

6

RESULTADOS EXPERIMENTAIS

6.1

Introdução

Nos capítulos anteriores foram apresentados algoritmos e heurísticas úteis para o processo de recomendação de produtos a clientes. Também foi apresentada a proposta de unir os algoritmos de Filtragem Colaborativa e o Método Fuzzy de Hosseinpour, de forma a criar o Filtro Híbrido Fuzzy, um sistema integrado capaz de incrementar as vantagens e reduzir as desvantagens daqueles algoritmos.

O objetivo deste capítulo é apresentar resultados experimentais destes algoritmos sendo aplicados tanto na base de dados sintética, criada artificialmente com base na teoria de marketing, dados do estatísticos da população brasileira e o método Fuzzy de Hosseinpour, bem como na base real *MovieLens* que é a mais utilizada no estudo de filtros colaborativos.

O Filtro Híbrido Fuzzy, o Filtro Fuzzy e o FC Categórico concebidos nessa tese, bem como FC Item-Item serão testados na base de dados sintética de modo a avaliar o quanto as suas performances são prejudicadas com relação a variações de esparsidade no banco de dados, característica comum em bancos de dados comerciais. Os testes no banco de dados sintético também permite analisar o comportamento de cada algoritmo que compõe o Filtro Híbrido Fuzzy individualmente de modo a compreender o seu comportamento e potencialidades.

Adicionalmente, o Filtro Fuzzy e o FC Item-Item são testados na base de dados real *MovieLens* para gerar uma comparação entre a acurácia de ambos os algoritmos diante de compras feitas por usuários reais.

Por fim, são feitos comentários sobre os resultados obtidos em todos os experimentos e sugeridas melhorias e trabalhos futuros na área.

6.2

Treinamento na Base de Dados Sintética

Os experimentos foram desenvolvidos em uma base com vinte mil clientes representando as características demográficas da população brasileira obtidas pelo senso demográfico do IBGE em 2002 e por dados da ABEP. Foram definidas quatorze categorias de itens. Para cada uma das categorias criadas, foi também desenvolvido um conjunto de especificações técnicas e seus respectivos posicionamentos. No total, foram criados 1000 itens para cada categoria, com especificações técnicas geradas aleatoriamente, formando uma ampla gama de posicionamentos.

Com base em dados médios de compras de brasileiros no mercado de e-commerce, foram definidas sete compras por usuário, levando em conta suas características e as especificações dos itens.

A vantagem de se utilizar a base de dados sintética está em se ter uma avaliação do usuário para todos os itens, o que não é possível se ter em uma base de dados real de mesmo porte, bem como a liberdade de variar a esparsidade, aumentando ou diminuindo o número de usuários com relação ao número de itens disponíveis.

Usuários foram aleatoriamente escolhidos para gerar cinco subconjuntos de treinamento e testes diferenciados, tendo sido necessário fazer validação cruzada. Os algoritmos de recomendação foram utilizados sobre os subconjuntos, gerando resultado de precisão, revocação e F1. Os resultados passaram por uma média aritmética gerando os resultados finais apresentados nesse capítulo.

Foi treinado o FC Item-Item (*Benchmark*) na base de compras de treinamento, utilizando similaridade cosseno entre os vetores. Esse treinamento foi feito para esparsidades diferentes da base de dados. Foi treinado também o FC Categórico usando similaridade cosseno nas mesmas condições. Além disso, os itens foram comparados via similaridade por números fuzzy de seus posicionamentos, para permitir o uso rápido do Filtro Fuzzy. Tendo sido treinado o FC Categórico e criados relações de similaridade entre os números fuzzy de todos os itens, o Filtro Híbrido Fuzzy passou a estar igualmente treinado.

O valor de esparsidade foi variado alterando a razão entre o número de usuários presentes no treinamento e o número de itens total, que é constante em

todos os experimentos. O treinamento foi feito com apenas 5.000 usuários na base com maior esparsidade, e com 20.000 na menos esparsa.

O algoritmo é avaliado para um usuário no conjunto de treinamento, gerando N recomendações para cada item comprado c pertencente a sua cesta de compras de treinamento (C_{treino}). O Top- N considera então $N \cdot C_{treino}$ itens sendo recomendados.

O sucesso da recomendação está em o(s) item(s) presentes na cesta de compras de teste (C_{teste}) estar(em) presente(s) também no Top- N . Na base de dados, foram feitas 6 compras de treinamento e 1 compra de teste para cada usuário. Essa característica das compras restringe o valor de N em 6. A razão desta restrição é que o algoritmo de recomendação criado utiliza todos os itens comprados para recomendar. No caso mais simples, o algoritmo fará apenas 1 recomendação para cada item comprado (6 itens comprados=6 recomendações) levando $N=6$. O valor de N aumenta em múltiplos de 6 pois se eu fizer duas recomendações por item comprado terei 12 recomendações ($N=12$) e assim por diante.

6.2.1

FC Item-Item (Colaborativo Item-para-Item)

As Figura 18 e 19 mostram a taxa de F1 para o Filtro Base avaliado para diversas esparsidades (valores de usuários) e valores de N referentes a TOP- N da base de dados sintética.

Os resultados mostram que o filtro colaborativo item-item (FC Item-Item) apresenta perda de performance na medida em que o valor de N aumenta. Isso não significa que dentro dos N itens retornados pelo recomendador não existam acertos, mas sim que existe muito maior chance de que nesses N itens ele recomende erroneamente, o que gera F1 causa uma performance decrescente. A performance dele é máxima para $N=12$, significando que com 2 itens comprados este recomendador recomenda 12 itens onde a maior parte é de interesse do comprador.

