



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO**

FABRÍCIO DIAS ALVES LEMOS

**UMA ABORDAGEM PARA INSERÇÃO DE SENSIBILIDADE AO CONTEXTO EM
TÉCNICAS DE RECOMENDAÇÃO BASEADAS EM CONHECIMENTO**

FORTALEZA, CEARÁ

2012

FABRÍCIO DIAS ALVES LEMOS

UMA ABORDAGEM PARA INSERÇÃO DE SENSIBILIDADE AO CONTEXTO EM
TÉCNICAS DE RECOMENDAÇÃO BASEADAS EM CONHECIMENTO

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação, da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientadora: Profa. Dra. Rossana Maria de Castro Andrade.

Co-orientador: Prof. Dr. Windson Viana de Carvalho.

FORTALEZA, CEARÁ

2012

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Ciências e Tecnologia

-
- L577a Lemos, Fabricio Dias Alves.
Uma abordagem para inserção de sensibilidade ao contexto em técnicas de recomendação baseadas em conhecimento. / Fabricio Dias Alves Lemos. – 2012.
102 f. : il. color., enc. ; 30 cm.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Departamento de Computação, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2011.
Área de Concentração: Ciência da Computação.
Orientação: Profa. Dra. Rossana Maria de Castro Andrade.
Coorientação: Prof. Dr. Windson Viana de carvalho.
1. Sistemas de multimídia. 2. Fotografia – técnicas digitais. 3. Multimídia interativa. I. Título.

FABRÍCIO DIAS ALVES LEMOS

UMA ABORDAGEM PARA INSERÇÃO DE SENSIBILIDADE AO CONTEXTO EM
TÉCNICAS DE RECOMENDAÇÃO BASEADAS EM CONHECIMENTO

Dissertação submetida à Coordenação do
Curso de Pós-Graduação em Ciência da
Computação, da Universidade Federal do
Ceará, como requisito parcial para a obtenção
do grau de Mestre em Ciência da Computação.

Aprovada em: ____ / ____ / _____

BANCA EXAMINADORA

Profa. Dra. Rossana Maria de Castro Andrade (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Windson Viana de Carvalho
Universidade Federal do Ceará (UFC)

Prof. Dr. Fernando Antônio Mota Trinta
Universidade Federal do Ceará (UFC)

Prof. Dr. Cláudio de Souza Baptista
Universidade Federal de Campina Grande (UFCG)

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, a quem sou grato por tudo o que tenho e conquistei.

Aos meus pais, Carlos Alberto e Maria das Graças, que prezam a educação como um dos valores mais importantes a serem buscados, e que sempre me dão todo o suporte necessário para o alcance dos meus objetivos.

À Evellyne, minha esposa, por todo o apoio, incentivo, compreensão e cumplicidade na superação de todas as dificuldades e desafios encontrados.

À professora Rossana, pela orientação, oportunidade e confiança oferecida e a quem sou grato pelo apoio dado durante toda minha formação acadêmica.

Ao professor Windson, pela valiosa ajuda e dedicação e indispensável direcionamento que foram de primordial importância a esse trabalho.

Aos professores Fernando Trinta e Cláudio Baptista pela presença na banca examinadora.

Ao João Marcelo, bolsista de iniciação científica, pela excelente ajuda dispensada na implementação do trabalho.

Ao Rafael Carmo, pela importante ajuda na geração dos indicadores estatísticos das avaliações do trabalho.

Ao colega André Fonteles, pela ajuda na construção da interface com o usuário da solução proposta.

À Secretaria da Fazenda, na pessoa do orientador de célula Pedro Peixoto, pela flexibilidade de horário dispensada para que eu pudesse atender os compromissos acadêmicos.

A todos os colegas do GREat, em especial à Nayane, Benedito, Ricardo e Adyson, pelo compartilhamento de experiências e auxílio no alcance dos objetivos do mestrado.

RESUMO

O principal desafio de um Sistema de Recomendação é conseguir identificar e recomendar itens que tenham maior chance de corresponder aos interesses de seus usuários que, em geral, possuem uma natureza bastante subjetiva e heterogênea. É importante, então, que os Sistemas de Recomendação possam recomendar itens personalizados a partir da identificação do perfil de cada usuário. Entretanto, muitas vezes, o perfil do usuário não é suficiente para que o sistema consiga identificar completamente seus interesses. A utilização do sistema em um contexto diferente do habitual pode acarretar em um resultado insatisfatório para a recomendação, sendo necessária sua adaptação ao novo contexto. Este trabalho propõe uma abordagem para inserção de sensibilidade ao contexto em Sistemas de Recomendação, a qual utiliza, além do contexto do usuário, o próprio contexto dos itens recomendados. Para tal, técnicas de recomendação baseadas em conhecimento são adaptadas para incluir regras de domínio com caráter contextual. A abordagem proposta é utilizada na concepção de um sistema de recomendação de fotos sensível ao contexto que explora o contexto do usuário e o contexto de geração das fotos para a realização da recomendação. Através do sistema desenvolvido, a abordagem é validada a partir de experimentos com usuários, nos quais são avaliados a qualidade da recomendação e os ganhos obtidos com a utilização das informações contextuais.

Palavras-Chave: Sistemas de Recomendação, Sensibilidade ao Contexto, Anotação de Fotos.

ABSTRACT

The main challenge of recommender systems is to be able to identify and recommend items that have a great chance of meeting the interests of their users, which generally have a very subjective and heterogeneous nature. It is important, then, that recommender systems could recommend personalized items from the identification of each user's profile. However, the user's profile is often not enough for the system to be able to completely identify the user's interests. The use of the system in a different context from the usual can cause an unsatisfactory result for the recommendation, requiring it to be adapted to the new context. This work proposes an approach for the adoption of context awareness in recommender systems, which combines the context of the recommended items to the user's context in the time of the recommendation. For this purpose, knowledge-base recommendation techniques are adapted to include domain rules with contextual characteristics. The proposed approach is applied in the development of a context awareness photo recommender system that exploits the user context and the context when the photo was created as a means to improve the recommendation. The photo recommender system is validated by user experiments in which the quality of the recommendation and the gains from the use of contextual information are evaluated.

Keywords: Recommender Systems, Context-Awareness, Photo Annotation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Técnicas de Recomendação e suas Fontes de Conhecimento, adaptado de (Burke, 2007)	21
Figura 2 – Recomendação Colaborativa do site Amazon.com.....	23
Figura 3 – Recomendação Baseada em Conteúdo do site <i>Last.fm</i>	24
Figura 4 – Recomendação de Elementos de Código Fonte (Antunes et al., 2012)	26
Figura 5 – Arquitetura em Camadas para Aplicações Sensíveis ao Contexto proposta em Ailisto et al., 2002.....	35
Figura 6 – Arquitetura de referência da linha de produto <i>MobiLine</i> (Marinho et al., 2010) ...	37
Figura 7 – Modelo de Contexto para Anotação de Documentos Multimídia (Viana et al., 2011)	39
Figura 8 – Telas da aplicação <i>Columbus</i> (Rost et al., 2011).....	46
Figura 9 – Processo de Recomendação do <i>C² Music</i> (Lee and Lee, 2007).....	48
Figura 10 – Passos para Recomendação com Pré-Filtro (Adomavicius and Tuzhilin, 2008)..	49
Figura 11 – Fluxo de execução da abordagem de recomendação	54
Figura 12 – Modelo de Contexto adotado para a Recomendação Contextual	56
Figura 13 – Padrão de Projeto <i>Composite</i> - (Gamma et al., 1995).....	62
Figura 14 – Diagrama de Classes da Modelagem do Cálculo de Similaridades	63
Figura 15 – Diagrama de Objetos para um Cálculo de Similaridade Simples	64
Figura 16 – Diagrama de Objetos para um Cálculo de Similaridade Composto.....	65
Figura 17 – Diagrama de Objetos para uma Hierarquia de Cálculo de Similaridade	66
Figura 18 – Fluxo de Telas da Aplicação Móvel.....	70
Figura 19 – Fluxo de Execução do <i>MMedia2U</i>	72
Figura 20 – Diagrama de Componentes da Solução	74
Figura 21 – Sistemática de Construção do <i>Gold Standard</i>	82
Figura 22 – Sistemática de Execução do Algoritmo.....	83
Figura 23 – Precisão Média Obtida pela Recomendação.....	85
Figura 24 – <i>Recall</i> Médio Obtido pela Recomendação	87
Figura 25 – <i>F-measure</i> Médio Obtido Pela Recomendação	88

LISTA DE TABELAS

Tabela 1 – Exemplo de Avaliações.....	29
Tabela 2 – Similaridade para Estações do Ano (Lee and Lee, 2007).....	48
Tabela 3 – Similaridade entre Atividades	73
Tabela 4 – Similaridade para o dia da semana	74
Tabela 5 – Contextos Utilizados nos Experimentos	81
Tabela 6 – Categorias de Itens Geradas após a Avaliação e Recomendação.....	84
Tabela 7 – Precisão Média Obtida pela Recomendação	86
Tabela 8 – <i>Recall</i> Médio Obtido pela Recomendação.....	87
Tabela 9 – <i>F-measure</i> Médio Obtido Pela Recomendação.....	88
Tabela 10 – Questionário: uma Pergunta por Experimento	90
Tabela 11 – Questionário: uma Pergunta por Usuário	90
Tabela 12 – <i>MMedia2U</i> e Trabalhos Relacionados.....	95
Tabela 13 – Publicações.....	95

LISTA DE FÓRMULAS

Fórmula 1 – Correlação de Pearson para Cálculo de Similaridade entre dois Usuários (Resnick et al., 1994).....	29
Fórmula 2 – Coeficiente de Dice na comparação de livros.....	29
Fórmula 3 – Similaridade entre um Item e a Preferência do Usuário - (Mcsherry, 2003).....	30
Fórmula 4 – Similaridade pela Proximidade de Valores Numéricos (Jannach et al., 2011)....	30
Fórmula 5 – Similaridade para Valores Representados em uma Hierarquia (Slimani et al., 2006)	31
Fórmula 6 – Similaridade entre um Item e a Preferência do Usuário (Mcsherry, 2003)	59
Fórmula 7 – Cálculo de Similaridade Global entre Usuário e Item.....	59
Fórmula 8 – Exemplo de Cálculo de Similaridade	60
Fórmula 9 – Cálculo de Similaridade em uma Hierarquia - (Slimani et al., 2006).....	72
Fórmula 10 – Cálculo de Similaridade para Valores Numéricos	73
Fórmula 11 – Cálculo da Precisão (Cleverdon et al., 1966).....	84
Fórmula 12 – Cálculo do <i>Recall</i> (Cleverdon et al., 1966)	86
Fórmula 13 – Cálculo do <i>F-measure</i> (Yang and Liu, 1999).....	87

LISTA DE ABREVIATURAS E SIGLAS

EXIF	<i>Exchangeable Image File Format</i>
GPS	<i>Global Positioning System</i>
GREat	Grupo de Redes de Computadores, Engenharia de Software e Sistemas
IPTC	<i>International Press Telecommunications Council</i>
OML	<i>Open Modeling Language</i>
W3C	<i>World Wide Web Consortium</i>
WIFI	<i>Wireless Fidelity</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	MOTIVAÇÃO	14
1.2	OBJETIVOS E CONTRIBUIÇÕES	15
1.3	ESTRUTURA DO DOCUMENTO	17
2	SISTEMAS DE RECOMENDAÇÃO	19
2.1	INTRODUÇÃO	19
2.2	TIPOS DE RECOMENDAÇÃO	21
2.2.1	<i>Recomendação Colaborativa</i>	22
2.2.2	<i>Recomendação Baseada em Conteúdo</i>	23
2.2.3	<i>Recomendação Baseada em Conhecimento</i>	24
2.3	VANTAGENS E LIMITAÇÕES DAS TÉCNICAS DE RECOMENDAÇÃO	26
2.4	CÁLCULO DE SIMILARIDADE	28
2.5	CONCLUSÃO	31
3	SENSIBILIDADE AO CONTEXTO	32
3.1	INTRODUÇÃO	32
3.2	SENSIBILIDADE AO CONTEXTO EM SISTEMAS MULTIMÍDIA	40
3.3	SENSIBILIDADE AO CONTEXTO EM SISTEMAS DE RECOMENDAÇÃO	41
3.4	TRABALHOS RELACIONADOS	43
3.4.1	<i>Columbus – Exploração em mobilidade de fotografias geolocalizadas</i>	45
3.4.2	<i>C²_Music</i>	47
3.4.3	<i>Sistema de Recomendação de Filmes</i>	49
3.5	CONCLUSÃO	50
4	ABORDAGEM PROPOSTA	51
4.1	VISÃO GERAL DA PROPOSTA	51
4.2	FLUXO DE EXECUÇÃO DA ABORDAGEM DE RECOMENDAÇÃO	53
4.3	MODELO DE CONTEXTO	55
4.4	CÁLCULOS DE SIMILARIDADE	59
4.5	MODELAGEM DO CÁLCULO DE SIMILARIDADE	60
4.6	CONCLUSÃO	66
5	MMEDIA2U	68
5.1	PRINCÍPIOS NORTEADORES DO <i>MMEDIA2U</i>	68
5.2	APLICAÇÃO MÓVEL	69
5.3	CÁLCULOS DE SIMILARIDADE	72

5.4	COMPONENTES DA SOLUÇÃO	74
5.5	CONCLUSÃO	76
6	EXPERIMENTO E AVALIAÇÃO.....	78
6.1	CONCEPÇÃO DA AVALIAÇÃO.....	78
6.2	CONSTRUÇÃO DO <i>GOLD STANDARD</i>	80
6.3	RESULTADOS QUANTITATIVOS.....	82
6.4	QUESTIONÁRIO COM OS USUÁRIOS	89
6.5	CONSTRUÇÃO DO REPOSITÓRIO DE FOTOS	90
6.6	CONCLUSÃO	91
7	CONCLUSÃO E TRABALHOS FUTUROS.....	93
7.1	RESULTADOS ALCANÇADOS.....	93
7.2	PUBLICAÇÕES.....	95
7.3	TRABALHOS FUTUROS	96
8	BIBLIOGRAFIA.....	97

1 INTRODUÇÃO

A constante evolução e a massificação do uso da informática são responsáveis por um número cada vez maior de dados sendo digitalizados, compartilhados e acessados para os mais diferentes propósitos. Estes dados podem gerar bases de conhecimento que nos auxiliariam em uma série de atividades cotidianas, tais como: a decisão de qual filme assistir no final de semana ou quais os pontos turísticos mais interessantes a serem visitados em uma cidade.

Com o progresso dos Sistemas de Informação, surgiram aplicações que extraem ao máximo o potencial dessas bases de conhecimento e procuram nos auxiliar nesses processos decisórios. Estes são chamados de Sistemas de Recomendação e tem como objetivo prover, de maneira acessível, recomendações de qualidade para uma abrangente comunidade de usuários (Jannach et al., 2011). Sistemas de Recomendação também ajudam a lidar com um dos mais importantes desafios no desenvolvimento de Sistemas de Informação, que é a sobrecarga de informações. Tais sistemas ajudam as pessoas na recuperação da informação (e.g., vídeos, programas de TV, rotas, imagens, pessoas, entre outras) que provavelmente coincidam com suas preferências e intenções. Sistemas de Recomendação tentam identificar os itens do conjunto de informação que tem uma maior chance de corresponder os desejos de seus usuários (Adomavicius and Tuzhilin, 2008). A qualidade de uma recomendação é medida pela sua capacidade de prever quais os itens que podem despertar maior interesse em seus usuários.

Uma evidência dos ganhos obtidos pelos Sistemas de Recomendação é o surgimento de diversas aplicações *on-line* que utilizam técnicas de recomendação para oferecer um conteúdo personalizado aos usuários e, potencialmente, aumentar sua satisfação com o uso do sistema. O site de compras *Amazon* (Linden et al., 2003) representa um exemplo deste tipo de sistema, recomendando itens similares aos presentes no carrinho de compras do usuário e que costumam ser comprados por outros usuários com preferências semelhantes.

Um importante ponto em Sistemas de Recomendação é relacionado ao contexto do usuário. O uso do sistema em um contexto diferente do costumeiro pode causar uma recomendação com resultado insatisfatório, já que preferências e intenções podem ser influenciadas pelo contexto (localização, trajetória, período do dia, atividade, entre outros) do usuário. O domínio de sistemas sensíveis ao contexto trata exatamente dos desafios da

incorporação do contexto do usuário durante sua execução. A sensibilidade ao contexto refere-se à capacidade que um sistema possui de detectar a situação do usuário e guiar o seu comportamento de acordo (Marinho et al., 2010).

1.1 Motivação

Por depender tanto das informações dos usuários quanto dos itens recomendados, Sistemas de Recomendação podem se beneficiar da sensibilidade ao contexto ao considerar não somente as características intrínsecas de cada entidade, mas também as características de seu contexto. Um site de recomendação de restaurantes, por exemplo, ao ser utilizado por um usuário que deseja jantar na próxima meia hora, pode adaptar sua recomendação para restaurantes que sejam próximos ao usuário, estejam abertos e com mesa disponível para a quantidade de pessoas que irão acompanhar o usuário.

Tradicionalmente, porém, o contexto destas informações, mesmo em alguns casos sendo relevante para uma melhor recomendação, não é considerado no processo (Adomavicius and Tuzhilin, 2008). Técnicas de recomendação sensíveis ao contexto já desenvolvidas para determinados domínios mostram que é possível o aperfeiçoamento de Sistemas de Recomendação com uso de informações contextuais. Entretanto, essas abordagens focam no contexto do usuário e não incorporam o contexto de geração ou o contexto da situação atual dos itens recomendados. Em (Adomavicius et al., 2005), por exemplo, os autores implementam um Sistema de Recomendação de filmes que considera o contexto do usuário (e.g., se este vai assistir o filme em casa ou no cinema), mas não considera o contexto do item recomendado (e.g., se o filme está sendo exibido em um cinema próximo ao usuário). Em geral, os sistemas de recomendação sensíveis ao contexto baseiam-se em técnicas de recomendações colaborativas e baseadas em conteúdo, tais técnicas, contudo, são suscetíveis ao problema do início frio (*cold start problem*). Esse problema se caracteriza pela inabilidade dos Sistemas de Recomendações realizarem uma recomendação satisfatória a um novo usuário, do qual o sistema ainda não tem o perfil traçado. Também acontece com novos itens adicionados ao repositório, os quais o sistema ainda não possui um histórico de uso.

A possibilidade de construção de repositórios de itens ricos em informações contextuais traz uma nova perspectiva para o desenvolvimento de Sistemas de Recomendação

sensíveis ao contexto. Esses repositórios permitem que regras de domínio sejam aplicadas às informações contextuais, tornando possível a realização de recomendações satisfatórias mesmo sem uma base histórica de uso do sistema. Essas regras de domínio podem ser incorporadas em técnicas de recomendação baseadas em conhecimento que, de posse das informações contextuais, recomendam os itens que melhor se adequem ao contexto do usuário. O sistema EMMA (Ganneau et al., 2008), por exemplo, coleta as informações de contexto do usuário associadas a utilização de diferentes tipos de interfaces gráficas em dispositivos móveis. Estas informações são, posteriormente, utilizadas para adaptar a interface de uso do sistema de acordo com o contexto do usuário. Em outro exemplo de uso de informações contextuais, Böhmer et al. (2010) registram o contexto de utilização de aplicações presentes no dispositivo móvel do usuário. Estas informações são, então, usadas para inferir as aplicações que o usuário deseja utilizar em seu contexto atual.

A concepção de repositórios de itens enriquecidos com informações contextuais, entretanto, não é trivial. Esses repositórios necessitam conter informações contextuais em um modelo semântico que permita a extração de conhecimento e a consequente inserção dessas informações no processo de recomendação.

1.2 Objetivos e Contribuições

Este trabalho possui dois objetivos principais:

- a) A elaboração de uma abordagem para incorporação de sensibilidade ao contexto em Sistemas de Recomendação; e
- b) A aplicação da abordagem proposta na implementação de um sistema móvel de recomendação de fotos sensível ao contexto.

A abordagem tem como finalidade prover ao usuário, transparentemente, recomendações mais próximas de seu interesse. A transparência decorre do fato do usuário não necessitar informar ao sistema todos os parâmetros utilizados na busca pelos itens recomendados. Os critérios de busca podem ser extraídos automaticamente a partir de regras de domínio aplicadas às informações contextuais.

Um dos pontos centrais da abordagem é conceber um algoritmo de recomendação que incorpore as informações contextuais. No processo de recomendação, o algoritmo deve utilizar informações contextuais capturadas do usuário e compará-las com as informações contextuais dos itens passíveis de recomendação. A abordagem proposta apoia-se na hipótese de que itens gerados em um contexto semelhante ao contexto atual do usuário são mais relevantes do que itens gerados em contextos com pouca similaridade ao seu.

Para atingir o objetivo são necessários ainda:

- a) A adoção de um modelo de contexto que torne possível a extração e o processamento das informações significativas para a recomendação;
- b) A elaboração de regras de cálculo de similaridade a serem aplicadas no processo de recomendação que façam uso das informações contextuais do modelo de contexto e de regras de domínio. As regras de cálculo de similaridade devem então ser utilizadas para selecionar aqueles itens que são mais relevantes ao usuário;
- c) A avaliação do sistema móvel de recomendação de fotos sensível ao contexto aplicando a abordagem proposta. A avaliação deve permitir também comprovar os ganhos no uso de informação contextual na recomendação.

O domínio escolhido para aplicação da abordagem proposta é o de recomendação de fotos, uma vez que no mesmo se observa a importância das informações contextuais na descrição dos itens.. Nesse domínio, as informações contextuais são importantes para que os usuários descrevam e recuperem as fotos geradas (Kennedy et al., 2007). Além disso, já existem propostas de construção de repositórios de fotos com informações contextuais associadas ao momento em que a foto foi gerada (Viana et al., 2011) (Figueirêdo et al., 2012).

O Sistema de Recomendação sensível ao contexto concebido é chamado de *MMedia2U (Mobile Media to You)*. Ele é composto de:

- a) Um sistema móvel, responsável pela aquisição do contexto do usuário e exibição das fotos recomendadas; e
- b) Um serviço central de recomendação, responsável por identificar, no repositório de fotos, aquelas que mais se adequem ao contexto atual do usuário.

O trabalho propõe-se, ainda, a realizar o estudo da influência de diferentes dimensões de informações contextuais na qualidade da recomendação realizada pelo sistema desenvolvido. Pretende-se, com isso, analisar a relevância relativa das informações contextuais presentes no modelo de contexto proposto. Esse objetivo é alcançado através da avaliação do *MMedia2U*. Na avaliação, a qualidade da recomendação é medida a partir da comparação das fotos recomendadas com fotos consideradas relevantes ao contexto do usuário. A atribuição de relevância às fotos é realizada a partir de experimentos com usuários, em que estes escolhem, a partir de um conjunto de fotos apresentadas, aquelas que mais lhes agradam em um determinado contexto.

As hipóteses analisadas nesse trabalho são:

- a) Hipótese I: é possível realizar recomendações satisfatórias de itens georeferenciados sem conhecimento prévio do usuário, considerando apenas seu contexto atual;
- b) Hipótese II: o contexto em que os itens foram gerados ou o seu contexto atual possui relevância na realização de recomendações; e
- c) Hipótese III: a utilização de um modelo de contexto considerando várias dimensões contextuais pode trazer ganhos em relação a uma recomendação que utiliza somente um atributo contextual (e.g., a localização).

1.3 Estrutura do Documento

Os próximos capítulos desta dissertação são descritos a seguir:

- a) O Capítulo 2 apresenta uma fundamentação sobre Sistemas de Recomendação, suas aplicações e desafios de desenvolvimento encontrados. As principais técnicas de recomendação e suas respectivas características, vantagens e desvantagens são discutidas;
- b) O Capítulo 3 aborda a Sensibilidade ao Contexto. Os principais conceitos são apresentados e seu uso é exemplificado a partir da apresentação de sistemas sensíveis ao contexto nos domínios de recomendação de notícias, músicas,

filmes, entre outros. No Capítulo 3 também são apresentados os trabalhos estritamente relacionados à proposta dessa dissertação e como Sistemas de Recomendação podem ser melhorados a partir do uso de informações contextuais;

- c) No Capítulo 4, a abordagem proposta nesta dissertação é apresentada. Neste capítulo é explicado como a Recomendação Baseada em Conhecimento pode ser estendida ao contemplar informações contextuais de usuários e itens;
- d) No Capítulo 5, o sistema desenvolvido como prova de conceito da abordagem proposta é apresentado;
- e) No Capítulo 6, é mostrado como a abordagem proposta foi validada a partir de experimentos. Os experimentos consistiram da montagem de uma base de itens anotados com informações contextuais e uma avaliação destes itens por parte de usuários. A recomendação do sistema foi então comparada com a avaliação feita pelos usuários;
- f) O Capítulo 7 conclui a dissertação, listando os resultados alcançados e as contribuições realizadas. São também enumerados os trabalhos futuros.

2 SISTEMAS DE RECOMENDAÇÃO

Este capítulo descreve as principais características de Sistemas de Recomendação. Os conceitos apresentados são importantes para o entendimento de como a proposta deste trabalho utiliza e estende as técnicas tradicionais de recomendação. A Seção 2.1 descreve a motivação para o surgimento dos Sistemas de Recomendação e apresenta alguns conceitos desta área de estudo. Na Seção 2.2, são caracterizadas as principais técnicas para a implementação de Sistemas de Recomendação. A Seção 2.3, por sua vez, apresenta as vantagens e as desvantagens de cada uma destas técnicas. A Seção 2.4 fornece exemplos de Cálculos de Similaridade e de como eles podem ser utilizados nas diversas técnicas de recomendação. Por fim, a Seção 2.5 faz a conclusão sobre os conceitos vistos neste capítulo.

2.1 Introdução

Tomar decisões e fazer escolhas sobre os mais variados assuntos são atos cotidianos da maioria das pessoas. Tais escolhas e decisões variam em complexidade e poder de influência e vão desde a escolha de um bom livro até a decisão de qual universidade escolher para um curso de pós-graduação. Em uma situação ideal, existiriam disponíveis todas as informações e experiências possíveis e necessárias para nos permitir fazer a melhor escolha. Contudo, geralmente, este não é o caso e processos decisórios são fontes de dúvidas e inseguranças em grande quantidade de ocasiões.

Esta insegurança é maior, principalmente nas escolhas mais complexas ou que possuem um maior impacto na vida do tomador de decisão. Para superar este problema, geralmente recorre-se a recomendações, que podem nos ajudar a tomar uma decisão mais consciente. As recomendações podem vir de pessoas que tenham mais experiência ou conhecimento em determinado assunto ou que simplesmente já tenham utilizado e avaliado alguma das opções disponíveis no processo de escolha. Podem ainda vir de amigos, de fontes especializadas (revistas, livros, entre outros) ou de serviços contratados.

