Figura 20 ilustra o gráfico da percentagem de acertos por usuário da lista de recomendação FC aplicada ao nome do programa. Observa-se que a configuração média e mínima apresentam comportamento muito semelhante, atingindo 100% de acerto nos dias 5 e 15, enquanto na configuração máxima, os valores são muito próximos a 0%.

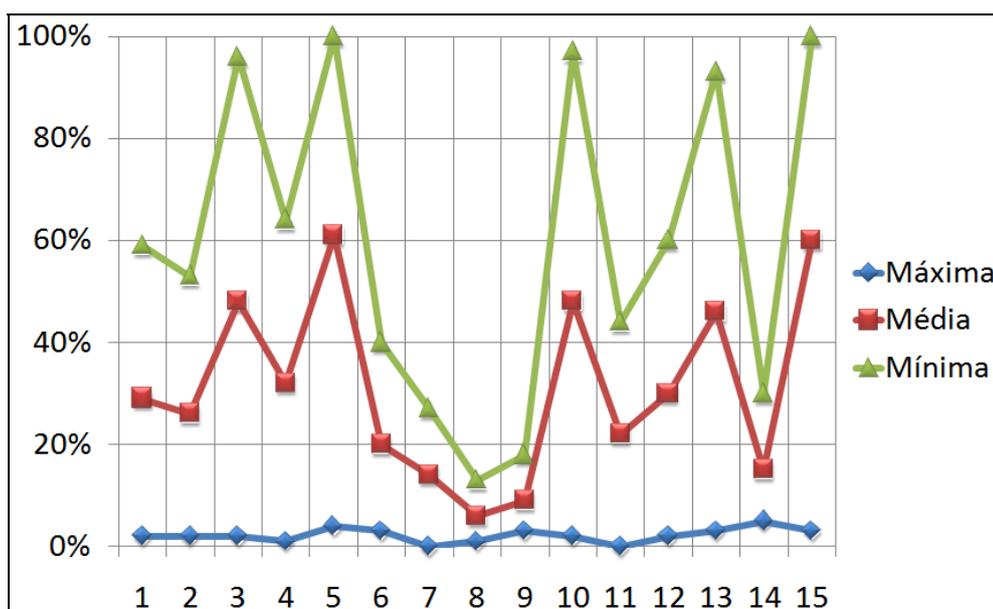


Figura 20 – Gráfico FC\_NP percentagem de acertos por usuário

## Filtragem Colaborativa: Atributo Categoria (FC\_C)

A Tabela 12 apresenta os resultados obtidos por dia e por usuário para a análise da lista de recomendação FC aplicada à categoria. A tabela é composta por duas colunas principais, uma classificada por Dia e outra classificada por Usuário. As colunas secundárias correspondem ao número de dias, ao número de usuários e à porcentagem de acertos correspondente a cada uma das configurações. Pela tabela observa-se que o sistema acertou em média 3% com a configuração máxima, 33% com a configuração média e 65% com a mínima. As Figuras 21 e 22 apresentam os gráficos gerados a partir dessa tabela.

**Tabela 12 – Porcentagem de acertos FC\_C**

<i>Dia</i>				<i>Usuário</i>			
<i>Dia</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>	<i>Usuário</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1	28%	21%	43%	1	3%	30%	60%
2	7%	29%	57%	2	0%	22%	44%
3	3%	33%	65%	3	2%	43%	86%
4	0%	0%	0%	4	1%	35%	70%
5	0%	31%	63%	5	5%	59%	100%
6	0%	33%	67%	6	3%	28%	56%
7	0%	41%	81%	7	0%	21%	43%
8	0%	31%	63%	8	3%	12%	23%
9	0%	35%	71%	9	3%	16%	33%
10	0%	47%	95%	10	4%	51%	100%
11	0%	47%	93%	11	3%	31%	61%
12	0%	38%	76%	12	5%	37%	74%
13	0%	23%	45%	13	3%	44%	87%
14	0%	46%	92%	14	5%	21%	43%
				15	6%	45%	90%
<i>Média</i>	3%	33%	65%	<i>Média</i>	3%	33%	66%
<i>Máximo</i>	28%	47%	95%	<i>Máximo</i>	5%	59%	100%
<i>Mínimo</i>	0%	0%	0%	<i>Mínimo</i>	0%	12%	23%

A Figura 21 ilustra o gráfico da porcentagem de acertos por dia da lista de recomendação FC aplicada à categoria. Observa-se que a configuração máxima mantém-se em 0% por vários dias. No dia 10, a configuração mínima e a média apresentam a maior porcentagem de acerto - 95% e no dia 4, a menor - 0%.

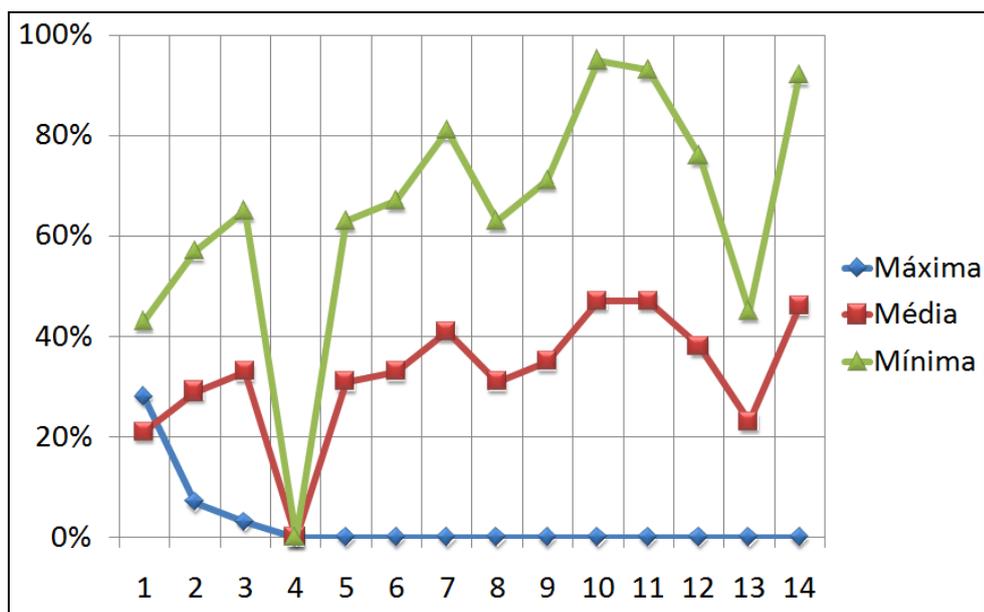


Figura 21 – Gráfico FC\_C percentagem de acertos por dia

A Figura 22 exibe o gráfico da percentagem de acertos por usuário da lista de recomendação da FC aplicada à categoria. Novamente, a configuração máxima apresenta comportamento próximo a 0%, enquanto as configurações média e mínima se assemelham; nos dias 5 e 10 atingem o valor máximo - 100%, e nos dias 2 e 7, o valor mais baixo - 0%.

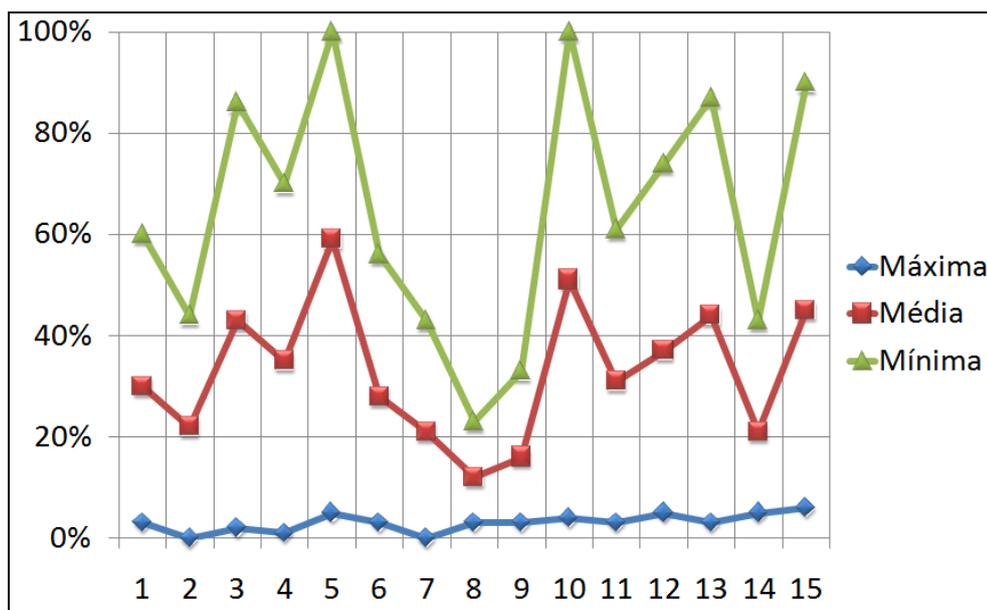


Figura 22 – Gráfico FC\_C percentagem de acertos por usuário

## Filtragem Baseada em Conteúdo: Atributo Nome do Programa (FBC\_NP)

A Tabela 13 apresenta os resultados obtidos por dia e por usuário para a análise da lista de recomendação da FBC aplicada ao nome do programa. A tabela é composta por duas colunas principais, uma classificada por Dia e outra classificada por Usuário. As colunas secundárias correspondem ao número de dias, ao número de usuários e à porcentagem de acertos para cada uma das configurações. Pela tabela observa-se que o sistema acertou em média 7% com a configuração máxima, 12% com a configuração média e 23% com a mínima. As Figuras 23 e 24 apresentam os gráficos gerados a partir dessa tabela.

**Tabela 13 – Porcentagem de acertos FBC\_NP**

<i>Dia</i>				<i>Usuário</i>			
<i>Dia</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>	<i>Usuário</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1	10%	15%	31%	1	3%	4%	9%
2	10%	17%	33%	2	2%	3%	6%
3	7%	11%	23%	3	9%	15%	31%
4	7%	13%	25%	4	6%	10%	20%
5	5%	8%	16%	5	11%	19%	37%
6	1%	1%	3%	6	6%	11%	21%
7	11%	18%	36%	7	3%	6%	11%
8	8%	14%	29%	8	7%	11%	21%
9	6%	11%	21%	9	2%	3%	61%
10	10%	16%	33%	10	9%	16%	31%
11	9%	15%	29%	11	8%	14%	29%
12	7%	13%	25%	12	11%	18%	36%
13	0%	0%	0%	13	9%	15%	30%
14	5%	9%	19%	14	6%	10%	20%
				15	12%	20%	40%
<i>Média</i>	<i>7%</i>	<i>12%</i>	<i>23%</i>	<i>Média</i>	<i>7%</i>	<i>12%</i>	<i>27%</i>
<i>Máximo</i>	<i>11%</i>	<i>18%</i>	<i>36%</i>	<i>Máximo</i>	<i>12%</i>	<i>20%</i>	<i>61%</i>
<i>Mínimo</i>	<i>0%</i>	<i>0%</i>	<i>0%</i>	<i>Mínimo</i>	<i>2%</i>	<i>3%</i>	<i>6%</i>

A Figura 23 exibe o gráfico da porcentagem de acertos por dia da lista de recomendação FBC aplicada ao nome do programa. As 3 configurações apresentam, em média, valores abaixo de 40%, atingindo 0% no dia 13. O valor mais

alto ocorre no dia 7 para as três configurações: máxima 11%, média 18% e mínima 36%. Em geral, as três configurações apresentam comportamento semelhante.