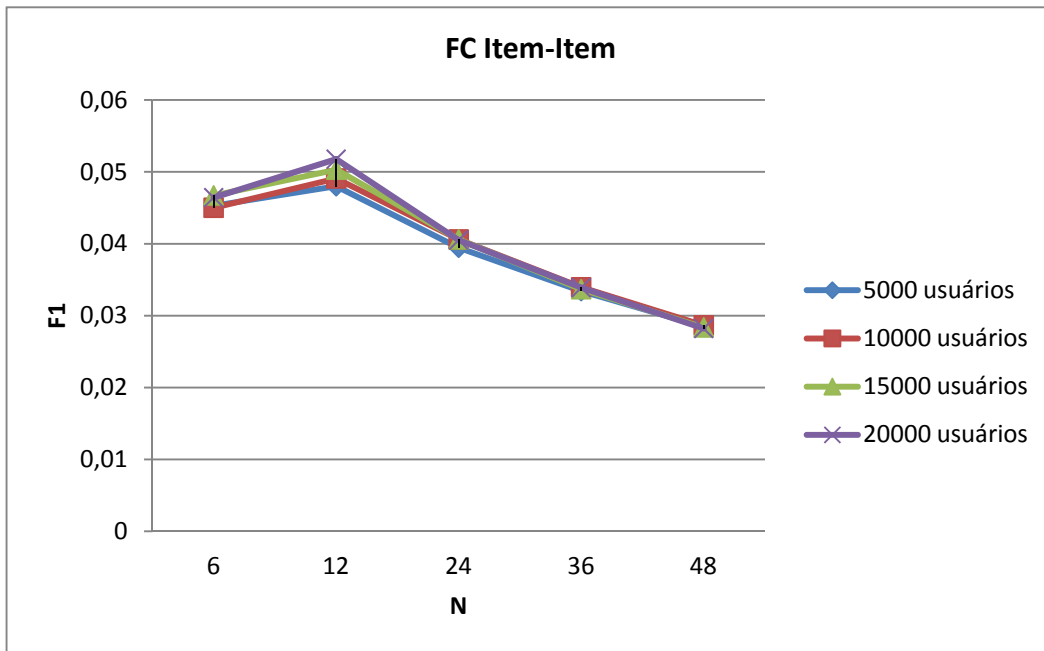


Figura 18- FC Item-Item variação com o valor de Top-N para diversos valores de esparsidade do banco de dados (número de usuários).

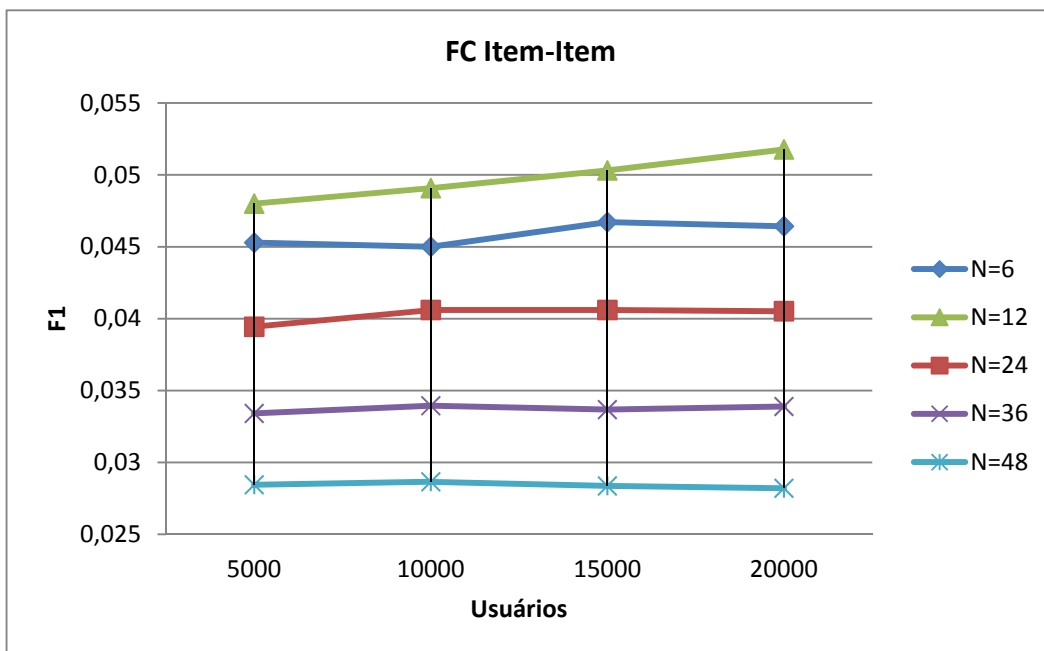


Figura 19- FC Item-Item variação com o valor de esparsidade (número de usuários) comparada em diversos valores de Top-N

Pode-se notar que na medida em que o número de usuários aumenta (diminuindo a esparsidade do banco de dados) o filtro colaborativo item-item possui melhora de performance, sentida mais força nos valores menores de N,

quando a performance F1 é maior. Esse resultado é esperado, uma vez que este filtro colaborativo apresenta como uma de suas deficiências a perda de performance com esparsidade, apesar de que, pelo que foi visto no capítulo 3, outros filtros colaborativos podem apresentar esta dificuldade com maior ou menor força.

6.2.2

Filtro Fuzzy (Algoritmo Fuzzy Individual)

Neste experimento considera-se que a categoria está sempre correta e se deixa que apenas o Filtro Fuzzy tome a decisão sobre o item. Nas Figuras 20 e 21 abaixo estão representadas as performances medidas por F1 para o Filtro Fuzzy avaliado para diversas esparsidades e valores de TOP-N.