Na década de 90, a evolução dos Sistemas de Informação, o número cada vez maior de informações em formato digital e a massificação do uso da informática serviram

como ambiente propício para o surgimento de aplicações que vieram para auxiliar nesses processos decisórios. Estes sistemas são chamados de Sistemas de Recomendação, que vêm, desde então, auxiliando em várias circunstâncias do dia a dia. Principalmente a partir do ano 2000, os Sistemas de Recomendação tem sido alvo de estudo e pesquisa das mais diversas áreas da Ciência da Computação, dentre elas: Inteligência Artificial, Mineração de Dados, Segurança e Computação Ubíqua (Adomavicius and Tuzhilin, 2008). Em 2006, os Sistemas de Recomendação ganharam também grande destaque mundial com o *Netflix Prize*¹, concurso que oferecia 1 milhão de dólares para quem melhorasse em pelo menos 10% a precisão do algoritmo de recomendação de filmes da empresa *Netflix*.

Tradicionalmente, os Sistemas de Recomendação lidam com dois tipos de entidade: usuários e itens. Os itens são os objetos que serão recomendados e, possivelmente, avaliados pelos usuários (Adomavicius and Tuzhilin, 2008). A principal responsabilidade de um Sistema de Recomendação é prever quais os itens que melhor correspondem às necessidades e preferências de um usuário. Pelo fato das preferências dos usuários possuírem um caráter majoritariamente subjetivo, a personalização é uma característica sempre presente em Sistemas de Recomendação.

Um sistema que sempre sugere os mesmos itens a todos os usuários (e.g., um sistema de comércio eletrônico que recomenda somente os itens mais vendidos) sem levar em consideração o perfil, histórico, característica, desejo particular ou situação de cada usuário, não pode ser considerado um Sistema de Recomendação (Jannach et al., 2011).

Um conceito importante em Sistemas de Recomendação é a avaliação que o usuário faz de um item. Esta avaliação é basicamente uma forma de capturar a preferência de um usuário específico para um determinado item. Ela pode ser graduada de diferentes formas, como, por exemplo, a partir de uma nota na escala de 1 a 5 (e.g., atribuindo uma quantidade de “estrelas”, como nas primeiras versões do *Youtube*²) ou somente uma avaliação “positiva” ou “negativa”.

A forma de obtenção da avaliação pode ser tanto explícita quanto implícita. Na obtenção explícita, por exemplo, o usuário pode preencher um formulário no qual fornece uma nota a cada item (e.g., atribuição de “estrelas” no *Youtube*). Na forma implícita, pelo comportamento do usuário, o sistema consegue inferir a avaliação, como, por exemplo, no

¹ <http://www.netflixprize.com/>

² <http://www.youtube.com/>

sistema online de recomendação de músicas *Last.fm*³, quando o usuário escuta constantemente determinada música, o sistema infere que ela tem uma avaliação positiva para aquele usuário.

2.2 Tipos de Recomendação

Sistemas de Recomendação podem se utilizar de diferentes técnicas e fontes de dados na realização de seu objetivo. De acordo com Jannach et al. (2011), as técnicas tradicionais de Sistemas de Recomendação são divididas em três paradigmas principais:

- a) Recomendação Colaborativa;
- b) Recomendação Baseada em Conteúdo;
- c) Recomendação Baseada em Conhecimento.

Uma das principais diferenças entre as técnicas está na fonte de conhecimento utilizada. Determinadas técnicas utilizam, como entrada, dados que não são necessários para outras técnicas. Os tipos de dados existentes ou a sua dificuldade de obtenção são fatores de grande influência na escolha da técnica de recomendação a ser utilizada. Burke (2007) categoriza as fontes de conhecimento utilizadas pelas abordagens listadas de acordo com a Figura 1. As subseções seguintes detalham o uso das fontes de conhecimento em cada técnica de recomendação.

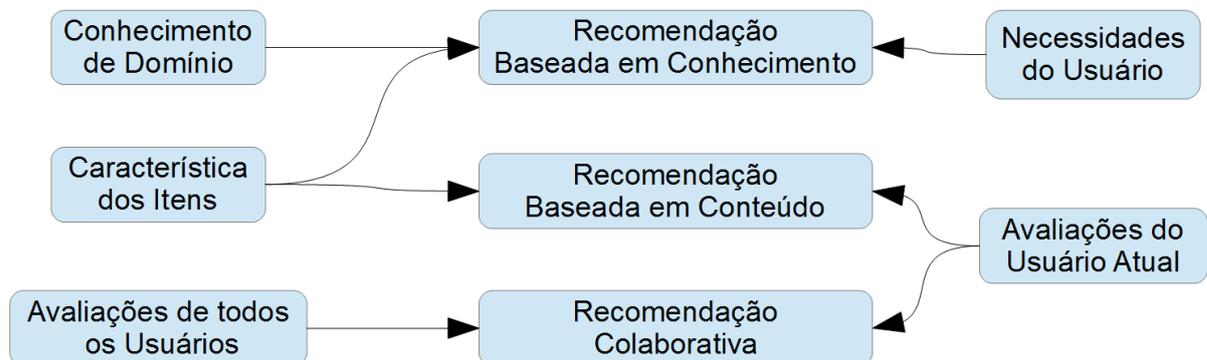


Figura 1 – Técnicas de Recomendação e suas Fontes de Conhecimento, adaptado de (Burke, 2007)

³ <http://www.last.fm/>

Estas abordagens de recomendação podem ser combinadas em Sistemas de Recomendação híbridos, ajudando a evitar certas limitações de cada método. As próximas subseções apresentam as características de cada abordagem.

2.2.1 Recomendação Colaborativa

É comum, quando uma pessoa busca opiniões, dar mais importância àquelas que veem de pessoas que conhecidamente têm preferências parecidas com as dela. Por exemplo, quando se deseja comprar algum álbum de músicas, uma pessoa procura opiniões de amigos que compartilham o mesmo gosto musical e tende a dar menos importância às opiniões daqueles que se sabe que o gosto musical não agrada. Se um produto foi avaliado positivamente por quem tem um gosto similar, é provável que também seja bem avaliado pela pessoa. Este é um processo colaborativo, já que se depende das avaliações de outras pessoas para ajudar na escolha.

Sistemas que incorporam a técnica da Recomendação Colaborativa utilizam desta mesma abordagem e partem do princípio de que usuários que demonstraram o mesmo interesse no passado irão, provavelmente, continuar com interesses similares no futuro. Ao se fazer uma recomendação a um usuário específico, dá-se prioridade aos itens avaliados como positivos por usuários com perfil similar.

A compatibilidade de interesses entre dois usuários é medida pela semelhança nas avaliações de itens que foram avaliados por ambos. Quanto mais avaliações parecidas dois usuários tiverem, mais similares eles serão no processo de recomendação. Para um Sistema de Recomendação de música, por exemplo, imaginemos que tenha que se decidir se uma determinada música deve, ou não, ser recomendada a um usuário. Um algoritmo de recomendação puramente colaborativo procuraria avaliações dessa música feita por usuários que tenham apresentado no passado avaliações similares ao do usuário em questão. Se um grande número dessas avaliações tiverem sido positivas, a música estará bem ranqueada para a recomendação.

A Figura 2 mostra um exemplo de recomendação colaborativa do site *Amazon.com*⁴. Nesse site, na página de descrição de um produto, são sugeridos outros produtos que foram comprados, por outros usuários, em conjunto com o item analisado.



Figura 2 – Recomendação Colaborativa do site Amazon.com

2.2.2 *Recomendação Baseada em Conteúdo*

Nos processos de decisão, é comum que, com o passar do tempo, as escolhas se concentrem em determinadas sub-categorias de itens. Ainda no exemplo de compra de álbuns de músicas, geralmente, escolhas se concentram em um subconjunto de gêneros musicais (e.g., um usuário com histórico de compra predominantemente de discos de Rock, Jazz e Blues). Tais preferências históricas podem ajudar na construção de um perfil do usuário e serem utilizadas no momento da recomendação. Ao utilizar esta abordagem, a Recomendação Baseada em Conteúdo parte do princípio de que um determinado usuário terá maior interesse em itens que tenham similaridade a itens avaliados, por ele mesmo, como positivos. Esta técnica se concentra no conteúdo dos itens porque é a partir dele que as características dos itens são extraídas e utilizadas para montagem do perfil do usuário.

Algoritmos deste paradigma dependem de informações extraídas dos itens passíveis de recomendação para que se consiga determinar quão similar um item é em relação a outro. No caso de um Sistema de Recomendação de fotos, um algoritmo puramente baseado

⁴ <http://amazon.com/>

em conteúdo, ao decidir se uma determinada foto será, ou não, recomendada ao usuário, procurará avaliações, feitas pelo mesmo usuário, de fotos similares à foto em questão. A similaridade poderia ser calculada utilizando as anotações da foto (e.g., *tags*) ou informações extraídas do seu conteúdo (e.g., textura, cores). Se a maioria das avaliações das fotos similares encontradas tiver sido positiva, a foto estará bem ranqueada para a recomendação.

A Figura 3 mostra um exemplo de *Recomendação Baseada em Conteúdo* do site *Last.fm*⁵. Nesse site, são recomendadas ao usuário bandas que tenham características similares a outras bandas que o mesmo usuário já escutou.

Figura 3 – Recomendação Baseada em Conteúdo do site *Last.fm*

2.2.3 *Recomendação Baseada em Conhecimento*

Pode-se adotar uma abordagem mais direta em um processo em que já se conhece, ou há meios de se extrair, o desejo do usuário e consegue-se, dentro do conjunto de itens e do conhecimento de suas características, identificar aqueles itens que mais se adequam ao desejo do usuário considerado. Essas preferências do usuário podem ser elicitadas de várias formas como, por exemplo, o usuário informa quais os tipos ou categorias de itens que mais lhe interessam (e.g., na recomendação de hotéis, o usuário informa a preferência por hotéis 3 estrelas, que possuam internet e café da manhã) ou indicam a preferência por itens semelhantes a algum outro (e.g., restaurantes de culinária francesa semelhantes a algum outro que ele tenha visitado). O Sistema de Recomendação deve, então, procurar em toda sua base, os itens que melhor correspondam às preferências informadas pelo usuário.

⁵ <http://www.last.fm/>

Este tipo de recomendação é comum em lojas virtuais, que possuem produtos divididos em categorias, cada uma delas possuindo um conjunto de características que podem ser utilizadas pelo usuário na escolha do produto que deseja comprar.

Na categoria de telefones celulares, por exemplo, o usuário pode escolher os produtos de acordo com sua adequação às características consideradas, por ele, relevantes. Exemplos de características para telefones celulares são: tamanho da tela, tipo de teclado, sistema operacional, marca, modelo, entre outras. Na realização da recomendação, o sistema pode adotar uma estratégia restritiva, em que, para cada valor de característica informado pelo usuário (e.g., celulares com tela de 3 polegadas), é criada uma restrição na consulta e são retornados somente os itens que atendam a todas as restrições.

Outro tipo de estratégia é baseado em similaridades, em que o usuário informa, por exemplo, as características que ele deseja maximizar (e.g., memória do celular) e as que deseja minimizar (e.g., preço). O sistema então verifica os produtos que mais se aproximam das características consideradas ideais pelo usuário (e.g., menor preço e maior quantidade de memória) e retorna uma lista ranqueada de acordo com o valor de similaridade calculada. O usuário pode ainda informar o peso da relevância dado a cada característica (e.g., a quantidade de memória é mais importante do que o poder de processamento) e o sistema os utilizar para melhor ranquear os itens que tenham maior similaridade nas características de maior relevância para o usuário.

Os itens passíveis de recomendação necessitam ser anotados com características objetivas para que possam ser avaliados quanto a sua adequação às preferências do usuário e deve-se possuir um grande conhecimento do domínio da aplicação para que regras de similaridade possam ser implementadas.

A Figura 4 mostra a tela de recomendação de elementos de código fonte da aplicação *SDiC* (Antunes et al., 2012). Nesta aplicação os elementos de código (e.g., métodos, classes, interfaces) são recomendados de acordo com o elemento sendo editado no momento pelo usuário (e.g., uma classe de código Java). São aplicadas regras de domínio que indicam quais os elementos que possuem uma maior associação com o elemento editado. As recomendações são apresentadas em uma “nuvem” de artefatos, exemplificada na Figura 4.

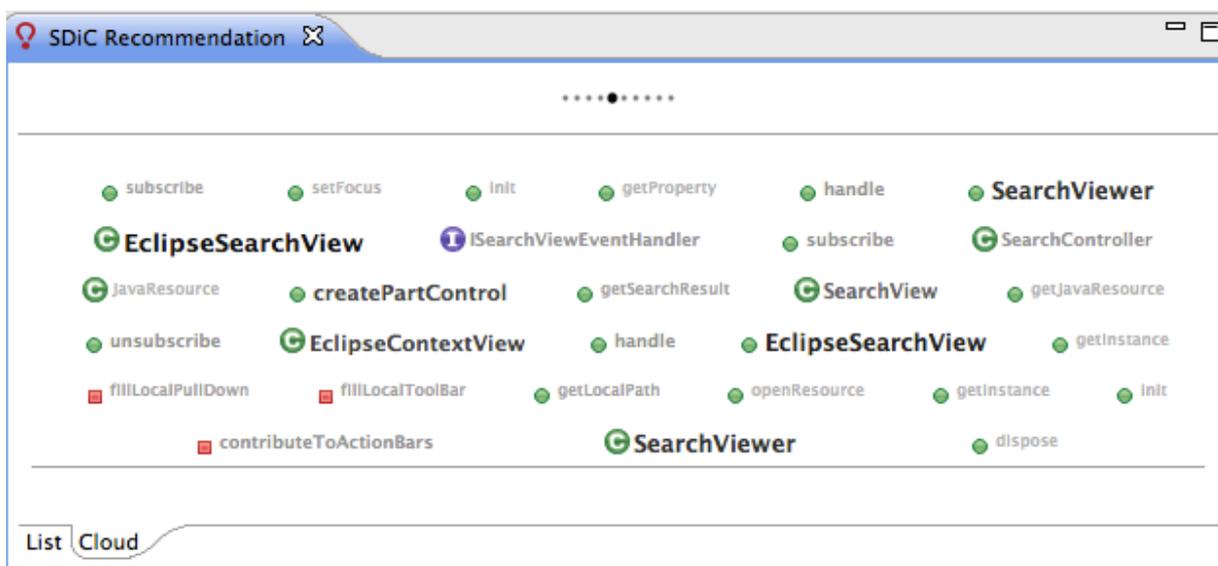


Figura 4 – Recomendação de Elementos de Código Fonte (Antunes et al., 2012)

2.3 Vantagens e Limitações das Técnicas de Recomendação

Cada técnica contém suas vantagens e limitações e entendê-las é de suma importância para avaliar a aplicabilidade das mesmas.

A Recomendação Colaborativa, por não utilizar conteúdo e nem propriedades dos objetos recomendados, possui como facilitador a não dependência de informações sobre os itens. Esta independência das informações dos itens torna mais simples a modelagem do sistema, principalmente quando grandes e heterogêneas bibliotecas de itens são utilizadas.

Por depender da presença de um histórico de avaliações, a eficiência da Recomendação Colaborativa é bastante dependente do tamanho de sua base de usuários e da quantidade de itens que cada usuário avaliou e geralmente só passa a ter resultados satisfatórios a partir de uma intensa utilização. Esta limitação ocasiona também o problema do início frio (*cold start problem*) que afeta tanto usuários novos do sistema, quanto novos itens adicionados à biblioteca de recomendação. Para o novo usuário, o sistema não pode depender das avaliações para o cálculo de similaridade, pois tal usuário ainda não realizou nenhuma avaliação. Similarmente, um item recém-adicionado à biblioteca ainda não foi avaliado por nenhum usuário, dessa forma ele será preterido no momento de geração do *ranking*. Estas limitações são ainda mais comprometedoras em domínios em que se demora a gerar uma extensa base histórica de avaliações como, por exemplo, em recomendações de imóveis, já que costumamos comprar um imóvel com bem menos frequência do que compramos outros

itens de menor durabilidade (Jannach et al., 2011). Outro problema desta abordagem está relacionado a matriz de avaliações (itens *vs* usuários) que é extremamente esparsa quando o número de usuários ou itens é muito grande.

A Recomendação Baseada em Conteúdo possui como vantagem a não necessidade de uma grande comunidade de usuários. Mesmo que o sistema só possua um único usuário, ele deve ser capaz de realizar a recomendação, desde que, para isso, tenha conhecimento sobre o perfil do usuário e meios de extrair e classificar o conteúdo de cada item. Porém, a caracterização adequada do perfil do usuário pode ser um processo demorado, muitas vezes necessitando que o usuário interaja com o sistema por diversas vezes. Este custo torna ineficiente a recomendação para novos usuários e os sistemas são afetados também pelo problema do início frio. Por valorizar bastante itens similares aos já avaliados positivamente pelo usuário, a Recomendação Baseada em Conteúdo corre ainda o risco da sobre-especialização, em que a recomendação se concentra majoritariamente em itens com características muito parecidas e não recomenda itens com outras características mas que possam também despertar o interesse do usuário. Este problema pode ser atenuado pela introdução de Diversidade (Ziegler et al., 2005) na lista de itens recomendados. Um conjunto de itens recomendados é considerado diverso se não há concentração de itens com características muito semelhantes. A Novidade também é importante para contornar este problema. A Novidade se caracteriza pela capacidade do sistema recomendar itens diferentes dos que já foram previamente vistos pelo usuário (Vargas and Castells, 2011).

A Recomendação Baseada em Conhecimento, ao realizar uma abordagem mais direta, em que recomenda-se os itens que melhor se adequam às preferências do usuário, não necessita da manutenção de um histórico de avaliações, já que o próprio usuário, na maioria das vezes, já informa explicitamente no que está interessado. A não necessidade do histórico de avaliações permite ao sistema fazer recomendações adequadas a usuários que estão utilizando o sistema pela primeira vez, não apresentando o problema do início frio.

A Recomendação Baseada em Conhecimento também permite que itens, mesmo que nunca tenham sido avaliados por outros usuários, sejam recomendados, bastando que para isso suas características sejam as mais adequadas ao desejo do usuário. Uma das limitações desta técnica é a necessidade, em muitas das ocasiões, da intervenção do usuário para que ele informe suas preferências. A descoberta das preferências ou expectativas do usuário representa um desafio em Sistemas de Recomendação baseados em conhecimento. Alguns sistemas utilizam questionários e formulários que devem ser preenchidos pelo usuário antes que a recomendação possa ser realizada (Felfernig et al., 2006). Outra forma de extrair as

preferências do usuário é através de críticas que este faz a itens recomendados, permitindo que o sistema ajuste a recomendação em um processo incremental (Burke, 2002). O ponto comum entre estas abordagens é que as preferências do usuário são licitadas através de interações deste com o sistema. Estas intervenções podem ser vistas, por alguns usuários, como incômodas e acabar desestimulando-o a utilizar o sistema (Burke, 2000).

Outra dificuldade que pode ser encontrada em Sistemas de Recomendação baseados em conhecimento é a necessidade de informações detalhadas sobre as características dos itens. Em um Sistema de Recomendação de hotéis, exemplos de características que devem ser levantadas são: faixa de preço, estrelas atribuídas, localização, disponibilidade, opções de lazer, acessibilidade, entre outras. Muitas dessas informações necessitam de intervenção humana para serem preenchidas, podendo tornar a criação e manutenção das bases de itens um processo bastante oneroso, principalmente em bases grandes. Além disso, em alguns domínios, o acesso a informações sobre os itens pode ser dificultado pela necessidade de integração de diversas fontes de conhecimento (e.g., bases externas).

2.4 Cálculo de Similaridade

Um conceito central das três técnicas de recomendação apresentadas na Seção 2.2 é a necessidade de se realizar cálculos de similaridade. Cada técnica utiliza cálculos de similaridade de maneiras diferentes.

Na Recomendação Colaborativa, em que recomendamos ao usuário itens avaliados positivamente por usuários semelhantes a ele, precisamos calcular a similaridade entre usuários. Existem na literatura várias técnicas de cálculo de similaridade entre usuários aplicáveis a este tipo de recomendação. O ponto comum entre elas é o grande peso que a avaliação feita pelos usuários tem no resultado da similaridade.

A Tabela 1 exemplifica um hipotético histórico de avaliações feitas por três usuários que deram notas entre 1 e 5 para alguns itens. Percebe-se facilmente que as avaliações do Usuário 1 são mais próximas das avaliações do Usuário 3 em comparação às avaliações do Usuário 2.

	Item 1	Item 2	Item 3	Item 4
Usuário 1	2	1	3	5
Usuário 2	5	4	1	2
Usuário 3	2	2	2	4

Tabela 1 – Exemplo de Avaliações

Técnicas de cálculo de similaridade, como o Coeficiente de Correlação de Pearson vista na Fórmula 1, podem ser utilizadas nas amostras de avaliações disponíveis para calcular o quão similar um usuário é de outro. Na Fórmula, P é o conjunto de itens, $r_{a,p}$ é a avaliação que o usuário a fez para o item p e \bar{r}_a é a média de todas avaliações feitas pelo usuário a . A técnica de Pearson retorna um valor entre -1 e +1 ao dividir a covariância das avaliações de dois usuários pelo produto de seus desvios padrão. Quanto mais alto o valor retornado, mais similares são os usuários.

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

Fórmula 1 – Correlação de Pearson para Cálculo de Similaridade entre dois Usuários (Resnick et al., 1994)

Na Recomendação Baseada em Conteúdo, em que recomenda-se ao usuário itens semelhantes a outros que ele avaliou positivamente, precisa-se calcular a similaridade entre itens. Neste tipo de recomendação, deseja-se saber a similaridade de um item não avaliado pelo usuário em relação aos itens já avaliados por ele. O *Coeficiente de Dice*, exemplificado na Fórmula 2, pode ser utilizado em alguns domínios, retornando um valor entre zero e um e quanto mais alto o valor retornado, maior a similaridade. Para calcular a similaridade entre livros ou artigos, por exemplo, pode-se levar em consideração o tamanho da interseção de suas palavras chave (Jannach et al., 2011). As palavras chave de documentos textuais podem ser informadas manualmente ou podem ser utilizados meios automáticos de extração.

$$2 \times \frac{|palavrasChave(livro_a) \cap palavrasChave(livro_b)|}{|palavrasChave(livro_a)| + |palavrasChave(livro_b)|}$$

Fórmula 2 – Coeficiente de Dice na comparação de livros

Na Recomendação Baseada em Conhecimento, desejamos selecionar aqueles itens que mais se adequam (tenham uma maior similaridade) às necessidades do usuário. Neste tipo

de recomendação, deve-se primeiramente extrair todos os requisitos do usuário, que podem ser de diferentes categorias. Em um Sistema de Recomendação de televisores, por exemplo, duas categorias de requisitos poderiam ser preço e tamanho do visor. Posteriormente, toda a biblioteca de itens é percorrida e, para cada item, os valores dos atributos de cada categoria são também extraídos. Após a extração dos valores, dependendo do domínio, existem várias medidas de similaridades que podem ser utilizadas. A Fórmula 3 mostra o cálculo de similaridade proposto por (Mcsberry, 2003) aplicado na comparação de um item às preferências do usuário.

$$\text{similaridade}(item, Pref) = \frac{\sum_{p \in Pref} w_p * \text{sim}_p(Pref, item)}{\sum_{p \in Pref} w_p}$$

Fórmula 3 – Similaridade entre um Item e a Preferência do Usuário - (Mcsberry, 2003)

A Fórmula 3 faz o somatório do cálculo de similaridade individual de cada tipo de característica do item contida no conjunto de preferências do usuário. A função sim_p é particular de cada característica e deve se adequar a sua semântica e forma de representação. O valor de cada similaridade individual pode ainda ser ponderado pelo coeficiente w_p e deve refletir o grau de importância que o usuário dá a determinada característica do item. Em atributos numéricos (e.g., tamanho do visor de uma televisão), em que a similaridade é definida por quão próximos são dois valores, a função sim_p pode seguir a Fórmula 4.

$$\text{sim}_c(U, I) = 1 - \frac{|V_c(U) - V_c(I)|}{\max(c) - \min(c)}$$

Fórmula 4 – Similaridade pela Proximidade de Valores Numéricos (Jannach et al., 2011)

Na Fórmula 4, $V_c(U)$ e $V_c(I)$ representam, respectivamente, os valores da característica c para o usuário e o item. Por sua vez, os valores $\max(c)$ e $\min(c)$ representam, respectivamente, o valor máximo e mínimo para o atributo comparado do característica c .

Em outros casos, em que a característica é representada por uma hierarquia (e.g., uma notícia jornalista pode ser mapeada na categoria Esporte → Futebol → Campeonato Carioca) a similaridade pode ser calculada pela distância dos elementos ao nó raiz, de acordo com a Fórmula 5 (Slimani et al., 2006). Nesta fórmula, N representa a distância entre o mais próximo nó ancestral comum de U e I em relação ao nó raiz da hierarquia. N_u e N_i representam a distância de U e I , respectivamente, ao nó raiz da hierarquia; $PF(U, I)$ é o fator de penalização para dois nós vizinhos.

$$\text{sim}_c(U, I) = \frac{2 * N}{N_u + N_i} * PF(U, I)$$

Fórmula 5 – Similaridade para Valores Representados em uma Hierarquia (Slimani et al., 2006)

Neste trabalho, foram exploradas diversas fórmulas de cálculo de similaridade, descritas na Seção 4.4 na comparação de informações contextuais tanto de usuários quanto de itens.

2.5 Conclusão

Este capítulo apresentou a motivação e as principais abordagens na construção de Sistemas de Recomendação. Esta categoria de sistema possui uma grande importância e apresenta muitas oportunidades de ganhos na experiência obtida por usuários de Sistemas de Informação. As principais técnicas de implementação de Sistemas de Recomendação foram apresentadas. Essas técnicas utilizam diferentes tipos de dados como entrada no momento da realização da recomendação, cada uma possuindo vantagens e desvantagens e cuja criticidade pode variar de acordo com o domínio da aplicação. Como apresentado, estas técnicas utilizam diversas variações de cálculo de similaridade que são adequadas a cada técnica e domínio de aplicação. No próximo capítulo será mostrado como os Sistemas de Recomendação podem se beneficiar da sensibilidade ao contexto e retornar aos usuários resultados mais adequados à situação em que se encontram.

3 SENSIBILIDADE AO CONTEXTO

Este capítulo apresenta uma visão geral sobre Sistemas Sensíveis ao Contexto e sua aplicação em Sistemas de Recomendação. O capítulo descreve como a utilização de informações de contexto podem beneficiar sistemas de informação em vários aspectos, tais como: a adaptação do comportamento e interface, descoberta de serviços, disponibilização de conteúdo e, também, a recomendação de itens.

A Seção 3.1 aborda a motivação para o surgimento de aplicações sensíveis ao contexto, os tipos de modelagem e representação de contexto mais comumente utilizados e fornece exemplos de infraestruturas de suporte de aplicações sensíveis ao contexto. As Seções 3.2 e 3.3 dão exemplos de sistemas sensíveis ao contexto no domínio de aplicações multimídia e de recomendação, respectivamente. A Seção 3.4 descreve as características dos trabalhos relacionados e como o trabalho proposto nesta dissertação se diferencia das aplicações estudadas. Por fim, a Seção 3.5 faz a conclusão do capítulo e uma breve descrição de como as técnicas descritas auxiliaram na elaboração da proposta de mestrado.