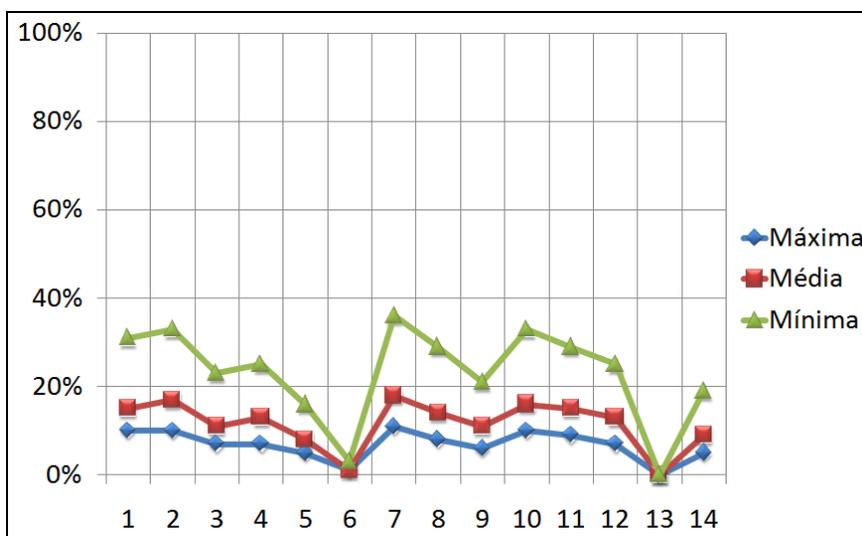


Figura 23 – Gráfico FBC\_NP porcentagem de acertos por dia

A Figura 24 ilustra o gráfico da porcentagem de acertos por usuário da lista de recomendação FBC aplicada ao nome do programa. No dia 9, a configuração mínima atinge 60%, o maior valor do gráfico. O menor valor para essa configuração ocorre no dia 2. Os dias 5 e 15 apresentam os maiores valores para as configurações máxima e média - 11% e 20% e os dias 2 e 9 apresentam os menores - 2%.

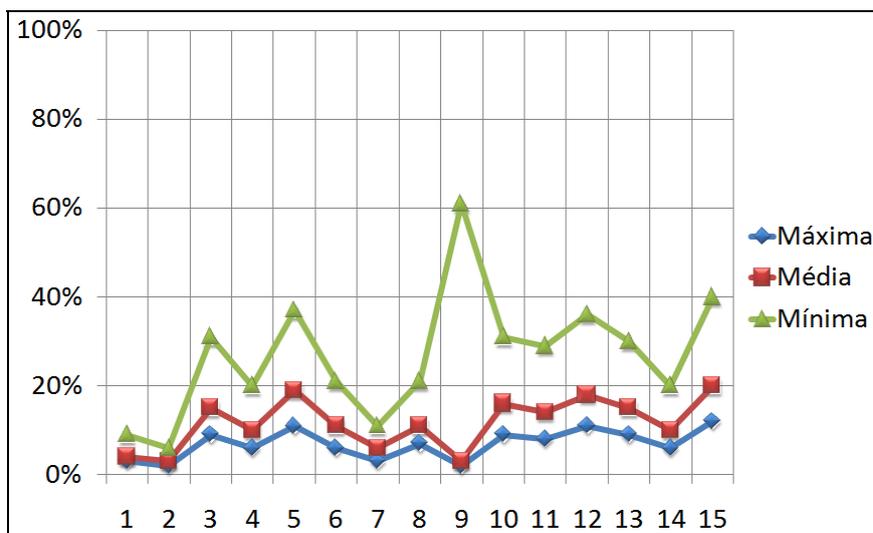


Figura 24 – Gráfico FBC\_NP porcentagem de acertos por usuário

## Filtragem Baseada em Conteúdo: Atributo Categoria (FBC\_C)

A Tabela 14 apresenta os resultados obtidos por dia e por usuário para a análise da lista de recomendação da FBC aplicada à categoria. A tabela é composta por duas colunas principais, uma classificada por Dia e outra classificada por Usuário. As colunas secundárias correspondem ao número de dias, ao número de usuários e à porcentagem de acertos para cada uma das configurações. Pela tabela, observa-se que o sistema acertou em média 5% com a configuração máxima, 8% com a configuração média e 17% com a mínima. As Figuras 25 e 26 apresentam os gráficos gerados a partir dessa tabela.

**Tabela 14 – Porcentagem de acertos FBC\_C**

<i>Dia</i>				<i>Usuário</i>			
<i>Dia</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>	<i>Usuário</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1	4%	7%	13%	1	3%	6%	11%
2	7%	11%	23%	2	3%	5%	10%
3	6%	10%	20%	3	7%	12%	24%
4	0%	0%	0%	4	4%	7%	14%
5	5%	9%	17%	5	8%	14%	27%
6	4%	7%	13%	6	4%	7%	14%
7	8%	13%	27%	7	2%	3%	6%
8	4%	7%	15%	8	2%	4%	7%
9	8%	9%	19%	9	2%	4%	7%
10	7%	12%	24%	10	9%	15%	30%
11	7%	11%	23%	11	6%	9%	19%
12	5%	9%	17%	12	5%	8%	16%
13	1%	2%	4%	13	7%	12%	24%
14	5%	9%	19%	14	4%	6%	13%
				15	8%	14%	27%
<i>Média</i>	5%	8%	17%	<i>Média</i>	5%	8%	17%
<i>Máximo</i>	8%	13%	27%	<i>Máximo</i>	9%	15%	30%
<i>Mínimo</i>	0%	0%	0%	<i>Mínimo</i>	2%	3%	6%

A Figura 25 apresenta o gráfico da porcentagem de acertos por dia da lista de recomendação FBC aplicada à categoria. As três configurações apresentam a porcentagem de acertos abaixo de 40%. A configuração mínima teve sua melhor taxa de acertos no dia 7 - 27%, a configuração média no dia 2 e máxima no dia 7,

8% e 13%. O dia 4 é o que apresenta a taxa de acertos mais baixa para as três configurações - 0%.

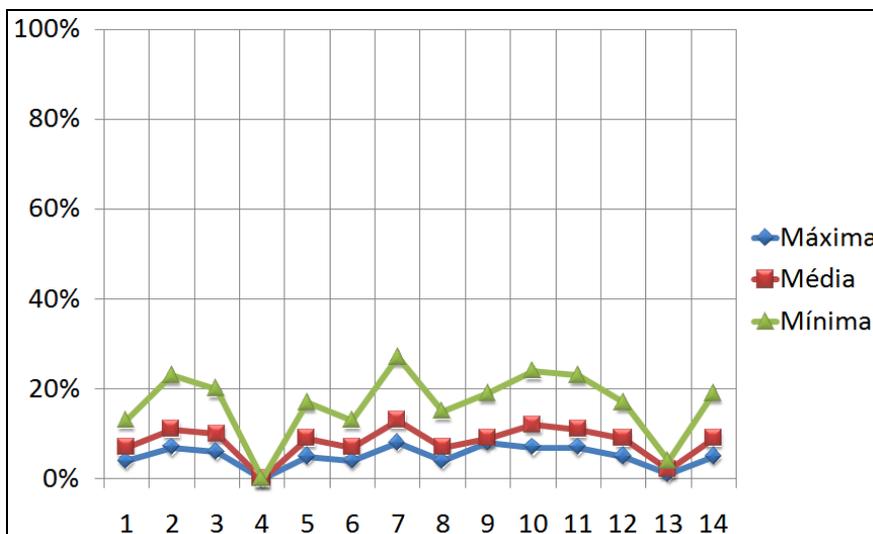


Figura 25 – Gráfico FBC\_C porcentagem de acertos por dia

A Figura 26 apresenta o gráfico da porcentagem de acertos por usuário da lista de recomendação FBC aplicada à categoria. As três configurações apresentam valores abaixo de 40%. Para a configuração máxima, o valor mais alto é 9% no dia 10 e o mais baixo é 2% nos dias 7, 8 e 9. Para a configuração média, o valor mais alto é 15% no dia 10 e o mais baixo é 3% no dia 7. Para a mínima, o valor mais alto é 30% no dia 10 e o mais baixo é 6% no dia 7.

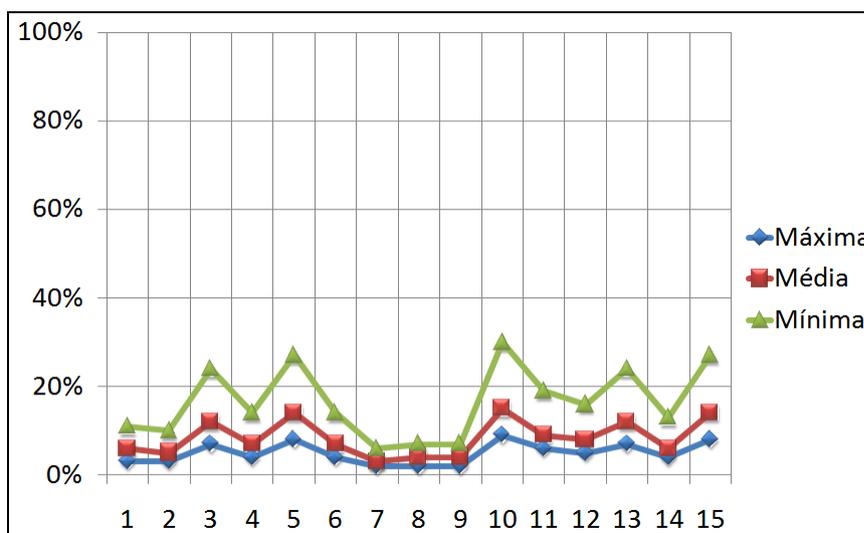


Figura 26 – Gráfico FBC\_C porcentagem de acertos por usuário

## Filtragem Híbrida: Atributo Nome do Programa (FH\_NP)

A Tabela 15 apresenta os resultados obtidos por dia e por usuário para a análise da lista de recomendação da FBC aplicada à categoria. A tabela é composta por duas colunas principais, uma classificada por Dia e outra classificada por Usuário. As colunas secundárias correspondem ao número de dias, ao número de usuários e à porcentagem de acertos para cada uma das configurações. Pela tabela, observa-se que o sistema acertou em média 8% com a configuração máxima, 13% com a média e 27% com a mínima. As Figuras 27 e 28 apresentam os gráficos gerados a partir dessa tabela.