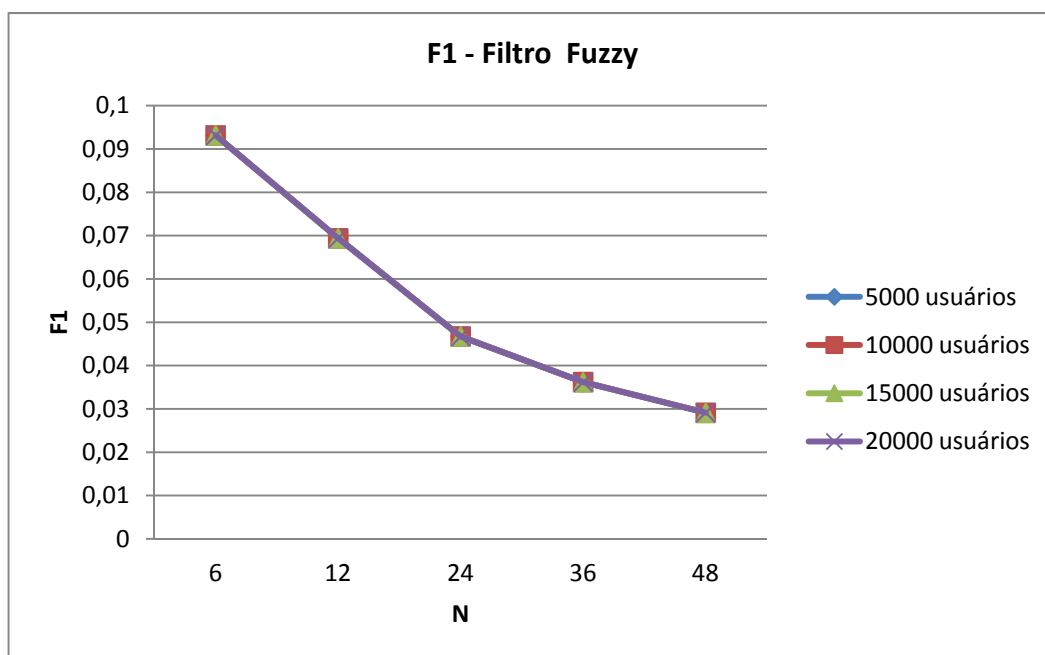


Figura 20- Filtro Fuzzy variação com o valor de Top-N para diversos valores de esparsidade do banco de dados (número de usuários).

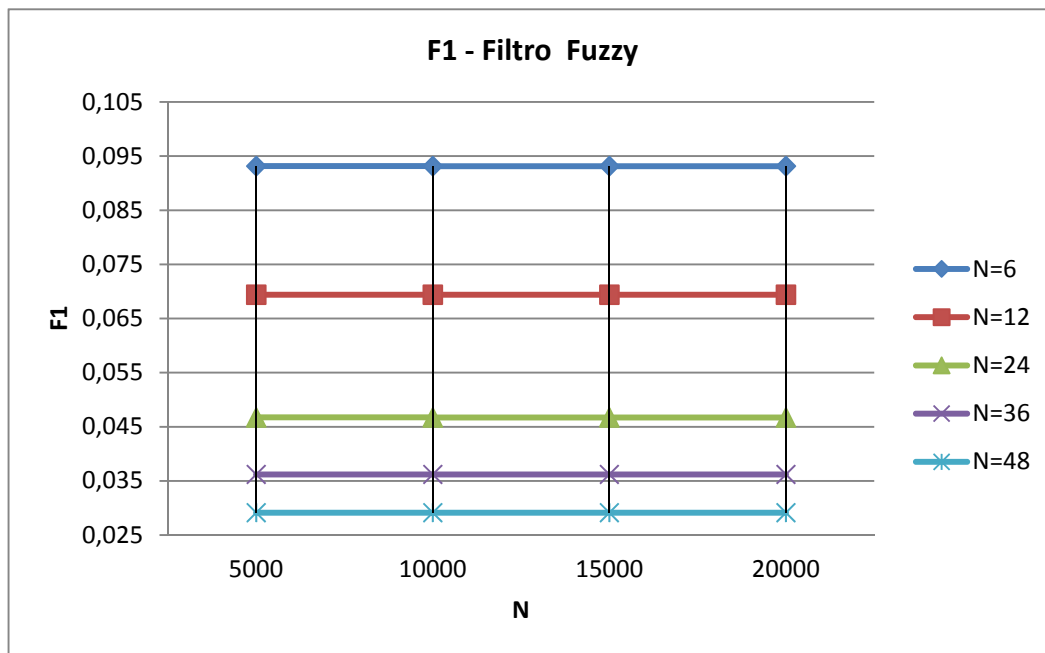


Figura 21- Filtro Fuzzy variação com o valor de esparsidade (número de usuários) comparada em diversos valores de Top-N

A observação da Figura 20 quando comparada a Figura 18 leva a duas constatações. A primeira de que o Filtro Fuzzy apresenta uma performance muito superior ao do FC Item-Item para os valores de N mais baixos, se aproximando aos valores do FC Item-Item para os maiores valores de N. Isso se evidencia por uma curva descendente constante e mais abrupta. Ao contrário do FC Item-Item não ocorre um máximo em N=12, sendo o máximo em N=6. Não se pode afirmar com esse dado que o Filtro Fuzzy terá performance superior ao FC Item-Item numa base real, pois a base de dados sintética, como já foi mencionado, pode ser tendenciosa, porém os dados comportamentais individuais de ambos algoritmos apontam que o Filtro Fuzzy tem performance melhor com um número menor de itens, o que já é um benefício válido. Cabe lembrar que o artigo de Hosseimpour indica uma taxa de performance F1 bastante elevada em seus experimentos.

