

Predição de Notas de Filmes com Lógica Fuzzy

Geovani Anacleto da Silva, Tulio Rodrigues Carreira

Resumo— Este trabalho se propõe a estudar e implementar um sistema de recomendações de filmes utilizando-se de lógica fuzzy para, a partir de médias de notas calculadas levando-se em consideração determinados aspectos, sugerir uma nota que seria proposta para o usuário em questão, e, assim, sugerir os filmes que ele mais aprovaria.

Keywords—*fuzzy, filmes, netflix prize competition*

I. INTRODUÇÃO

A LÓGICA fuzzy, também conhecida como difusa, é uma lógica multivalorada onde os valores lógicos das variáveis podem ser representados por qualquer número real entre 0 e 1 (falso e verdadeiro, respectivamente). A possibilidade de os valores não serem completamente verdadeiros ou completamente falsos permite que se lide com o conceito de verdades parciais *ahlawat*. Com este tipo de lógica, é possível avaliar conceitos não-quantificáveis, como a avaliação de temperatura (quente, morno, frio, etc.), a avaliação da distância entre dois pontos (próximo, muito próximo, longe, muito longe, etc.), e a avaliação do nível de conteúdo de um recipiente (vazio, parcialmente cheio, cheio, etc.). No geral, a lógica fuzzy é útil por sua capacidade de modelar problemas da vida real um tanto quanto imprecisos de raciocínio por serem descritos na linguagem natural. Uma outra aplicação da lógica fuzzy se vê na atribuição de notas (baixa, média, alta) para filmes, problema este que se trata do objeto de estudo deste artigo.

A Netflix, uma provedora multinacional de filmes e séries de televisão via streaming, propôs, em 2006, uma competição com o intuito de premiar o algoritmo de filtragem colaborativa que melhor superasse o próprio algoritmo da Netflix, sendo capaz de prever as notas que seus usuários dariam para novos filmes, utilizando, como base, os filmes por eles já assistidos [?]. Em Setembro de 2009, o prêmio de US\$1,000,000.00 foi concedido a uma equipe que superou o algoritmo da Netflix em 10,06% [?]. Nos vários algoritmos propostos pelos competidores, foram utilizadas técnicas diversas, como modelos de regressão, modelos de vizinhança, fatoração de matrizes, lógica fuzzy, e assim por diante.

Mais detalhadamente, o Netflix Prize fornece uma série de arquivos de textos (os conjuntos de dados para treinamento) que contêm triplas com o identificador do filme, o identificador do usuário, a nota fornecida e a data. As quatro variáveis de entrada do sistema fuzzy dizem respeito a informações como nota da data, nota do filme, nota do usuário e a nota de similaridade, todas a serem elucidadas neste trabalho.

Uma implementação de algoritmo capaz de estimar as notas dadas pelos usuários que envolva lógica fuzzy é de muita pertinência para testar aplicações práticas da lógica difusa, tendo em vista que converte dados exatos (o valor de uma nota: 2, por exemplo) em dados mais próximos da realidade (2 é uma nota predominantemente baixa, mas também um

pouco média). Na solução proposta por [?], projeto-base deste trabalho, existem quatro variáveis de entrada, que, ao serem avaliadas em conjunto, retornam um único valor de resposta: a nota estimada para o filme que determinado usuário daria na data em questão.

Neste contexto, ressalta-se que este trabalho tem por objetivo a implementação de um sistema inteligente que se utiliza de lógica fuzzy para prever as notas que os usuários atribuem para filmes com o objetivo de oferecer recomendações precisas, mostrando a importância de conceitos presentes na lógica fuzzy para a solução de problemas não-binários e inexatos da vida real. Para isso, propõe-se a modelagem e a implementação de um sistema inteligente que se utiliza da biblioteca *jFuzzyLogic* do Java para a conversão de dados de entrada em dados fuzzy para o fornecimento de, como resposta final, valores vertiginosamente mais próximos da realidade.

Algoritmos com mineração de dados são muito relevantes em aplicações provedoras de serviços, por terem o intuito de trazer os melhores resultados para seus usuários. Realizar tal feito com a ajuda de lógica fuzzy é um estudo útil para verificar se esta se trataria de uma boa técnica de solução do problema proposto pela Netflix.

Para tal feito, estruturou-se o trabalho primeiramente expondo conceitos fundamentais da lógica fuzzy aplicados na implementação do sistema; seguidamente, é detalhado o trabalho-base deste projeto, juntamente com a modelagem do sistema, a análise dos resultados provenientes da implementação do sistema e, por fim, as conclusões obtidas no processo.