**Tabela 15 – Porcentagem de acertos FH\_NP**

<i>Dia</i>				<i>Usuário</i>			
<i>Dia</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>	<i>Usuário</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1	10%	15%	31%	1	3%	4%	9%
2	11%	19%	38%	2	2%	3%	6%
3	8%	13%	26%	3	8%	14%	29%
4	10%	17%	35%	4	6%	10%	20%
5	5%	9%	17%	5	11%	19%	37%
6	6%	10%	20%	6	6%	11%	21%
7	12%	21%	42%	7	3%	6%	11%
8	9%	15%	31%	8	5%	9%	17%
9	7%	12%	25%	9	2%	3%	6%
10	9%	15%	31%	10	9%	16%	31%
11	9%	15%	29%	11	8%	14%	29%
12	9%	16%	32%	12	11%	18%	36%
13	0%	0%	0%	13	9%	15%	30%
14	5%	9%	19%	14	6%	10%	20%
				15	12%	20%	40%
<i>Média</i>	8%	13%	27%	<i>Média</i>	7%	11%	23%
<i>Máximo</i>	12%	21%	42%	<i>Máximo</i>	12%	20%	40%
<i>Mínimo</i>	0%	0%	0%	<i>Mínimo</i>	2%	3%	6%

A Figura 27 apresenta o gráfico da porcentagem de acertos por dia da lista de recomendação híbrida aplicada ao nome do programa. Observa-se, pelo gráfico, que a porcentagem de acertos fica abaixo dos 40%. O valor mínimo é atingido no dia 13,

para as três configurações - 0%. Os valores máximos são atingidos no dia 7 - 12%, 21% e 42%, respectivamente, para as configurações máxima, média e mínima.

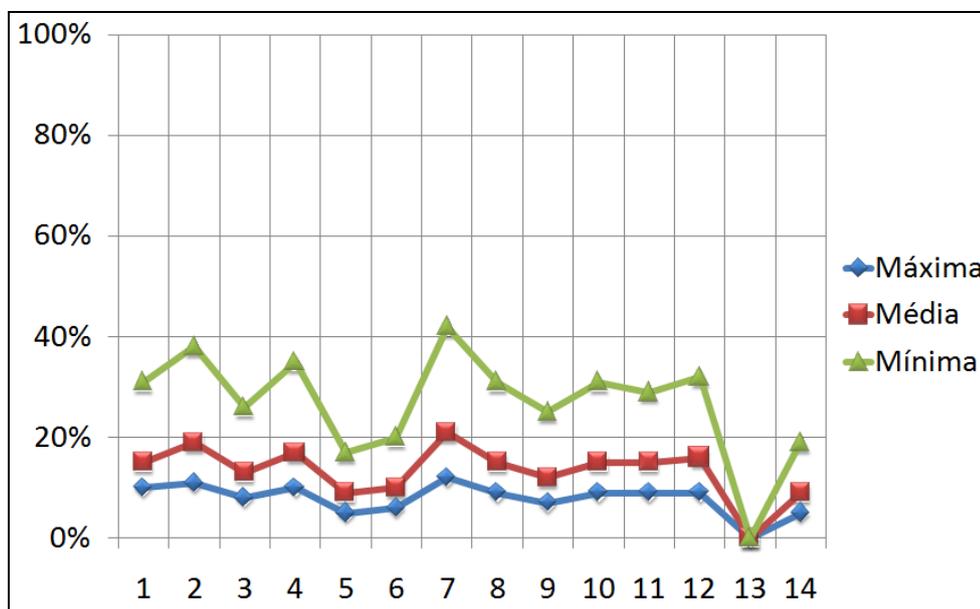


Figura 27 – Gráfico FH\_NP porcentagem de acertos por dia

A Figura 28 apresenta o gráfico da porcentagem de acertos por usuário da lista de recomendação híbrida aplicada ao nome do programa. O menor valor, próximo a 0%, ocorre no dia 9 para a configuração máxima e média e no dia 2, para as três. Os maiores valores são atingidos no dia 15 - 12%, 20% e 40%, respectivamente, para as configurações máxima, média e mínima.

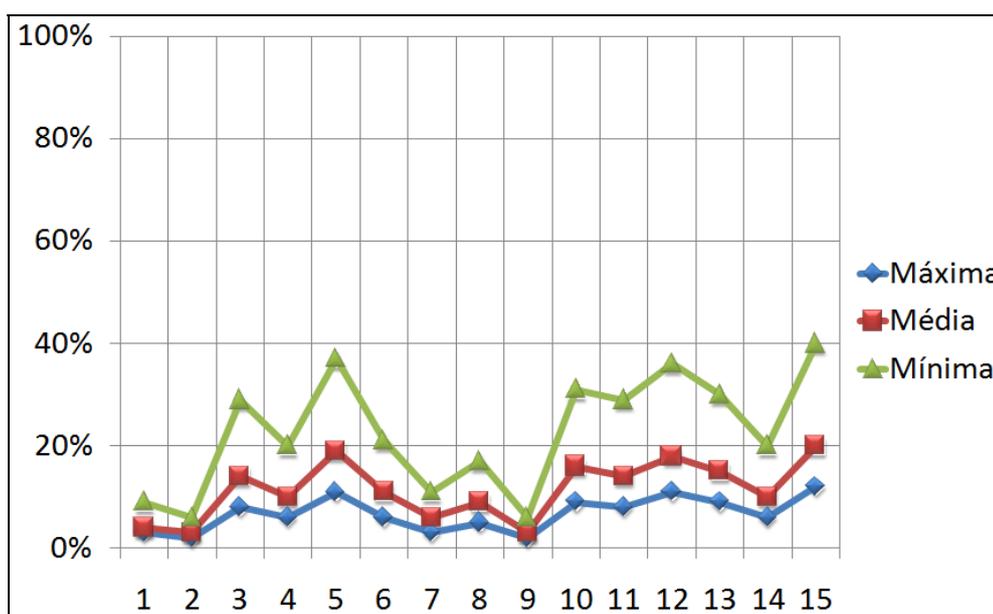


Figura 28 – Gráfico FH\_NP porcentagem de acertos por usuário

## Filtragem Híbrida: Atributo Categoria (FH\_C)

A Tabela 16 apresenta os resultados obtidos por dia e por usuário para a análise da lista de recomendação da FBC aplicada à categoria. A tabela é composta por duas colunas principais, uma classificada por Dia e outra classificada por Usuário. As colunas secundárias correspondem ao número de dias, ao número de usuários e à porcentagem de acertos para cada uma das configurações. Pela tabela, observa-se que o sistema acertou em média 5% com a configuração máxima, 8% com a média e 17% com a mínima. As Figuras 29 e 30 apresentam os gráficos gerados a partir dessa tabela.

**Tabela 16 – Porcentagem de acertos FH\_C**

<i>Dia</i>				<i>Usuário</i>			
<i>Dia</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>	<i>Usuário</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1	4%	7%	13%	1	3%	6%	11%
2	7%	11%	23%	2	3%	5%	10%
3	6%	10%	20%	3	7%	12%	24%
4	0%	0%	0%	4	4%	7%	14%
5	5%	9%	9%	5	8%	14%	27%
6	4%	7%	7%	6	4%	7%	13%
7	8%	13%	13%	7	2%	3%	6%
8	4%	7%	7%	8	2%	4%	8%
9	5%	9%	9%	9	2%	4%	7%
10	7%	12%	12%	10	9%	15%	30%
11	7%	11%	11%	11	6%	9%	19%
12	5%	9%	9%	12	5%	8%	16%
13	1%	2%	2%	13	7%	12%	24%
14	5%	9%	9%	14	4%	6%	13%
				15	8%	14%	27%
<i>Média</i>	5%	8%	10%	<i>Média</i>	5%	8%	17%
<i>Máximo</i>	8%	13%	23%	<i>Máximo</i>	9%	15%	30%
<i>Mínimo</i>	0%	0%	0%	<i>Mínimo</i>	2%	3%	6%

A Figura 29 apresenta o gráfico da porcentagem de acerto por dia da lista de recomendação da FH aplicada à categoria. As configurações máxima e média apresentam seus valores mais altos no dia 7 - 8% e 13%, respectivamente, enquanto a configuração mínima atinge 23% no dia 2. As três configurações apresentam os seus valores mais baixos no dia 4 - 0%.

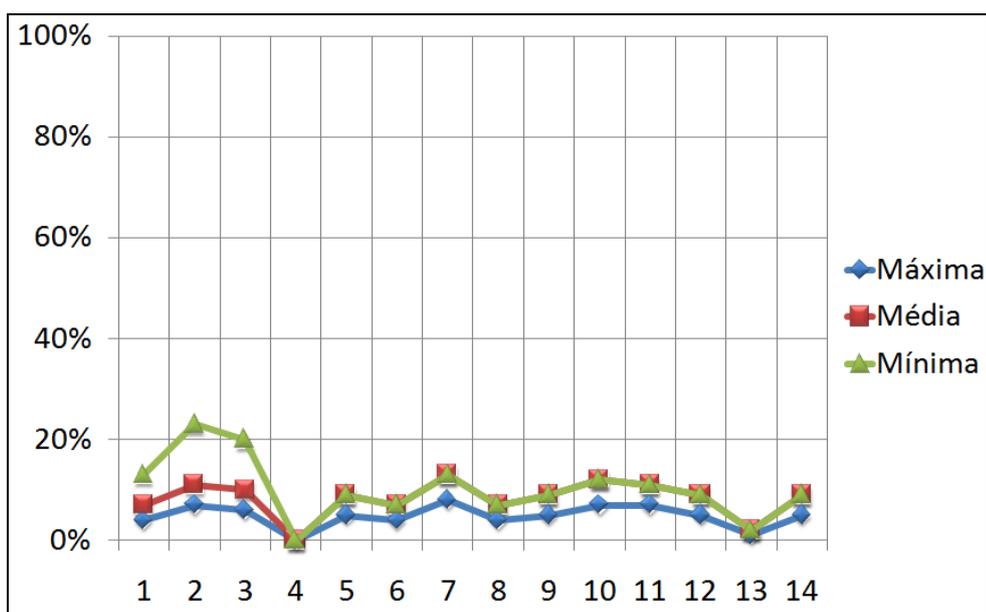


Figura 29 – Gráfico FH\_C porcentagem de acertos por dia

A Figura 29 apresenta o gráfico da porcentagem de acertos por usuário da lista de recomendação da FH aplicada à categoria. As configurações média e mínima apresentam os seus valores mais baixos no dia 7 - média 3% e mínima 6%, enquanto a configuração máxima apresenta o valor mais baixo nos dias 7, 8 e 9 – 2%. Os valores mais altos são 9%, 15% e 30%, respectivamente, para a máxima, a média e a mínima, no dia 10.