No Filtro Fuzzy, ao contrário do que se observa no FC Item-Item, o cálculo de similaridade independe das avaliações do usuário sobre cada item presente na base de dados. O Filtro Fuzzy depende apenas do cálculo de compactação próxima entre a média dos posicionamentos de itens comprados pelo usuário e do item a ser recomendado. Assim, a esparsidade da base de dados não é relevante e isso se reflete na Figura 21 pela invariância com relação ao número de usuários.

Percebe-se claramente que o número de usuários não produz qualquer efeito sobre a acurácia do algoritmo.

Cabe lembrar que o algoritmo teve seu desempenho avaliado quando operando de forma isolada, levando em conta um acerto de categoria em todas as recomendações. O Filtro Fuzzy para uma base de dados com itens de diversas categorias não poderia ser usado sozinho, mas foi feita uma análise desta forma para entender seu comportamento isolado de forma a ser possível tomar conclusões no comportamento do Filtro Híbrido Fuzzy, ao qual ele irá compor.

6.2.3

FC Categórico (Filtro Colaborativo de Categorias)

A seguir os experimentos do Filtro Colaborativo Categórico sobre a base de dados sintética. Nas Figuras 22 e 23 abaixo estão representadas as performances medidas por F1 para o FC Categórico avaliado para diversas esparsidades e valores de TOP-N.

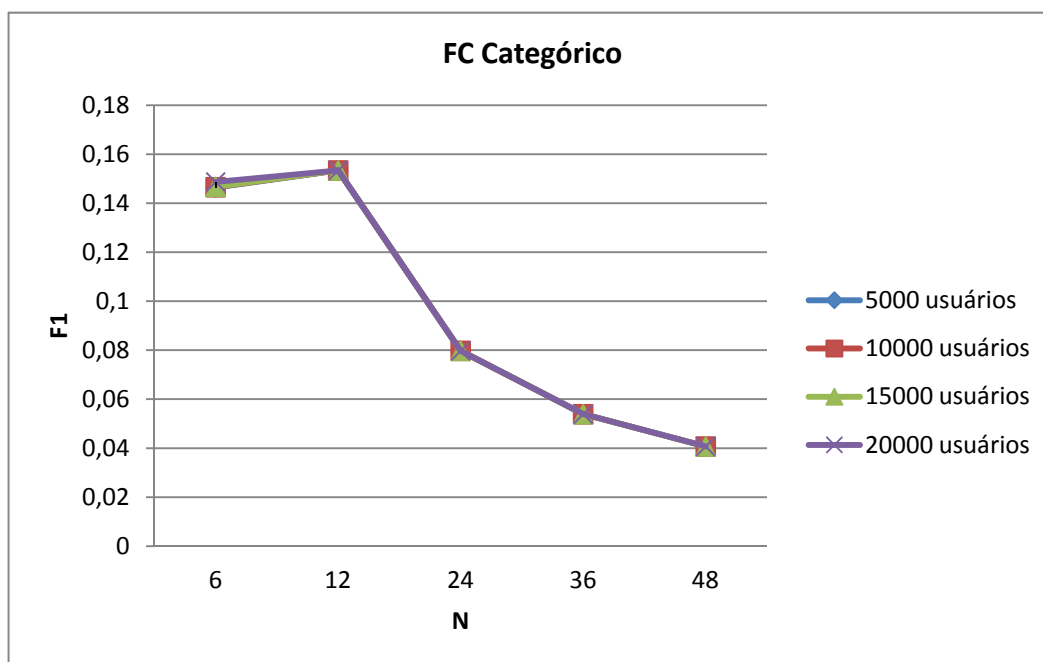


Figura 22- FC Categórico em variação com o valor de Top-N para diversos valores de esparsidade do banco de dados (número de usuários).

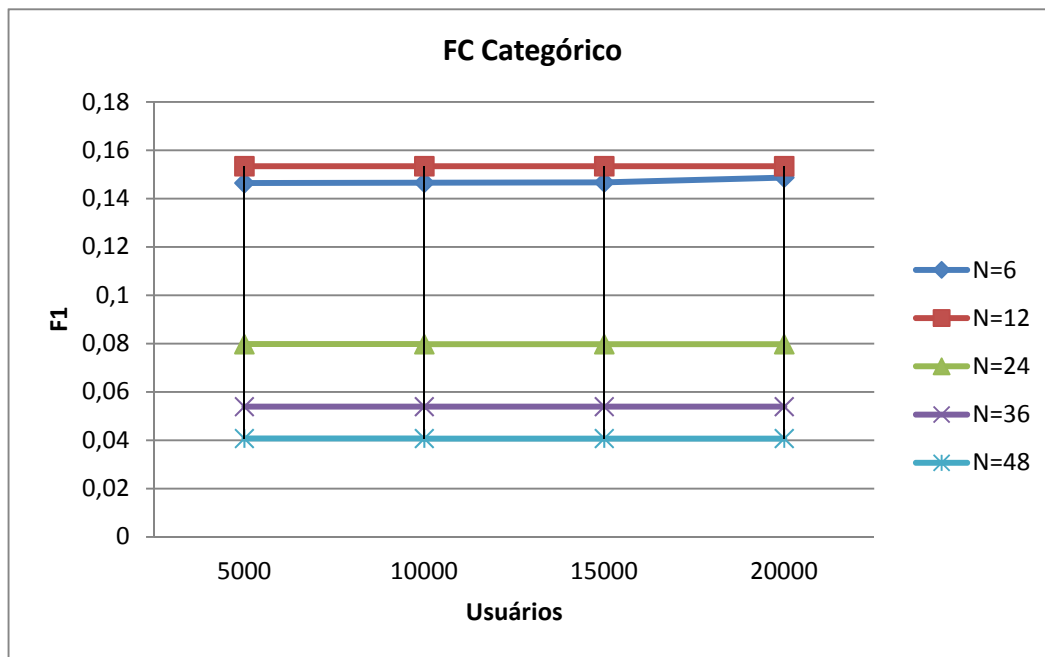


Figura 23- FC Categórico em variação com o valor de esparsidade (número de usuários) comparada em diversos valores de Top-N

A curva do FC Categórico apresentada na Figura 22 apresenta o mesmo contorno que o FC Item-Item apresentado na Figura 18, o que era de se esperar já que o FC Categórico também é um filtro colaborativo baseado em memória item-item, com a diferença de que os itens são categorias. O FC Categórico apresentou o valor máximo de F1 para N=12 demonstrando que o filtro colaborativo baseado em itens de fato tem uma performance maior quando utiliza dois itens como base para sua recomendação.