II. CONCEITOS FUNDAMENTAIS

Os principais conceitos da Lógica Fuzzy que foram utilizados no trabalho são apresentados nesta seção. Todos os conceitos abaixo foram descritos com base em [?].

A. Conceitos de Lógica Fuzzy

1) *Conjunto Fuzzy*: Um conjunto fuzzy A definido em um universo U é caracterizado por uma função de pertinência que mapeia todos os elementos do universo U no intervalo $[0,1]$.

2) *Função de pertinência*: Classifica elementos observados não como meros pertencentes ou não pertencentes de um conjunto, mas sim de acordo com seu respectivo grau de participação (pertinência) no conjuntos sendo levados em consideração. Um determinado elemento pode, por exemplo, participar de dois ou até mais conjuntos em graus diferentes. Dessa forma, as funções de pertinência são o conceito base de funcionamento da lógica Fuzzy. Tais funções são representadas por $u(x)$.

3) *Funções de pertinência contínuas*: Função matemática que pode ter uma entre diversas funções matemáticas. A função gaussiana, presente nas variáveis de entrada da modelagem, tende a zero quando há valores muito maiores ou muito menores que a média.

4) *Grau de pertinência*: Quantifica o pertencimento de determinado elemento em um conjunto a ser considerado, conforme as possibilidades abaixo, em que A representa um elemento e U representa o universo.

$u(x)$ valendo zero indica a não pertinência de A em U;

$u(x)$ valendo 1 indica a pertinência total de A em U;

$u(x)$ entre 0 e 1 indica a pertinência parcial de A em U.

5) *Operador de Implicação Fuzzy*: Regras que seguem o formato se x, então y. A função de pertinência da implicação, dada por $u(R)(x,y)$, retorna o grau de relacionamento entre as premissas (antecedente) e as conclusões (consequente).

6) *Relações Binárias Fuzzy*: Dados dois conjuntos CRISP, X e Y, uma relação fuzzy sobre eles é qualquer conjunto fuzzy que obedece a máxima: $u(R): X \times Y \rightarrow [0][1]$, onde $u(R)$ representa a função de pertinência

7) *Regras de inferência fuzzy*: Levam ao raciocínio aproximado (processo onde conclusões são obtidas a partir de premissas). Especificam conclusões obtidas através de informações conhecidas.

B. Sistema de Inferência Fuzzy

1) *Interface de Fuzzificação*: Como primeira etapa, é importante que se adquira conhecimento acerca do processo que se quer controlar. Assim, a interface de fuzzificação é o passo responsável por analisar o problema, definir variáveis linguísticas, criar conjuntos fuzzy e definir funções de pertinência.

2) *Base de Regras*: Tem por objetivo representar, de forma sistemática, a maneira como o controlador humano gerencia o sistema sob sua supervisão. Em outras palavras, a base de regras é o conjunto de implicações fuzzy que definirão o comportamento do sistema de controle. As regras incondicionais são as que mantêm maior prioridade por não dependerem de fatores adversos. A princípio, deve haver quantas regras se for necessário para que se cubram todas as combinações entre as variáveis.

3) *Motor de Inferência*: Trata-se do núcleo do controlador fuzzy. É capaz de simular tomadas de decisão baseando-se nos conceitos de lógica fuzzy. Também consegue deduzir ações de controle, relacionando regras de inferência com algumas variáveis.

4) *Interface de Defuzzificação*: Responsável por produzir um resultado mensurável na lógica de CRISP a partir dos conjuntos fuzzy e dos graus de pertinência de cada um. O valor da variável de saída pelas regras fuzzy deve ser traduzido para um valor Crisp e inteligível ao usuário.

III. DOMÍNIO E TRABALHOS RELACIONADOS

A. Domínio do Sistema de Controle Fuzzy

Um sistema de recomendação combina várias técnicas para selecionar itens personalizados com base no interesse dos

usuários e no contexto em que estão inseridos. Empresas como Amazon, Netflix, Google e Spotify são reconhecidas pelo uso intensivo de sistemas de recomendação, com os quais obtêm grande vantagem competitiva. O Netflix Prize, competição proposta pela Netflix na busca de um algoritmo melhor para a predição das preferências de seus usuários, tinha como objetivo final a melhoria do sistema de recomendação da Netflix.

Os sistemas de recomendação tornaram-se um tema extremamente atrativo, graças à propagação do e-commerce e o advento do consumo em dispositivos móveis, promovendo itens de consumo direcionados a determinados públicos-alvo. Tais sistemas se baseiam obrigatoriamente em itens sendo expostos e oferecidos ao usuário. Em outro caso, perderiam seu sentido [?].