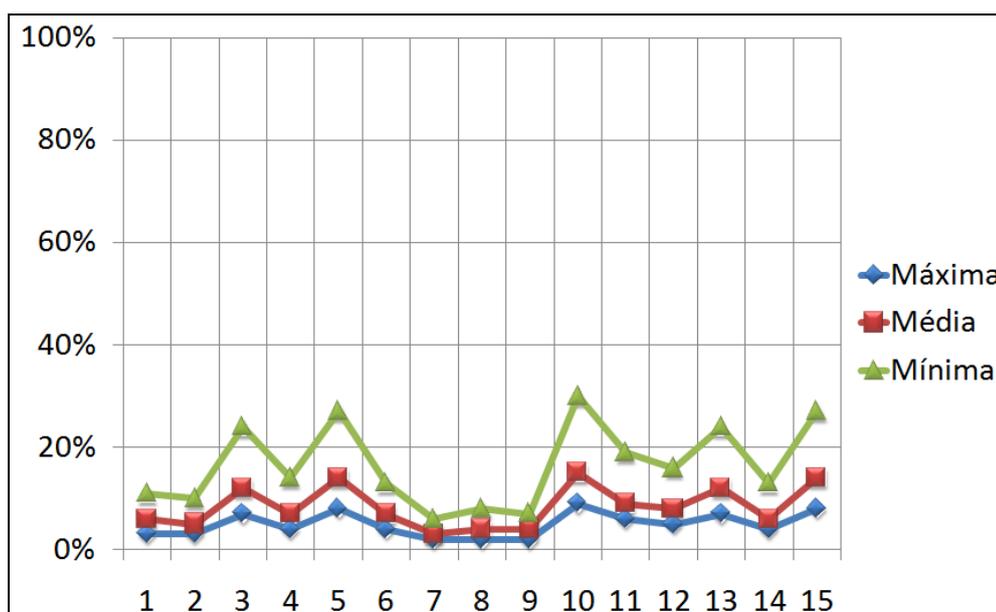


Figura 30 – Gráfico FH\_C porcentagem de acertos por usuário

## Médias

As Tabelas 17 e 18 apresentam as médias obtidas para as seis listas, sendo que a primeira delas está organizada por dia e a segunda, por usuário. Pela Tabela 17, observa-se que, com a configuração mínima, a porcentagem de acertos foi de 65%, enquanto que, com a configuração máxima, ela atingiu 8%. A configuração média atingiu no máximo 33%. A média da porcentagem de acertos por dia para as três configurações atingiu 5%, 18% e 34%, respectivamente. Os menores valores foram o de 2% para a configuração máxima, o de 8% para a média e o de 10% para a mínima.

**Tabela 17 – Média da porcentagem de acertos por dia**

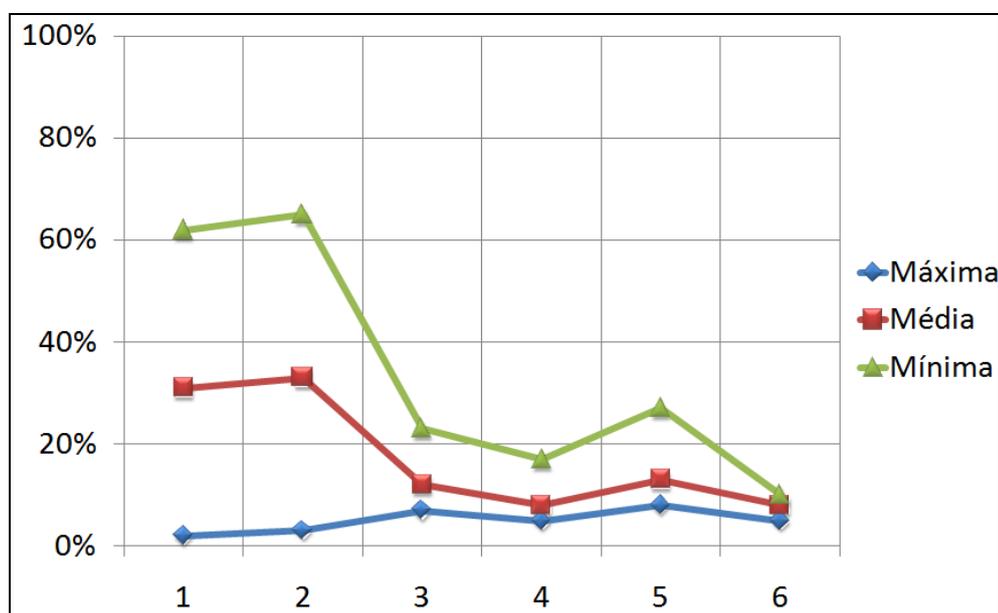
<i>Dia</i>			
<i>Lista</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1 – FC – Nome Programa	2%	31%	62%
2 – FC – Categoria	3%	33%	65%
3 – FBC – Nome Programa	7%	12%	23%
4 – FBC – Categoria	5%	8%	17%
5 – FH – Nome Programa	8%	13%	27%
6 – FH – Categoria	5%	8%	10%
<i>Média</i>	5%	18%	34%
<i>Máximo</i>	8%	33%	65%
<i>Mínimo</i>	2%	8%	10%

Pela Tabela 18, observa-se que, com a configuração mínima, a porcentagem de acertos foi de 66%, enquanto que, com a configuração máxima, ela atingiu 7%. A configuração média atingiu no máximo 33%. A média para as três configurações atingiu 5%, 18% e 35%, respectivamente. Os menores valores foram o de 2% para a configuração máxima, 8% para a média e 17% para a mínima. As Figuras 31 e 32 apresentam os gráficos gerados a partir dessas tabelas.

**Tabela 18 – Média da porcentagem de acertos por usuário**

<i>Usuário</i>			
<i>Lista</i>	<i>Máxima</i>	<i>Média</i>	<i>Mínima</i>
1 – FC – Nome Programa	2%	31%	62%
2 – FC – Categoria	3%	33%	66%
3 – FBC – Nome Programa	7%	12%	27%
4 – FBC – Categoria	5%	8%	17%
5 – FH – Nome Programa	7%	11%	23%
6 – FH – Categoria	5%	8%	17%
<i>Média</i>	5%	17%	35%
<i>Máximo</i>	7%	33%	66%
<i>Mínimo</i>	2%	8%	17%

A Figura 31 apresenta o gráfico da média da porcentagem de acertos por dia para as seis listas. A configuração mínima obteve as melhores porcentagens de acerto e a máxima, as piores, os quais ficaram abaixo dos 20%. Para as três configurações, as listas 1 e 2 obtiveram os melhores resultados.



**Figura 31 – Gráfico: Média da porcentagem de acertos por dia**

A Figura 32 apresenta o gráfico da média da porcentagem de acertos por usuário para as seis listas. Para as configurações média e mínima, os maiores valores estão indicados na linha 2 – FC – Categoria, 33% e 66%, enquanto que,

para a configuração máxima, o maior valor está indicado na linha 3 – FBC – Nome Programa, 7%.

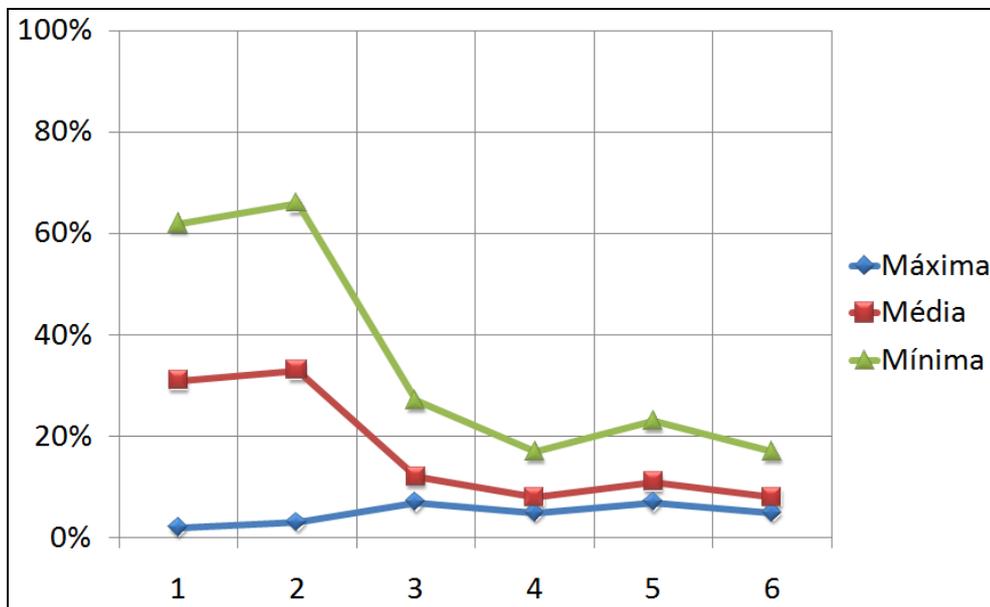


Figura 32 – Gráfico: Média da porcentagem de acertos por usuário

## Comparativos

A Tabela 19 apresenta a média da porcentagem de acertos por nome do programa e é dividida por dia e por usuário.

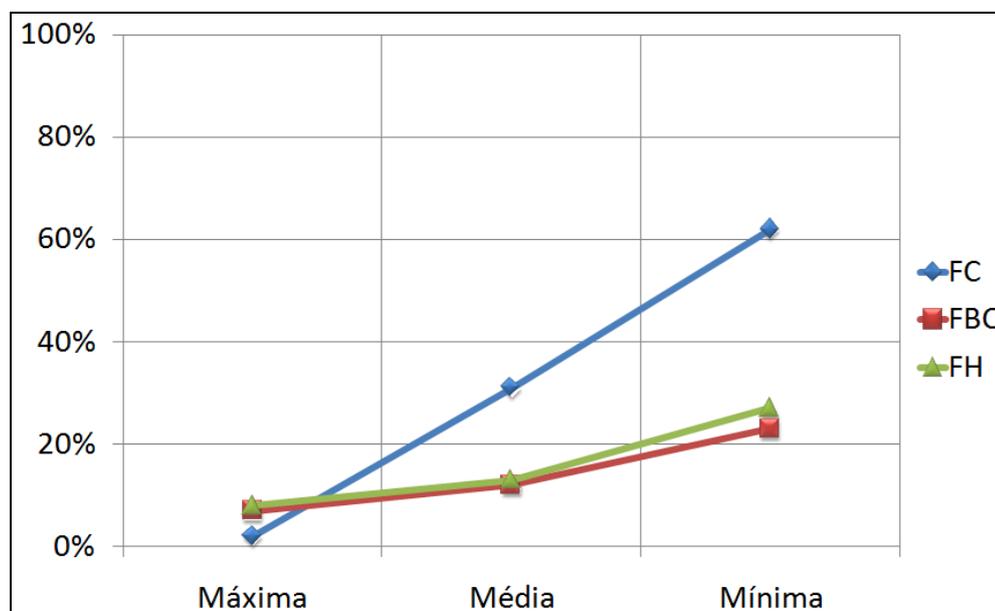
Tabela 19 – Média da porcentagem de acertos por nome do programa

Dia				Usuário			
Lista	Máxima	Média	Mínima	Lista	Máxima	Média	Mínima
1 – FC	2%	31%	62%	1 – FC	2%	31%	62%
2 – FBC	7%	12%	23%	2 – FBC	7%	12%	27%
3 – FH	8%	13%	27%	3 – FH	7%	11%	23%
<i>Média</i>	6%	19%	37%	<i>Média</i>	5%	17%	35%
<i>Máximo</i>	8%	31%	62%	<i>Máximo</i>	7%	31%	62%
<i>Mínimo</i>	2%	12%	23%	<i>Mínimo</i>	2%	11%	23%