A primeira vista, a Figura 23 demonstra que, da mesma forma que o Filtro Fuzzy, o FC Categórico não apresenta variação com a diminuição da esparsidade do banco de dados. Por outro lado, percebe-se que para N=6 existe uma pequena, porém visível melhora de performance do algoritmo de FC Categórico com a redução da esparsidade. Como foi dito, o FC Categórico é um filtro colaborativo baseado em itens com a diferença que os itens são categorias. Ao contrário do que ocorre no FC Item-Item, onde cada um dos usuários comprou 6 itens dos 14.000 itens disponíveis (total de 120.000 compras no caso de 20.000 usuários, base menos esparsa), no caso das categorias as 6 compras serão sobre 14 categorias apenas, aumentando em muito a chance de que cada usuário tenha uma boa análise pelo menos metade das categorias. Isso significa que com o mesmo

número de usuários, a base de dados sintética para o FC Categórico é muito menos esparsa do que a mesma base de dados sintética para o FC Item-Item.

Como em geral, mesmo em bases de dados reais, o número de categorias existentes de itens é muito inferior ao número de usuários, pode-se considerar que para a aplicação onde este filtro está sendo proposto ele é invariante a esparsidade.

6.2.4

Filtro Híbrido Fuzzy

A seguir os experimentos do Filtro Híbrido Fuzzy sobre a base de dados sintética. Nas figuras 24 e 25 abaixo estão representadas as performances medidas por F1 para o Filtro Híbrido Fuzzy avaliado para diversas esparsidades e valores de TOP-N.

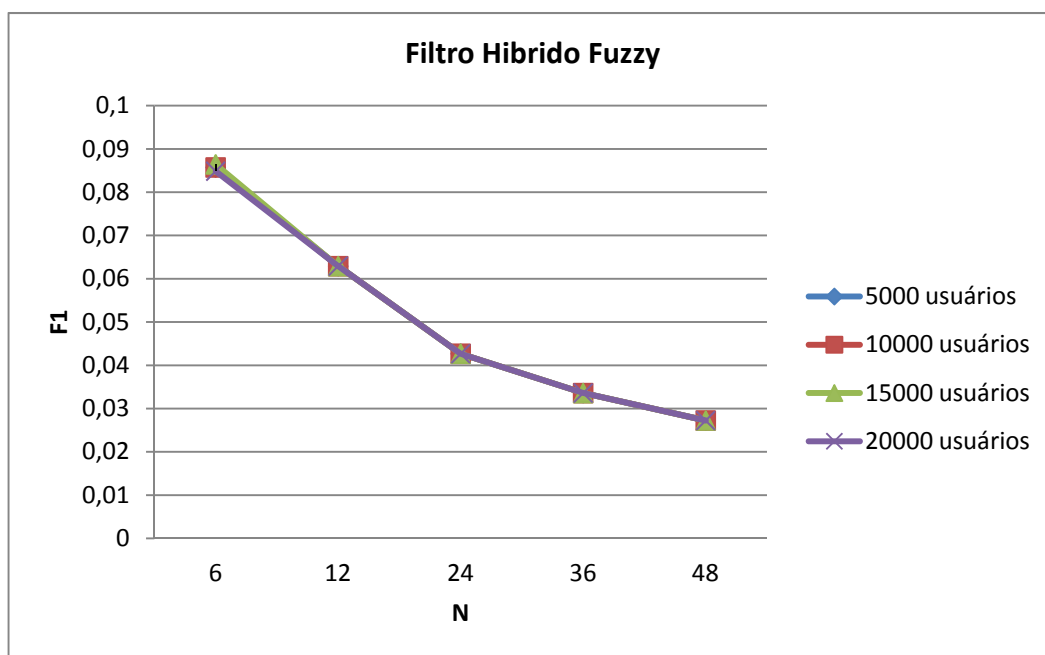


Figura 24- Filtro Híbrido Fuzzy em variação com o valor de Top-N para diversos valores de esparsidade do banco de dados (número de usuários).

