
A ESTRUTURA DE FATORES DA SERENDIPIDADE COMO UM CONSTRUTO EM SISTEMAS DE RECOMENDAÇÃO

PREPRINT

Andre Paulino de Lima

Universidade de São Paulo

São Carlos, SP, Brazil

andre.p.lima@usp.br

26 de março de 2021

RESUMO

Nos últimos anos, a serendipidade tem recebido muita atenção da comunidade de pesquisa em sistemas de recomendação, devido ao consenso crescente do seu papel na satisfação do usuário e no combate aos efeitos de filtros invisíveis. Apesar disso, ainda não há consenso a respeito da definição mais adequada para serendipidade, nem como medir sua presença em recomendações. O objetivo deste estudo é esclarecer a estrutura de fatores que contribuem para a serendipidade de recomendações por meio de uma análise fatorial aplicada ao conjunto de dados Serendipity 2018, que tem se firmado como o principal conjunto de dados da área para estudos de serendipidade. Os resultados da análise exploratória sugerem que a adoção de um modelo reflexivo que combina contribuições dos fatores de relevância e surpresa se adequa melhor aos dados do que um modelo que envolve relevância, novidade e imprevisibilidade.

Palavras-chave Sistemas de Recomendação · Serendipidade · Surpresa · Novidade · Imprevisibilidade · Análise Fatorial Exploratória · Conjunto de Dados Serendipity 2018

1 Introdução

Tradicionalmente, os sistemas de recomendação empregam a avaliação dada pelo usuário a um conjunto de itens para induzir modelos de preferência do usuário. Idealmente, esses modelos capturam aspectos relativamente estáveis da atitude do usuário em relação à categoria de itens recomendados pelo sistema [13, 2]. O sucesso da tarefa de indução da preferência se evidencia pela consistência, ao longo do tempo, de avaliações positivas dadas pelos usuários aos itens que lhe foram recomendados pelo sistema de recomendação. Entretanto, com a alegação de que sistemas de recomendação atuam na formação de filtros invisíveis (*filter bubbles*, em Inglês) [5, 20], as estratégias de recomendação, que buscavam primeiramente obter um grau adequado de satisfação do usuário, passaram também a fomentar a ampliação de sua preferência por meio da recomendação de itens que são qualitativamente distintos dos itens mais bem avaliados pelo usuário [19, 11]. Recomendações que atendem a esses dois requisitos são qualificadas como serendipitosas [8, 15].

Nos últimos anos, o interesse da comunidade pelo tema resultou em inúmeros estudos sobre serendipidade, mas, apesar disso, a geração sistemática de recomendações serendipitosas permanece sendo uma meta desafiadora. Em boa medida, isso se deve às dificuldades em definir e mensurar o grau em que a serendipidade é embutida em novas recomendações. A literatura registra múltiplas definições para serendipidade e diferentes autores postulam modelos teóricos que divergem

no tipo e número de fatores que determinam o grau de serendipidade embutida em uma recomendação [3, 15]. De abordagem empírica, há estudos conduzidos com usuários nos quais os autores buscam avaliar a serendipidade em recomendações por meio de questionários submetidos ao usuário [23, 14]. Entretanto, divergências metodológicas entre esses estudos torna difícil a comparação entre abordagens e entre resultados obtidos, de forma que o consenso a respeito dos principais fatores determinantes da serendipidade permanece limitado.

Recentemente, um estudo com usuários do sistema de recomendação de filmes *MovieLens* resultou na publicação de um conjunto de dados sobre serendipidade [14]. Um aspecto promissor desse conjunto de dados (*Serendipity 2018*) é que o número de respostas coletadas ($N = 2.150$) é suficientemente grande para a condução de uma análise fatorial com nível adequado de qualidade. No presente estudo, esse conjunto de dados é explorado para esclarecer a questão sobre os fatores determinantes da serendipidade por meio da aplicação de análise fatorial.

O restante deste documento está organizado da seguinte forma: a Seção 2 apresenta brevemente as diferentes noções de serendipidade, traça um paralelo entre o estado atual da pesquisa em serendipidade com a pesquisa em usabilidade que tem sido conduzida pela comunidade de interação humano-computador, e aponta os benefícios que a análise fatorial trouxe para o estudo de usabilidade como motivação para este estudo. Na Seção 3, a metodologia de coleta e características distribucionais do conjunto de dados *Serendipity 2018* são apresentadas. A técnica de análise fatorial exploratória é brevemente descrita na Seção 4; os resultados da sua aplicação são descritos nas Seções 4.1 e 4.2. A discussão dos resultados é desenvolvida na Seção 5 e o trabalho é concluído na Seção 6, onde se apresentam as conclusões finais, limitações e oportunidades de continuidade deste estudo.