**Tabela 20 – Média da porcentagem de acertos por categoria**

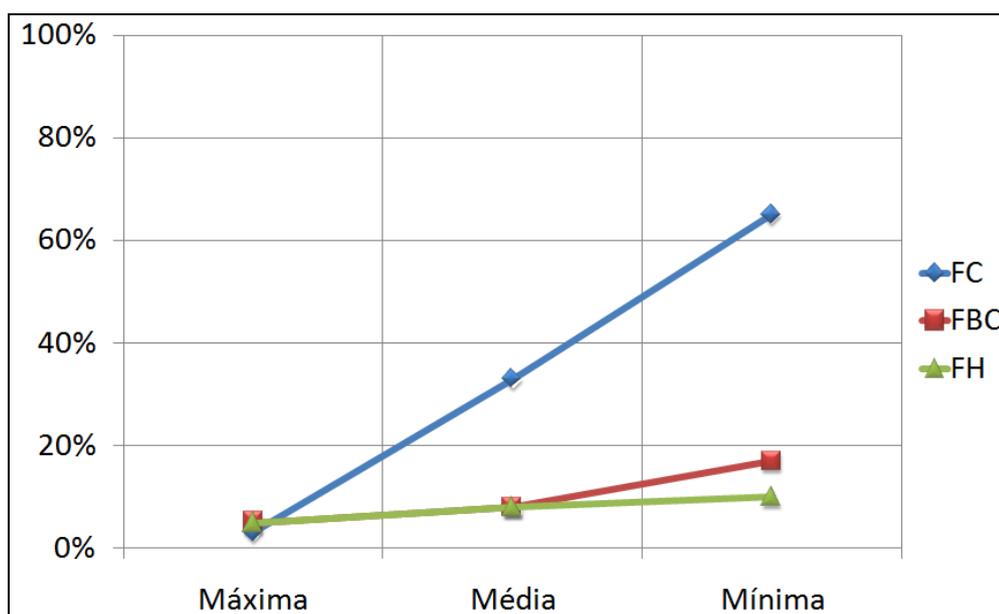
Dia				Usuário			
Lista	Máxima	Média	Mínima	Lista	Máxima	Média	Mínima
1 – FC	3%	33%	65%	1 – FC	3%	33%	66%
2 – FBC	5%	8%	17%	2 – FBC	5%	8%	17%
3 – FH	5%	8%	10%	3 – FH	5%	8%	17%
Média	4%	16%	31%	Média	4%	16%	33%
Máximo	5%	33%	65%	Máximo	5%	33%	66%
Mínimo	3%	8%	10%	Mínimo	3%	8%	17%

A Tabela 20 apresenta a média da porcentagem de acertos por categoria e é dividida por dia e por usuário.



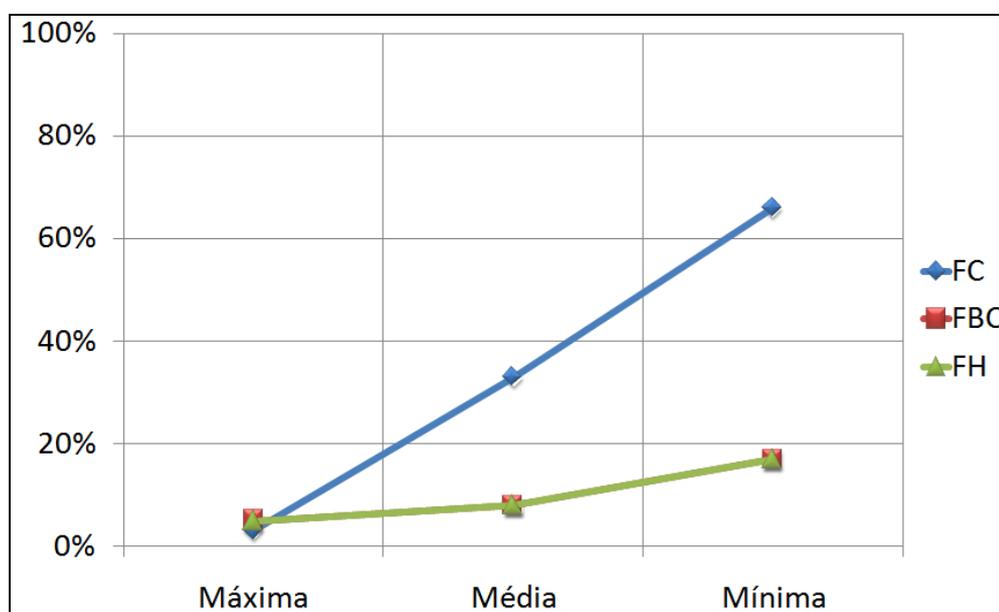
**Figura 33 – Gráfico: comparação do nome do programa por dia**

A Figura 33 apresenta o gráfico da comparação entre as listas FC, FBC e FH para a média da porcentagem de acertos por nome do programa, classificada por dia.



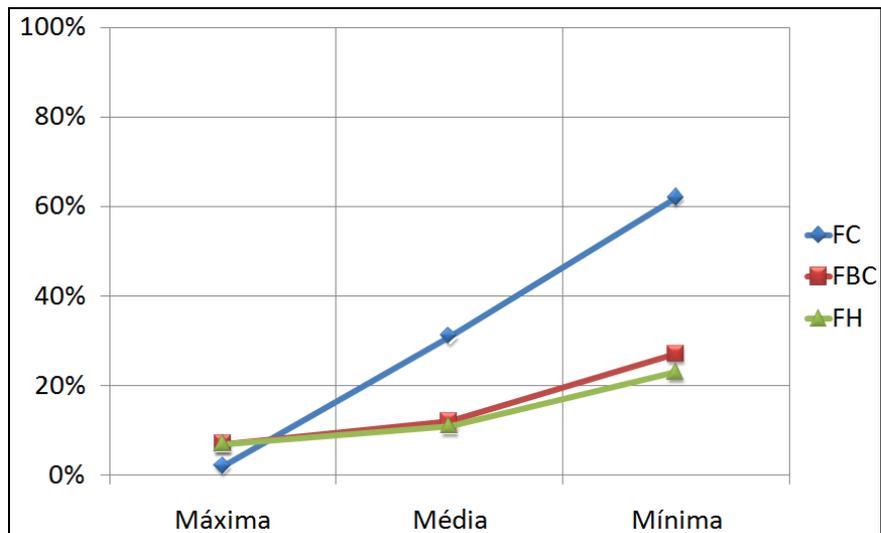
**Figura 34 – Gráfico: comparação da categoria por dia**

A Figura 34 apresenta o gráfico da comparação entre as listas FC, FBC e FH para a média da porcentagem de acertos por nome do programa, classificada por usuário.



**Figura 35 – Gráfico: comparação do nome do programa por usuário**

A Figura 35 apresenta o gráfico da comparação entre as listas FC, FBC e FH para a média porcentagem de acertos por categoria, classificada por dia.



**Figura 36 – Gráfico: comparação da categoria por usuário**

A Figura 36 apresenta o gráfico da comparação entre as listas FC, FBC e FH para a média porcentagem de acertos por categoria, classificada por usuário.

### 5.3 Observações

Este capítulo apresentou a descrição do sistema de recomendação BiPODiTVR, explicitando detalhes de sua arquitetura e de sua implementação. O BiPODiTVR é um sistema desenvolvido para o contexto da TVDPI brasileira e é dividido em duas partes, cliente e servidor. No desenvolvimento da parte cliente, foram utilizadas tecnologia nacional, uma máquina virtual com a implementação de referência do *middleware* Ginga-NCL e suas respectivas linguagens de programação, NCL e LUA. Já na parte servidor, foram utilizados Java, JPA, MySQL e Tomcat.

O BiPODiTVR é ainda um sistema de recomendação híbrido que utiliza FBC e FC para gerar as recomendações. Além disso, usa as técnicas do Cosseno e dos Vizinhos mais próximos para encontrar os itens similares ao perfil do usuário e os itens similares, de outros usuários, a um usuário alvo.

Foram observadas algumas peculiaridades nos gráficos gerados e uma análise foi feita a respeito delas. BiPODiTVr utiliza o histórico do usuário e o EPG

para gerar as recomendações, além dos parâmetros de dia e hora de solicitação. Se o usuário permanecer um tempo, um mês por exemplo, sem utilizar o sistema, quando voltar a utilizá-lo, o sistema utilizará o histórico armazenado, de um mês atrás, para gerar as recomendações. O sistema pode até não acertar logo no primeiro dia de uso, pois estará recomendando programas de acordo com o perfil passado do usuário. Entretanto, as preferências do usuário com o decorrer do tempo podem mudar, mesmo se o usuário utilizar o sistema com frequência. Por este motivo, o sistema descobre as preferências do usuário e gera os perfis. A partir do momento que o usuário volta a utilizar o sistema, este descobrirá quais são as novas preferências do usuário e começará a gerar recomendações de acordo com elas. Os perfis antigos são importantes para avaliar a evolução e mudanças de comportamento e preferências dos usuários.

Programas que não são transmitidos todos os dias não influenciam na geração das recomendações, justamente pelo emprego dos parâmetros no momento da geração dos perfis e das recomendações. BiPODiTVR existe para auxiliar o usuário de TVD Portátil e permitir que ele usufrua da melhor forma o seu tempo de visualização de programas. É por este motivo que BiPODiTVR utiliza como parâmetros o dia e o horário de solicitação e busca no EPG somente os programas pertinentes ao que foi solicitado – programas que estejam sendo transmitidos naquele dia com uma hora de antecedência ou após uma hora do horário de solicitação.

O cenário em que os testes foram aplicados é específico. Considerou-se que os usuários assistiram à TV todos os dias, o sistema gerou perfis e, por fim, gerou recomendações em 3 configurações (5, 10 e 20) para 3 tipos diferentes de filtragens (FBC, FC e FH). Em um outro cenário, os resultados poderiam ter sido diferentes, talvez para pior, ou para melhor, dependendo dos parâmetros e atributos adotados. Não foi possível realizar testes em outros cenários, mas estes são inclusos nos trabalhos futuros. Além disso, a forma de avaliação de acertos considera as recomendações geradas para um dia em particular e as visualizações do usuário do dia anterior a este.

O que se observa nos gráficos é que, quando há porcentagens de acerto muito baixas, como ocorre nos dias 4, 6 e 13, em todas as filtragens, o usuário

não assistiu à nenhum – ou um número muito baixo relativo ao número de recomendações geradas – programa ou categoria recomendada. Portanto, das 5, 10 ou 20 recomendações geradas, o usuário não assistiu à nenhuma – ou um número muito baixo relativo ao número de recomendações geradas – o que acarreta esses valores baixos quando aplicado o cálculo da porcentagem de acerto.

Outra peculiaridade foi encontrada nas recomendações geradas para o atributo Nome do Programa. Observando-se os resultados gerados e apresentados nos gráficos foi possível concluir que este tipo de recomendação é muito específica e limitada, pois utiliza e recomenda o nome exato do programa. Apesar de existirem muitos programas, cada um é único no EPG, independente de sua frequência de transmissão, o nome do programa será sempre único. Portanto, se o sistema recomenda 5 nomes de programas e o usuário não assistiu à nenhum daqueles programas específicos, quando aplicado o cálculo da porcentagem de acertos, o resultado será negativo.

Já as recomendações geradas com o atributo categoria abrangem todos os programas que fazem parte daquela categoria e permitem maior número de opções. O usuário pode até não ter assistido àquele programa em particular, mas há probabilidade dele ter assistido à um programa que faz parte de alguma das categorias recomendadas. Supondo que o sistema tenha recomendado 5 programas que pertencem a 3 das categorias preferenciais do usuário, é provável que ele tenha assistido à algum programa dentro dessas categorias – independente de qual seja o nome do programa – em algum momento do seu passado.