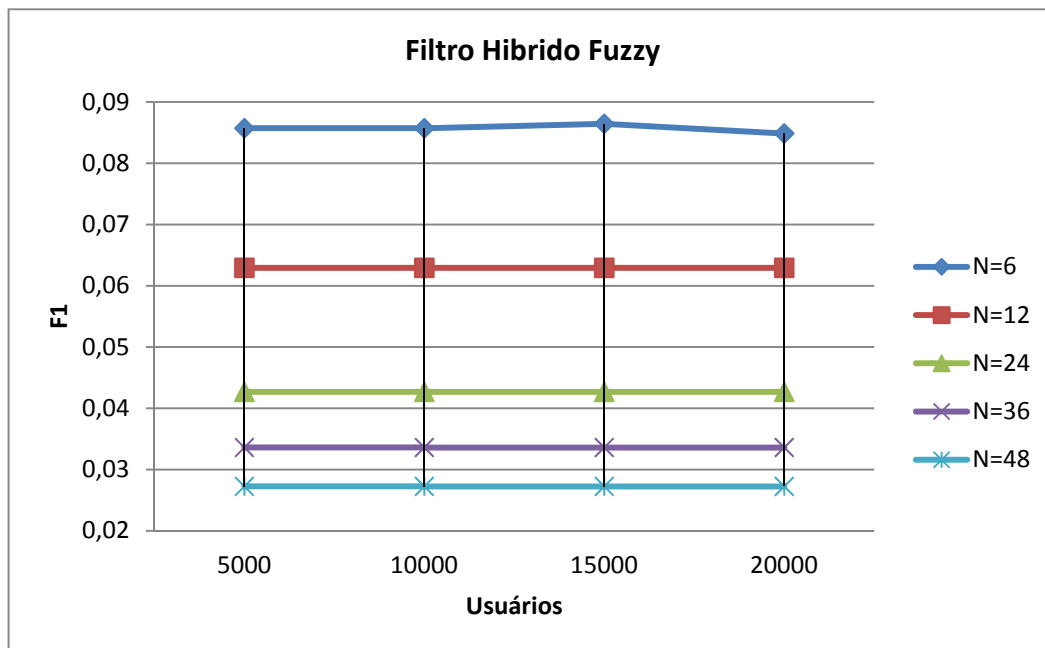


Figura 25- Filtro Híbrido Fuzzy em variação com o valor de esparsidade (número de usuários) comparada em diversos valores de Top-N

Percebe-se pela análise da Figura 24 que o Filtro Híbrido Fuzzy apresenta uma curva semelhante ao do Filtro Fuzzy, que o compõe, sem apresentar o valor máximo em $N=12$ como os dois Filtros Colaborativos baseados em memória (FC Item-Item e FC Categórico) apresentaram. O Filtro Híbrido Fuzzy, da mesma forma que o Filtro Fuzzy, apresenta performance máxima com um menor número de itens avaliados, sendo esta uma vantagem com relação ao FC Item-Item.

Da mesma forma que a comparação de performance entre o FC Item-Item e o Filtro Fuzzy pode ser arriscada (pois a base de dados sintética pode favorecer o Filtro Fuzzy) a comparação de performance entre o FC Item-Item e o Filtro Híbrido Fuzzy pode não ocorrer em bases reais, porém os dados apontam que o Filtro Híbrido Fuzzy apresentou desempenho melhor. Por outro lado, comparando o Filtro Híbrido Fuzzy com o Filtro Fuzzy, o primeiro apresenta um desempenho ligeiramente inferior ao segundo, que o compõe. Esse resultado era esperado, uma vez que o acerto no Filtro Híbrido Fuzzy requer um acerto conjunto tanto no FC Categórico como do Filtro Fuzzy. No entanto, como o FC Categórico apresenta desempenho elevado, ele pouco afeta o Filtro Híbrido Fuzzy.

Recorde-se que, caso o FC Categórico ofereça a recomendação errada, o Filtro Fuzzy não terá chance de recomendar o item certo, uma vez que todas as

suas recomendações estarão em uma categoria que não foi comprada. Caso o FC Categórico acerte a recomendação e o Filtro Fuzzy erre, o item será da categoria certa, mas não será o item que o usuário deseja comprar naquela categoria. O Filtro Fuzzy utilizado na base de dados sintética foi adaptado para que sempre fosse acertada a categoria e sozinho, não poderia sequer ser testado em uma base de dado com várias categorias.

Analisando a Figura 25 percebe-se a invariância a esparsidade, conforme era esperado do Filtro Híbrido Fuzzy por ser composto de dois algoritmos que não variam com a esparsidade. Note que para $N=6$ o Filtro Híbrido Fuzzy possui uma pequena variância a escala, igual ao FC Categórico, significando que essa invariância depende de que o número de categorias seja muito inferior ao número de usuários, como é o caso no domínio do problema que está sendo tratado.

6.2.5

Comparações entre os Algoritmos

Apresentam-se a seguir, nas Figuras 26 e 27, os valores da medida F1 para o FC Item-Item, FC Categórico, Filtro Fuzzy e para o Filtro Híbrido Fuzzy, de forma a permitir a comparação de ambos os algoritmos para diversos valores de esparsidade da base de dados.

Utilizou-se o valor de Top-6, por ser este o menor valor de N experimentado, para avaliar como os algoritmos são afetados pela esparsidade no número de usuários.