3.1 Introdução

A evolução dos dispositivos computacionais tornou possível que aparelhos antes limitados a um ambiente fixo (e.g., desktops) pudessem ser substituídos por aparelhos móveis (e.g., laptops, PDAs, smartphones e tablets). A mobilidade do usuário foi um dos principais ganhos advindos com a portabilidade e acarretou em uma maior dinamicidade dos ambientes em que os usuários utilizam os dispositivos computacionais e suas aplicações. Essa grande disponibilidade de recursos computacionais se aproxima da visão de computação ubíqua prevista por Weiser (1993). A Computação Ubíqua se caracteriza pelo “melhora do uso de computadores, tornando-os disponíveis por todo o ambiente físico sendo, porém, praticamente invisíveis ao usuário” (Weiser, 1993, tradução nossa). A mobilidade e o uso de sensores capazes de capturar o contexto ao seu redor possibilitou a construção de sistemas que podem adaptar seu comportamento de acordo com o contexto em que são executados.

De acordo com Dey et al. (2000),

“Contexto é qualquer informação que pode ser utilizada para caracterizar a situação de uma entidade. Uma entidade é uma pessoa, lugar ou objeto que é considerado relevante para interação entre o usuário e a aplicação, incluindo o próprio usuário e a aplicação” (Dey and Abowd, 2000, tradução nossa).

Sistemas sensíveis ao contexto, por sua vez, são caracterizados como “sistemas que usam o contexto para disponibilizar informações relevantes e/ou serviços para o usuário, onde a relevância depende da ação do usuário” (Dey and Abowd, 2000, tradução nossa). A sensibilidade ao contexto é vista como a capacidade de perceber a situação do usuário em vários aspectos, e adaptar o comportamento do sistema em conformidade (Marinho et al., 2010). Assim, o termo "sensível ao contexto" está associado com os sistemas que guiam o seu comportamento de acordo com seu contexto de uso sem a necessidade de intervenção explícita do usuário.

Nesse cenário, a sensibilidade ao contexto tornou-se um campo bastante relevante e tem sido utilizada para aumentar a eficiência e usabilidade de sistemas, particularmente, nas aplicações para dispositivos móveis (Baldauf et al., 2007). Um sistema de anotação de fotos sensível ao contexto, por exemplo, pode considerar como elementos contextuais de uma foto: a localização do usuário (e.g., coordenadas GPS e endereço), a informação temporal de criação da foto (e.g., dia da semana, ano, estação do ano) e dados sobre as pessoas próximas no instante em que a foto foi tirada.

A localização do usuário é uma das características de contexto mais utilizadas, principalmente nos primeiros sistemas sensíveis ao contexto, porém existem diversos outros tipos de informações contextuais que também podem ser consideradas e cuja utilização depende do domínio e do objetivo da aplicação (Schmidt et al., 1999). No domínio de aplicações multimídia, por exemplo, Viana et al. (2008) definem cinco dimensões para a representação de informações contextuais:

- a) Espacial: endereço, coordenadas geográficas, entre outras;
- b) Temporal: horário, dia da semana, estação do ano, entre outras;
- c) Social: se o usuário está acompanhado da família, amigos ou se está sozinho, entre outras;
- d) Computacional: poder de processamento do dispositivo, carga da bateria, entre outras;
- e) Espaço-temporal: temperatura, condição de luminosidade, entre outras.

A sensibilidade ao contexto é de essencial importância em sistemas móveis, pois é a partir dela que estes sistemas podem se adaptar às mudanças de situação do ambiente (Wang et al., 2004). A incorporação de contexto em aplicações pode beneficiar suas funcionalidades em diversas formas:

- a) Adaptando a interface do sistema e o seu comportamento de execução;
- b) Aumentando a relevância do conteúdo disponibilizado ao usuário;
- c) Ajudando na descoberta de serviços do ambiente ou remotos; e
- d) Oferecendo contexto e semântica às atividades desempenhadas pelos usuários.

Este tipo de aplicação sensível ao contexto motivou o surgimento de propostas de infraestruturas de suporte a esses sistemas que pudessem melhor contemplar suas necessidades específicas e melhor implementar suas funcionalidades com o intuito de prover reuso de código e facilitar a evolução destas aplicações. Geralmente, sistemas sensíveis ao contexto seguem uma arquitetura, em linhas gerais, uniforme. Aplicações móveis ou sensores ambiente capturam informações iniciais que correspondem a uma descrição de baixo nível do contexto (e.g., coordenadas *GPS*). Em seguida, o sistema acessa outras fontes de dados ou agrega informações dos vários sensores para inferir informações contextuais de alto nível (e.g., um endereço, o nome do amigo próximo), realizando um processo de enriquecimento contextual. Mecanismos de inferência também podem ser usados para derivar informações de acordo com regras pré-estabelecidas (e.g., regras aplicadas a ontologias). Finalmente, o contexto resultante é explorado pela tarefa sensível ao contexto do sistema, tais como, adaptação do sistema (Gu et al., 2004) e anotação de fotos (Viana et al., 2011).

Ailisto et al. (2002) propõem uma arquitetura em camadas para a implementação de uma aplicação sensível ao contexto, mostrada na Figura 5. Nesta arquitetura, as responsabilidades são divididas em cinco camadas, permitindo a modularização dos componentes e uma possível reutilização e compartilhamento dos serviços implementados.

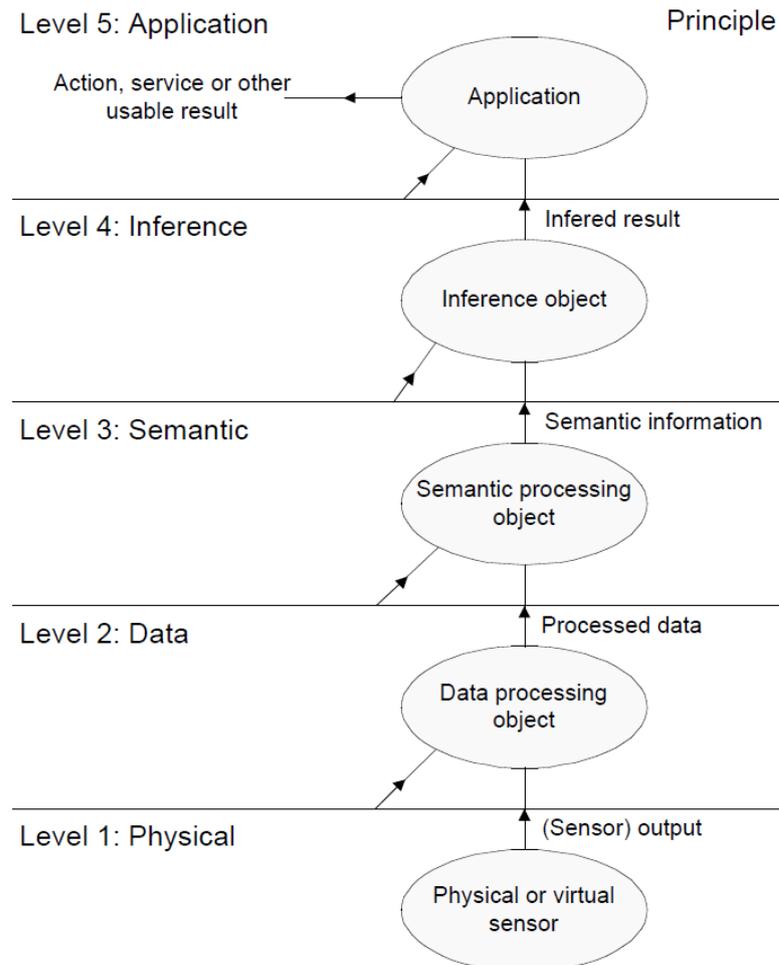


Figura 5 – Arquitetura em Camadas para Aplicações Sensíveis ao Contexto proposta em Ailisto et al., 2002

As cinco camadas possuem as seguintes responsabilidades:

1. **Camada Física** - Responsável por recuperar as informações do ambiente através de sensores. Por exemplo, o sensor *Bluetooth* do celular do usuário recupera os sinais enviados por dispositivos próximos;
2. **Camada de Dados** – Responsável por processar os dados adquiridos da Camada Física. Por exemplo, o celular analisa os sinais recebidos pelo sensor Bluetooth e recupera o identificador de cada dispositivo próximo;
3. **Camada de Semântica** – Responsável por dar significado aos dados obtidos. Por exemplo, o sistema reconhece dois dos dispositivos próximos, um sendo um laptop do local de trabalho e outro sendo do diretor da empresa em que o usuário trabalha;
4. **Camada de inferência** – Aplica regras de inferência a partir dos dados coletados para, por exemplo, identificar situações em que o usuário se encontra.

Por exemplo, o sistema infere, por ser horário comercial e pelos dispositivos próximos, que o usuário provavelmente está em uma reunião de trabalho; e

5. **Camada de Aplicação** – Responsável por adaptar a aplicação a partir das informações de contexto coletadas. Por exemplo, identificando que o usuário está em uma reunião de trabalho, o sistema coloca o celular no modo silencioso e manda um SMS a todos que ligarem para aquele celular, informando o motivo do usuário não poder atender às ligações.

Uma arquitetura para aplicações sensíveis ao contexto deve, ainda, contemplar requisitos específicos da aplicação sendo construída. Pinheiro et al. (2008), por exemplo, ao tornar personalizável um sistema sensível ao contexto, associa o elemento Perfil aos elementos de contexto. O objetivo da associação é permitir que outras características do usuário, que não somente as informações de contexto, sejam consideradas no processo de adaptação da aplicação

A arquitetura de referência da linha de produto *MobiLine* (Marinho et al., 2010), por outro lado, adota uma abordagem orientada a serviços, que facilita a decomposição da aplicação em partes menores atômicas. A Figura 6 mostra a arquitetura de referência do *MobiLine*. Nela estão presentes quatro camadas:

- a) **Camada de Sensores de Contexto:** fornece uma abstração sobre como as informações contextuais são capturadas para, posteriormente, serem manipuladas pelas camadas superiores;
- b) **Camada de Gestão de Contexto:** é responsável por armazenar, no Repositório de Contexto, as informações de contexto providas pela camada de Sensores de Contexto;
- c) **Camada de Serviço:** responsável por gerenciar o ciclo de vida dos serviços e possibilitar a comunicação entre serviços e aplicações; e
- d) **Camada de Aplicação:** é formada pelas aplicações móveis sensíveis ao contexto.

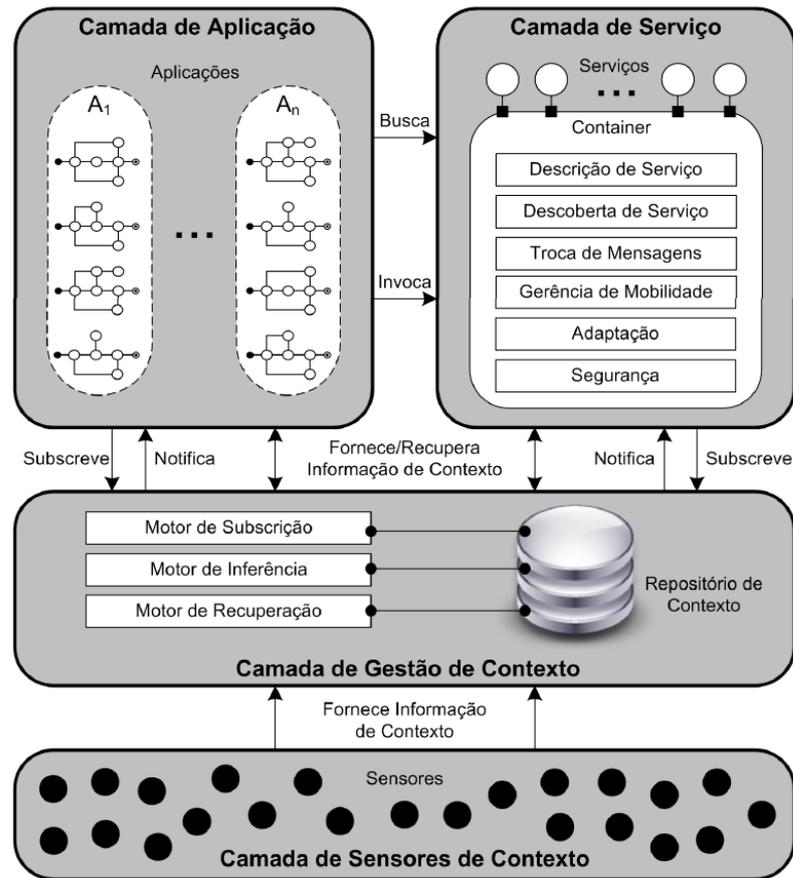


Figura 6 – Arquitetura de referência da linha de produto *MobiLine* (Marinho et al., 2010)

Esses dois exemplos (Pinheiro et al., 2008 e Marinho et al., 2010) de arquitetura são uma evolução à proposta genérica da arquitetura vista na Figura 5. De toda forma, porém, a necessidade de divisão de responsabilidades para a aquisição e tratamento das informações contextuais permanece, substancialmente, a mesma.

Uma parte fundamental do processo de desenvolvimento de aplicação ou sistema sensível ao contexto é definir quais informações devem compor o contexto, visto que os elementos que o descrevem dependem da tarefa do sistema e da capacidade do mesmo em observar estas informações (Pinheiro et al., 2008) (Viana et al., 2008). Esta definição está associada à criação de um modelo de contexto no qual se determinam os elementos que compõem a sua descrição e como ele deverá ser representado (e.g., usando ontologias, *XML*, entre outras) e armazenado, permitindo que os dados contextuais possam ser processados por uma máquina (Baldauf et al., 2007). Strang e Linnhoff-popien (2004) enumeram seis abordagens relevantes de modelagem de contexto:

- a) **Chave-valor:** o contexto é descrito a partir de um conjunto de atributos chave-valor (e.g., Localização:Brasil). Possui como vantagem a simplicidade,

porém a falta de uma estrutura mais elaborada limita os algoritmos que podem ser aplicados;

- b) **Linguagem de marcação:** o contexto é descrito através de alguma linguagem de marcação (e.g., *XML*). Permite a utilização de *profiles XML*, utilização de um vocabulário mais rico (e.g., hierarquia de elementos) e validação parcial a partir de restrições da marcação (e.g., *XML schema*);
- c) **Modelos Gráficos (e.g., *UML*):** podem permitir uma boa definição de níveis estruturais e derivar código fonte. Geralmente não possuem um alto nível de formalidade e são mais utilizados por humanos em atividade de modelagem do que pelo sistema em sua execução;
- d) **Modelos Orientados a Objetos:** utilizam o paradigma da Orientação a Objetos para obter benefícios como encapsulamento e reusabilidade;
- e) **Modelos Lógicos:** modelam o contexto a partir de fatos, expressões e regras. A partir das regras, novos fatos e expressões podem ser derivados. Possuem um alto nível de formalismo e uma maior dificuldade de lidar com informações contextuais incompletas e ambíguas; e
- f) **Modelos Baseados em Ontologias:** são ideais para representação de conceitos e relações. Possuem uma estrutura de dados facilmente entendida por computadores e um nível adequado de formalismo;

Modelos baseados em ontologias foram considerados por Strang et. al (Strang and Linnhoff-popien, 2004) como os que melhor atendem os requisitos de composição distribuída, validação parcial, tratamento de informações incompletas e ambíguas, formalismo e aplicabilidade nos ambientes existentes. Bettini et al. (2010) fazem um estudo similar, em que aponta como pontos fortes de modelo baseado em ontologias:

- a) Capacidade de lidar com informações heterogêneas, provindas de diversas fontes de informação;
- b) Facilidade na representação de relacionamentos e dependências entre os diversos tipos de contexto; e
- c) Capacidade de realização de inferências, gerando novos fatos a partir das informações existentes.

Entre as desvantagens apontados por Bettini et al., (2010) no uso de ontologias na modelagem de informações contextuais estão a ausência de um mecanismo para tratamento da

variação das informações contextuais no decorrer do tempo e meios deficientes de expressar a qualidade da informações e tratar informações imperfeitas. A partir da criação de uma ontologia, Viana et. al (2011) propõem um modelo de contexto para anotação de documentos multimídia (Figura 7) que representa o contexto do usuário no momento de geração do conteúdo. Neste modelo, uma foto é associada a um Contexto, que é caracterizado por uma coleção de Elementos Contextuais que podem descrever informações computacionais, temporais, espaciais, espaço-temporais e sociais.

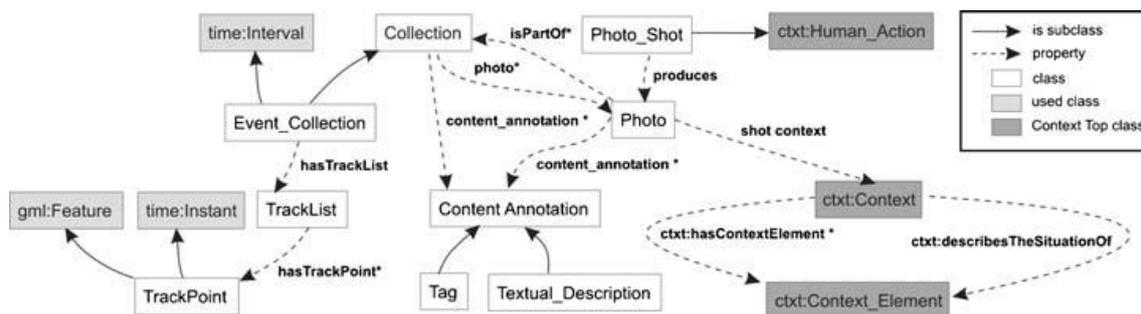


Figura 7 – Modelo de Contexto para Anotação de Documentos Multimídia (Viana et al., 2011)

Quanto a aquisição de informações contextuais, segundo os autores Adomavicius e Tuzhilin (Adomavicius and Tuzhilin, 2008), as informações contextuais podem ser adquiridas de várias maneiras, dentre elas:

- a) **Explicitamente:** é pedido ao usuário que forneça a informação requerida (e.g., preenchendo um formulário ou respondendo algumas perguntas);
- b) **Implicitamente:** a informação é adquirida sem a necessidade de interação com o usuário. Podem, por exemplo, ser utilizados sensores (e.g., *GPS* para o contexto de localização) ou alguma informação presente nos dados da operação (e.g., intervalo de tempo em que determinada atividade é executada); e
- c) **Inferência:** dados estatísticos ou métodos de mineração de dados podem ser utilizados para inferir o contexto. Como vimos no exemplo dado para descrever a arquitetura da Figura 5, a localização e os dispositivos próximos podem ser utilizados para inferir a situação em que o usuário se encontra.

A aquisição do contexto também pode se beneficiar da utilização de serviços web que possibilitam uma maior integração e colaboração entre sistemas. Para descoberta da atividade desempenhada pelo usuário, por exemplo, pode-se utilizar serviços de agenda

eletrônica como o *Google Calendar*⁶ que disponibiliza as informações da agenda de seus usuários através de serviços web. A qualidade e granularidade requeridas de uma informação de contexto podem também ser diferentes de acordo com seu uso. Para um sistema de bilhetes aéreos, por exemplo, pode ser suficiente saber qual a cidade em que o usuário se encontra. Para um sistema turístico, uma localização mais precisa, como endereço e coordenadas geográficas atuais, pode ser necessária.

3.2 Sensibilidade ao Contexto em Sistemas Multimídia

No domínio de aplicações multimídia, a noção de contexto e, principalmente, a sua exploração, é ligeiramente diferente da abordagem vista na Seção 3.1. A Sensibilidade ao Contexto é mais do que um simples mecanismo de adaptação. Esta distinção foi inicialmente analisada em alguns trabalhos que estudaram o comportamento dos usuários para organizar e encontrar fotos (Kennedy et al., 2007). Na verdade, a maioria das informações referidas pelas pessoas sobre suas lembranças de imagens consistem em aspectos relacionados ao seu contexto no momento em que uma foto foi tirada (quando, onde, com quem, entre outros). A popularização de dispositivos móveis equipados com sensores de localização e GIS (Sistemas de Informação Geográficos) forneceram a tecnologia e os dados necessários para desenvolver sistemas multimídia capazes de reunir as informações de contexto desejadas. A sensibilidade ao contexto é aplicada no domínio multimídia para: anotação, indexação, compartilhamento e recomendação de conteúdo multimídia. No decorrer desta seção, são apresentados alguns sistemas multimídia e sensíveis ao contexto. Uma pesquisa mais detalhada dessas abordagens é descrita em Viana et al. (2011) e Braga et al (2011).

PhotoGeo (Figueirêdo et al., 2012), por exemplo, é um sistema que utiliza informações contextuais do ambiente e do momento de geração de fotografias para anotar, classificar e organizar fotos tiradas em situações de mobilidade. O algoritmo é composto de quatro etapas. Na primeira e na segunda etapas, as fotos são agrupadas por localidade e data, respectivamente. Na terceira etapa, possíveis erros no agrupamento são corrigidos a partir de aplicação de regras de domínio. Na quarta etapa, as informações de contexto são utilizadas

⁶ <http://www.google.com/calendar/>

para classificar as fotos de acordo com o evento em que elas foram geradas. No final do processo, os eventos atribuídos às fotos contém as seguintes informações: Nome do Evento, Local de realização do evento, Cidade, Unidade da Federação, País, Data de Início, Duração (e.g., Vila Galé Hotel – Fortaleza – Ceará – Brazil – 05/16/2008). Com a utilização destas informações contextuais, o sistema não somente diminui o esforço do usuário ao organizar automaticamente suas coleções de fotos, como também permite que as fotos sejam mais facilmente pesquisadas e recuperadas.

3.3 Sensibilidade ao Contexto em Sistemas de Recomendação

Sistemas de acesso à informação que não levam em consideração o contexto de uso possuem limitações quanto a sua funcionalidade que, dependendo do domínio, podem trazer resultados insatisfatórios ao usuário. Tais sistemas retornam sempre o mesmo resultado, independente da localização, atividade, pessoas próximas e humor do usuário e independente também da data e horário de uso.

A Sensibilidade ao Contexto em Sistemas de Recomendação foi motivada a partir de pesquisas que reconhecem a dependência das necessidades de longo prazo do usuário em relação ao tempo, localização, e qualquer outra informação sobre o ambiente físico em torno do usuário (Adomavicius et al., 2005) (Boutemedjet, 2010). Ela apresenta um nível adicional de personalização, considerando a influência do ambiente externo ao usuário em seu/sua apreciação dos produtos ou itens. Por depender tanto das informações dos usuários quanto dos itens recomendados, Sistemas de Recomendação podem se beneficiar da sensibilidade ao contexto ao considerar não somente as características intrínsecas de cada entidade, mas também as características de seu contexto. Hoje em dia, a evolução dos dispositivos móveis e das tecnologias de sensores residenciais proporciona uma melhor caracterização do contexto do usuário, permitindo que esta informação seja utilizada no aperfeiçoamento das recomendações.

O contexto do usuário já é explorado em Sistemas de Recomendação. As informações contextuais sobre os itens, porém, ainda permanecem pouco utilizadas em processos de recomendação. Pode-se atribuir esta pouca utilização a dois fatores principais:

- a) A Recomendação Colaborativa, uma das mais populares técnicas de recomendação, não faz uso de todo o conjunto de características dos itens recomendados;
- b) A escassez de bibliotecas de itens ricas em informações contextuais. Somente recentemente, dispositivos e softwares capazes de gerar tais conteúdos ganharam relativa popularidade.

Em processos de recomendação, quanto mais acesso a informações relevantes do usuário, maior a possibilidade de oferecimento de conteúdo personalizado. Quanto mais informações existentes dos itens passíveis de recomendação, maior a possibilidade de escolher aqueles que mais se adequem às necessidades dos usuários. As informações contextuais de usuários e itens oferecem, portanto, um ganho potencial para Sistemas de Recomendação. As informações de contexto do usuário geralmente são utilizadas em dois momentos. O momento mais relevante é o instante em que o usuário está utilizando o sistema. Nesse momento a informação é utilizada para retornar aqueles itens que mais se adequam ao contexto de uso (e.g., recomendar restaurantes próximos ao usuário). O segundo momento, comum em sistemas colaborativos, ocorre na avaliação de itens por usuários. O sistema, além de armazenar o score da avaliação, armazena também o contexto em que o item foi consumido ou utilizado para que possa, dependendo do score atribuído, ser recomendado a outros usuários em contextos similares.

As informações de contexto dos itens podem ser registradas a partir de duas técnicas, dependendo da natureza do item. Uma das técnicas é mais adequada a itens que podem ser gerados em diferentes contextos. Uma vez que o item foi gerado, seu contexto não muda, sendo relevante, portanto, somente o contexto de geração. Isto se aplica, por exemplo, a conteúdos multimídias (e.g., fotos, vídeos) gerados em situações de mobilidade, nas quais pode-se registrar a localização, o instante, os objetos e as pessoas próximas em que o conteúdo foi gerado. A outra técnica exige que as informações de contexto sobre os itens sejam atualizadas com determinada frequência. Ela pode ser utilizada para itens em que o contexto pode mudar e influenciar o interesse que o usuário tem sobre o item. Em um hipotético sistema de recomendações de restaurantes, por exemplo, para a recomendação em tempo real, seria interessante que o sistema mantivesse atualizada a situação do restaurante (e.g., aberto ou fechado, lotado ou com mesas disponíveis, tempo de espera, entre outras). Dessa forma, a recomendação possuiria mais chances de ser adequada às expectativas e necessidades do usuário.

Técnicas de recomendação sensíveis ao contexto já desenvolvidas mostram a viabilidade da implementação de Sistemas de Recomendação sensíveis ao contexto (por exemplo, notícias, filmes, música, serviços) (Lee and Lee, 2007). Adomavicius et al. (Adomavicius and Tuzhilin, 2008), por exemplo, implementaram um Sistema de Recomendação de filmes que leva em consideração o contexto do usuário (por exemplo, se ele vai assistir o filme em casa ou no cinema) e atestaram melhorias na precisão e no *recall* das recomendações. Apesar destes bons resultados iniciais, muitos problemas permanecem no uso da sensibilidade ao contexto em Sistemas de Recomendação, tais como:

- a) Como adquirir contexto de maneira precisa?
- b) Quais dimensões contextuais influenciam na recomendação de forma significativa?
- c) Como calcular similaridade de contexto?
- d) Como avaliar Sistemas de Recomendação sensíveis ao contexto?

3.4 Trabalhos Relacionados

A abordagem proposta nesta dissertação busca incorporar a sensibilidade ao contexto em Sistemas de Recomendação a partir da adaptação de técnicas de recomendação baseadas em conhecimento. Ao se utilizar de conhecimento de domínio, tanto o contexto do usuário quanto dos itens são explorados.

Adomavicius et al. (Adomavicius and Tuzhilin, 2008) propõem uma classificação dos sistemas sensíveis ao contexto em duas categorias.