Concluindo, a FBC e a FH obtiveram resultados pouco positivos, mostrando que essas filtragens, para o cenário em que foi aplicado, não são totalmente aplicáveis. A FBC, por sua vez, dentro do cenário testado, foi a que obteve melhores resultados e, portanto, poderia ser a única filtragem a ser utilizada pelo sistema.

# 6

## TRABALHOS RELACIONADOS

---

Este capítulo apresenta seis trabalhos relacionados. Os dois primeiros foram desenvolvidos pelo grupo de pesquisa em Personalização e Privacidade na Internet do Departamento de Computação da UFSCar (Universidade Federal de São Carlos) e os demais são relacionados a sistemas de recomendação para TVDPI. Na última seção, apresentam-se as as observações finais.

### 6.1 Recommender TV: Suporte ao Desenvolvimento de Aplicações de Recomendação para o Sistema Brasileiro de TV Digital

No trabalho de Ávila (2010), desenvolveu-se um novo módulo na implementação de referência do *middleware* Ginga, que possibilita suporte ao desenvolvimento de sistemas de recomendação.

O *RecommenderTV* tem como principal objetivo permitir a aquisição de dados através do oferecimento de recursos de mineração de dados e da disponibilização de acesso a essas informações. Assim, os desenvolvedores de sistemas de recomendação podem concentrar seus esforços em detalhes de interface e navegabilidade. Os serviços ofertados pelo módulo *RecommenderTV* são opcionais o que flexibiliza o uso desses serviços para o contexto do sistema proposto.

No *RecommenderTV*, todas as ações do usuário, assim como, horário, dia da semana, provedor de serviço, serviço, gênero e subgênero são persistidas no SQLite<sup>6</sup> – um banco de dados para sistemas embarcados. Também é necessário que o *RecommenderTV* acesse as informações constantes nas tabelas SI (Sistemas de Informação), EIT (Tabela de informação de eventos) e SDT (Tabela de descrição de serviço), fornecidas pelo módulo *Data Processing* do *middleware* Ginga, para auxiliar na geração das recomendações. O *RecommenderTV* também consulta o módulo

---

<sup>6</sup> SQLite<sup>6</sup> é uma biblioteca em linguagem C que implementa um banco de dados SQL embutido. ([www.sqlite.org](http://www.sqlite.org))

*Context Manager* do *middleware* Ginga antes de realizar a mineração de dados, a fim de verificar o clock, a quantidade de memória e o uso da CPU. (ABNT 15603-2, 2009)

O sistema é dividido em *Recommender Engine* e *Recommender API*. O módulo *Recommender Engine* coleta, armazena, minera dados, extrai padrões e cuida da manutenção da base, comunica-se com os módulos do Ginga, é invisível às aplicações de sistemas de recomendação e é dividido em 5 agentes. O *Local agent* monitora e detecta os dispositivos de interação do usuário com o set-top-box, persiste as ações, formata e calcula o tempo de visualização e também gerencia o tamanho da base de dados. O *Scheduler Agent* dispara o processo de mineração sempre que os recursos necessários estiverem disponíveis. Para solicitar uma operação de mineração de dados, o módulo utiliza horário, CPU e memória disponível e para obter tais informações comunica-se com o módulo *Context Manager*. O *Mining agent* escalona quais algoritmos serão executados de acordo com os recursos computacionais, encapsula também os algoritmos de mineração de dados e acessa o histórico de visualização e as tabelas EIT e SDT. O *Filter agent* filtra os dados resultantes dos módulos *Mining Agent* e *Data Agent* e gerencia a inserção e a remoção das recomendações.

O *Recommender API* é visível aos sistemas de recomendação e permite que esses ofereçam serviços personalizados. O módulo executa na camada de apresentação da implementação de referência do *middleware* Ginga. As chamadas aos métodos C++ fornecidos pela *Recommender API* devem ser feitas na linguagem Lua.

A simulação do sistema utilizou um gerador de carrossel, desenvolvido por Ávila, que transmitiu os dados através da recepção dos fluxos de transporte via protocolo *User Datagram Protocol* (UDP). No lado do telespectador, o set-top-box foi simulado por um MiniPC que executava a implementação de referência do *middleware* Ginga com o módulo *RecommenderTV*. O *middleware* Ginga foi configurado para localizar cada provedor de serviço e um endereço IP foi atribuído a cada um dos provedores que compunham o ambiente de simulação.

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

Lucas (2009) propõe e desenvolve um sistema de recomendação para ambientes multiusuário baseado em um modelo de personalização, o RepTVD. O objetivo principal do sistema consiste em identificar os padrões de comportamento e tendências dos membros de um grupo.

Mediante uma arquitetura residente ao set-top-box, o sistema visa a recomendar conteúdo pela utilização de mineração de dados, de acordo com os padrões de comportamento descobertos no uso da televisão e através da adoção do algoritmo Apriori. O modelo de personalização para TVD pode ser mapeado para utilizar diferentes técnicas ou algoritmos de forma a recomendar conteúdo, definindo as fases necessárias pra isso, assim como a interação dessas. O modelo é composto de oito fases:

Fases 1, 2 e 3: definem as estratégias de como coletar e armazenar as informações;

Fases 4 e 5: definem como as informações devem ser processadas, a fim de descobrir informações úteis do grupo de telespectadores;

Fases 6 e 7: definem como transformar essas informações em recomendações;

Fase 8: define a forma de apresentação das sugestões aos telespectadores.

A arquitetura do RePTVD foi desenvolvida como uma aplicação que permanece residente no STB e é composta por três módulos, descritos a seguir.

O módulo metadados analisa os dados da tabela SI e extrai as informações importantes para o sistema. Os dados coletados são armazenados em arquivos ARFF<sup>7</sup> (*Attribute-Relation File Format*) para serem minerados com a biblioteca

---

<sup>7</sup> Um ARFF é um arquivo de texto ASCII que descreve uma lista de instâncias que compartilham um conjunto de atributos. (<http://www.cs.waikato.ac.nz/~ml/weka/arff.html>)

WEKA<sup>8</sup>. O RePTVD utiliza a tabela SDT, que disponibiliza as informações referentes ao provedor do serviço, a tabela EIT, que contém informações da grade de programação e a tabela TDT, que obtém a data e o horário do sistema.

O módulo perfil do grupo de telespectadores coleta informações de visualização implicitamente, a qual se dá pelo monitoramento da troca de canais e do tempo de visualização. O módulo utiliza o dia, o horário e o período como parâmetros para a previsão e a efetuação da recomendação. O arquivo ARFF armazena as informações dos dois módulos e é atualizado de forma contínua e autônoma. Um segundo arquivo ARFF armazena a categoria, o período do dia e o tempo de visualização e é utilizado para efetuar recomendações em tempo real.

O módulo recomendador tem duas funções. Na primeira, deve-se capturar informações e descobrir padrões/regras que permitam a melhor compreensão dos dados, através da utilização do algoritmo Apriori. Na segunda, usa-se filtragem colaborativa para comparar as preferências descobertas a partir do conteúdo da grade de programação e cria-se uma lista dinâmica de recomendação, processo que ocorre a cada 24h. A lista gerada contém 4 recomendações, sendo duas novas, que seguem o padrão, e duas mais comuns aos telespectadores.

Também são fornecidas 4 recomendações em tempo real. A lista de recomendações é gerada quando a TV é ligada. O sistema verifica data/hora atuais, realiza uma busca no arquivo ARFF e aplica um algoritmo de filtragem que retorna as recomendações pelo período do dia.

Vale ressaltar que o RePTVD não é apenas uma adaptação de um sistema de recomendação da WEB para ser utilizado na TVD, mas também um sistema baseado nas suas características particulares, pelo fato de não estar atrelado a nenhum vídeo e sua existência ser independente dos programas audiovisuais em transmissão.

---

<sup>8</sup> Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. (<http://www.cs.waikato.ac.nz/ml/weka/>)

### 6.3 Content Personalization for Mobile TV Combining Content-Based and Collaborative Filtering

O trabalho desenvolvido por Einarsson (2009) emprega uma combinação das filtragens FBC e FC e usa valores de similaridades *user-user* e item-item pré-calculados para identificar os *k* vizinhos mais próximos. A fim de avaliar o sistema, desenvolveu-se um protótipo com uma interface web simples na plataforma IPTV<sup>9</sup> e utilizou-se um conjunto de dados limitados.

O processo de produção de recomendações nesse sistema pode ser dividido em três fases. Na primeira, ocorre a construção de matrizes de similaridade através de avaliações de usuários utilizadas para calcular os valores de similaridade usuário-usuário e através dos atributos de item, usados para calcular valores de similaridade item-item. Na segunda, calculam-se as avaliações previstas através da utilização de técnicas de filtragem baseada em conteúdo e filtragem colaborativa. Na terceira, ocorrem o pós-filtro das avaliações previstas, a qualificação das previsões responsáveis e a formalização das recomendações.

A arquitetura desse sistema inclui Web browser, aparelho celular, TV, internet, informações do site IMDB<sup>10</sup>, o núcleo do sistema, que contém um parser de atributos, um website em PHP<sup>11</sup>, um agente de recomendação, uma pilha de processos, um gerador de previsões e bases de dados para as avaliações dos usuários, informações dos itens e dos usuários e similaridades entre itens e usuários.

As principais características funcionais do sistema são as as matrizes de similaridade, a matriz de avaliação, a pilha de processos, o log de avaliação, os usuários e os convites, os filmes, os atributos do filme e o contador de atributos.

As matrizes de similaridade armazenam os dados que refletem quão iguais são os usuários e também os valores da similaridade item-item.

---

<sup>9</sup> IPTV usa o protocolo IP Internet Protocol como meio de transporte do conteúdo e transmissão de sinais televisivos.

<sup>10</sup> The Internet Movie Database: Banco de dados com informações sobre filmes, disponível na internet.

<sup>11</sup> Hypertext Preprocessor. [www.php.net/](http://www.php.net/)

A matriz de avaliação consiste das avaliações reais e das avaliações previstas de todos os itens do sistema para cada usuário e não admite espaços vazios. A pilha de processos é responsável pela organização e pela atualização constante das avaliações previstas e de outras ações cruciais durante o período em que o sistema está sendo executado e é implementada como uma lista FIFO (*First In First Out*).

O log de avaliações ajuda a verificar qual a diferença entre a avaliação real e a prevista e, assim, é possível averiguar quão bem o sistema funciona. As previsões antigas, mesmo que substituídas pelas novas, permanecem armazenadas nesse log, para fins de verificação.

As informações de autenticação dos usuários também são armazenadas e cada um deles pode convidar mais 20 pessoas para utilizar o sistema.

Para o método de FBC em item, o conteúdo dos atributos é usado para calcular todas as similaridades dos itens e, os itens semelhantes aos do usuário destino que tenha gostado no passado são usados para produzir recomendações.

O protótipo desenvolvido apresenta as seguintes funcionalidades: apresentação das avaliações, informações do conteúdo dos itens, informações estatísticas das avaliações, apresentação das recomendações e envio de convites.

## **6.4 Mobile TV Targeted Advertisement and Content Personalization**

Este trabalho tem como foco serviços personalizados para TVDPI e aborda novas formas de fornecimento de conteúdo de mídia, o que motiva a melhora dos negócios e enfatiza as possibilidades de personalização.