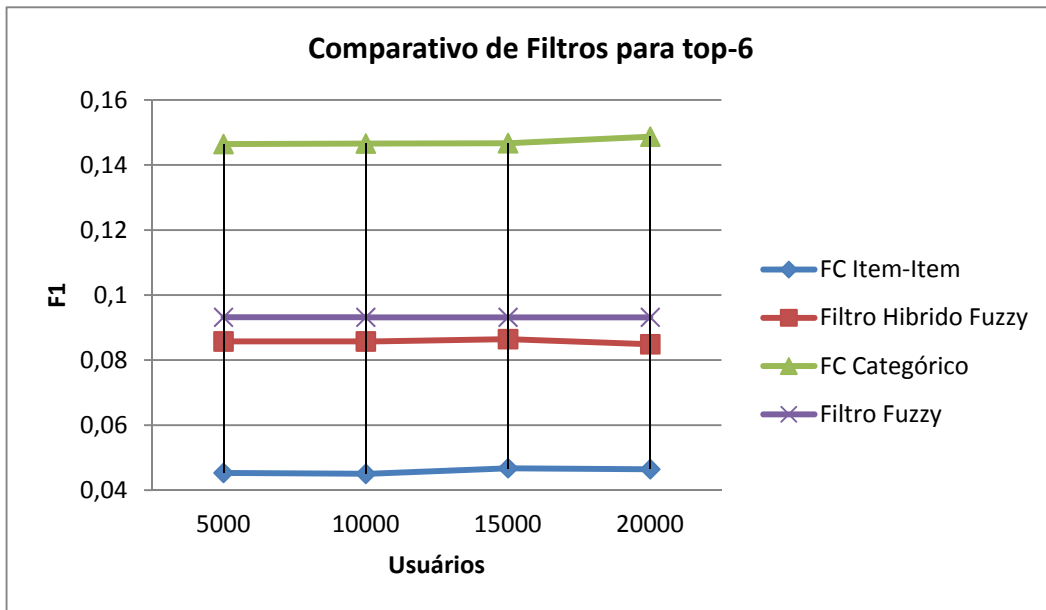


Figura 26 – Comparação entre recomendadores para Top-6

Para avaliar como os algoritmos são afetados pelo aumento do top-N, foram utilizados 20.000 usuários, valor para o qual o algoritmo base apresenta melhores resultados.

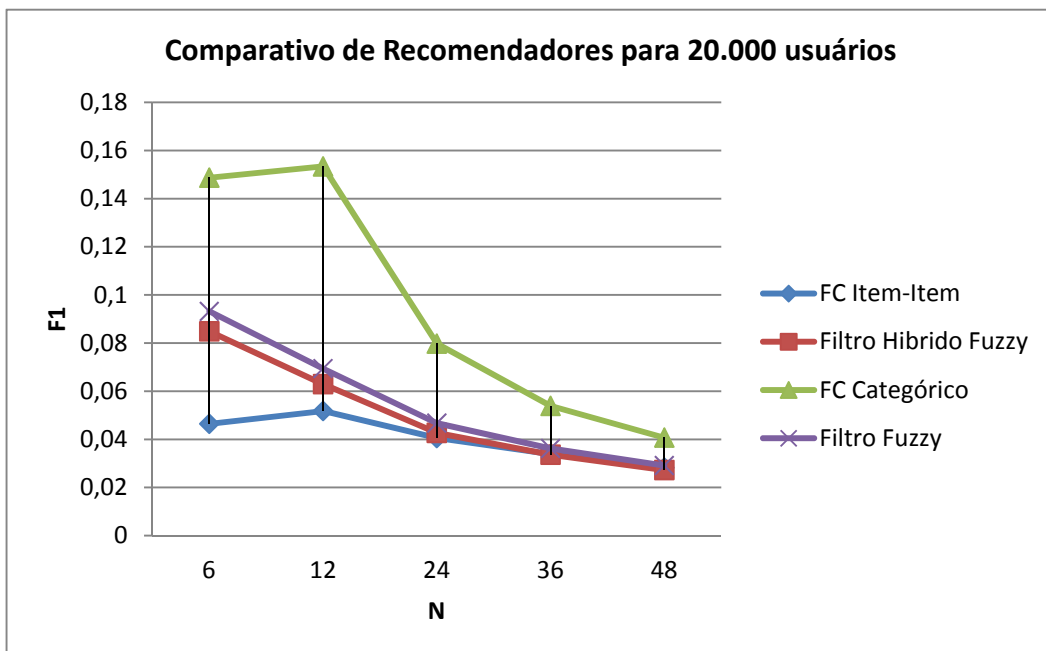


Figura 27 – Comparação entre recomendadores para 20.000 usuários

Observa-se, avaliando as Figuras 26 e 27, que o desempenho do Filtro Híbrido Fuzzy foi superior ao do C Item-Item em dois quesitos importantes: ele

não é afetado negativamente por bases de dados mais esparsas e apresenta alto desempenho para valores baixos de N . Recorde-se que, que quanto maior o valor de N , maior a chance do usuário não observar a recomendação. Mesmo para valores altos de N , o Filtro Híbrido Fuzzy apresenta resultados semelhantes aos do FC Item-Item.

6.3

Resultados com a Base de Dados Real

Apresentam-se, nesta seção, os resultados de experimentos conduzidos com a base de dados *MovieLens*. Esta base possui apenas uma categoria (filmes), de modo que foram comparados apenas o FC Item-Item e o Filtro Fuzzy.

Esta base de dados possui um número constante de usuários e, portanto, não caberia efetuar análise do efeito da esparsidade nos algoritmos. De qualquer forma, experimentos com base de dados sintética demonstraram que apenas o FC Item-Item seria afetado.

Na base *MovieLens*, cada usuário avalia os itens (filmes) com um número entre 1 e 5. Foram consideradas apenas avaliações de valor igual ou superior a 4.

Comparado à base sintética, no qual cada usuário avalia apenas 7 itens, na base *MovieLens* o usuário avalia todos os filmes gerando um número muito maior de avaliações por usuário. As compras dos usuários foram divididas em 80% para o conjunto de treinamento e 20% para teste. Foram criadas cinco destas bases para fins de validação cruzada.

Os resultados dos algoritmos utilizando as medidas de precisão, revocação e F1 são apresentados na Figura 28 a seguir para diversos valores de top- N .

Observa-se que o Filtro Fuzzy apresenta desempenho superior ao do FC Item-Item. Sendo a base de dados *MovieLens* uma base de dados real extensamente usada para avaliar o desempenho de recomendadores, esse resultado demonstra que o Filtro Fuzzy, assim como na base de dados sintética, apresenta resultados superiores ao do FC Item-Item, que é um benchmark para comparação de recomendadores. O fato de não ter sido possível testar o Filtro Híbrido Fuzzy na base *MovieLens* (por falta de categorias) não reduz o mérito deste resultado, uma vez que, conforme foi visto na base sintética, o FC Categórico apresenta elevada

taxa de acerto e afeta de maneira pouco significativa os resultados do Filtro Fuzzy.