Na primeira se encaixam os sistemas que utilizam as informações contextuais como critério de busca com o objetivo de filtrar os itens mais indicados ao contexto do usuário. São exemplos de sistemas dessa categoria:

- a) *COMPASS* (Van Setten et al., 2004): guia turístico em que a recomendação dos pontos de interesse leva em consideração a localização atual do usuário (e.g., um museu próximo à localização do usuário tem mais chances de ser recomendado do que outro mais distante);

- b) *Kaleido Photo* (Xia et al., 2010): realiza a anotação contextual das fotos e permite que o usuário filtre os álbuns de acordo com categorias pré-estabelecidas (e.g., evento em que a foto foi gerada). Neste sistema, o usuário tem que informar explicitamente os filtros a serem utilizados, não ocorrendo uma adaptação contextual; e
- c) *Columbus* (Rost et al., 2011): detalhado na Subseção 3.4.1, é uma aplicação móvel mais simples que mostra fotos geolocalizadas geradas perto da localização do usuário.

Na segunda categoria estão os sistemas que utilizam as informações contextuais no momento em que o usuário avalia algum item. Além das avaliações e características dos itens e dos usuários, os sistemas sensíveis ao contexto dessa categoria levam em consideração também as informações contextuais (e.g., localização, companhia). Exemplos de sistemas dessa categoria são:

- a) O trabalho descrito por Adomavicius et al. (Adomavicius and Tuzhilin, 2008): visto na Subseção 3.4.3, ao registrar o momento em que o usuário assistiu algum filme, armazena tanto a nota dada pelo usuário ao filme quanto em qual contexto o filme foi assistido;
- b) *C²_Music* (Lee and Lee, 2007): pertencente ao domínio de recomendação de conteúdo multimídia, o sistema, visto na Subseção 3.4.2, incorpora informações contextuais na recomendação de músicas. De uma maneira geral, o comportamento deste sistema é similar ao Sistema de Recomendação de filmes, em que na avaliação de uma música é incorporado o contexto no qual ela foi escutada; e
- c) *Epaper* (Maidel et al., 2008): Sistema de Recomendação de notícias jornalísticas e, mesmo não contemplando sensibilidade ao contexto, utiliza uma ontologia para modelar as notícias em uma hierarquia de acordo com o conteúdo abordado (e.g., Esporte → Futebol → Copa do Mundo). Realizando uma recomendação baseada em conteúdo, o sistema traça o perfil do usuário a partir do seu histórico de leitura e recomenda aqueles itens mais similares aos que estão presentes no histórico levantado. A avaliação da eficiência da recomendação do *Epaper* foi comprovada com a realização de experimentos

com usuários, que eram requisitados a graduar, numa escala de 0 a 5, seu interesse pelos itens.

Os Sistemas de Recomendação sensíveis ao contexto estudados que adotam em seu modelo mais de uma dimensão contextual, porém, não consideram o contexto de criação ou utilização dos itens passíveis de recomendação. Estes sistemas também não realizam uma recomendação baseada em conhecimento. Tais sistemas dependem, portanto, de uma extensa base histórica em que usuários em contextos semelhantes ao usuário atual tenham utilizado e avaliado uma grande quantidade de itens. Sem essa base histórica, estes sistemas não conseguem relacionar o usuário atual com nenhum contexto de utilização de itens, não sendo possível a realização de uma recomendação contextual. Outro problema ocorre quando, mesmo na existência de uma extensa base histórica, algum item novo é adicionado à coleção. Como este item ainda não foi utilizado, não é possível recomendá-lo de acordo com técnicas de recomendação colaborativa.

3.4.1 Columbus – Exploração em mobilidade de fotografias geolocalizadas

Columbus (Rost et al., 2011) é uma aplicação móvel que permite que os usuários explorem o ambiente ao seu redor a partir de fotos geolocalizadas e apresentadas em um mapa. Ao se locomover com seu dispositivo móvel, o usuário tem a oportunidade de visualizar as fotos que foram geradas em locais próximos. O estudo teve como finalidade a observação de como a exploração do ambiente ao redor pode ser influenciada pela observação de fotos tiradas, no mesmo ambiente, por outros usuários. A Figura 8 corresponde a duas telas da aplicação. Na tela da esquerda, as fotos próximas ao usuário são marcadas no mapa por pontos vermelhos. O usuário pode posicionar um cursor na tela e no canto inferior são mostradas as miniaturas das fotos próximas à posição do cursor. A distância entre o local em que a foto foi gerada e o local em que se encontra o usuário foi o único critério utilizado para a marcação das fotos no mapa: as fotos próximas são exibidas e as geradas em locais distantes não são exibidas.

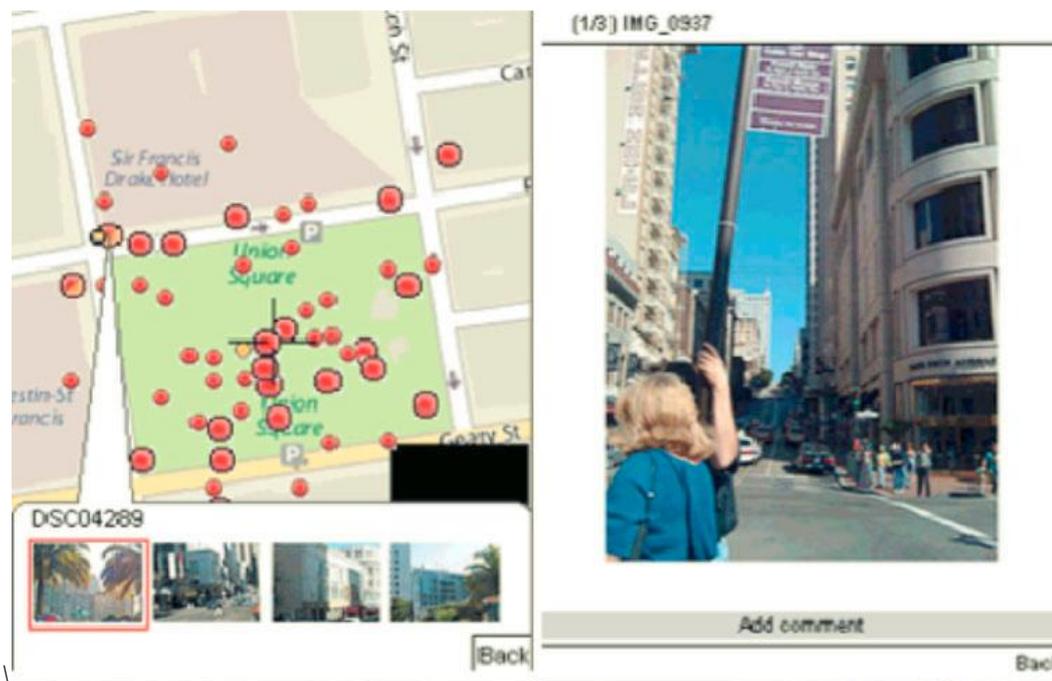


Figura 8 – Telas da aplicação *Columbus* (Rost et al., 2011)

Em um primeiro experimento, com três participantes que conheciam o local de exploração, foi constatado que os usuários apreciaram o fato da aplicação mostrar somente fotos próximas à sua localidade. Os participantes demonstraram menor interesse em fotos que faziam parte de grandes agrupamentos (clusters) no mapa, já que tais fotos geralmente representam imagens, geradas por um só ou por vários usuários, dos mesmos objetos ou pontos de observação.

Em um segundo experimento, com oito usuários que não conheciam o local de exploração, as fotografias exibidas ajudaram os usuários a aprender mais sobre o ambiente ao redor. Alguns usuários, ao identificarem ao seu redor algum objeto exibido na fotografia, puderam orientar-se melhor na utilização do mapa. Estes usuários tiveram também uma maior satisfação quando, para algumas fotografias, puderam facilmente alinhar seu ponto de vista com a da paisagem retratada. Em alguns experimentos, as fotografias serviram de estímulo para que os usuários visitassem certas localidades e vissem onde determinada fotografia foi gerada ou que procurassem ao seu redor o objeto retratado pela fotografia, permitindo que encontrassem pontos de interesse que não encontrariam caso não estivessem utilizando o sistema. Nos dois experimentos, os usuários apresentaram certa frustração com erros de anotação de localização de algumas imagens, acarretando que algumas fotos aparecem no mapa em local diferente do qual elas foram realmente geradas. As fotografias influenciaram também as respostas dadas pelos usuários quando perguntados sobre as características do local visitado.

Apesar de não ser um Sistema de Recomendação sensível ao contexto complexo, por utilizar simplesmente a localização como filtro, os resultados iniciais do *Columbus* apontam que usuários podem se interessar pelo uso de fotografias geolocalizadas capturadas por outros usuários.

3.4.2 C^2_Music

C^2_Music (Lee and Lee, 2007) é uma aplicação de recomendação, sensível ao contexto, de músicas. Ao recomendar músicas aos seus usuários, a aplicação leva em consideração, além dos dados demográficos e de comportamento, as informações contextuais do usuário. O estudo tem como objetivo a averiguação do ganho que a sensibilidade ao contexto pode trazer na recomendação. A recomendação colaborativa foi escolhida como estratégia de recomendação: para um usuário específico. O C^2_Music recomenda as músicas que usuários similares escutaram em contextos similares. No cálculo de similaridade entre contextos, o sistema utiliza dados de localização dos usuários e a situação meteorológica no momento em que estes escutavam as músicas e no momento corrente do usuário que receberá a recomendação.

A Figura 9 mostra o processo de recomendação do C^2_Music . Na primeira etapa, o contexto do usuário é recuperado. Na segunda etapa, são buscados os usuários que escutaram músicas em um contexto similar ao do usuário corrente. Na etapa seguinte, são identificados, entre aqueles selecionados nas etapas anteriores, aqueles que tem uma maior similaridade com o usuário corrente. Na última etapa são recomendadas as músicas que foram mais escutadas pelos usuários identificados.

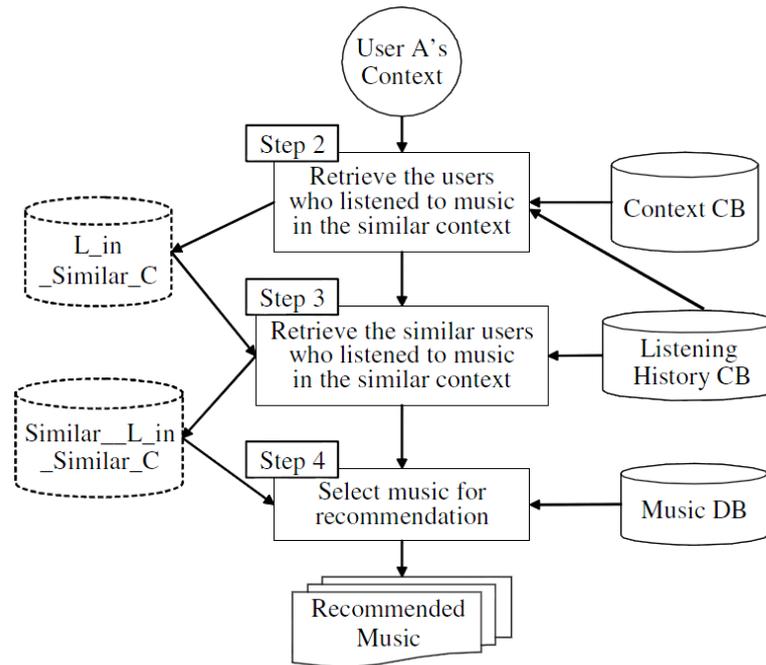


Figura 9 – Processo de Recomendação do C^2_Music (Lee and Lee, 2007)

Para o cálculo de similaridades entre usuários, o sistema utiliza o histórico de músicas escutadas. A Tabela 2 apresenta um exemplo de cálculo de similaridade temporal da proposta, no caso, os escores de similaridade atribuídos na comparação entre duas estações do ano.

	Primavera	Verão	Outono	Inverno
Primavera	1	0.2	0.5	0.2
Verão	0.2	1	0.2	0
Outono	0.5	0.2	1	0.2
Inverno	0.2	0	0.2	1

Tabela 2 – Similaridade para Estações do Ano (Lee and Lee, 2007)

O sistema foi avaliado a partir da validação das músicas recomendadas utilizando-se uma base histórica obtida de uma empresa de *streaming* de música da Coreia, contendo músicas escutadas por 659 clientes durante 6 meses. 60% da base foi utilizada como dado de treinamento e 40% como dado de teste. Para efeitos comparativos foram geradas duas versões do sistema: uma utilizando e outra não utilizando as informações de contexto. A precisão (Cleverdon et al., 1966) média da versão que não utilizou os dados contextuais foi de 0.461. A precisão da versão sensível ao contexto foi de 0.542. Com o ganho de 8% obtido com a utilização das informações contextuais, os autores concluíram que a precisão da recomendação pode ser melhorada a partir da utilização das informações contextuais de localização, data, horário e condições meteorológicas.

3.4.3 Sistema de Recomendação de Filmes

Adomavicius et al (2005) implementaram um sistema sensível ao contexto de recomendação de filmes. Neste sistema, filmes que foram assistidos em contexto similar ao que o usuário se encontra possuem maiores chances de serem recomendados. O sistema utilizou uma base histórica de avaliações em que, além do score que cada usuário atribuiu ao filme, foi registrado também o contexto no qual o filme foi assistido. As informações contextuais consideradas foram: tempo (dia da semana ou final de semana), lugar (no cinema ou em casa), companhia (sozinho, com amigos, com namorado(a) ou com familiares). O sistema utilizou uma abordagem híbrida e em etapas. Na primeira etapa, o sistema filtra as avaliações de filmes que foram assistidos em contexto similar ao o usuário se encontra. Na segunda etapa o sistema utiliza as avaliações filtradas para realizar a recomendação utilizando técnicas de recomendação colaborativa.

A Figura 10 exibe os passos do processo desse Sistema de Recomendação. No processo de filtragem, no primeiro passo, são consideradas as informações de Usuários, Itens, Contexto e Escore (*rating*). No segundo passo, com os dados filtrados, a recomendação colaborativa não utiliza as informações contextuais, permitindo que técnicas tradicionais, não sensíveis ao contexto, possam ser utilizadas.

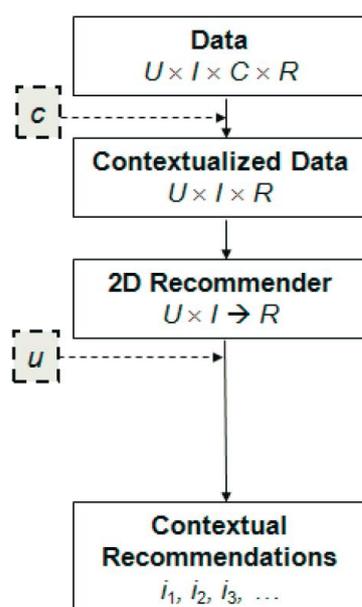


Figura 10 – Passos para Recomendação com Pré-Filtro (Adomavicius and Tuzhilin, 2008)

O sistema foi validado a partir de uma base histórica contendo 1457 avaliações de 202 filmes realizadas por 62 usuários. A abordagem utilizou uma validação *10-fold*, com 90%

da base histórica utilizada para treinamento do algoritmo e 10% para a validação da recomendação. O indicador utilizado foi o *F-measure* (Yang and Liu, 1999). O resultado obtido foi comparado com uma versão do sistema que não realiza o pré-filtro, não utilizando, assim, as informações de contexto. O *F-measure* calculado na versão não sensível ao contexto foi 0.463 e na versão sensível ao contexto foi 0.526. A sensibilidade ao contexto, assim, trouxe um ganho de 0.063 e, de acordo com os testes de significância realizados pelos autores, ficou constatado um ganho estatisticamente significativo obtido na utilização de informações contextuais na base histórica utilizada.

3.5 Conclusão

A Sensibilidade ao Contexto pode trazer ganhos concretos para sistemas de informação, de maneira especial para sistemas móveis e de recomendação. É a partir da sensibilidade ao contexto que a abordagem proposta neste trabalho adapta as técnicas tradicionais de recomendação. Este capítulo mostrou os conceitos centrais de Sensibilidade ao Contexto apresentando abordagens de modelagem de contexto e exemplos de arquiteturas desenvolvidas para atender às necessidades deste tipo de sistema. Foram estudados trabalhos relacionados que também utilizam sensibilidade ao contexto para alcançar o objetivo de uma melhor recomendação. No próximo capítulo será detalhada a proposta deste trabalho e como ele utiliza as técnicas aqui estudadas para oferecer ganhos em processos de recomendação.

4 ABORDAGEM PROPOSTA

As possibilidades de melhoria de Sistemas de Recomendação ao utilizar informações contextuais são notórias e todo o seu potencial (e.g., incorporação do contexto dos itens recomendados) ainda não foi explorado. Nesse capítulo é apresentada a proposta desta dissertação que tem como objetivo propor uma abordagem para a incorporação de sensibilidade ao contexto em Sistemas de Recomendação que explorem repositório de itens associados com informações contextuais. O intuito da abordagem é de, transparentemente, prover ao usuário recomendações mais próximas de seu interesse e de seu contexto corrente.

A Seção 4.1 expõe os ganhos que são esperados com a proposta deste trabalho, cita as abordagens de captura de contexto dos itens recomendados e enumera os passos previstos na construção de um Sistema de Recomendação sensível ao contexto. A Seção 4.2 apresenta o fluxo de execução da recomendação, descrevendo todos os passos envolvidos desde a aquisição do contexto até a exibição dos itens recomendados.

O processo de criação de um modelo de contexto é estudado na Seção 4.3, enumerando as entidades contextuais consideradas na prova de conceito e a importância de uma modelagem utilizando-se conceitos padronizados. A Seção 4.4 descreve o algoritmo de geração do *ranking* de itens e a criação de cálculos de similaridade, conceito central em Sistemas de Recomendação baseados em conhecimento.

Na Seção 4.5, a modelagem e construção da solução são analisadas. São mostrados a importância e os meios de construção de um repositório de cálculos de similaridade que podem ser reutilizados e combinados de acordo com o Sistema de Recomendação a ser desenvolvido. A Seção 4.6 realiza a conclusão deste capítulo.

4.1 Visão Geral da Proposta

Como visto no capítulo 3, o avanço da sensibilidade ao contexto possibilitou o surgimento de aplicações que, ao gerar conteúdo, disponibilizassem também, na forma de metadados, o contexto sob o qual o item foi gerado. Este recurso tem possibilitado o surgimento de repositórios de itens anotados com informações contextuais ricas em semântica

e metadados (Viana et al., 2011) (Figueirêdo et al., 2012) que facilitam a organização, o gerenciamento e a busca do conteúdo anotado. A abordagem proposta neste trabalho se utiliza dessas informações na realização de recomendações sensíveis ao contexto.

Nesse trabalho é esperado que, utilizando o contexto de geração dos itens, consiga-se superar limitações de técnicas de recomendação sensíveis ao contexto já conhecidas e que não se beneficiam de um repositório de itens rico em contexto e semântica. Um repositório com estas características oferece a oportunidade de uso de técnicas de Recomendação Baseada em Conhecimento. Este tipo de técnica foi escolhido pois permite a aplicação de regras de domínio que melhor exploram as informações disponíveis a partir da base de itens e do contexto do usuário.

A aplicação destas regras não depende de uma base histórica de avaliações dos itens, o que facilita que bases recentes, construídas por aplicações sensíveis ao contexto, sejam exploradas em Sistemas de Recomendação. Vários domínios podem se beneficiar dessa abordagem para implantar as primeiras versões dos seus sistemas de recomendação sensíveis ao contexto sem a necessidade de possuir avaliações para os itens em contextos de uso diferentes.

Como não depende de técnicas de Recomendação Colaborativas e Recomendação Baseada em Conteúdo, a solução não sofre do problema do início frio e permite que itens recém-adicionados ao repositório sejam também recomendados. A incorporação de contexto nessa técnica de recomendação permite ainda que novos usuários sejam contemplados com recomendações logo em suas primeiras interações com o sistema, não necessitando que seja, previamente, extraído o seu perfil ou adquirido seu histórico de uso. Basta, para isso, que se possua o contexto do usuário, o sistema aplique regras de domínio na escolha dos itens que melhor se adequem a sua situação.

A sensibilidade ao contexto ajuda, também, a superar algumas das dificuldades de implementação de um Sistema de Recomendação baseado em conhecimento. Ao utilizar a sensibilidade ao contexto, este trabalho adota o contexto do usuário como meio de inferência de suas preferências. Com um usuário a procura de indicação de um restaurante, por exemplo, o sistema pode adotar a localização do usuário como critério de preferência e indicar somente restaurantes que fiquem a um determinado raio de distância do usuário. Pode, também, em um cenário mais complexo, verificar o tempo previsto de deslocamento do usuário aos restaurantes situados em sua região. Para isto poderia, além da distância física, utilizar, dentre outras, informações de trânsito e a situação dos transportes públicos (e.g., tempo de espera de um ônibus). Ao permitir a diminuição da necessidade de interação entre o usuário e o sistema,

a solução proposta pode ser utilizada para melhorar a usabilidade dos Sistemas de Recomendação e trazer uma experiência mais satisfatória a seus usuários.

É importante ressaltar que a adoção da técnica de Recomendação Baseada em Conhecimento na abordagem proposta não impede que um sistema híbrido de recomendação seja construído a *posteriori*, no qual, avaliações dos usuários sobre os itens ou similaridade entre itens sejam utilizados para melhorar ainda mais a recomendação. Entretanto, não faz parte do escopo da abordagem proposta a criação desses modelos híbridos de recomendação sensível ao contexto.

Os passos definidos nesta proposta para a elaboração de um Sistema de Recomendação sensível ao contexto são:

- a) Definição de um modelo de contexto que contemple as mais importantes informações contextuais relevantes para o sistema;
- b) Extração, e possível enriquecimento, do contexto do usuário que utiliza o sistema;
- c) Acesso a um repositório de itens com informações de contexto;
- d) Elaboração de cálculos de similaridade para que sejam escolhidos os itens com maiores chances de corresponder às preferências do usuário; e
- e) Escolha do peso de cada cálculo de similaridade para que os atributos mais importantes possam ter uma maior influência no resultado da recomendação.

Estas etapas são apresentadas nas próximas seções e aplicadas a um sistema desenvolvido para a prova de conceito (descrito no Capítulo 5). Para não haver perda de generalidade, são levantados outros possíveis cenários de aplicação da abordagem proposta e como utilizá-la em outros domínios.

4.2 Fluxo de Execução da Abordagem de Recomendação

A abordagem de recomendação proposta foca no uso de sistemas de recomendação por usuários em situação de mobilidade. Assim, a recomendação foi projetada seguindo uma arquitetura orientada a serviços cliente-servidor para computação móvel

baseada em *RESTful Web Services* (Fielding, 2000). Nesta arquitetura há dois participantes principais:

- a) **Servidor de aplicação:** responsável pelo processamento dos dados e onde é executado o algoritmo de recomendação e a busca de itens; e
- b) **Dispositivo móvel:** responsável unicamente pela aquisição de contexto, com dados brutos e não trabalhados, e pela interface de interação com o usuário.

Esta limitação de responsabilidade do dispositivo móvel permite que dispositivos de baixo processamento possam fazer uso do serviço. A divisão de responsabilidades objetiva um menor acoplamento, já que permite que o serviço seja reutilizado por outras aplicações. Procurou-se também facilitar a manutenibilidade, pois se espera que as manutenções ocorram principalmente no código do serviço de recomendação, que concentra a maior parte das responsabilidades, o que diminui a necessidade de se atualizar o software embarcado.

A Figura 11 representa uma visão geral do fluxo de execução da solução proposta.

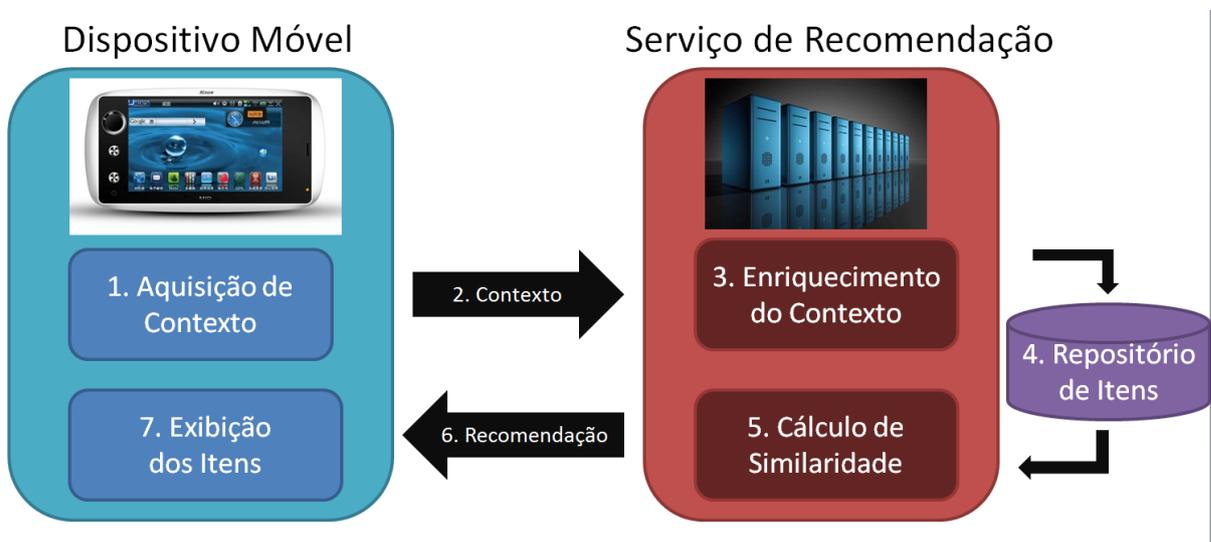


Figura 11 – Fluxo de execução da abordagem de recomendação

No passo 1, a aplicação de recomendação executada no dispositivo móvel é responsável por recuperar o contexto do usuário. Alguns tipos de informações (e.g., localização, hora e data) podem ser adquiridos a partir de sensores e componentes (e.g., *GPS* e relógio). Outros tipos dependem de informações passadas pelo usuário (e.g., o usuário informa à aplicação qual a atividade sendo desempenhada).

No segundo passo, a aplicação móvel acessa, a partir de uma chamada *HTTP*, o *Web-Service* disponibilizado pelo serviço de recomendação, informando o contexto atual do usuário.

O serviço de recomendação, hospedado em um servidor *WEB*, ao receber a requisição, realiza o enriquecimento de contexto do usuário no passo 3, incorporando novas informações a partir dos dados enviados. Exemplos de enriquecimento são a extração de informações da data (e.g., dia da semana, mês, entre outras) e das coordenadas de localização (e.g., bairro e cidade).

No passo 4, os metadados dos itens armazenados no repositório são pesquisados e comparados com o contexto do usuário passado. No passo 5, é elaborado um *ranking* de acordo com os cálculos de similaridade implementados e os itens que possuem um contexto com maior similaridade ao do usuário são enviadas ao dispositivo móvel. O *ranking*, retornado no passo 6, possui a *URL* pela qual os detalhes sobre os itens podem ser recuperados. Na exibição dos itens, passo 7, além do conteúdo, o usuário pode também ter acesso às mesmas informações de contexto utilizadas pelo sistema.