Os autores definem o sistema como o núcleo de um novo modelo de mercado, no qual os usuários recebem conteúdos adaptados ao seu perfil. O sistema utiliza duas abordagens da personalização, a que incide sobre a configuração do próprio conteúdo, como fontes, cores, etc. e a que incide sobre o conteúdo propriamente dito, como as preferências dos usuários. Além disso, a personalização ocorre tanto no lado servidor, quanto no lado cliente.

O principal objetivo do sistema é substituir o conteúdo comum por um conteúdo personalizado e adaptado, de forma a ser mais atraente para os usuários. Devido a isso, esse sistema permite tanto a recepção de TV por broadcast, quanto por streaming multimídia.

A arquitetura desse sistema conta com um servidor de conteúdo personalizado, que faz a atribuição e a gestão do conteúdo conforme as preferências do usuário e também com os dispositivos portáteis, que coletam as informações dos usuários e as enviam para o servidor.

O funcionamento do sistema ocorre em várias etapas. Primeiramente, o usuário deve fazer o *login* e, assim, os seus dados de conexão são enviados ao servidor. Após o *login* ter sido efetuado com sucesso, o usuário começa a assistir TV, conforme a tecnologia que dispõe o aparelho (*broadcast* ou *streaming*).

O servidor envia o conteúdo personalizado, escolhido com base nas preferências para o dispositivo portátil do usuário, para ser armazenado previamente à sua exibição. Todas as ações do usuário no dispositivo são monitoradas, armazenadas e enviadas ao servidor, enquanto ele assiste TV.

Na primeira vez em que o usuário acessa a aplicação no dispositivo portátil, ele deve preencher um formulário com as suas preferências, que servirá como perfil inicial e no qual o sistema se baseará. Ao receber os dados coletados, o servidor atualiza o perfil do usuário previamente informado por ele. Dessa forma, o sistema mescla a coleta de dados, usando tanto a abordagem explícita, quanto a abordagem implícita.

Quando o usuário decide não mais assistir TV, a aplicação envia os dados de conexão para o servidor e o dispositivo é, então, desconectado do servidor. O sistema também permite que provedores de conteúdo acessem o servidor e recuperem informações do usuário, dados estatísticos e de consumo. (FEDERICO ET AL, 2009).

## 6.5 ZapTV: Personalized User-Generated Content for Handheld Devices in DVB-H Networks

O ZapTV é um sistema que permite personalizar o conteúdo gerado pelo usuário para dispositivos portáteis em redes móveis DVB-H, oferecendo serviços de valor agregado como acesso multimodal (Web e Celulares), canal de retorno, anotação de vídeo, compartilhamento e distribuição personalizada de conteúdo.

O ZapTV envolve, além da tecnologia provida pelo DVB-H, outras tecnologias como a *TV-Anytime*<sup>12</sup>, para descrever os atributos do conteúdo multimídia, o que permite que o usuário encontre, navegue e gerencie conteúdos a partir de uma ampla variedade de fontes (Internet, TVD, etc) e também que possa compartilhar e distribuir diversos tipos de conteúdo multimídia (jogos, músicas, etc.) levando em consideração os perfis dos usuários.

Usa-se tecnologias emergentes da Web 2.0 que realçam a cooperação e o compartilhamento de informações e permitem a criação de comunidades virtuais que estabelecem ligações entre os usuários que partilham características e interesses comuns. Tecnologias envolvidas na Web Semântica melhoram o acesso automático e o compartilhamento de recursos, fornecendo processos de raciocínio que descobrem relações entre recursos anotados semanticamente.

Dentre as principais funcionalidades do ZapTV destaca-se uma Rede Social para que os usuários enviem, publiquem, anotem, visualizem, classifiquem e sugiram os conteúdos gerados por eles próprios a outros. Outra funcionalidade é a difusão personalizada de conteúdo, que ocorre tanto explícita (para o site e para o dispositivo portátil), quanto implicitamente, sendo que essa última é um *engine* inteligente de personalização (FH) que seleciona automaticamente os conteúdos de interesse para cada usuário, considerando os metadados da *TV-Anytime*, o nível de audiência, a classificação e o histórico de visualização do usuário.

O Planejamento de difusão de canais temáticos fornece conteúdos diversificados para uma faixa etária, para um gênero específico para um tema

---

<sup>12</sup> TV-Anytime é um conjunto de especificações para o controle da entrega de conteúdo multimídia para usuários de DVR's, digital vídeo recorder. ([www.tv-anytime.org/](http://www.tv-anytime.org/))

específico, etc. A aplicação cliente nos receptores portáteis provê capacidades interativas para visualizar, baixar, aceitar recomendações e utilizar o canal de retorno (2G/3G) para classificar conteúdos, enviar o histórico de visualização e acessar serviços. ZapTV transmite do EPG para os dispositivos portáteis, assim como alertas de recomendação e o gerenciamento da capacidade de interação.

O ZapTV procura melhorar a recomendação através da utilização de um *engine* de personalização inteligente que combina mecanismos tradicionais de filtragem de informação com processos de raciocínio semântico. Ele foi modelado nos princípios de participação e compartilhamento entre os usuários da Web 2.0. A forma de geração, compartilhamento, classificação e anotação de conteúdos ocorrem colaborativamente, o que facilita a tarefa de busca desses conteúdos. (SOLLA ET AL, 2008)

## **6.6 A Lightweight Mobile TV Recommender: Towards a One-Click-To-Watch Experience**

O trabalho foi desenvolvido para o sistema DVB-H e aborda as características particulares do ambiente dos dispositivos portáteis e segue o padrão OMA-BCAST.

Os autores identificaram alguns requisitos para os Sistemas de Recomendação dedicados a esse ambiente como: escalabilidade, latência de resposta, flexibilidade para padrões correntes de transmissão, proteção à privacidade do usuário, entre outros. O sistema de recomendação enquadra-se na categoria de sistemas com filtragem baseada em conteúdo e emprega mineração de texto e uma interface simples com o usuário. Além disso, ele aceita linguagem natural como entrada de texto, assim como quatro valores que refletem as preferências do usuário para comédia, ação, terror e erotismo.

Nesse sistema, a recomendação ocorre da forma descrita a seguir. Primeiramente, são extraídos os textos (a) e, em seguida, procuram-se as emoções contidas no texto (b). As distâncias entre os temas são computadas (c) e um índice é calculado para cada entrada (d). Retorna-se uma lista de programas ordenada por

esse índice (e). Os dados do EPG, ao serem atualizados, passam novamente pelas fases (a) e (b) e são armazenados juntamente com os dados computados em um banco de dados, tornando-se um EPG DataSet (conjunto de dados de EPG) semanticamente enriquecido e denominado SED (Semantic EPG Data Set). O EPG é, então, enviado como um fragmento XML no protocolo FLUTE.

A arquitetura do sistema integra, na arquitetura existente do DVB-H, uma fonte de guia de serviços e um servidor de geração de guia de serviços (SED Generation). A fonte de guia de serviços envia o EPG para o SED Generation que, por sua vez, gera e encapsula o EPG DataSet com o conteúdo de áudio e vídeo que é, então, enviado para broadcast e chega aos dispositivos portáteis. Os autores também identificaram que o SED Generation precisa gerar os SEDs pelo menos uma vez para todos os clientes.

O fluxo do processo ocorre como descrito a seguir. O sistema de recomendação processa as descrições dos programas no EPG com uma semana de antecedência. As SEDs incorporadas ao EPG resultam desse processamento e são continuamente transmitidas através de um carrossel de dados baseado no protocolo FLUTE. Quando o cliente recebe um EPG atualizado, extraem-se as SEDs que são armazenadas em disco e as SEDs desatualizadas são excluídas. O usuário pode ajustar suas preferências ou aceitar as configurações armazenadas quando desejar ver TV e, assim, as recomendações são computadas para as SEDs armazenadas. Por fim, o usuário assiste ao programa apropriado. (BÄR ET AL, 2008)

## **6.7 Observações Finais**

Os dois primeiros trabalhos relacionados (6.1 e 6.2) foram desenvolvidos para TVD em residências e fazem uso de um set-top-box, diferentemente dos outros trabalhos citados, que focam em TVDPI.

Ambos os trabalhos citados nas seções anteriores fornecem soluções para a personalização e a sobrecarga de informação na TVD. Em 6.6, o mecanismo do sistema de recomendação aplica mineração de textos, filtragem baseada em

conteúdo e necessita que o usuário informe alguns dados, ao passo que em 6.5 o mecanismo é mais sofisticado; utilizando filtragem de informação híbrida e raciocínio semântico, identifica o usuário de forma explícita e implícita. Vale ressaltar que é necessário a utilização de login em todos os trabalhos comentados e os diferenciais de 6.4, em relação a todos os outros, são a propaganda personalizada e a recepção de conteúdos tanto por *streaming* quanto por *broadcast*.

O trabalho de 6.3 utiliza FH no núcleo do sistema de recomendação, aplicando cálculos matemáticos de similaridades entre usuários e itens, a fim de evitar problemas como: esparsidade, primeira avaliação, *cold-start*, novos itens e usuários no sistema, entre outros que surgem ao se utilizar uma única técnica.

No BiPODiTVR, todo o processamento para a obtenção das recomendações é executado no servidor remoto, que é livre das limitações de hardware existentes nos dispositivos portáteis. Além disso, o BiPODiTVR utiliza FC e FBC para a geração das recomendações com o auxílio das técnicas do Cosseno e dos Vizinhos mais Próximos, faz a coleta dos dados implicitamente e necessita da identificação apenas no momento da solicitação das recomendações.

# 7

## CONCLUSÕES E TRABALHOS FUTUROS

---

O principal objetivo desta dissertação consistiu na proposta de um sistema de recomendação para a TVDPI. Esse sistema permite a oferta de personalização aos usuários da TVDPI. Esse objetivo foi alcançado com a concepção do BiPODiTVR, o que demandou o cumprimento de algumas etapas.

Primeiramente, realizou-se um estudo sobre a TVDPI, verificando-se a viabilidade do desenvolvimento de um servidor de recomendação e uma aplicação cliente.

A etapa seguinte consistiu do estudo das técnicas de recomendação, o qual envolveu a pesquisa sobre filtragem de informação, mineração de dados e outros métodos. Finalizadas as etapas de pesquisa, iniciou-se a modelagem da arquitetura do BiPODiTVR e do banco de dados utilizado, além da simulação manual do sistema. Em seguida, a última etapa tratou da validação/implementação do sistema. O BiPODiTVR foi implementado, testado e validado em um ambiente compatível com as normas do sistema brasileiro de TVDPI e os resultados obtidos demonstram a eficácia e a aplicabilidade do sistema.

A limitação do sistema desenvolvido se restringe ao fato de ser necessário utilizar a internet para fazer a recomendação. Caso o dispositivo portátil não tenha suporte à uma forma de conexão à internet, o usuário não conseguirá utilizar o BiPODiTVR.

Para a realização de trabalhos futuros, espera-se incluir novos algoritmos para a descoberta de perfil e a geração de recomendações, aumentando a qualidade e também o número de acertos do sistema. Realizar testes em mais cenários de uso para obter novas informações a respeito da funcionalidade do sistema e então melhorá-lo. Uma interface web pode ser desenvolvida para acesso por parte dos

usuários, permitindo que acessem às informações de seu perfil, podendo até mesmo editá-las. Ela também permitirá gerar estatísticas, gráficos, etc. sobre os usuários, programas e provedores de serviços.