Sistemas de recomendação podem ser categorizados em basicamente três tipos [?] conforme abaixo:

- Baseados em conteúdo: recomendam produtos que sejam semelhantes aos que foram consumidos pelo usuário no passado.
- De filtragem colaborativa: recomendam produtos que pessoas semelhantes ao usuário em questão preferiram no passado através da análise da vizinhança. Uma desvantagem é requerer um grande número de informações sobre o usuário e sua vizinhança para funcionar corretamente.
- Sistemas híbridos (combinam as duas abordagens acima mencionadas)

O trabalho de base para este estudo implementou um sistema de recomendação de filtragem colaborativa por considerar os vizinhos do usuário ao estimar suas avaliações acerca de determinados filmes [?]. Já neste estudo, a modelagem do sistema foi alterada de forma que somente são levadas em consideração as informações a respeito do próprio usuário; dessa forma, o sistema proposto é baseado em conteúdo.

Esta alteração foi necessária para um maior enfoque na compreensão e aplicação dos conceitos de lógica fuzzy ao invés de procurar elaborar algoritmos mais complexos que fugiriam do tema proposto.

B. Trabalhos Relacionados

O trabalho que serviu de base para o desenvolvimento do sistema de recomendações em questão é do tipo de filtragem colaborativa com lógica fuzzy, utilizando-se de dados de vizinhos para a predição da avaliação dos usuários [?]. Ele lida com quatro variáveis de entrada, que, ao serem avaliadas em conjunto, retornam um único valor de resposta: a nota estimada para o filme que determinado usuário daria na data em questão.

IV. LÓGICA FUZZY MODELANDO O SISTEMA

A proposta é, de forma geral, estimar uma nota que o usuário daria para determinado filme na data informada, com base nos dados de treinamento, que contém uma série de arquivos de textos, cada um com diversas triplas com o identificador do filme, o identificador do usuário, a nota fornecida e a data.

A. Variáveis Linguísticas e seus Conjuntos Fuzzy

Existem quatro variáveis de entrada, que, ao serem avaliadas em conjunto, retornam um único valor de resposta: a nota estimada para o filme que determinado usuário daria na data em questão.

As quatro variáveis de entrada são:

- Média da Data: considera a média de notas que o filme recebeu na data em questão
- Média do Filme: considera a média de notas que o filme recebe em geral
- Média do Usuário: considera a média de notas que o usuário atribui a filmes do mesmo gênero que o filme a ser avaliado
- Média do Gênero: considera a média de notas que filmes do gênero em questão recebem

Obs.: a variável de entrada que diz respeito à média do gênero foi elaborada para substituir a média de similaridade, inicialmente proposta pelo projeto-base, porque fugiria do propósito de estudo voltado a lógica fuzzy e suas aplicações.

A variável de saída é a Média Estimada, e, claramente, assim como as variáveis de entrada, é do tipo REAL e pode ter valores entre 1 e 5, como acontece na faixa de pontuação dos filmes da Netflix.

As funções de fuzzificação de todas as cinco variáveis (as quatro de datas e também a variável de saída) são iguais, todas possuindo 3 conjuntos fuzzy: nota baixa, nota média e nota alta. Além disso, todas são gaussianas.

B. Base de Regras

importante que, a princípio, haja quantas regras se for necessário para que se cubram todas as combinações entre as variáveis de entrada. Dessa forma, ao combinar as quatro variáveis de entrada alternando seus conjuntos fuzzy, obtiveram-se, no total, 81 regras. Algumas foram incluídas abaixo para análise. Foi levado em consideração que quando a maioria das quatro notas de entrada tiver um valor igual, muito provavelmente a nota estimada também terá esse valor fuzzy. Foi considerado também que a maioria das notas residirão na faixa média de valores (exemplo, 2 notas baixas, 1 média e 1 alta, fará com que o resultado final seja nota média).