A contribuição deste trabalho se concentra principalmente nas etapas 5, Cálculo de Similaridade e 6, Recomendação, em que o algoritmo de recomendação é aplicado utilizando-se os cálculos de similaridade e os itens melhor ranqueados são recomendados ao usuário. Espera-se que este fluxo, por ser implementado em uma arquitetura genérica, possa atender necessidades que não são exclusivas de aplicações de recomendação de conteúdo multimídia (domínio escolhido para a aplicação de prova de conceito). Dessa forma, ele poderá também ser utilizado em aplicações de recomendação de outros domínios.

4.3 Modelo de Contexto

Como visto na Seção 3.1, a definição de um modelo de contexto é fundamental na elaboração de um sistema sensível ao contexto. Esta definição corresponde à primeira das etapas propostas neste trabalho, enumeradas na Seção 4.1. No modelo de contexto, estão definidas todas as informações contextuais consideradas relevantes pelo algoritmo de recomendação. Além dos dados, as suas relações entre si e entre outros objetos do domínio da aplicação precisam também ser definidos. O modelo deve contemplar as informações contextuais do usuário no momento de utilização da aplicação. Ele deve também contemplar

as informações contextuais dos itens recomendados, seja o contexto de geração ou o contexto que representa a situação atual do item.

O modelo de contexto proposto neste trabalho teve como base o *Context Top* (Viana et al., 2011) e está representado na Figura 12. O contexto é composto por um conjunto de entidades que descrevem as pessoas, objetos e lugares consideradas relevantes para aplicações sensíveis ao contexto, de acordo com a definição de contexto vista na Seção 3.1.

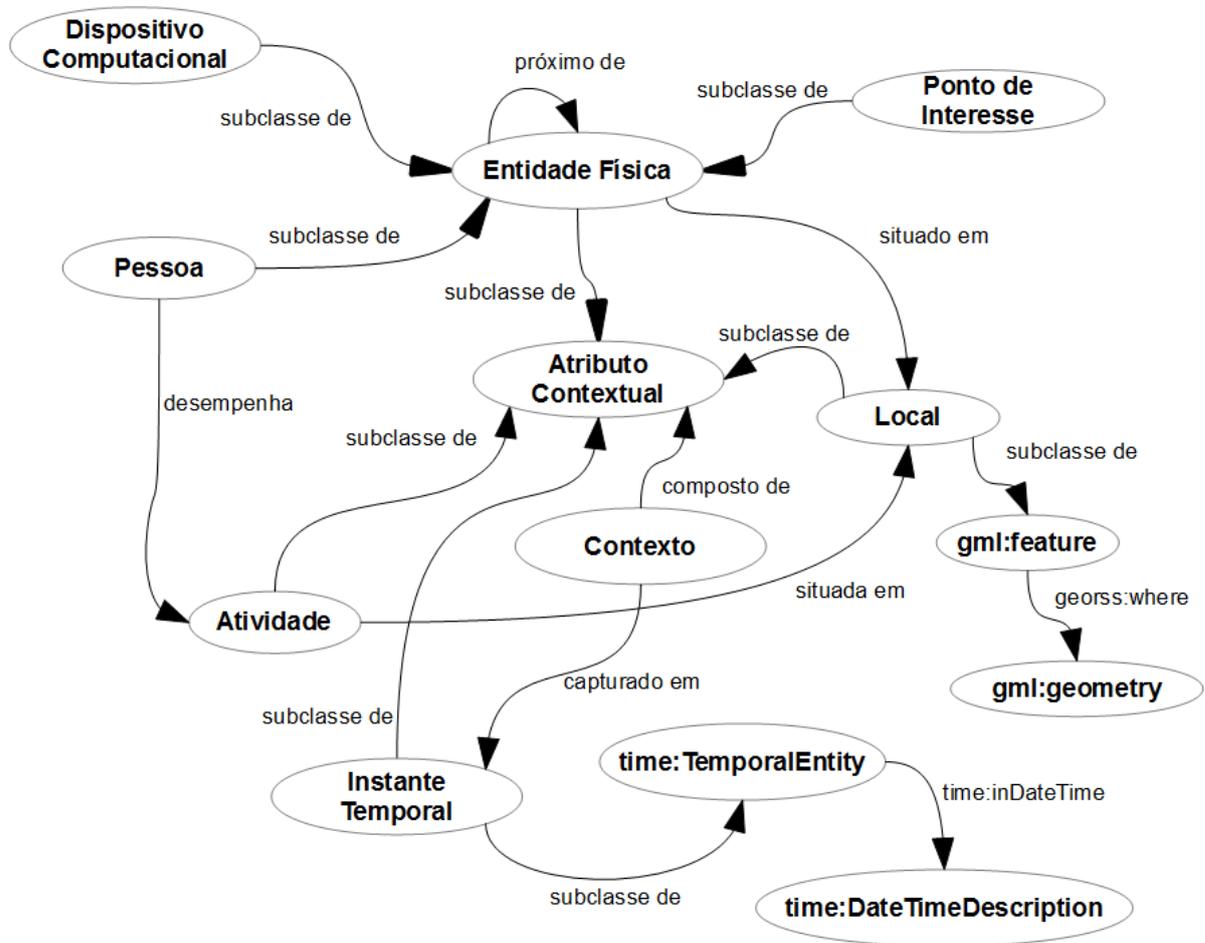


Figura 12 – Modelo de Contexto adotado para a Recomendação Contextual

As entidades contextuais presentes no modelo são:

- a) **Contexto:** representa o contexto de um usuário ou item, de acordo com a definição de (Dey and Abowd, 2000) vista na Seção 3.1
- b) **Atributo Contextual:** são os elementos que podem ser utilizados para compor o contexto de uma determinada entidade.

- c) **Instante Temporal:** representa o momento em que o contexto é capturado.
- d) ***time:TemporalEntity*:** classe da *Time Ontology*⁷. Representa um instante ou intervalo temporal.
- e) ***time:DateTimeDescription*:** classe da *Time Ontology*. Contém as propriedades do Instante Temporal, tais como dia, mês, ano, entre outras.
- f) **Entidade Física:** representa uma pessoa ou objeto do mundo físico que pode compor o contexto de uma outra entidade a partir de relacionamentos de proximidade ou localização.
- g) **Pessoa:** representa tanto o usuário atual da aplicação como o indivíduo que utilizou ou gerou os itens do repositório;
- h) **Atividade:** mapeia a atividade sendo desempenhada pelo usuário da aplicação e a atividade que estava sendo desempenhada pelo autor de conteúdo passível de recomendação no momento de sua geração;
- i) **Dispositivo Computacional:** é o dispositivo móvel utilizado pelo usuário da aplicação;
- j) **Ponto de Interesse:** elementos registrados e associados aos itens que possivelmente despertem o interesse dos usuários (e.g., pontos turísticos, atividades sendo desempenhadas, entre outras);
- k) **Local:** ponto geográfico em que uma Entidade Física pode estar situada ou uma Atividade foi ou está sendo desempenhada.
- l) ***gml:Feature*:** classe da *Geography Markup Language - gml*⁸. Classe genérica que representa entidades físicas.
- m) ***gml:Geometry*:** classe da *Geography Markup Language - gml*. Define a localização e a região de uma entidade.

Entre as diferenças entre o modelo de contexto proposto e o *ContextTop* estão:

- A diminuição do número de subclasses diretas de Elemento Contextual. Foram adicionados relacionamentos de subclasse entre as próprias dimensões contextuais utilizadas (e.g., Dispositivo Computacional e Pessoa são subclasses de Entidade Física)

⁷ <http://www.w3.org/TR/owl-time/>

⁸ <http://www.opengeospatial.org/standards/gml/>

- Foram modelados novos relacionamentos entre os atributos contextuais (e.g., Pessoa desempenha uma Atividade que é situada em um Local, relacionamento de proximidade entre Entidades Físicas)

O modelo possui as dimensões espacial (localização geográfica e pontos de interesse), social (pessoa e atividade sendo desenvolvida), temporal (data e hora) e computacional (dispositivo móvel). Estas dimensões são importantes na representação de elementos contextuais em diferentes domínios e em especial em aplicações multimídia (Viana et al., 2011). O contexto dos itens considerado é o contexto em que outro usuário utilizou ou criou o item (e.g., o contexto em que se apresentava o indivíduo que criou uma fotografia). Este contexto é comparado em uma associação direta (utilizando-se os mesmos atributos) ao contexto do usuário, o que permite que tanto o contexto dos itens quanto o contexto do usuário sejam mapeados com os mesmos atributos contextuais.

Os elementos que descrevem o contexto são dependentes das tarefas executadas pelo sistema e da capacidade do mesmo de observar estas informações (Viana et al., 2011). O modelo aqui apresentado não se propõe a ser fechado, mas sim um ponto de partida para os desenvolvedores de Sistemas de Recomendação que devem estendê-lo ou customizá-lo para o domínio do sistema em questão.

No entanto, alguns cuidados foram tomados para que a adaptação para um possível reaproveitamento fosse facilitada. O modelo não possui nenhuma dependência a entidades específicas do domínio do sistema de prova de conceito (e.g., fotografias, cálculos de similaridade, algoritmos de recomendação, entre outras). Ao invés de depender destas entidades, são estas que se referenciam a elementos do modelo, como será visto na Seção 4.5, em um relacionamento unidirecional. Os elementos também são definidos reaproveitando-se conceitos mapeados e definidos em uma linguagem de representação de conhecimento (*OWL-DL*⁹). Na representação são utilizadas ontologias e linguagens de modelagem (e.g., *Time Ontology* e *Geography Markup Language - gml*) definidas por consórcios de padronização (e.g., *W3C*¹⁰). Esta aderência a padrões facilita a utilização de elementos do modelo por outras aplicações que estejam aderentes aos mesmos padrões.

⁹ <http://www.w3.org/TR/owl-features/>

¹⁰ World Wide Web Consortium - <http://www.w3.org/>

4.4 Cálculos de Similaridade

O cálculo de similaridade é utilizado no sistema para que sejam recuperados aqueles itens que tenham sido criados ou que estejam em um contexto o mais similar ao contexto do usuário atual do sistema. O algoritmo foi desenvolvido adaptando-se a técnica de recomendação baseada em conhecimento vista na Fórmula 6.

$$\text{similaridade}(item, Pref) = \frac{\sum_{p \in Pref} w_p * \text{sim}_p(Pref, item)}{\sum_{p \in Pref} w_p}$$

Fórmula 6 – Similaridade entre um Item e a Preferência do Usuário (Mcsberry, 2003)

A Fórmula 6 faz o somatório do cálculo de similaridade individual de cada tipo de característica do item contida no conjunto de preferências do usuário. A função sim_p é particular de cada característica e deve se adequar a sua semântica e forma de representação. O valor de cada similaridade individual pode ainda ser ponderado pelo coeficiente w_p e deve refletir o grau de importância que o usuário dá a determinada característica do item.

A adaptação da abordagem proposta consiste, principalmente, em adicionar a sensibilidade ao contexto ao considerar o contexto do usuário como indicativo de suas preferências e o contexto dos itens como parte do conjunto de suas características. Dessa forma, a similaridade global é calculada entre o contexto do usuário U e o contexto de um item I utilizando a Fórmula 7.

$$\text{similaridade}(U, I) = \sum_{c \in CONT} w_c * \text{sim}_c(U, I)$$

Fórmula 7 – Cálculo de Similaridade Global entre Usuário e Item

Na Fórmula 7, a similaridade é calculada sem a necessidade de uma base de treinamento. Neste caso, c é um atributo pertencente a uma das dimensões do modelo de contexto utilizado (e.g., localização); w_c é o peso de influência do atributo c (e.g., localização tem peso 50%); e sim_c é a função de similaridade para o atributo c . São recomendados ao usuário U aqueles itens que possuam o maior valor da similaridade calculada. O peso de influência dos atributos contextuais não é alterado dentro do escopo de uma comparação (e.g., localização teria peso 50% para todos os itens) e sempre todos os atributos de contexto são considerados no somatório. Dessa forma, o denominador da Fórmula 6 foi omitido na

Fórmula 7 já que sempre terá o mesmo valor para todos os itens e não influenciaria no *ranking*.

Em um hipotético sistema de recomendação de restaurantes, em que a recomendação é realizada considerando-se a localização e a companhia (e.g., se está sozinho ou com os filhos) do usuário, teria que se atribuir o peso da similaridade para essas duas informações contextuais. Tem-se também que se definir as regras para a atribuição da similaridade para o contexto utilizado. Pode-se, por exemplo, atribuir a similaridade de localização de acordo com a distância física do local em que se encontra o usuário até o local do restaurante. A similaridade para companhia pode ser atribuída de acordo com as características do restaurante (e.g., similaridade 1 caso o usuário esteja com os filhos pequenos e o restaurante possua espaço apropriado para crianças).

Para exemplificar o uso da Fórmula 7, atribuindo-se peso de 70% para o contexto de localização e peso de 30% para companhia e supondo que o contexto do restaurante comparado tenha similaridade de localização igual 0,5 e similaridade de companhia de 0,8, a similaridade global tem valor 0,59. O valor do cálculo de similaridade é obtido pela Fórmula 8.

$$\text{similaridade}(U, I) = (0,7 * 0,5) + (0,3 * 0,8) = 0,59$$

Fórmula 8 – Exemplo de Cálculo de Similaridade

Cada dimensão do modelo de contexto deve ter um cálculo de similaridade distinto. Esses cálculos dependem do domínio da aplicação do Sistema de Recomendação, como pode ser visto na Seção 5.3, em que são apresentados os cálculos de similaridade implementados para a prova de conceito dessa dissertação.

4.5 Modelagem do Cálculo de Similaridade

Como visto nos capítulos e subseções anteriores, existem diversas fórmulas para se calcular a similaridade entre usuários e itens. As variações podem ocorrer nos atributos a serem comparados, na estratégia de cálculo de similaridade para um atributo específico ou no peso que se dá para a similaridade do atributo na Fórmula 7. Os atributos, os cálculos de similaridade e os pesos utilizados dependem do Sistema de Recomendação sendo

desenvolvido. Um Sistema de Recomendação de restaurantes, por exemplo, provavelmente precisaria de cálculos de similaridade distintos do de um Sistema de Recomendação de fotos. A solução proposta, para não perder a generalidade, necessita, portanto, possibilitar a inserção e remoção de cálculos de similaridade. Esta flexibilidade permite que a solução possa ser reutilizada em Sistemas de Recomendação de diferentes domínios. A preocupação com a flexibilidade e a evolução dos Sistemas de Recomendação já foi apontada por outros pesquisadores (Santos and Ferraz, 2011).

O primeiro requisito para a flexibilização é que as classes de cálculo de similaridade sejam separadas em um componente específico independente de domínio. Desta forma, ao se reutilizar a solução, somente este componente e suas dependências teriam que ser importadas. Este passo foi alcançado com a criação do componente *great-similarities*, detalhado na Subseção 5.4 (página 74). Outra necessidade é a abstração, para a classe de recomendação, de toda a complexidade de instanciação dos objetos de cálculo de similaridade. A partir de uma interface única, a classe de recomendação deve ser capaz de obter a similaridade entre as entidades passadas. Esta abstração permite que o cálculo de similaridade possa ser modificado sem que se precise alterar o processo de recomendação.

Para tratar dessas necessidades, as classes de cálculo de similaridade foram modeladas em uma hierarquia e utilizadas na recomendação de acordo com o padrão de projeto *Composite* (Gamma et al., 1995). Este padrão tem aplicabilidade na “representação ‘parte-todo’ de uma hierarquia de objetos” e quando “o cliente deve ignorar a diferença entre composições de objetos e objetos individuais, tratando uniformemente todos os objetos em uma estrutura composta” (Gamma et al., 1995, tradução nossa). A estrutura do padrão pode ser visto na Figura 13.

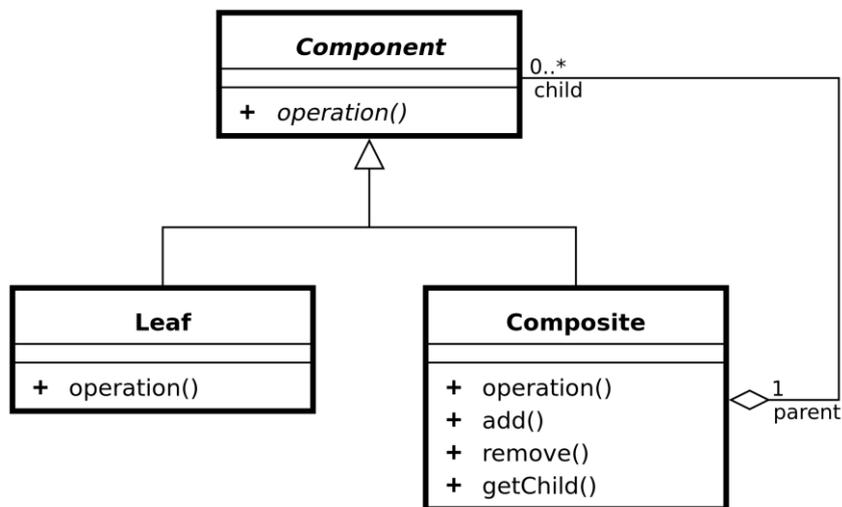


Figura 13 – Padrão de Projeto *Composite* - (Gamma et al., 1995)

Os participantes e suas responsabilidades do padrão *Composite* são:

- Component:** declara a operação, a partir de um método abstrato, a ser implementada por todas as subclasses. É por meio de uma instância dessa classe que o cliente realiza as operações necessárias;
- Leaf:** define o comportamento dos objetos “folha”, que não possuem filhos. É este tipo de classe que de fato implementa as regras para a realização da operação desejada
- Composite:** armazena uma coleção de objetos “folha” e, ao receber uma requisição do cliente, redireciona a operação para os objetos “folha” que fazem parte de sua composição.

A modelagem das similaridades segue esta mesma estrutura e pode, com algumas classes exemplo, ser vista na Figura 14.

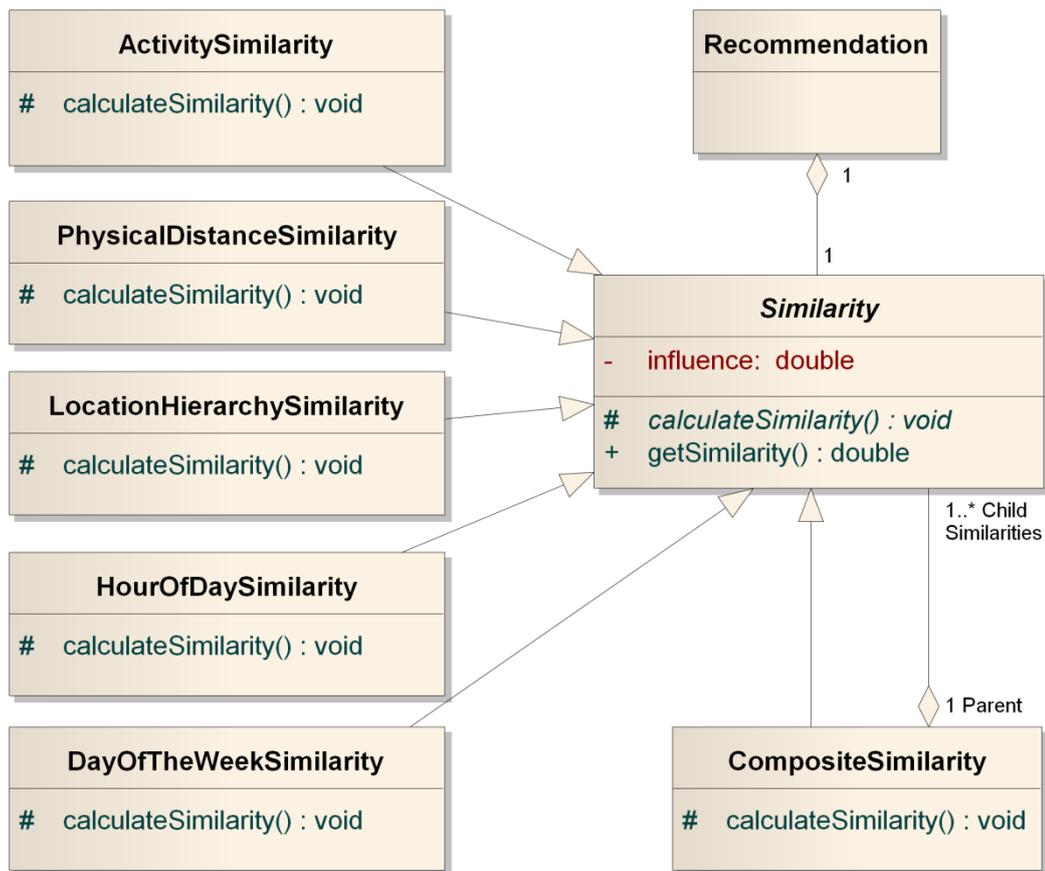


Figura 14 – Diagrama de Classes da Modelagem do Cálculo de Similaridades

Os participantes e suas responsabilidades são:

- a) **Similarity**: declara a operação *calculateSimilarity* a ser implementada por suas subclasses. Esta operação realiza o cálculo de similaridade entre duas instâncias, uma pertencente ao usuário e a outra a um determinado item, de atributos de contexto. Os atributos de contexto comparados são específicos de cada subclasse. A operação *getSimilarity* somente retorna a similaridade calculada. Cada instância armazena o valor do peso de influência no cálculo de similaridade (e.g., localização possui peso 50%), mapeado no atributo *influence*, da dimensão contextual sendo comparada. Esta classe é a representação da classe *Component* do padrão *Composite*;
- b) **ActivitySimilarity, PhysicalDistanceSimilarity, etc.**: implementam o cálculo de similaridade para a dimensão contextual e estratégia de cálculo escolhida. Representam os objetos “folha” (*Leaf*) do padrão *Composite*;
- c) **CompositeSimilarity**: armazena uma coleção de classes “folha” de cálculo de similaridade, permitindo que a coleção possa ser utilizada uniformemente,

através de uma única instância da classe *CompositeSimilarity*. Sua implementação do método *calculateSimilarity* retorna o somatório das similaridade de cada classe da coleção. O somatório é ponderado pelo peso de influencia de cada objeto “folha”. Representa a classe *Composite* do padrão homônimo;

- d) **Recommendation**: utiliza a classe *Similarity* para ordenar decrescentemente os itens de acordo com sua similaridade com o contexto do usuário. Os primeiros itens da lista são os recomendados. A similaridade é calculada através de uma instância da classe *Similarity*. Esta classe representa o cliente do padrão *Composite*.

Um Sistema de Recomendação em que a Localização é a única dimensão contextual considerada e em que a similaridade é calculada pela distância física entre os itens e o usuário teria, em tempo de execução, o modelo de cálculo de similaridade instanciado de acordo com o diagrama de objetos da Figura 15. Neste cenário, a classe *PhysicalDistanceSimilarity*, sendo uma subclasse de *Similarity*, pode ser utilizada diretamente pela classe *Recommendation*.

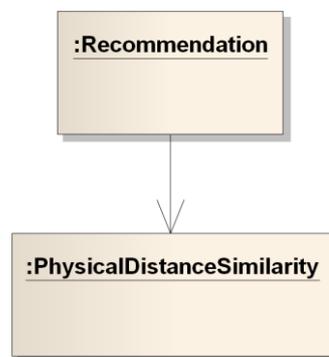


Figura 15 – Diagrama de Objetos para um Cálculo de Similaridade Simples

Em outro sistema, que utilize um modelo mais complexo de cálculo de similaridade, considerando mais dimensões contextuais, o modelo poderia ser instanciado de acordo com o diagrama de objetos da Figura 16. Neste cenário são utilizadas três dimensões contextuais, com cálculos de similaridade para distância física, atividade e horário do dia. Em uma similaridade múltipla, a classe *Recommendation* se relaciona com uma instância da classe *CompositeSimilarity*, que também é uma subclasse de *Similarity*. O mais importante nesse cenário é que a classe *Recommendation*, por se relacionar com somente uma instância de *Similarity*, não é ciente da complexidade do cálculo de similaridade adotado. A mesma

implementação de *Recommendation* pode ser utilizada nos dois cenários, permitindo uma maior facilidade na reutilização da abordagem proposta neste trabalho.

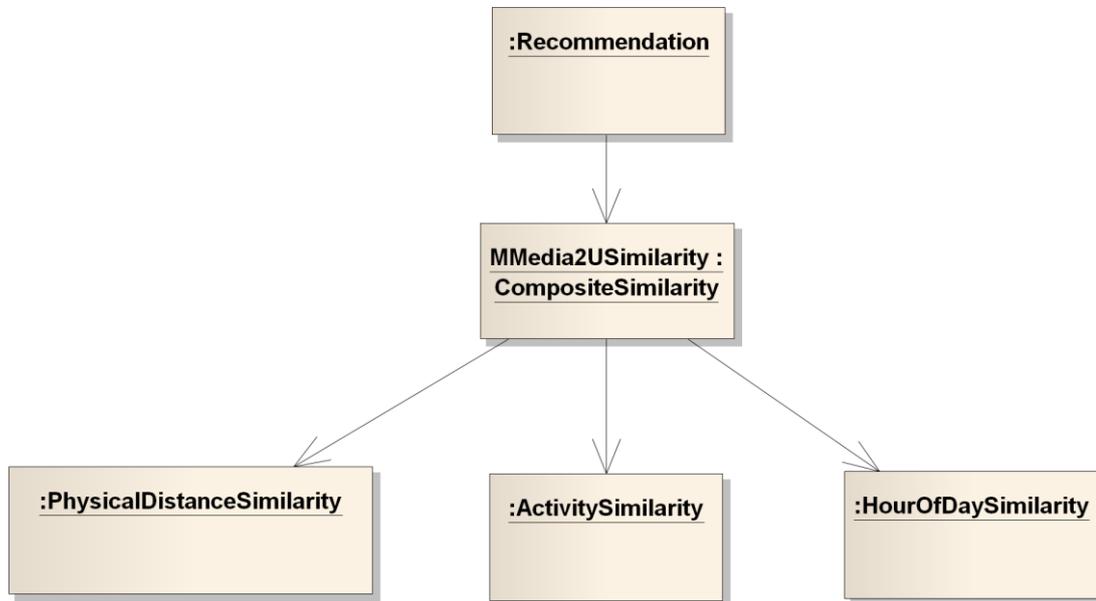


Figura 16 – Diagrama de Objetos para um Cálculo de Similaridade Composto

Pode-se ainda adotar um cenário mais complexo, em que se deseja agrupar objetos de cálculo de similaridade em conjuntos lógicos de acordo com a dimensão contextual em que atuam. Um possível cenário é exemplificado na Figura 17. O agrupamento de cálculos de similaridade (e.g., *PhysicalDistanceSimilarity* e *LocationHierarchySimilarity* agrupados em um mesmo *CompositeSimilarity*) permite uma divisão lógica de responsabilidades e que os pesos de influencia sejam atribuídos não mais de uma maneira isolada (individualmente a cada objeto “folha”), mas a toda uma hierarquia, através da instância da *CompositeSimilarity* utilizada.