Como o BiPODiTVr usa FC, FBC e FH, é possível criar um tipo de rede social na TVDPI, mostrando aos usuários aqueles que são mais parecidos com eles e permitindo que entrem em contato uns com os outros. Tudo isso poderá ainda ser exibido, além de na web, também na tela do dispositivo. A forma de identificação do usuário também pode ser melhorada com a utilização do OpenID. O OpenID é um sistema que permite que a identificação do usuário seja feita por qualquer servidor que tenha suporte ao protocolo. Um único login pode ser utilizado para acessar vários sites diferentes. (OpenID, 2010)

Pretende-se também incluir como parâmetros para a descoberta de preferências do usuário a classificação indicativa dos programas (faixa etária) e a sinopse. Com isso, uma gama maior de recomendações pode ser fornecida ao usuário. No caso da sinopse, poderiam ser descobertos, por exemplo, os atores de cinema preferidos e, então, recomendar filmes que tenham a participação desses atores. Muitas outras preferências do usuário podem ser descobertas pela sinopse do programa e o BiPODiTVR pretende explorar essas opções.

# REFERÊNCIAS BIBLIOGRÁFICAS

3GPP. **3rd Generation Partnership Project**. Disponível em: <<http://www.3gpp.org/>>. Acessado em: Maio/2010

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). NBR 15603-2. Televisão Digital Terrestre – Multiplexação e serviços de informação (SI) Parte 2: Estrutura de dados e definições da informação básica de SI. Versão 3, 21/09/2009, 129 páginas

ADOMAVICIUS, Gediminas; TUZHILIN, Alexandre. **Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions**. IEEE Transactions on Knowledge and Data Engineering, New York, v. 17, n. 6, p. 734-749, June. 2005.

ALENCAR, Marcelo Sampaio de. Televisão Digital. São Paulo. Érica. 2007. 352p.

ÁVILA, Paulo Muniz de. **RecommenderTV: Suporte ao Desenvolvimento de Aplicações de Recomendação para o sistema brasileiro de TV Digital**. 87 páginas. Departamento de Computação. Universidade Federal de São Carlos (UFSCar), São Carlos/SP, 01/2010.

BÄR, Arian; BERGER, Andreas; EGGER, Sebastian; SCHATZ, Raimund. **A Lightweight Mobile TV Recommender: Towards a One-Click-to-Watch Experience**. In Proceedings 6th European Interactive TV Conference, p.142-147, Salzburg, Áustria, 03-04/07/2008.

BALABANOVIC, Marko; SHOHAM, Yoav. **FAB: Content-Based, Collaborative Recommendation**. In: Communications of the ACM, New

York, USA, v. 40, n. 3, p. 66-72, March 1997.

BMCO. **Broadcast Mobile Convergence**. Disponível em: <<http://www.bmcoforum.org/>>. Acessado em: Julho de 2010.

BORGES, Fábio; SOUZA, Raquel. **Códigos Reed-Solomon**. In: XXXI CNMAC, Congresso Nacional de Matemática Aplicada e Computacional. Belém, Amazonas, 8-11/09/2008.

CAMPOS, Manoel. **NCLua SOAP**. Disponível em: <<http://manoelcampos.com/tvd/nclua-soap/>>. Acessado em: Março de 2011.

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de Dados: Funcionalidades, Técnicas e Abordagens**. Maio/2002, PUC/Rio, Rio de Janeiro, Brasil. 35 páginas. Relatório Técnico.

Cristo, M. **Sistemas de Recomendação, Métodos e Avaliação**. 81 slides. 2009.

DAB. **Digital Audio Broadcasting**. Disponível em: <<http://www.worlddab.org/>>. Acessado em: Julho de 2010.

DIBEG. **Digital Broadcasting Expert Group**. Disponível em: <<http://www.dibeg.org/>>. Acessado em: Maio de 2010.

DMB. **Digital Multimedia Broadcast**. Disponível em: <<http://www.worlddab.org/>>. Acessado em: Maio de 2010.

DVB. **Digital Video Broadcast**. Disponível em: <[www.dvb.org](http://www.dvb.org)>. Acessado em: Maio de 2010.

EINARSSON, O. P. Content **Personalization for Mobile TV Combining**

**Content-Based and Collaborative Filtering.** Master Thesis. Center for Information and Communication Technologies. Technical University of Denmark. August 22, 2007.

FEDERICO, Menéndez; CEDRÓN-Fernández, Iago; ÁLVAREZ, José Manuel; NÚÑEZ, José Luiz; URIBE, Silvia. **Mobile TV Targeted Advertisement and Content Personalization.** 16th International Workshop Conference on Systems, Signals and Image Processing, Chalkida, Greece, 18-19/06/2009.

FLUTE. **File Delivery over Unidirectional Transport.** Disponível em: <<http://www.faqs.org/rfcs/rfc3926.html>>. Acessado em: Julho de 2010.

GATTO, Elaine Cecília; ZORZO, Sérgio Donizetti. **Sistema de Recomendação para TV Digital Portátil Interativa Brasileira.** In: 8th International Information and Telecommunication Technologies Symposium (I2TS). Florianópolis, Santa Catarina, Brazil. 09-11/12/2009.

GATTO (a), Elaine Cecília; ZORZO, Sérgio Donizetti. **Aplicando Filtragem Colaborativa aos dados de TV fornecidos pelo IBOPE.** Relatório Técnico. Universidade Federal de São Carlos, Departamento de Computação, São Carlos/SP, Junho/2010.

GATTO (b), Elaine Cecília; ZORZO, Sérgio Donizetti. **Aplicando Filtragem Baseada em Conteúdo aos dados de TV fornecidos pelo IBOPE.** Relatório Técnico. Universidade Federal de São Carlos, Departamento de Computação, São Carlos/SP, Junho/2010.

GATTO (c), Elaine Cecília; ZORZO, Sérgio Donizetti. **Application of recommendation techniques for Brazilian Portable Interactive Digital TV.** In: 17th International Conference on Systems, Signals and Image Processing (IWSSIP). Niterói, Rio de Janeiro, Brazil. 17-19/06/2010 (c).

GINGA-J. Disponível em: <<http://dev.openginga.org/>>. Acessado em Junho de 2010.

GINGA-NCL. Disponível em: <[www.gingancl.org.br/](http://www.gingancl.org.br/)>. Acessado em: Maio de 2010.

GMIT. **Galway – Mayo Institute of Technology**. Disponível em: <<http://www.gmit.ie/>>. Acessado em: Agosto de 2010

LASER. **Light Weight Application Scene Representation**. Disponível em: <[http://www.mpeg-laser.org/html/techSection\\_technicalOverview.htm](http://www.mpeg-laser.org/html/techSection_technicalOverview.htm)>. Acessado em: Maio/2010

LASTFM. Disponível em: <[www.lastfm.com](http://www.lastfm.com)>. Acessado em: Setembro de 2010.

LUA. Disponível em: <[www.lua.org/portugues.html](http://www.lua.org/portugues.html)>. Acessado em: Agosto de 2010.

LUCAS, Adriano dos Santos. **Personalização para Televisão Digital utilizando a estratégia de Sistema de Recomendação para ambiente multiusuário**. 103 páginas. Dissertação de Mestrado. Departamento de Computação. Universidade Federal de São Carlos (UFSCar), São Carlos/SP, 05/2009.

LOSS, Márcio Rogério. **TV digital e a invasão da privacidade**. Sorocaba/SP. Editora Minelli, 2008.

MAXIMINO, Allison Patrick. **Comparação de Algoritmos de aprendizado de máquina para classificação de padrões em uma base de dados médicos**. Ribeirão Preto. 2007. Centro Universitário Barão de Mauá. Monografia de Conclusão de Curso. 48 páginas. Disponível em: <>. Acessado em:

Fevereiro/2009.

MEDIAFLO. **Media Forward Link Only**. Disponível em: <<http://www.mediaflo.com/>>. Acessado em: Maio de 2010.

NCL. **Nested Context Language**. Disponível em: <[www.ncl.org.br/](http://www.ncl.org.br/)>. Acessado em: Agosto de 2010.

OMA-BCAST. **Open Mobile Alliance – Mobile Broadcast Services Enabler Suite**. Disponível em: <<http://www.openmobilealliance.org/default.aspx>>. Acessado em: Maio de 2010.

OPENID. Disponível em: <<http://openid.net/>>. Acessado em: Setembro de 2010.

RESNICK, P.; VARIAN, H. R. **Recommender Systems**. Communications of the ACM, New York, vol. 40, n. 3, p. 77-87, March 1997.

RNP. **Middleware**. Disponível em: <<http://www.rnp.br/noticias/2006/not-060926.html>>. Acessado em: Junho de 2010.

SANTOS, Alexandro, et. al. **TV Digital Móvel**. Revista WebMobile, Rio de Janeiro/RJ, Ano 1, 6.ed., p. 47-54. 12/2005-01/2006.

SEBASTIAN, Egger; JORDÂNIA, Norbert; SCHATZ, Raimund; WAGNER, Siegfried. **Mobile TV becomes Social – Integrating Content with Communications**. In: International Conference on Information Technology Interfaces (ITI'07), 29th, 2007/06/25-28, Cavtat, Croatia. Proceedings. IEEE Catalog, 2007/08/08, p. 263-270.

SCHNEIDER, Luís Felipe. **Mineração de Dados (Data Mining). Tópicos Avançados em Modelos de Banco de Dados**. Universidade Federal do Rio

Grande do Sul. 44 slides.

SOARES (a), Luiz Fernando Gomes; RODRIGUES, Rogério Ferreira. **Nested Context Model 3.0 Part 1**. PUC. Rio de Janeiro/RJ. 2005.

SOARES (b), Luiz Fernando Gomes; BARBOSA, Simone Diniz Junqueira. **Programando em NCL 3.0. Desenvolvimento de aplicações para o middleware Ginga. TV Digital e Web**. Rio de Janeiro/RJ. Editora Elsevier. 2009.

SOFTWARE PÚBLICO. **Ginga wiki**. Disponível em:  
<[http://www.softwarepublico.gov.br/dotlrn/clubs/ginga/one-community?page\\_num=3](http://www.softwarepublico.gov.br/dotlrn/clubs/ginga/one-community?page_num=3)>. Acessado em: Junho de 2010

SOLLA, Alberto Gil; FERNÁNDEZ; Yolanda Blanco; ARIAS, José J. Pazos; CABRER, Manuel Ramos; NORES, Martín López. **ZapTV: Personalized User-Generated Content for Handheld Devices in DVB-H Mobile Newtorks**. In: Proceedings 6th European Interactive TV Conference, p.193-203, Salzburg, Áustria, 03-04/07/2008.

TORRES, Roberto. **Personalização na Internet**. Novatec Editora. 2004. 158p.

YEUN, Chan Yeob. **Mobile TV Technologies**. In: International Conference on Information and Communication Technology (ICICT'07), 5th, 2007/12/16-18, Dhaka, Bangladesh. Proceedings. IEEE Catalog, 21/03/2008, p. 2-9.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2nd Edition, Morgan Kaufmann, 525 pages, June 2005.