RULE 1 : IF mediaGenero IS baixa AND mediaUsuario IS baixa AND mediaFilme IS baixa AND mediaData IS baixa THEN notaEstimada IS baixa;

RULE 2 : IF mediaGenero IS baixa AND mediaUsuario IS baixa AND mediaFilme IS baixa AND mediaData IS media THEN notaEstimada IS baixa;

RULE 13 : IF mediaGenero IS media AND mediaUsuario IS media AND mediaFilme IS baixa AND mediaData IS baixa THEN notaEstimada IS baixa;

RULE 14 : IF mediaGenero IS media AND mediaUsuario IS baixa AND mediaFilme IS media AND mediaData IS baixa THEN notaEstimada IS baixa;

RULE 15 : IF mediaGenero IS media AND mediaUsuario IS baixa AND mediaFilme IS baixa AND mediaData IS media THEN notaEstimada IS baixa;

RULE 38 : IF mediaGenero IS media AND mediaUsuario IS media AND mediaFilme IS baixa AND mediaData IS alta THEN notaEstimada IS media;

RULE 39 : IF mediaGenero IS media AND mediaUsuario IS media AND mediaFilme IS alta AND mediaData IS baixa THEN notaEstimada IS media;

RULE 40 : IF mediaGenero IS media AND mediaUsuario IS alta AND mediaFilme IS baixa AND mediaData IS media THEN notaEstimada IS media;

C. Definições de Funcionamento do Sistema de Inferência

A T-norma utilizada para o operador fuzzy E é o MIN, de forma que o MAX fica implicitamente definido para o operador OU [?].

O método usado para a ativação do consequente é o MIN. A ativação é uma operação que define o método (no caso, MIN) usado no processo de implicação das regras [?].

O método usado para a acumulação é o MAX. A acumulação é uma operação que permite a combinação de resultados de uma variável quando as regras estão sendo consideradas no resultado final [?].

O método COG (Centro de Área) é o usado para a defuzzificação. Com este método, a saída é o valor do universo que divide a área sob a curva da função de pertinência em duas partes iguais (massas iguais). O COG encontra um ponto que representa o centro de gravidade do conjunto fuzzy resultante da agregação [?].

V. CENARIOS E ANÁLISE DOS RESULTADOS

O sistema fuzzy para recomendações tem que estar pronto para diversos cenários, uma vez que cada pessoa tem gostos bem diferentes uma das outras, embora sejam influenciadas pela opinião da sociedade em assuntos culturais

Assim sendo, estamos considerando não só os gostos do usuário, representado pela média das notas dadas pelo usuário a filmes parecidos, mas também consideramos as notas que os filmes receberam dos outros, o que representa a influência da opinião externa ao gosto do usuário.

Vamos descrever brevemente dois cenários possíveis:

A. Cenário 1: Preferência

Neste cenário, o filme apela muito para o gosto do usuário, porém não é muito popular. Como as variáveis de mediaUsuario e mediaGenero são altas, o filme mesmo assim tem uma boa recomendação ao usuário.

B. Cenário 2: Moda ou Classico

Neste caso, o usuário não se interessa pelo filme, porém ele está tendo excelentes notas naquele dia ou tradicionalmente bem avaliado por todos. Sendo assim suas variáveis mediaData e mediaMovie são altas, e ele acaba se tornando uma sugestão forte ao usuário, que imagina-se, pode ser influenciado por outras opiniões

VI. CONCLUSÃO

A lógica fuzzy permite a realização de modelagens de problemas da vida real de forma mais precisa com relação à lógica tradicional, tendo em vista que, nos problemas reais, os valores e constantes não se apresentam de forma tão fixa como o esperado pela lógica clássica, havendo inclusive a participação de qualificadores e modificadores, como "muito", "pouco", "parcialmente", etc.

Os conceitos da lógica fuzzy podem ser aplicados em diversas áreas que envolvam um processamento mais complexo de grandes quantidades de dados, como em Inteligência Artificial e Mineração de Dados. Na implementação de sistemas de recomendação, por exemplo, a lógica fuzzy se faz útil por permitir uma aproximação mais sutil do valor esperado ao se levar em conta bases de dados de treinamento.

VII. REFERÊNCIAS

ahlawat Ahlawat, Nishant, Ashu Gautam, and Nidhi Sharma (International Research Publications House 2014) "Use of Logic Gates to Make Edge Avoider Robot." International Journal of Information Computation Technology (Volume 4, Issue 6; page 630) ISSN 0974-2239 (Retrieved 27 April 2014)
 netflix Netflix Prize. Disponível em <http://www.netflixprize.com/>
 winner The BellKor Solution to the Netflix Grand Prize. Disponível em http://www.netflixprize.com/assets/GrandPrize2009_BellKor.pdf
 movieRating A movie rating prediction system of user propensity analysis based on collaborative filtering and fuzzy system. Disponível em <http://ieeexplore.ieee.org/document/5277415/>
 marioleto aulas da graa recommend https://www.ibm.com/developerworks/br/local/data/sistemas_recomendacao/ implementando