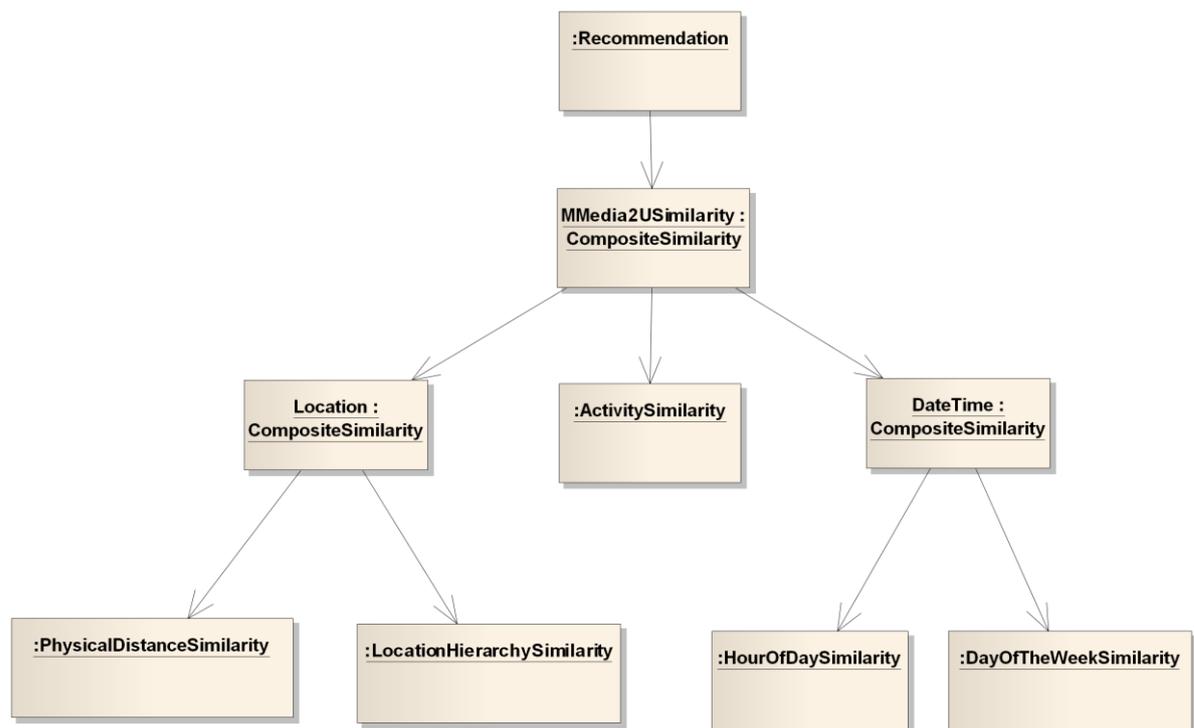


Figura 17 – Diagrama de Objetos para uma Hierarquia de Cálculo de Similaridade

A instanciação dos objetos que compõem a estrutura de cálculo de similaridade precisa ainda ser abstraída para que a classe *Recommendation* não necessite instanciar individualmente cada classe de similaridade. Esse processo é realizado utilizando o padrão *Factory Method* (Gamma et al., 1995), cuja intenção é “definir uma interface para a criação de um objeto, deixando que subclasses decidam que classe instanciar” (Gamma et al., 1995, tradução nossa). Com esta solução é esperado que o componente *great-similarities* possua uma coleção de classes de cálculo de similaridades que podem ser reutilizadas de acordo com o domínio da aplicação de recomendação sendo criada. Esta reutilização se dá com a implementação de *Factory Methods* que retornam a instância de *Similarity* com a estrutura de similaridades a ser utilizada.

4.6 Conclusão

Nesse capítulo foi apresentada a proposta de inserção de contexto em Sistemas de Recomendação. A proposta, ao utilizar regras de domínio, a partir de cálculos de similaridade entre o contexto do usuário e o contexto dos itens, adapta a recomendação baseada em conhecimento e oferece meios de superar limitações, vistas na seção 2.3, presentes em outras

técnicas de recomendação. Na elaboração da abordagem foi também proposto um modelo de contexto com as representações das informações utilizadas. Os cálculos de similaridade, que implementam as regras de domínio, foram projetados de maneira a facilitar o reaproveitamento da abordagem em diferentes domínios de aplicações. O próximo capítulo apresenta o *MMedia2U*, sistema implementado como prova de conceito da abordagem vista nesse Capítulo 4.

5 *MMEDIA2U*

Este capítulo apresenta o sistema desenvolvido como prova de conceito da abordagem proposta, o *MMedia2U*: um sistema sensível ao contexto de recomendação de fotos. A Seção 5.1 aborda os princípios que nortearam a escolha do domínio da prova de conceito e qual o público alvo do sistema. A Seção 5.2 descreve as características, o fluxo da aplicação móvel e as tecnologias utilizadas para sua implementação. Na Seção 5.3, são apresentados os cálculos de similaridade implementados para cada informação contextual utilizada na recomendação. A Seção 5.4 mostra os componentes envolvidos na implementação do serviço de recomendação e suas respectivas responsabilidades. Por fim, a Seção 5.5 contém a conclusão e as considerações finais do capítulo.

5.1 Princípios Norteadores do *MMedia2U*

A recomendação de conteúdo multimídia foi escolhida para o sistema de prova de conceito, que faz a recomendação de fotos anotadas com informações contextuais e permite que o usuário acesse os itens recomendados a partir de seu dispositivo móvel. Este domínio é interessante devido ao grande número de fotos geradas e disponibilizadas publicamente em sites de compartilhamento como *Flickr*¹¹, *Picasa Web*¹² e *Panoramio*¹³. Muitas dessas fotos são geradas por dispositivos móveis ou câmeras eletrônicas que armazenam a localização, data, horário e possivelmente outras informações contextuais. Estas fotos podem ainda ter as informações de contexto enriquecidas a partir da utilização de sistemas como o *PhotoGeo* (Figueirêdo et al., 2012), que enriquece os metadados das fotos com informações contextuais. O sistema pode também se beneficiar de fotos geradas por aplicações que, já no momento de geração da foto, registram, na forma de metadados, o contexto sob o qual a foto foi gerada (Viana et al., 2011).

¹¹ <http://flickr.com/>

¹² <http://picasaweb.google.com/>

¹³ www.panoramio.com/

O sistema é chamado *MMedia2U* e nele usuários recebem recomendações de fotos geradas em contexto similar ao seu. A similaridade é calculada em função de três dimensões contextuais (espacial, social e temporal). O sistema tem como público alvo dois tipos de usuários. O primeiro tipo são aqueles que estão em um contexto não habitual (e.g., turistas visitando um local pela primeira vez) e que possam aproveitar as fotos recomendadas para ter referências de atividades ou novos locais a serem explorados. O segundo grupo contém usuários que já tenham estado em um contexto semelhante e que podem ter nas fotos recomendadas uma nova visão e ponto de vista da situação em que se encontram. Os autores Rost et al., (2011) já observaram que fotos georeferenciadas podem influenciar de forma positiva e lúdica a exploração do espaço por essas duas categorias de usuários.

5.2 Aplicação Móvel

A aplicação móvel foi desenvolvida na plataforma *Android*¹⁴ e é compatível com aparelhos que possuem a versão 2.2 (*Froyo*) ou superior. A apresenta o fluxo de telas da aplicação. Ao ser iniciado, o sistema recupera as coordenadas do usuário e a data e hora local a partir do *GPS* e relógio do dispositivo. Na primeira tela, o usuário informa a atividade que está desempenhando (e.g., observando paisagens naturais, praticando esporte, entre outras) a partir de opções já pré-definidas ou utilizando uma entrada textual livre.

¹⁴ <http://developer.android.com/>

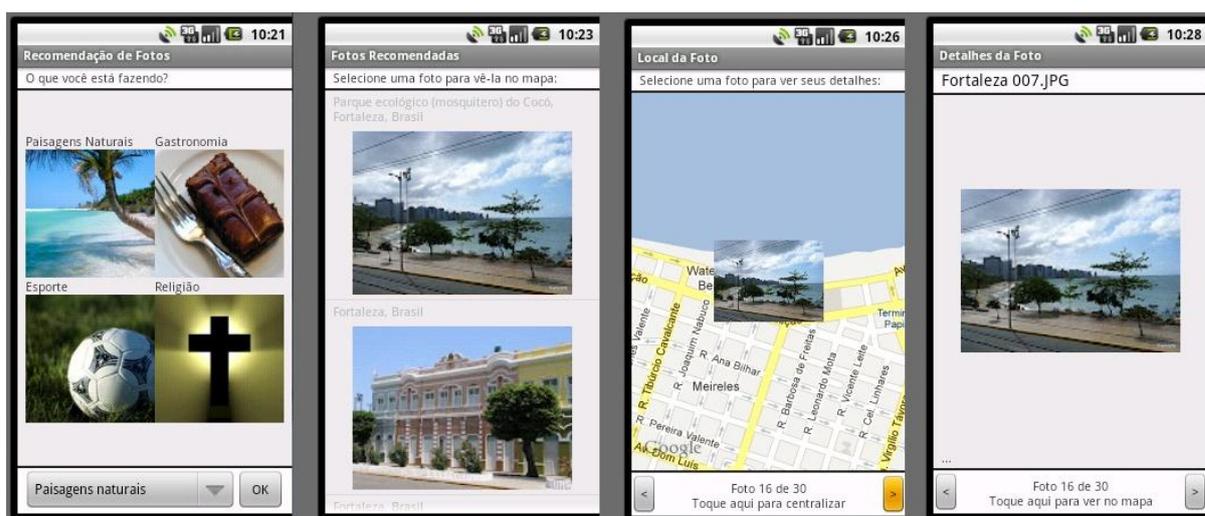


Figura 18 – Fluxo de Telas da Aplicação Móvel

A aplicação agrupa as informações adquiridas e as envia ao *Web-Service* de recomendação a partir de uma chamada *HTTPS*. O serviço de recomendação realiza o enriquecimento das informações contextuais enviadas e retorna a lista de imagens que foram geradas em um contexto semelhante ao do usuário, de acordo com o processo visto na Seção 4.2. A lista de imagens é retornada em um arquivo *XML*, exemplificado no Apêndice A.

Cada elemento na lista do arquivo *XML* representa uma fotografia. A lista é ordenada decrescentemente de acordo com o *ranking* de cada foto: as primeiras fotos são as mais próximas ao contexto do usuário. O conteúdo da fotografia não é retornado na primeira requisição. Ao invés dele, cada elemento do arquivo *XML* possui *URLs* pela qual a fotografia e a sua miniatura podem ser recuperadas. Esta *URL* corresponde ao endereço original da fotografia, remetendo ao repositório de imagens do *Google Picasa*, cujo banco de imagens foi utilizado nesse sistema de prova de conceito. Esta abordagem permite que a base do serviço de recomendação tenha a necessidade de armazenar e retornar somente os metadados das fotografias, o que oferece economia de espaço em disco no servidor de aplicação e largura de banda para o dispositivo móvel. Permite também que a resposta, pelo menor tamanho, seja dada tempestivamente e que o conteúdo das imagens seja recuperado por requisições paralelas. Os outros dados retornados no *XML* são utilizados para oferecer ao usuário maiores informações (e.g., descrição, tags, título) sobre as fotografias. Por padrão, são dez elementos retornados e essa quantidade pode ser ajustada por um parâmetro na requisição, dando oportunidade para que o tamanho do retorno seja adaptado ao tamanho da tela e à largura de banda do dispositivo.

A segunda tela da Figura 18 corresponde à listagem das miniaturas das fotografias retornadas pelo serviço de recomendação. A tela apresentada na figura contém somente duas fotografias, mas outras estão disponíveis se o usuário rolar a tela para baixo. Caso o usuário deseje ver detalhadamente alguma fotografia, após o clique em sua miniatura, o sistema posiciona a fotografia no mapa onde também está representada a localização atual do usuário. Ao clicar na miniatura do mapa, a quarta tela é exibida com a fotografia ocupando toda a extensão da tela do dispositivo em que é permitida a realização de zoom e exibição dos dados (e.g., descrição) da fotografia.

Estendendo o fluxo de execução da abordagem proposta, vista na Figura 11 (página 54), para o domínio de recomendação de fotos, a Figura 19 mostra o fluxo de execução do *MMedia2U*. No primeiro passo, a localização é adquirida pelo *GPS* ou pelo sensor *Wifi*. As informações de data e hora são recuperadas do relógio e calendário do dispositivo e a Atividade é recuperada a partir do formulário da primeira tela da Figura 18. No segundo passo, as informações de contexto Localização, Data, Hora e Atividade são enviadas ao *Web Service* de Recomendação. O serviço de recomendação acessa o repositório de fotos, que contém os links para as imagens armazenadas no *Picasa* e calcula a similaridade de acordo com a composição de objetos *Similarity* vistos na Figura 16 da Seção 4.5. As informações sobre as fotos recomendadas são retornadas em um *XML* com os metadados e as fotos selecionadas pelo usuário exibidas no dispositivo móvel.

Nesta fórmula, N representa a distância entre o mais próximo nó ancestral comum de U e I em relação ao nó raiz da hierarquia; e N_u e N_i representam a distância de U e I , respectivamente, ao nó raiz da hierarquia; $PF(U,I)$ é o fator de penalização para dois nós vizinhos. A similaridade para a atividade de interesse é calculada comparando-se a atividade de interesse ou ocasião em que se encontra o usuário e a atividade ou ocasião em que a foto foi gerada. Algumas das atividades mapeadas em fotos do *MMedia2U* e suas respectivas similaridades encontram-se na Tabela 3. Os valores foram atribuídos de acordo com a relação semântica de cada conceito representado.

	Compras	Festas	Lazer	Esporte
Compras	1	0	0	0
Festas	0	1	0,5	0
Lazer	0	0,5	1	0,5
Esporte	0	0	0,5	1

Tabela 3 – Similaridade entre Atividades

Em atributos numéricos, em que a similaridade é definida por quão próximos são dois valores, a Fórmula 10 é utilizada.

$$\text{sim}_c(U,I) = 1 - \frac{|V_c(U) - V_c(I)|}{\max(c) - \min(c)}$$

Fórmula 10 – Cálculo de Similaridade para Valores Numéricos

Na Fórmula 10, $V_c(U)$ e $V_c(I)$ representam os valores do contexto c para o usuário e o item, respectivamente; $\max(c)$ e $\min(c)$ representam, respectivamente, o valor máximo e mínimo para o atributo comparado do contexto c .

A similaridade entre datas pode comparar os diversos atributos relativos ao momento em que a foto foi tirada e o instante em que se encontra o usuário. Alguns atributos comparados foram: hora do dia, dia da semana e mês do ano. Os atributos de data foram comparados individualmente para analisar a influência particular de cada um (e.g., a semelhança entre a hora do dia teria maior influência na escolha do usuário do que a semelhança do dia da semana?). A similaridade entre o mês do ano pode ser calculada pela Fórmula 10, adaptando-a a valores cíclicos (e.g., a distância entre janeiro e dezembro é 1 ao invés de 11). A similaridade entre dias da semana pode ser calculada pela diferença de dias entre os dias da semana representados por duas datas ou pode-se, por exemplo, utilizar a Tabela 4 como critério. Da mesma forma, a similaridade entre o horário pode ser calculada

pela quantidade de minutos que separam as horas ou levando em consideração se pertencem, ou não, ao mesmo turno do dia (Lee and Lee, 2007).

Comparação entre as datas da foto e do usuário	Similaridade
Mesmo dia da semana	1
Ambos finais de semana ou dia de semana	0.5
Períodos da semana distintos	0

Tabela 4 – Similaridade para o dia da semana

5.4 Componentes da Solução

O algoritmo de recomendação foi implementado na plataforma *Java Enterprise Edition*¹⁵ (JEE) versão 6. Em sua modelagem, implementação e aplicação na prova de conceito, foram gerados componentes que agregam os artefatos de implementação de acordo com sua responsabilidade. Além de uma maior organização do código gerado, a componentização permite um maior reaproveitamento da solução. Procurou-se, principalmente, separar os artefatos que são específicos do sistema de prova de conceito, que possuem um reaproveitamento mais limitado, dos artefatos que podem, potencialmente, ser reutilizados em outros domínios que adotem a abordagem proposta. A Figura 20 mostra o diagrama dos componentes elaborados na solução.

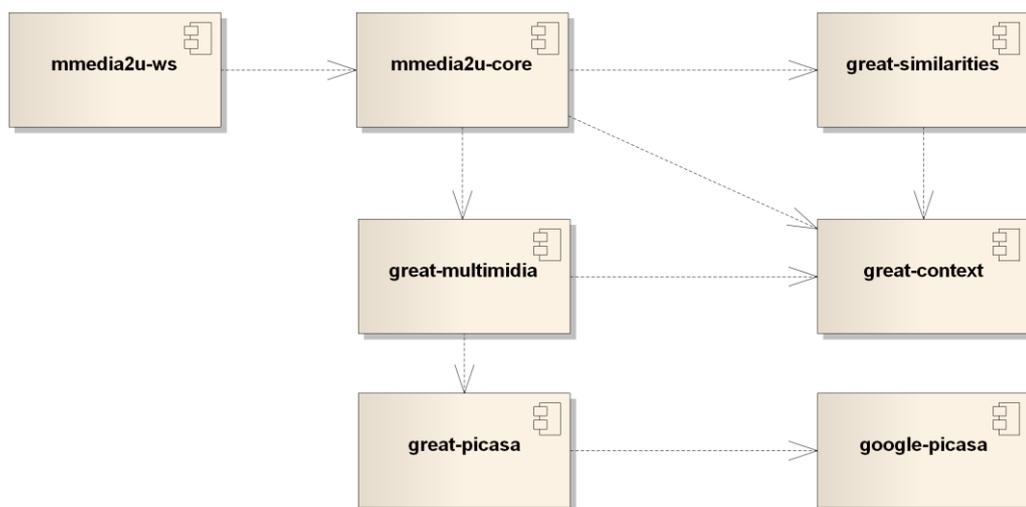


Figura 20 – Diagrama de Componentes da Solução

¹⁵ <http://www.oracle.com/technetwork/java/javaee/index.html>

Os componentes produzidos foram:

- a) ***mmedia2u-ws***: implementa o *web-service* disponibilizado para a aplicação móvel. O *web-service* segue uma arquitetura *Rest* (Fielding, 2000) que, dentre outras, provê vantagens relacionadas à escalabilidade e a simplicidade de implementação dos serviços e dos clientes. A adoção de *web-services*, por trazer independência de plataforma, também permite que aplicações móveis de outras plataformas (e.g., *iPhone*) venham a consumir o mesmo serviço. Este componente implementa somente as interfaces de entrada e saída do serviço e utiliza o *mmedia2u-core* para a obtenção das fotos a serem recomendadas;
- b) ***mmedia2u-core***: implementa as regras de negócio da solução a partir da coordenação com outros componentes. A implementação das regras em um componente específico permite que este seja utilizado na implementação da solução em outras arquiteturas (e.g., aplicação desktop);
- c) ***great-similarities***: implementa as regras de cálculo de similaridade. O cálculo de similaridade para cada dimensão contextual é encapsulado em um objeto específico de acordo com a estratégia utilizada. Para o atributo de localização, por exemplo, foram implementadas duas classes, uma para cálculo da similaridade utilizando-se distância física e outra para o cálculo utilizando-se a hierarquia Bairro → Cidade → UF... da região. Este componente serve de base para a construção de um repositório de objetos de cálculos de similaridade, visto na Seção 4.5;
- d) ***great-context***: este componente contém as abstrações utilizadas na representação do modelo de contexto. As classes implementadas (e.g., Localização, Pessoa, entre outras) trazem para a orientação a objetos as representações utilizadas na ontologia do modelo. Os componentes *mmedia2u-core*, *great-similarities* e *great-multimedia* fazem uso deste componente para associar informações de contexto às suas entidades de negócio (e.g., Usuário, Itens, Similaridade). Devido ao fato do modelo de contexto utilizar ontologias padronizadas é esperado que este componente possa ser reutilizado por outras aplicações que utilizem as mesmas ontologias;
- e) ***great-multimedia***: o componente é responsável pela implementação de serviços de aquisição de conteúdo multimídia. Também é responsável pela extração e armazenamento das propriedades dos itens a partir dos metadados

disponibilizados pelos repositórios utilizados e pelas informações contidas nos cabeçalhos (e.g., *EXIF*¹⁶ e *IPTC*¹⁷). A partir da associação com o componente *great-context*, as informações extraídas são utilizadas para a instanciação do modelo de contexto de cada item. A abstração oferecida por este componente permite que itens de diversas fontes (e.g., *Picasa*, *Flickr*), que possuem distintos meios de aquisição e estruturação dos metadados, possam ser utilizados a partir de uma representação única. Este componente não contém nenhuma regra relativa à recomendação de itens e é esperado que possa ser reutilizado por outras aplicações de aquisição e gerenciamento de itens multimídia;

- f) ***great-picasa***: realiza a comunicação com o *web-service* do *Picasa* para a aquisição dos metadados das fotografias a serem recomendadas. O conteúdo da fotografia não é recuperado durante o processo de cálculo de similaridade, pois nem todas serão recomendadas. A partir da URL da imagem, disponibilizada como metadado por este componente, o dispositivo móvel é capaz de recuperar o conteúdo da imagem. Este componente preenche os parâmetros da busca de acordo com os dados provenientes do contexto do usuário e faz a consulta utilizando o componente *google-picasa*;
- g) ***google-picasa***¹⁸: é um dos componentes desenvolvidos pela Google para facilitar a implementação de clientes para seus *web-services*.

5.5 Conclusão

Esse capítulo apresentou a forma como a abordagem, vista no Capítulo 4, foi aplicada na elaboração de um sistema, desenvolvido como prova de conceito. O *MMedia2U* foi construído a partir da utilização de conceitos e práticas como: ontologias, orientação a objetos, padrões de projeto e componentização. É esperado que as práticas propostas e utilizadas possam também ser reutilizada em Sistemas de Recomendação sensíveis ao

¹⁶ http://www.cipa.jp/english/hyoujunka/kikaku/pdf/DC-008-2010_E.pdf

¹⁷ <http://www.iptc.org/IPTC4XMP/>

¹⁸ <https://developers.google.com/picasa-web/>

contexto de outros domínios. A Tabela 12, no Capítulo 7, relaciona as características do *MMedia2U* com as características dos sistemas relacionados vistos na Seção 3.4. No próximo capítulo serão visto os resultados obtidos na validação da prova de conceito a partir da experimentação e da análise dos resultados obtidos na recomendação de fotos.

6 EXPERIMENTO E AVALIAÇÃO

A abordagem de incorporação de sensibilidade ao contexto em Sistemas de Recomendação proposta neste trabalho foi validada a partir de experimentos realizados com o sistema de prova de conceito *MMedia2U*. Os experimentos serviram como uma análise objetiva sobre o impacto da introdução de sensibilidade ao contexto na qualidade da recomendação. Foram também realizados questionários com os participantes dos experimentos para levantamento de questões subjetivas que possam influenciar na experiência que o usuário teria com um Sistema de Recomendação sensível ao contexto no qual os itens estão anotados com informações contextuais.

O restante deste capítulo está estruturado da seguinte forma: na Seção 6.1 é descrita a estratégia escolhida para a validação da recomendação. A Seção 6.2 explica o processo de construção do *gold standard* utilizado para a verificação da qualidade da recomendação. Os resultados objetivos obtidos são mostrados na Seção 6.3 e na Seção 6.4 são descritas as perguntas e as respostas obtidas com o questionário aplicado aos participantes dos experimentos. Na Seção 6.5 é mostrado como foi efetuado o processo de construção do repositório de fotos utilizadas nos experimentos e a Seção 6.6 faz a conclusão deste capítulo, com uma análise final dos resultados obtidos.

6.1 Concepção da Avaliação

O objetivo principal de um Sistema de Recomendação sensível ao contexto é recomendar itens que despertem o interesse do usuário tendo como informação adicional o contexto no qual ele está inserido. Entretanto, avaliar um Sistema de Recomendação não é uma tarefa das mais simples, pois a relevância do item recomendado tem um caráter pessoal muito forte e, muitas vezes, complexo de ser medido. Esta dificuldade é potencializada em sistemas em que a falta de dados históricos de avaliações de referência requerem que os itens recomendados sejam avaliados manualmente, tornando estudos em larga escala muito onerosos e difíceis de serem reproduzidos (Siersdorfer and Sizov, 2009).

No caso de um Sistema de Recomendação sensível ao contexto, a complexidade aumenta ainda mais, pois é mais difícil variar os contextos possíveis em situações reais (i.e., diversas localizações, situações diárias, atividades, entre outras).

Para avaliar o grau de sucesso da abordagem proposta, utilizaram-se métricas de análise da qualidade da recomendação através de experimentos com usuários. As métricas utilizadas foram as medidas de precisão, *recall* (Cleverdon et al., 1966) e *F-measure* (Yang and Liu, 1999), que são as mais populares na avaliação de sistemas de recuperação de informação (Herlocker et al., 2004).

Outro propósito do experimento foi avaliar a qualidade da recomendação ao se considerar diversas combinações de pesos de influência para cada dimensão contextual da fórmula do cálculo de similaridade (Fórmula 7). A utilização de diferentes pesos permite medir a relevância que cada dimensão contextual possui na qualidade da recomendação (e.g., uma recomendação que considera somente a dimensão de localização é melhor do que uma que considera somente a dimensão temporal?).

A determinação dos pesos que regulam a importância de cada dimensão contextual pode ser feita manualmente (sem treinamento) ou usando uma base de treinamento (histórico de avaliações). Neste trabalho, avaliou-se as duas abordagens. Ao utilizar uma base de treinamento, usou-se como abordagem o método dos Mínimos Quadrados (Björck, 1996), que é um modelo matemático e estatístico que tem por objetivo a solução de sistemas de equações lineares a partir da estimativa de parâmetros. O algoritmo de recomendação foi também avaliado em diferentes contextos, variando-se a localização, período do dia, dia da semana e atividade das situações passadas aos participantes.

Devido ao grande número de recomendações necessárias para contemplar as diferentes combinações de contextos e pesos de influência para todos os usuários, a análise *offline* (Herlocker et al., 2004) foi escolhida como forma de avaliação da qualidade da recomendação. Neste tipo de análise, que é uma contrapartida a experimentos em campo, a recomendação é comparada com registros históricos de avaliações de itens por usuários. A qualidade da recomendação é medida pela proporção, entre os itens recomendados, de itens que tiveram boa avaliação pelos usuários.

Não é necessário, porém, para o domínio de sistemas alvo da abordagem proposta, que a recomendação seja exaustiva, em que são retornados todos os itens que por ventura possam despertar interesse do usuário. Neste tipo de sistema, tenta-se superar o problema da sobrecarga de informação a partir de uma seleção de conteúdo. É admissível, portanto, que em uma seleção mais restritiva, em que recomendasse uma ínfima proporção de itens frente ao

extenso tamanho do repositório, que alguns itens relevantes sejam preteridos em relação a outros também considerados relevantes.