Os resultados obtidos neste capítulo devem ser avaliados sobre o ponto de vista dos objetivos originais deste documento, que era de criar um novo algoritmo capaz de aproveitar os conceitos de posicionamentos oriundos da teoria de marketing para gerar recomendações de maior qualidade e sobrepujar as deficiências dos algoritmos de recomendação utilizados no mercado. O Filtro Híbrido Fuzzy resultante desta pesquisa apresentou um desempenho superior ao do FC Item-Item, tanto na base de dados sintética quanto na base real (*MovieLens*), gerando bons resultados para poucos e muitos itens recomendados e invariância à esparsidade da base de dados.

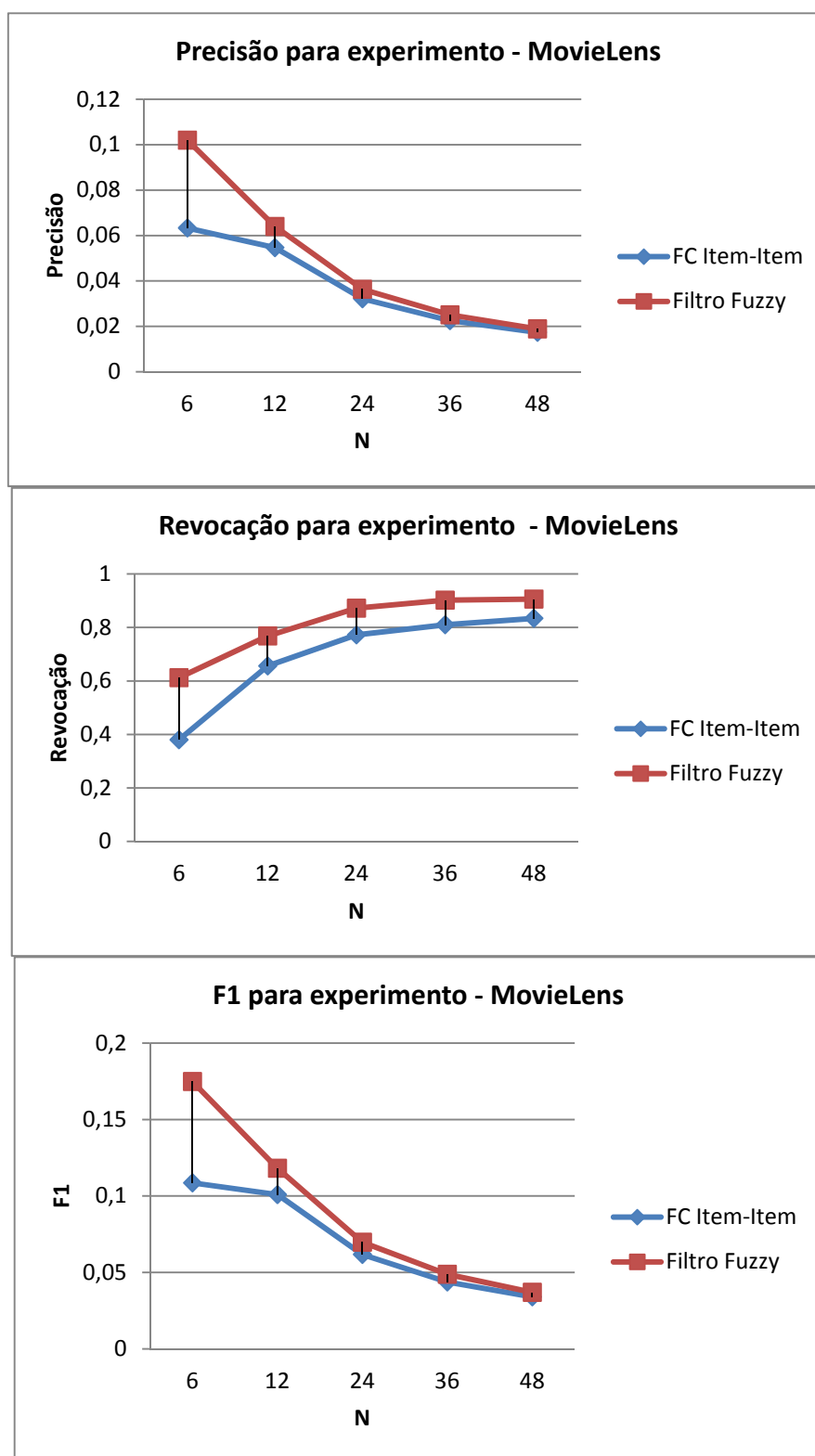


Figura 28- Comparação entre Filtro Fuzzy e FC Item-Item – Precisão, Revocação e F1 na base de dados *MovieLens*

Além disto, o Filtro Híbrido Fuzzy apresenta características únicas como a resolução de problemas de inicialização no caso de itens novos. Isso se deve ao fato de que todo item novo, adicionado na base de dados, possui por obrigações as especificações da categoria, que por sua vez se transformam em posicionamentos que são os únicos parâmetros usados pelo Filtro Fuzzy para sua recomendação. Como o FC Categórico não treina itens, e sim categorias (nas quais mesmo os itens novos se enquadram) este também não afeta o Filtro Híbrido Fuzzy na escolha de itens novos. O Filtro Híbrido Fuzzy, porém, ainda apresentará problemas com usuários novos, uma vez que ele depende da existência de pelo menos uma compra para recomendar.

Também é um algoritmo do qual é possível parcialmente rastrear os motivos pelo quais os resultados foram obtidos, uma vez que os números fuzzy trazem uma maior compreensão lingüística aos resultados. Por exemplo, pode-se constatar dos resultados que certos usuários comprem sempre itens com preço “Muito Baixo” ou de Design “Muito Alto”. De um modo geral, o algoritmo proposto foi capaz de atender os objetivos apresentados.