Em um Sistema de Recomendação de fotos, por exemplo, em uma base de milhares de itens, a lista de todas aquelas que possam despertar o interesse do usuário seria possivelmente muito extensa para que seja recomendada completamente. Além disso, uma lista extensa de recomendação não é adequada em dispositivos móveis devido às limitações de tamanho de tela, largura de banda e memória destes dispositivos.

6.2 Construção do *Gold Standard*

Uma avaliação *offline* objetiva e repetível de um Sistema de Recomendação requer a existência de uma base histórica de avaliações de itens por usuários (Herlocker et al., 2004). Estas avaliações devem capturar o grau de preferência subjetiva de um usuário específico para um determinado item. Ao se verificar a qualidade de uma recomendação, os itens recomendados são comparados com as avaliações contidas na base histórica. A qualidade é medida pela semelhança do conjunto de itens recomendados para um determinado usuário em relação ao conjunto dos itens melhor avaliados pelo mesmo usuário. O histórico de avaliações serve, então, como um *Gold Standard* sob o qual a recomendação é submetida. As métricas utilizadas na medição da qualidade da recomendação são vistas na Seção 6.3.

Devido à inexistência de uma base histórica de avaliações de fotos anotadas com informações contextuais, foi necessária a construção do *Gold Standard* a ser utilizado. O método de construção do *Gold Standard* consistiu em experimentos realizados com 13 usuários divididos da seguinte forma:

- a) 10 homens e 3 mulheres;
- b) 4 estudantes e 9 profissionais, quase todos da área de computação;
- c) Idade entre 21 e 35 anos.

Em cada experimento um determinado contexto era passado ao usuário. Um conjunto de fotos era exibido ao usuário e era pedido que este selecionasse aquelas fotos que despertariam seu interesse caso estivessem inseridos no contexto sugerido. Os contextos

utilizados nos experimentos são listados na Tabela 5. Os locais são pontos turísticos da orla de Fortaleza.

No contexto de número 1, por exemplo, o usuário estaria na Praia de Iracema, pelo período da manhã em um dia de semana, visitando os pontos turísticos e patrimônios históricos da região ao seu redor.

Número Contexto	Local	Atividade / Interesse	Período
1	Praia de Iracema	Visitação a pontos turísticos e patrimônios históricos	Dia de semana, pela manhã
2	Dragão do Mar	Festas	Sábado a noite
3	Feirinha da Beira Mar	Compras	Dia de semana, pela manhã
4	Praia de Iracema	Turismo Religioso	Dia de semana, pela manhã
5	Praia de Iracema	Visita e apreciação de paisagens naturais	Domingo, pela manhã
6	Feirinha da Beira Mar	Lazer	Domingo, pela manhã
7	Feirinha da Beira Mar	Gastronomia local	Dia de semana, a noite
8	Beira Mar, quilômetro zero	Prática de esportes	Dia de semana, pela manhã

Tabela 5 – Contextos Utilizados nos Experimentos

A Figura 21 ilustra a sistemática de construção do *Gold Standard*. As fotos escolhidas pelos usuários passaram a compor o *Gold Standard* e serviram de base histórica sob a qual a recomendação do *MMedia2U* foi comparada. Em cada experimento (i.e., um par contexto – usuário), uma média de 100 fotos foi apresentada aos usuários, das quais 20 foram tiradas em contexto semelhante ao passado ao usuário e 80 diferiam, em diferentes dimensões, do contexto apresentado (e.g., atividade igual, localização bem distinta). Cinco dos 13 usuários não conheciam o lugar escolhido como localização dos contextos experimentados. Alguns dos usuários avaliaram as oito coleções de fotos, outros somente um subconjunto delas com um total de 66 avaliações.

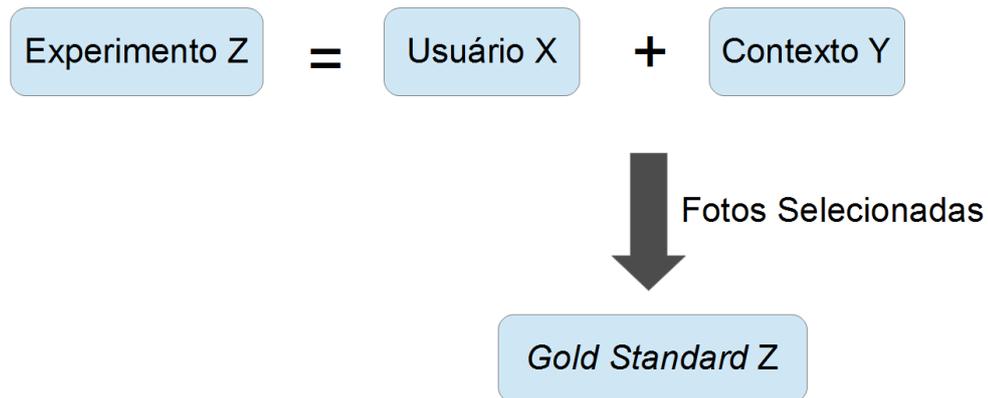


Figura 21 – Sistemática de Construção do *Gold Standard*

6.3 Resultados Quantitativos

Objetivou-se com os experimentos avaliar as hipóteses já mencionadas no Capítulo 1, as quais são aplicadas no domínio de recomendação de fotos:

- a) Hipótese I: é possível realizar recomendações satisfatórias de fotos georeferenciadas sem conhecimento prévio do usuário, considerando apenas seu contexto atual; mapeado
- b) Hipótese II: o contexto em que as fotos foram tiradas possui relevância na realização de recomendações; e
- c) Hipótese III: a utilização de um modelo de contexto considerando várias dimensões contextuais pode trazer ganhos em relação a uma recomendação que utiliza somente um atributo contextual (e.g., a localização).

Para responder a estes questionamentos, o algoritmo foi executado com diferentes pesos para cada dimensão contextual. Para cada *Gold Standard* gerado na etapa de avaliação o algoritmo foi executado e sua taxa de acerto para cada indicador (precisão, *recall* e *F-measure*) medida. Para cada combinação de pesos de influência das dimensões contextuais a taxa média de acerto do algoritmo foi calculada e o resultado utilizado para verificar a validade das hipóteses levantadas. A sistemática de execução do algoritmo está ilustrada na Figura 22. As setas representam as entradas para os processos executados. A Recomendação é realizada a partir de um Contexto e uma Combinação de Pesos de influência das dimensões

contextuais. A medição do resultado de acerto foi calculada a partir da comparação da recomendação com o *Gold Standard*.

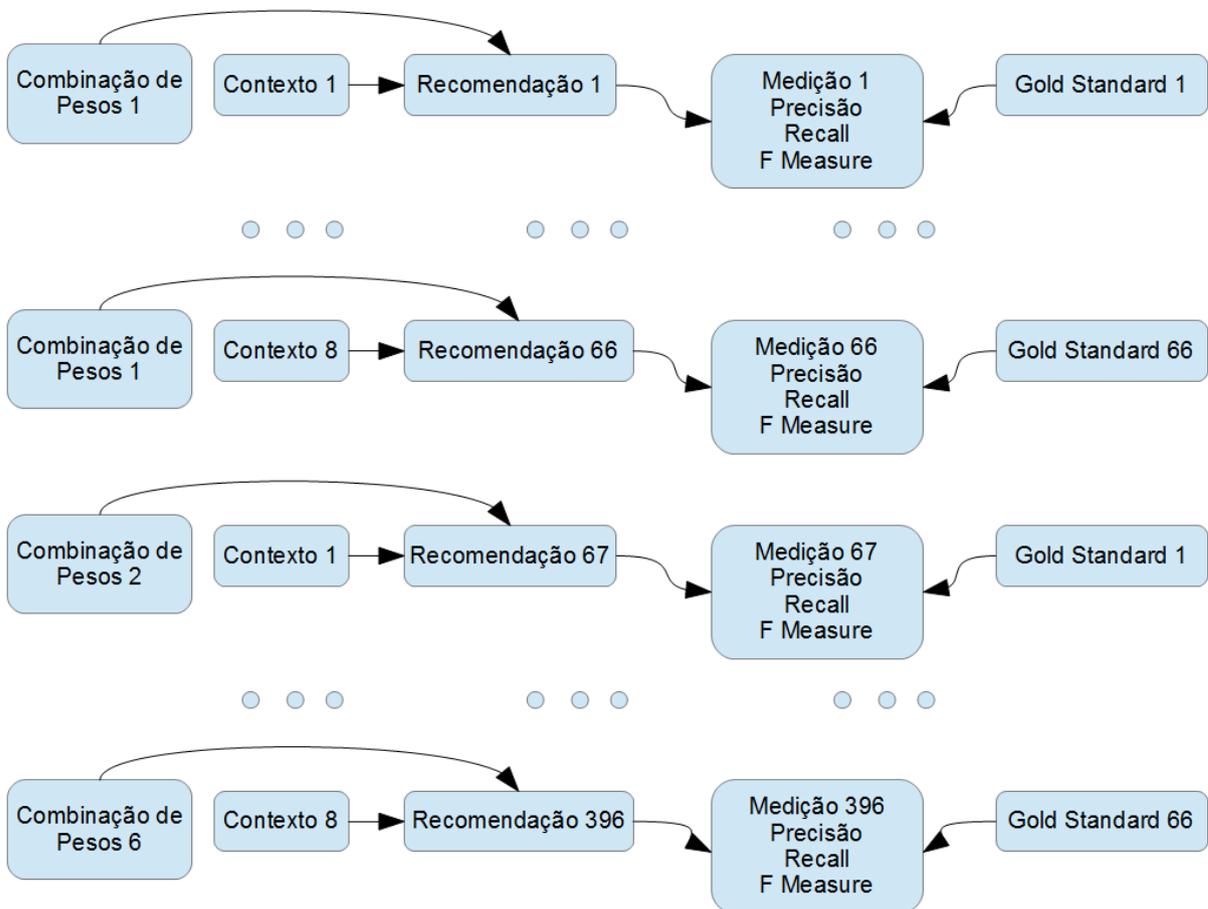


Figura 22 – Sistemática de Execução do Algoritmo

A variação dos pesos de influência das dimensões contextuais permitiu verificar a validade das hipóteses. O *Gold Standard* funcionou ao mesmo tempo como base para validação do sistema na versão sem treinamento (na qual os pesos do cálculo de similaridade foram atribuídos manualmente) e como base histórica na versão do sistema com treinamento (na qual os pesos do cálculo de similaridade foram atribuídos pelo método dos Mínimos Quadrados).

Na versão sem treinamento, foram feitas 4 combinações. Duas delas consideraram somente uma das dimensões contextuais (e.g. só a localização é considerada). Uma delas atribuiu-se pesos iguais às três dimensões contextuais. Outra combinação (Combinação I) baseada em observações empíricas sobre a importância das dimensões contextuais para o gerenciamento de fotos pessoais também foi realizada. Nela foi atribuído o dobro de importância da Atividade em relação à Localização e o quádruplo em relação aos atributos

temporais. Esta proporção levou em consideração a influencia de cada dimensão contextual na identificação de fotos por usuários em estudo feito por (Ames and Naaman, 2007).

Na versão com base histórica, utilizou-se os pesos obtidos a partir do treinamento do algoritmo em que, para um contexto, as avaliações feitas nos outros contextos foram utilizadas como treinamento pelo método *n-fold* (Jannach et al., 2011). A obtenção dos pesos nos dados de treinamento foi feita por regressão linear utilizando o método dos Mínimos Quadrados. Com a utilização de treinamento para a obtenção dos pesos pelo Mínimos Quadrados, vale ressaltar que o histórico de avaliações não armazena nenhuma informação de perfil do usuário e também que a utilização de treinamento pelo *n-fold* não utiliza nenhum conhecimento prévio do usuário, não invalidando, assim, a Hipótese I.

A escolha aleatória foi utilizada para efeitos comparativos com as combinações de pesos, já que não foi possível encontrar outros algoritmos de recomendação de fotos sensíveis ao contexto.

Como mencionado anteriormente, para avaliar os resultados do algoritmo utilizou-se as medidas de precisão, *recall* (Cleverdon et al., 1966) e *F-measure* (Yang and Liu, 1999). Neste tipo de avaliação, para cada recomendação, os itens são separados em quatro categorias: itens relevantes (aqueles avaliados positivamente pelo usuário), itens não relevantes (aqueles avaliados negativamente pelo usuário), itens selecionados para recomendação e itens não selecionados para recomendação. Seja a quantidade de itens de cada categoria definida segundo a Tabela 6.

	Selecionados para Recomendação	Não Selecionados para Recomendação	Total
Relevantes	N_{rs}	N_{rn}	N_r
Irrelevantes	N_{is}	N_{in}	N_i
Total	N_s	N_n	N

Tabela 6 – Categorias de Itens Geradas após a Avaliação e Recomendação

A precisão é calculada pela proporção dos itens relevantes recomendados (N_{rs}) em relação ao conjunto total de itens selecionados para recomendação (N_s). Ela é medida pela Fórmula 11.

$$Precisão = \frac{N_{rs}}{N_s}$$

Fórmula 11 – Cálculo da Precisão (Cleverdon et al., 1966)

A Tabela 7 e o gráfico da Figura 23 mostram a média da precisão das execuções do algoritmo de recomendação em relação a 4 tamanhos de listas de recomendação (*Top 3*, *Top 5*, *Top 10* e *Top 20*). Na Tabela 7, vemos que a média da precisão do algoritmo sem treinamento, atribuindo pesos iguais a todas as medidas de similaridade foi de 0.54 para o *Top 5* (5 itens recomendados). Utilizando os pesos obtidos pelo método dos Mínimos Quadrados, a precisão do *Top 5* foi de 0.58. Recuperando-se as fotos de maneira aleatória, sem o *ranking* gerado pela recomendação, a média da precisão foi de 0.28. As duas últimas linhas da tabela mostram a média da precisão ao se utilizar o cálculo de similaridade de somente uma das dimensões contextuais.

Dentre os experimentos sem base de treinamento, a Combinação I obteve os melhores resultados. É uma característica da métrica de Precisão que à medida que o tamanho da lista de fotos recomendadas aumenta, a tendência é que a proporção de recomendações corretas diminua. É salutar ressaltar que a lista de recomendação não deve ser extensa para uma visualização em telas pequenas como a dos dispositivos móveis. A precisão aleatória chega a ter um valor expressivo devido ao fato de alguns usuários terem classificado muitas fotos relevantes para determinados contextos. Um usuário específico, por exemplo, classificou como relevante metade das fotos que lhe foram apresentadas em um dos contextos.

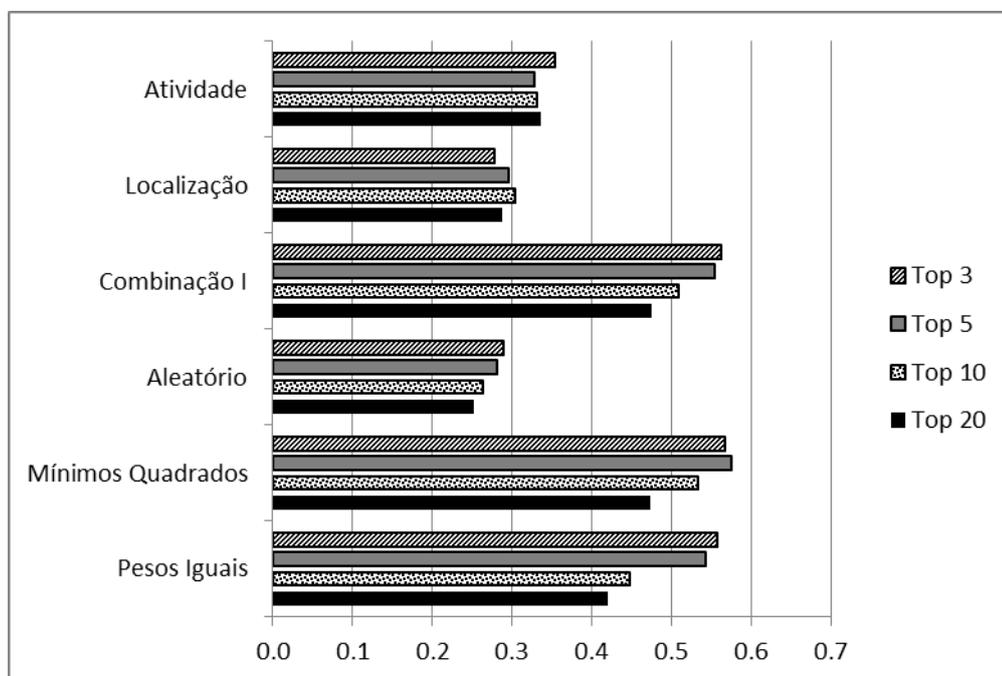


Figura 23 – Precisão Média Obtida pela Recomendação

	Top 3	Top 5	Top 10	Top 20
Pesos Iguais	0.5572	0.5433	0.4478	0.4201
Mínimos Quadrados	0.5672	0.5761	0.5328	0.4739
Aleatório	0.2886	0.2806	0.2627	0.2515
Combinação I	0.5622	0.5552	0.509	0.4746
Localização	0.2786	0.2955	0.3030	0.2873
Atividade	0.3532	0.3284	0.3313	0.3358

Tabela 7 – Precisão Média Obtida pela Recomendação

Comparando-se os resultados obtidos sem treinamento (Combinação I e Pesos Iguais) em relação ao experimento aleatório vemos que é possível recomendar com uma precisão bem maior do que a escolha aleatória (corroborando com as Hipóteses I e II). Além disso, o ganho em relação ao aleatório não foi tão acentuado ao considerar somente uma dimensão contextual (e.g., somente localização ou somente atividade), nos levando a crer que um modelo de contexto completo é essencial para uma boa recomendação contextual (Hipótese III).

Outra métrica utilizada foi o *recall*, que representa a proporção dos itens relevantes recomendados (N_{rs}) em relação ao conjunto total de itens relevantes (N_r). O *recall* é calculado pela Fórmula 12.

$$Recall = \frac{N_{rs}}{N_r}$$

Fórmula 12 – Cálculo do *Recall* (Cleverdon et al., 1966)

Mesmo a precisão individualmente sendo a medida mais apropriada para sistemas que não se propõem a retornar todos os objetos potencialmente relevantes (Herlocker et al., 2004) a taxa de *recall* foi estudada para que sejam obtidos mais indicadores sobre as variações na qualidade da recomendação na variação das dimensões contextuais utilizadas e de seus respectivos pesos de influência. A Figura 24 e a Tabela 8 mostram a média dos valores do *recall* medida em cada combinação. Os resultados encontrados estão consistentes com os obtidos na precisão, em que o resultado da Combinação I e Mínimos Quadrados foi consideravelmente superior ao Aleatório e superior aos que consideram somente uma dimensão contextual. O baixo valor de *recall* medido, principalmente no *Top 3* e *Top 5*, já era

esperado devido, principalmente, a pequena quantidade de itens recomendados frente a quantidade de itens considerados relevantes pelos usuários na geração do *Gold Standard*.

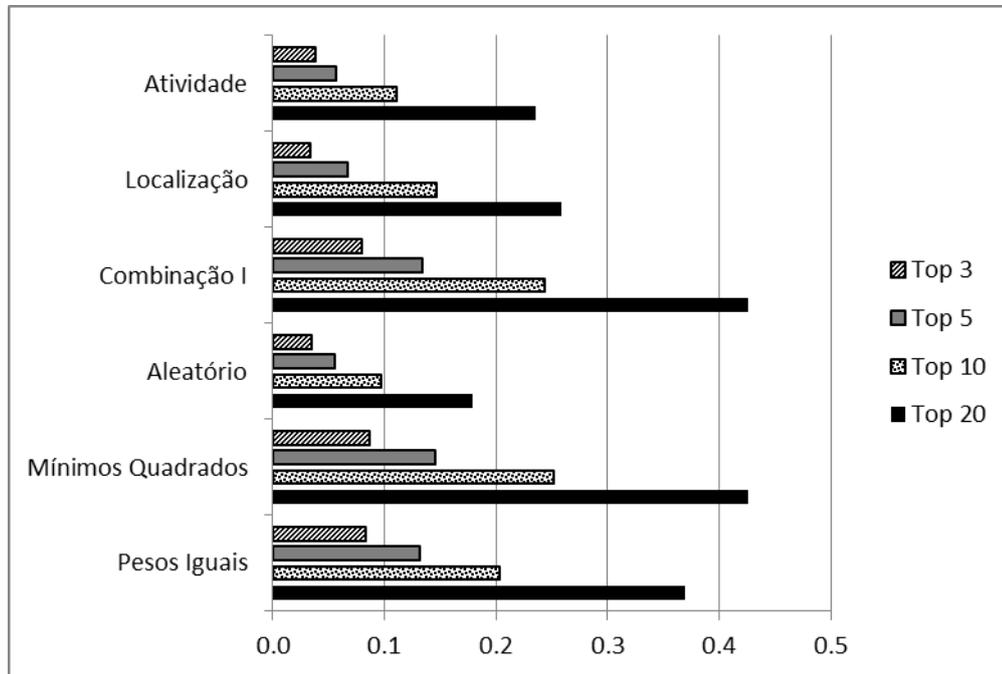


Figura 24 – Recall Médio Obtido pela Recomendação

	Top 3	Top 5	Top 10	Top 20
Pesos Iguais	0.0837	0.1320	0.2031	0.3700
Mínimos Quadrados	0.0861	0.1454	0.2521	0.4269
Aleatório	0.0352	0.0550	0.0973	0.1793
Combinação I	0.0800	0.1337	0.2435	0.4263
Localização	0.0329	0.0665	0.1469	0.2589
Atividade	0.0382	0.0562	0.1107	0.2360

Tabela 8 – Recall Médio Obtido pela Recomendação

Os valores de precisão e o *recall* possuem tendência de crescimento inversas. A medida que o tamanho da lista de itens recomendados aumenta, a expectativa é que a precisão diminua e o *recall* aumente. A medida que o tamanho da lista diminui, o comportamento inverso é esperado. Por este motivo adotou-se também o *F-measure* como medida. O *F-measure* agrega os valores da precisão e *recall* e é calculado pela Fórmula 13.

$$FMeasure = \frac{2 * Precisão * Recall}{Precisão + Recall}$$

Fórmula 13 – Cálculo do *F-measure* (Yang and Liu, 1999)

A Figura 25 e a Tabela 9 mostram o valor do *F-measure* (média harmônica) para os seis pesos utilizados. Pesos Iguais, Combinação I e Mínimos Quadrados obtiveram os melhores resultados para o *Top 3* e *Top 5*. Na recomendação *Top 10* e *Top 20* a Combinação I e os Mínimos Quadrados tiveram um maior destaque. Para o *Top 20*, por exemplo, os Mínimos Quadrados, com *F-measure* de 0.40 foi duas vezes melhor do que o resultado Aleatório, com *F-measure* 0.19.

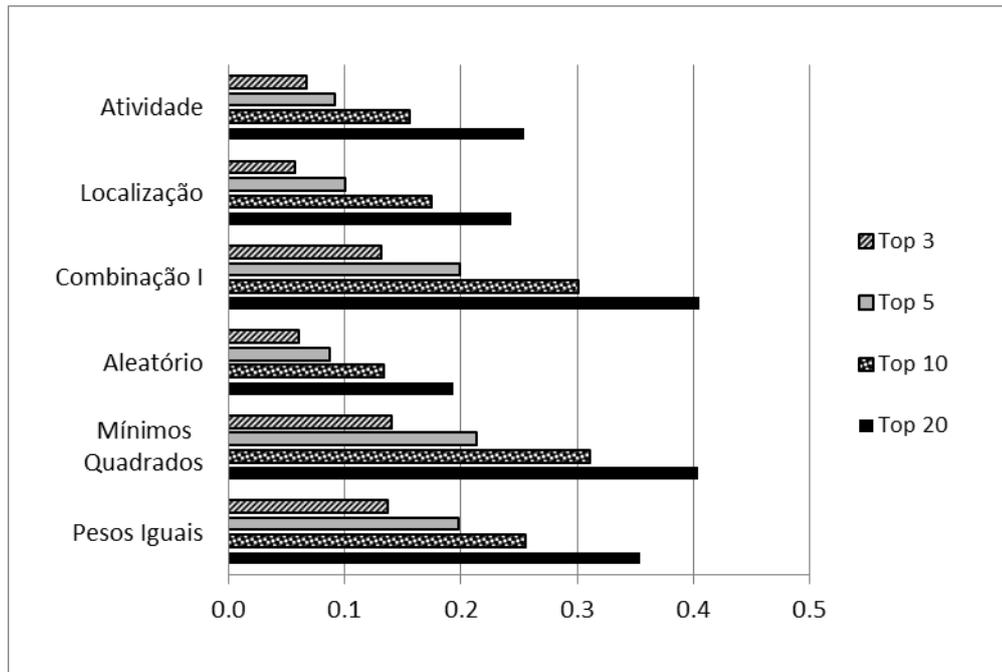


Figura 25 – *F-measure* Médio Obtido Pela Recomendação

	Top 3	Top 5	Top 10	Top 20
Pesos Iguais	0.1369	0.1976	0.2560	0.3546
Mínimos Quadrados	0.1408	0.2136	0.3112	0.4039
Aleatório	0.0601	0.0875	0.1332	0.1934
Combinação I	0.1309	0.1988	0.3005	0.4050
Localização	0.0566	0.0999	0.1747	0.2436
Atividade	0.0667	0.0919	0.1561	0.2540

Tabela 9 – *F-measure* Médio Obtido Pela Recomendação

Para dar maior validade aos resultados alcançados, a utilização de um teste estatístico foi realizada. Foi utilizado o método de análise de variância *Kruskal-Wallis*. Esse método é indicado para a análise dos dados obtidos pois não faz nenhuma restrição quanto a distribuição ou variância dos valores de entrada. O ganho na precisão dos mínimos quadrados em relação ao resultado aleatório possui um nível de significância (nível-p) de $1.04e^{-7}$. Por sua vez, o ganho na precisão da Combinação I em relação ao resultado aleatório possui um nível de significância de $4.33e^{-7}$. Os baixos valores encontrados permitem concluir que os

resultados possuem grande significância estatística e que os ganhos na precisão não foram decorrência do acaso.

Os valores obtidos também mostraram que a utilização de apenas um atributo (localização ou atividade, por exemplo) não é significativamente melhor que a escolha aleatória. Por fim, a comparação entre os resultados da Combinação I e do Mínimos Quadrados resultou em diferenças de performance sem significância estatística, o que pode indicar a não necessidade de treinamento para melhoria da precisão do Sistema de Recomendação.

6.4 Questionário com os Usuários

Em cada etapa dos experimentos, um questionário era aplicado aos usuários para que fossem investigados fatores relevantes na implementação de um Sistema de Recomendação contextual de fotos.

A Tabela 10 contém as perguntas e o percentual de cada resposta. Um dos fatores a ser investigado era o da própria relevância de um sistema móvel de recomendação de fotos e no caso da existência de tal sistema, se seria interessante a recomendação de fotos tiradas em um contexto parecido ao que se encontra o usuário. Nos experimentos, a cada novo contexto passado ao usuário, lhe era perguntado se, nesse contexto específico, ele gostaria de receber recomendações de fotos tiradas em um contexto parecido. O grau de interesse por um Sistema de Recomendação de fotos foi de 74% das ocasiões apresentadas. Os motivos variaram dentre aqueles que, em um contexto específico, não teriam interesse na recomendação. Algumas respostas foram de que a recomendação não seria interessante naquele contexto porque o usuário já conhecia bem a localização dada. Outros disseram que em uma atividade como festas, não teriam o interesse de acessar nenhum tipo de aplicativo móvel e que a recomendação seria interessante, possivelmente, em um período posterior. Além das perguntas aplicadas a cada experimento, no final de todo o processo de seleção de fotos foi perguntado aos usuários sua familiaridade com a utilização de aplicativos móveis com sensores de localização e sobre o seu interesse em receber recomendações contextuais de fotos.

O percentual de cada resposta pode ser visto na Tabela 11. Quando perguntados se, de maneira geral sem considerar um contexto específico, a recomendação de fotos geradas em contextos similares seria interessante, o grau de interesse foi de 100%.

Outro ponto investigado foi a relevância relativa de cada dimensão contextual. A cada usuário foi perguntado qual a dimensão contextual, entre localização, atividade de interesse e data e hora, eles prefeririam que as fotos recomendadas mais se aproximassem das deles. Oito dos usuários responderam que acham a proximidade da localização da foto o fator mais importante para despertar seu interesse. Cinco dos usuários responderam que a associação da atividade retratada pela foto com a atividade que ele está desempenhando no momento é o fator mais importante. Nenhum usuário achou a similaridade entre data e hora o fator mais importante.

Pergunta	Resposta Afirmativa	Resposta Negativa
Você conhece o local da situação apresentada?	59%	41%
A atividade apresentada desperta seu interesse?	71%	29%
Nas situações apresentadas, você acharia interessante acessar algum sistema que recomendasse fotos de situações semelhantes?	74%	26%

Tabela 10 – Questionário: uma Pergunta por Experimento

Pergunta	Resposta Afirmativa	Resposta Negativa
Nas situações apresentadas ou em qualquer outra situação de mobilidade, você se interessaria em receber recomendações de fotos que foram tiradas em situações parecidas?	100%	0%
Você possui smartphone com sensores de Localização (GPS)?	77%	23%
Já usou algum aplicativo móvel baseado em localização?	62%	38%

Tabela 11 – Questionário: uma Pergunta por Usuário

6.5 Construção do Repositório de Fotos

O *MMedia2U* possui um repositório de fotos para recomendação, que precisam ter informações contextuais para que sejam comparadas com o contexto dos usuários a quem serão recomendadas. Precisa-se saber, por exemplo, a localização em que a foto foi tirada e a atividade de interesse do autor da foto no momento de sua geração. Em um primeiro

momento, esperava-se utilizar diretamente bases de fotos de aplicações Web 2.0, tais como *Flickr* e *Picasa Web*. Entretanto, foi percebido que muitas das fotos destas bases não contêm todas as informações necessárias para o preenchimento do modelo de contexto e que também apresentam erros nas informações presentes nos metadados. Para a realização dos experimentos, foi construído, portanto, um repositório de fotos com as informações necessárias a partir de fotos recuperadas do *Picasa Web*, utilizando seus *web-services* para consulta de fotos georreferenciadas e disponibilizadas publicamente. Além do conteúdo das fotos, os *web-services* retornam os metadados no formato XML. Tais metadados contêm as informações utilizadas para alimentar o modelo de contexto das fotos com as informações de localização e data e hora. Para que estas informações possam ser lidas por visualizadores de fotos, elas precisam ser incorporadas ao arquivo da foto nos cabeçalhos *IPTC* e *EXIF*. A inserção, nos arquivos de cada foto, dos metadados recuperados do *Picasa Web* foi feita utilizando-se a biblioteca de manipulação de metadados de imagens *Apache Sanselan*¹⁹.

Após a incorporação dos metadados nas imagens, estes foram enriquecidos, manualmente, para que pudessem ser melhor explorados no algoritmo de similaridade, exposto na Seção 4.4. Exemplos de enriquecimento são a inferência da atividade de interesse a partir da descrição ou legendas das fotos e do dia da semana de acordo com a sua data de geração. Espera-se que com a evolução dos sistemas de gerenciamento de conteúdo multimídia, o esforço de enriquecimento e incorporação de metadados nas imagens seja bastante reduzido.

6.6 Conclusão

Este capítulo mostrou os resultados obtidos com a utilização da abordagem de recomendação contextual proposta neste trabalho. A recomendação foi realizada em um repositório de fotos com informações contextuais e foi verificado o grau de adequação das fotos recomendadas em relação às fotos consideradas relevantes pelos usuários. O levantamento das fotos relevantes foi realizado a partir de experimentos em que foi construído um *Gold Standard* para avaliação do algoritmo. A partir dos resultados quantitativos

¹⁹ <http://commons.apache.org/imaging/>

alcançados a partir dos experimentos com o *MMedia2U* (vistos na Seção 6.3), as três hipóteses levantadas foram validadas. As respostas dadas ao questionário também trazem indicativos da importância da aplicação da similaridade contextual em um processo de recomendação. No próximo capítulo é feita a conclusão do trabalho e levantados os trabalhos futuros.

7 CONCLUSÃO E TRABALHOS FUTUROS

A sensibilidade ao contexto possui reconhecida importância na construção de Sistemas de Recomendação. Porém, mesmo com o crescente avanço das técnicas de recomendação sensível ao contexto, algumas limitações ainda precisam ser superadas. Uma dessas limitações se refere a existência de meios de se explorar em todo potencial repositórios de itens com informações contextuais em um modelo semântico que permita a aplicação de regras de domínio. Essa dissertação propôs uma abordagem de recomendação sensível ao contexto que, aplicando uma recomendação baseada em conhecimento, explora as informações contextuais dos itens passíveis de recomendação com o objetivo de recomendar ao usuário aqueles que melhor se adequem ao seu contexto. No restante desse capítulo, a Seção 7.1 descreve os resultados alcançados. A Seção 7.2 apresenta as publicações originadas a partir dessa dissertação e a Seção 7.3 expõe os trabalhos futuros previstos.

7.1 Resultados Alcançados

Na abordagem proposta nesse trabalho, técnicas de recomendação baseadas em conhecimento foram adaptadas com a incorporação da sensibilidade ao contexto, permitindo que itens, mesmo sem nunca terem sido avaliados, possam ser recomendados somente a partir do seu contexto de geração.

A abordagem permite, também, que usuários, mesmo sem histórico de uso do sistema, recebam recomendações de acordo com seu contexto atual. Ao compor as características dos itens a partir de suas informações de contexto, a abordagem supera algumas das dificuldades de composição de um Sistema de Recomendação baseado em conhecimento.

No domínio foco da validação da abordagem (i.e., gerenciamento de fotos), essas informações já são geradas automaticamente por sistemas de geração de conteúdo sensíveis ao contexto (Viana et al., 2011) e sistemas de organização e busca de fotos baseados em informações contextuais dos itens (Figueirêdo et al., 2012). A abordagem proposta

comprovou que as informações geradas ou indexadas por estes tipos de sistemas podem ser utilizadas em processos de recomendação, diminuindo-se os custos de manutenção de uma base de itens rica em semântica e informações contextuais.

Na elaboração da proposta foi também construído um modelo de contexto que permite a representação de informações contextuais em sistemas de informação. Um algoritmo de recomendação e fórmulas de cálculo de similaridade para um conjunto de dimensões contextuais também foram propostos. Durante a modelagem da abordagem, as responsabilidades da recomendação e do cálculo de similaridade foram separadas para que tanto o algoritmo de recomendação quanto as fórmulas de similaridade pudessem ser mais facilmente reutilizados em outros sistemas.

Dentre as principais contribuições deste trabalho destaca-se ainda a elaboração de um sistema de recomendação que segue a abordagem proposta e a avaliação do impacto do uso de informações contextuais no domínio de aplicação escolhido (i.e., recomendação de fotos), detalhados a seguir.

O *MMedia2U*, um sistema móvel de recomendação sensível ao contexto de fotos, foi desenvolvido nesta dissertação como prova de conceito para realizar a recomendação de fotos que foram geradas em contexto similar ao do usuário do sistema. O sistema foi construído em uma arquitetura cliente-servidor, em que a aplicação móvel captura o contexto do usuário, acessa remotamente o serviço de recomendação e exibe as fotos recomendadas.

A recomendação nesta prova de conceito foi validada a partir da construção de um *Gold Standard* e os valores dos indicadores de qualidade obtidos pelo algoritmo nos permitiram concluir que, para os dados utilizados, a sensibilidade ao contexto pode trazer ganhos na recomendação de fotos se comparado a uma lista aleatória. É importante destacar que mesmo combinações de peso que não utilizaram treinamento (a Combinação I e pesos iguais, vistas na Seção 6.3) obtiveram resultados satisfatórios, o que permite a utilização do sistema mesmo sem uma base histórica de utilização. Esse resultado indica uma redução do inconveniente do início frio de Sistemas de Recomendação tradicionais. Além disso, o questionário aplicado aos participantes dos experimentos sugere que sistemas desta natureza despertam o interesse de usuários.

O código fonte do *MMedia2U* está disponível em <https://github.com/fabriciolemos/mmedia2u>.

A Tabela 12 mostra como as características do *MMedia2U* podem ser comparadas com as características dos trabalhos relacionados, vistos na Seção 3.4.

Sistema	Recomendação	Sensível ao contexto	Contexto Incorporado na Recomendação	
			Contexto do Usuário	Contexto dos Itens
<i>COMPASS</i>	Baseado em Conteúdo	Sim	Localização	Localização
<i>Kaleido Photo</i>	Baseado em Conteúdo	Não	Nenhum	Localização, Data, Contatos
<i>Columbus</i>	Baseada em Conhecimento	Sim	Localização	Localização
Adomavicius et al.	Colaborativa	Sim	Localização, Dia da Semana, Companhia	Nenhum
<i>C2_Music</i>	Colaborativa	Sim	Localização, Data, Condições Meteorológicas	Nenhum
<i>MMedia2U</i>	Baseada em Conhecimento	Sim	Localização, Atividade, Data e Horário	Localização, Atividade, Data e Horário

Tabela 12 – *MMedia2U* e Trabalhos Relacionados

7.2 Publicações

A partir do trabalho desenvolvido nessa dissertação, foram originadas três publicações, descritas na Tabela 13.

Título	Autores	Conferência
Um Sistema de Recomendação de Fotos Sensível ao Contexto	Fabício Lemos, Rafael Carmo, Windson Viana e Rossana Andrade	Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), São Paulo 2012
Towards a Context-Aware Photo Recommender System	Fabício Lemos, Rafael Carmo, Windson Viana e Rossana Andrade	Context-Aware Recommender Systems Workshop, Dublin 2012
Uma Proposta para o Desenvolvimento de Sistemas de Recomendação de Itens Anotados com Informações Contextuais.	Fabício Lemos, Windson Viana e Rossana Andrade	Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), Florianópolis 2011

Tabela 13 – Publicações

7.3 Trabalhos Futuros

É importante ressaltar que os resultados da experimentação não podem ser generalizados (dependência dos usuários e da base de fotos utilizadas), entretanto, funcionam como um bom indicativo de qualidade do protótipo de recomendação. A análise dos resultados alcançados foi limitada ao tamanho da base de fotos utilizada (655 fotos, com 335 de duas regiões turísticas de Fortaleza), pois o aumento da base poderia diminuir a precisão do algoritmo ou apresentar problemas de concentração de fotos em determinados lugares. Bases de fotos colaborativas, como o *Flickr* e o *Picasa*, poderão apresentar este problema. Nesses casos, novos filtros de busca e algoritmos de clusterização poderiam ser utilizados para amenizar os desafios. Possibilitar que os usuários acrescentem palavras de interesse como filtro de busca, pode também auxiliar para o aumento da precisão das recomendações.

Com o objetivo de melhorar o conteúdo do repositório de fotos do *MMedia2U*, é necessário realizar o aumento da base a partir da incorporação de fotos enriquecidas, possivelmente usando uma evolução do sistema *CoMMeDiA* (Viana et al., 2011). Essa evolução garantiria uma maior exatidão das informações anotadas contextualmente e a sua aquisição no repositório de forma automática. O aumento da base de fotos do *MMedia2U* permitiria também a avaliação do sistema com usuários (turistas ou não) em situação de mobilidade real em vários lugares de Fortaleza.

Um importante trabalho futuro seria avaliar o impacto da abordagem em outros domínios de sistemas nos quais o contexto de uso pode ser capturado e, potencialmente, influenciar as preferências do usuário. Alguns domínios candidatos são: recomendação de aplicativos, recomendação de modificações da interface gráfica de um dispositivo móvel, recomendação de artefatos de implementação de código, entre outros.

A adoção de um modelo híbrido de recomendação, adotando-se técnicas de recomendação colaborativas e baseadas em conteúdo, pode também ser utilizada na evolução da abordagem proposta. A recomendação colaborativa permitiria uma melhoria na recomendação ao considerar avaliações de itens feitas por outros usuários. A recomendação baseada em conteúdo permitiria o aprendizado das características dos itens que mais agradam ao usuário.

8 BIBLIOGRAFIA

- Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems* 23, 103–145. Available at: <http://portal.acm.org/citation.cfm?doid=1055709.1055714>.
- Adomavicius, G., and Tuzhilin, A. (2008). Context-aware recommender systems. *Proceedings of the 2008 ACM conference on Recommender systems - RecSys '08*, 335. Available at: <http://portal.acm.org/citation.cfm?doid=1454008.1454068>.
- Ailisto, H., Alahuhta, P., Haataja, V., Kyllönen, V., and Lindholm, M. (2002). Structuring context aware applications: Five-layer model and example case. *Proceedings of the Workshop on Concepts and Models for Ubiquitous Computing*, 1–5. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.2282&rep=rep1&type=pdf>.
- Ames, M., and Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 971–980. Available at: <http://dl.acm.org/citation.cfm?id=1240772> [Accessed October 10, 2012].
- Antunes, B., Cordeiro, J., and Gomes, P. (2012). An approach to context-based recommendation in software development. In *ACM conference on Recommender systems (RecSys)* (New York, NY, USA). Available at: <http://doi.acm.org/10.1145/2365952.2365986>.
- Baldauf, M., Dustdar, S., and Rosenberg, F. (2007). A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing* 2, 263–277. Available at: <http://www.inderscience.com/link.php?id=14070>.
- Bettini, C., Brdiczka, O., Henriksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing* 6, 161–180. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1574119209000510> [Accessed July 27, 2011].
- Björck, Å. (1996). *Numerical methods for least squares problems* (Philadelphia, PA: SIAM Press) Available at: <http://link.aip.org/link/?SJMAEL/31/2400/1>.
- Boutemedjet, S. D. Z. (2010). Using Images in Context-Aware Recommender Systems. In *1st International Workshop on Adaptation, Personalization and Recommendation in the Social-Semantic Web* Available at: <http://ir.ii.uam.es/apresw2010/papers/apresw2010-paper04.pdf> [Accessed June 2, 2012].

- Braga, R. B., Carvalho, W. V. de, Andrade, R. M. C., and Martin, H. (2011). CAPTAIN: Uma ferramenta para geração de diários de bordo digitais usando anotação contextual e conteúdo multimídia. In *Simpósio Brasileiro de Sistemas Multimídia e Web WebMedia*.
- Burke, R. (2007). Hybrid Web Recommender Systems. *The Adaptive Web 4321*, 377 – 408. Available at: <http://portal.acm.org/citation.cfm?id=1768211>.
- Burke, R. (2002). Interactive Critiquing for Catalog Navigation in E-Commerce. *Artificial Intelligence Review 18*, 245–267.
- Burke, R. (2000). Knowledge Based Recommender System. *Encyclopedia of Library and Information Systems 69*, 175–186.
- Böhmer, M., Prinz, M., and Bauer, G. (2010). Contextualizing mobile applications for context-aware recommendation. In *Adjunct Proceedings of Pervasive* Available at: <http://www.dfki.de/~mabo04/publications/Boehmer-PERVASIVE10-ContextualizingMobileApps.pdf> [Accessed October 12, 2012].
- Cleverdon, C. W., Mills, J., and Keen, M. (1966). Factors determining the performance of indexing systems. *Aslib Cranfield Research Project Cranfield England Vol 2*, 37–59. Available at: <https://dspace.lib.cranfield.ac.uk/handle/1826/863>.
- Dey, A. K., and Abowd, G. D. (2000). Towards a better understanding of context and context-awareness. In *CHI 2000 workshop on the what, who, where, when, and how of context-awareness* (Citeseer), pp. 1–6. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.4833&rep=rep1&rep_type=pdf [Accessed April 24, 2011].
- Deza, M. M., and Deza, E. (2009). *Encyclopedia of Distances* (Springer Berlin Heidelberg) Available at: <http://www.springerlink.com/index/10.1007/978-3-642-00234-2>.
- Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2006). An Integrated Environment for the Development of Knowledge-Based Recommender Applications. *International Journal of Electronic Commerce 11*, 11–34. Available at: <http://mesharpe.metapress.com/openurl.asp?genre=article&id=doi:10.2753/JEC1086-4415110201>.
- Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. *Building 54*, 162. Available at: <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>.
- Figueirêdo, H. F. D., Lacerda, Y. A., Paiva, A. C. de, Casanova, M. A., and Baptista, C. D. S. (2012). PhotoGeo: a photo digital library with spatial-temporal support and self-annotation. *Multimedia Tools and Applications 59*, 279–305. Available at: <http://dx.doi.org/10.1007/s11042-011-0745-x>.
- Gamma, E., Helm, R., Johnson, R., and Vlissides, J. (1995). *Design Patterns* A.-W. P. Co, ed. (Addison Wesley) Available at: <http://www.amazon.co.uk/exec/obidos/ASIN/0201633612/citeulike-21>.

- Ganneau, V., Calvary, G., and Demumieux, R. (2008). Learning Key Contexts of Use in the Wild for Driving Plastic User Interfaces Engineering. In Computer, Forbrig P And Paterno F, ed. (Springer-Verlag), pp. 271–278. Available at: <http://portal.acm.org/citation.cfm?id=1459880>.
- Gu, T., Pung, H. K., and Zhang, D. Q. (2004). A middleware for building context-aware mobile services. In Vehicular Technology Conference 2004 VTC 2004Spring 2004 IEEE 59th (Ieee), pp. 2656–2660. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1391402.
- Herlocker, J. L., Konstan, J. a., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22, 5–53. Available at: <http://portal.acm.org/citation.cfm?doid=963770.963772>.
- Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2011). *Recommender Systems: An Introduction* (Cambridge University Press) Available at: <http://www.worldcat.org/title/recommender-systems-an-introduction/oclc/645789647>.
- Kennedy, L., Naaman, M., Ahern, S., Nair, R., and Rattenbury, T. (2007). How flickr helps us make sense of the world: context and content in community-contributed media collections. In *Proceedings of the 15th international conference on Multimedia (ACM)*, pp. 631–640. Available at: <http://portal.acm.org/citation.cfm?id=1291384>.
- Lee, J. S., and Lee, J. C. (2007). Context awareness by case-based reasoning in a music recommendation system. *Ubiquitous Computing Systems*, 45–58. Available at: <http://www.springerlink.com/index/N9761104686H55R0.pdf> [Accessed April 24, 2011].
- Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7, 76–80. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1167344>.
- Maidel, V., Shoval, P., Shapira, B., and Taieb-Maimon, M. (2008). Evaluation of an ontology-content based filtering method for a personalized newspaper. *Proceedings of the 2008 ACM conference on Recommender systems - RecSys '08*, 91. Available at: <http://portal.acm.org/citation.cfm?doid=1454008.1454024>.
- Marinho, F. G., Costa, A. L., Lima, F. F. P., Neto, J. B. B., Filho, J. B. F., Rocha, L., Dantas, V. L. L., Andrade, R. M. C., Teixeira, E., and Werner, C. (2010). An Architecture Proposal for Nested Software Product Lines in the Domain of Mobile and Context-Aware Applications. *Brazilian Symposium on Software Components, Architectures and Reuse*, 51–60. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5631701> [Accessed August 22, 2011].
- Mcsherry, D. (2003). Similarity and Compromise. In *Case-Based Reasoning Research and Development*, K. Ashley and D. Bridge, eds., pp. 1–15. Available at: <http://www.springerlink.com/content/yfhu15b1vrxfqya6>.
- Pinheiro, M. K., Villanova-Oliver, M., Gensel, J., Berbers, Y., and Martin, H. (2008). Personalizing Web-Based Information Systems through Context-Aware User Profiles. In

- Conference on Mobile Ubiquitous Computing Systems Services and Technologies, pp. 231–238. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4641341>.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens : An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, R. Furuta and C. M. Neuwirth, eds. (ACM), pp. 175–186. Available at: <http://portal.acm.org/citation.cfm?id=192844.192905&type=series>.
- Rost, M., Cramer, H., and Holmquist, L. E. (2011). Mobile exploration of geotagged photographs. *Personal and Ubiquitous Computing*. Available at: <http://www.springerlink.com/index/10.1007/s00779-011-0433-x> [Accessed September 5, 2011].
- Santos, J., and Ferraz, C. (2011). Uma Arquitetura Flexível de Suporte à Execução de Algoritmos de Recomendação de Programas em TV Conectada. In *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*.
- Schmidt, A., Beigl, M., and Gellersen, H.-W. (1999). There is more to context than location. *Computers & Graphics* 3, 421–491.
- Van Setten, M., Pokraev, S., and Koolwaaij, J. (2004). Context-aware recommendations in the mobile tourist application COMPASS. In *Adaptive Hypermedia and Adaptive Web-Based Systems* (Springer), pp. 515–548. Available at: <http://www.springerlink.com/index/YK2M700K15KMDEJ2.pdf> [Accessed April 24, 2011].
- Siersdorfer, S., and Sizov, S. (2009). Social recommender systems for web 2.0 folksonomies. In *ACM conference on Hypertext and hypermedia - HT '09* (New York, New York, USA: ACM Press), p. 261. Available at: <http://portal.acm.org/citation.cfm?doid=1557914.1557959>.
- Slimani, T., Yaghlane, B. B., and Mellouli, K. (2006). A New Similarity Measure based on Edge Counting. *Engineering and Technology*, 34–38.
- Strang, T., and Linnhoff-popien, C. (2004). A Context Modeling Survey. In *Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp*.
- Vargas, S., and Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. *Proceedings of the fifth ACM conference on Recommender systems RecSys 11*, 107, 109. Available at: <http://dl.acm.org/citation.cfm?doid=2043932.2043955>.
- Viana, W., Bringel Filho, J., Gensel, J., Villanova-Oliver, M., and Martin, H. (2008). PhotoMap: from location and time to context-aware photo annotations. *Journal of Location Based Services* 2, 211–235. Available at: <http://www.tandfonline.com/doi/abs/10.1080/17489720802487956>.

- Viana, W., Miron, A. D., Moiscuc, B., Gensel, J., Villanova-Oliver, M., and Martin, H. (2011). Towards the semantic and context-aware management of mobile multimedia. *Multimedia Tools and Applications* 53, 391–429. Available at: <http://www.springerlink.com/index/10.1007/s11042-010-0502-6> [Accessed June 2, 2012].
- Wang, X. H., Gu, T., Zhang, D. Q., and Pung, H. K. (2004). Ontology based context modeling and reasoning using OWL. In *IEEE Annual Conference on Pervasive Computing and Communications Workshops*, 2004. (Ieee), pp. 18–22. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1276898>.
- Weiser, M. (1993). Some computer science issues in ubiquitous computing. In *Communications of the ACM (ACM)*, pp. 75–84. Available at: <http://portal.acm.org/citation.cfm?doid=159544.159617>.
- Xia, S., Gong, X., Wang, W., Tian, Y., Yang, X., and Ma, J. (2010). Context-Aware Image Annotation and Retrieval on Mobile Device. *2010 Second International Conference on Multimedia and Information Technology*, 111–114. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5474266> [Accessed April 12, 2011].
- Yang, Y., and Liu, X. (1999). A re-examination of text categorization methods. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval SIGIR 99 pages*, 42–49. Available at: <http://portal.acm.org/citation.cfm?doid=312624.312647>.
- Ziegler, C.-N., McNee, S. M., Konstan, J. a., and Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th international conference on World Wide Web - WWW '05*, 22. Available at: <http://portal.acm.org/citation.cfm?doid=1060745.1060754>.

APÊNDICES

APÊNDICE A Exemplo de arquivo XML enviado pelo serviço de recomendação do MMedia2U

```

<photos>
  <photo>
    <imageUrl>https://lh6.googleusercontent.com/-
24z0MnUzn9k/SkJm6E_iL2I/AAAAAAAABx0/g1eSNlM2ONs/s963/06+2007-10-
03++001%2C+Parque+ecol%C3%B3gico+%28mosquitero%29+do+Coc%C3%B3%2C+Fortaleza%
2C+Brasil.jpg
    </imageUrl>
    <thumbnailUrl>http://lh5.ggpht.com/-
24z0MnUzn9k/SkJm6E_iL2I/AAAAAAAABx0/g1eSNlM2ONs/s144/06%2525202007-10-
03%252520%252520001%25252C%252520Parque%252520ecol%2525C3%2525B3gico%252520%
252528mosquitero%252529%252520do%252520Coc%2525C3%2525B3%25252C%252520Fortal
eza%25252C%252520Brasil.jpg
    </thumbnailUrl>
    <alternateLink>https://picasaweb.google.com/102657583327319429948/Bra
sil2?authkey=c776KzYMBWw#5350952455228239714
    </alternateLink>
    <title>06 2007-10-03 001, Parque ecológico (mosquitero) do Cocó,
Fortaleza, Brasil.jpg
    </title>
    <description>Parque ecológico (mosquitero) do Cocó, Fortaleza, Brasil
    </description>
    <latitude>-3.718394</latitude>
    <longitude>-38.5433949</longitude>
    <timeTaken>2007-10-04T10:53:40.000-04:00</timeTaken>
    <tags>férias, Fortaleza</tags>
  </photo>
  <photo>
    <imageUrl>https://lh5.googleusercontent.com/-
SWPeVRja9g8/SkJlWlIUMCI/AAAAAAAABsA/nnc1JP5Jh04/s963/01+2007-09-
07++001%2C+Fortaleza%2C+Brasil.jpg
    </imageUrl>
    <thumbnailUrl>http://lh3.ggpht.com/-
SWPeVRja9g8/SkJlWlIUMCI/AAAAAAAABsA/nnc1JP5Jh04/s144/01%2525202007-09-
07%252520%252520001%25252C%252520Fortaleza%25252C%252520Brasil.jpg
    </thumbnailUrl>
    <alternateLink>https://picasaweb.google.com/102657583327319429948/Bra
sil2?authkey=c776KzYMBWw#5350951184860721186
    </alternateLink>
    <title>01 2007-09-07 001, Fortaleza, Brasil.jpg</title>
    <description>Fortaleza, Brasil</description>
    <latitude>-3.718394</latitude>
    <longitude>-38.5433949</longitude>
    <timeTaken>2007-09-07T10:51:53.000-04:00</timeTaken>
    <tags>viagem</tags>
  </photo>
  ...
</photos>

```