2 Revisão da literatura

A literatura de sistemas de recomendação registra diversas definições para serendipidade [11, 15, 9, 8]. Com efeito, o consenso a respeito de como mensurar a serendipidade ou seus fatores componentes, ainda é limitado [14, 3]. De acordo com uma das primeiras definições, “uma recomendação serendipitosa é aquela que ajuda o usuário a encontrar um item **surpreendentemente interessante**, o qual talvez não tivesse sido encontrado pelo usuário por seus próprios meios” [8] (nosso grifo). Nessa definição informal, a serendipidade depende tanto de qualidades objetivas do item recomendado (i.e. qualidades que tornam o item atraente ou repulsivo ao usuário) quanto de características subjetivas do usuário alvo da recomendação (i.e. preferência por itens com certas qualidades em detrimento de outras). Em outras palavras, a serendipidade é concebida como uma qualidade secundária, no sentido proposto pelo perspectivismo científico [6, p. 14], que emerge da interação do usuário com um item. Essa concepção da serendipidade como um fenômeno interacional permite salientar algumas semelhanças com o conceito de usabilidade da literatura da área de interação humano-computador, para o qual também se registram (1) múltiplas definições, (2) múltiplos modelos de medição, e (3) consenso limitado a respeito da definição e dos modelos de medição mais adequados [16]. Na verdade, como aponta Tractinsky [22], o consenso limitado sobre usabilidade é, ao mesmo tempo, efeito e reflexo da concorrência de múltiplas definições que, por sua vez, induzem diferentes modelos de medição. Desta forma, esforços devem ser direcionados no sentido de buscar a convergência das definições (e seus modelos de medição). Com base nas semelhanças apontadas, a serendipidade é enquadrada no presente estudo como um construto hipotético, a exemplo da abordagem recentemente apontada para dirimir os problemas operacionais em torno da usabilidade [22], com o objetivo de reunir evidências que possam promover a convergência das definições de serendipidade.

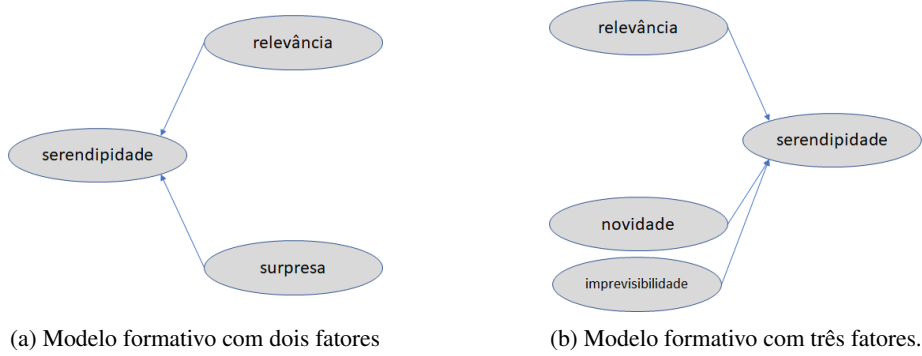
Vale ressaltar que, em estudos sobre sistemas de recomendação que seguem a definição de serendipidade proposta por Herlocker et. al. [8], a serendipidade corresponde a um construto operacionalizado por um modelo formativo¹ que combina os fatores de relevância e surpresa [10, 3], como ilustra a Figura 1a. Alternativamente, há estudos que retratam a serendipidade por meio de um modelo formativo no qual a relevância, a novidade e a imprevisibilidade se combinam [15, 9, 18], como mostra a Figura 1b. Entretanto, como apontam Kaminskas e Bridge [11], há uma confusão terminológica em torno das definições de surpresa, novidade e imprevisibilidade, de modo que a sobreposição

¹Empregamos o conceito de modelo formativo descrito por Bollen e Lennox [1], no qual o fator latente não é postulado como uma entidade com existência real, e sim uma grandeza abstrata que proporciona valor explicativo em alguma aplicação.

conceitual entre esses fatores é significativa. Com o intuito de (1) promover uma interpretação fiel às definições de novidade e imprevisibilidade empregadas na coleta do conjunto de dados explorado na próxima seção e, ao mesmo tempo, (2) tornar essas definições comparáveis com uma definição de surpresa que é condutiva à consiliência com as ciências cognitivas, são apresentadas a seguir as definições desses conceitos formuladas como narrativas centradas no (sujeito) participante:

- Novidade é o efeito percebido pelo usuário que é causado por sua exposição a um item com o qual ele tem pouca ou nenhuma familiaridade [14].
- Imprevisibilidade (*unexpectedness*, na literatura em Inglês) é o efeito percebido pelo usuário que decorre da sua exposição a um item que viola suas expectativas em algum grau [14].
- Surpresa é o efeito percebido pelo usuário que decorre de sua exposição a um item que possui qualidades que se salientam por serem distintas das qualidades dos itens aos quais o usuário foi anteriormente exposto [10, 3].

Figura 1: Dois modelos formativos empregados para descrever a serendipidade em sistemas de recomendação.



3 O conjunto de dados *Serendipity 2018*

Como mencionado, o conjunto de dados *Serendipity 2018* contém 2.150 respostas coletadas junto aos usuários do sistema Movielens, e cada resposta registra a avaliação do usuário a respeito de nove questões submetidas aos usuário por meio de um questionário, mais a avaliação (*rating*) que o usuário atribuiu a um determinado filme [14]. A seguir são descritos os processos empregados para recrutamento de participantes e para amostragem de filmes. Essas descrições são seguidas por uma sumarização estatística dos dados antes e depois da realização das tarefas de pré-processamento.

3.1 Recrutamento de participantes

O processo de recrutamento considerou como candidatos todos os usuários que atendiam aos seguintes requisitos: (R1) usuário submeteu cinco ou mais avaliações de filmes (*ratings*) no período entre 30 de dezembro de 2016 e 30 de março de 2017; (R2) pelo menos cinco dessas avaliações são favoráveis (*rating* igual ou superior a 3,5 estrelas); e (R3) ao menos cinco dessas avaliações favoráveis foram submetidas depois do período de carência, que se refere ao primeiro mês depois da data de ativação do usuário no sistema. Esses critérios de inclusão foram estabelecidos com o intuito de selecionar usuário ativos (R1) e que já tinham concluído o referido período de carência (R3), no qual é esperado que o novo usuário registre avaliações de filmes assistidos no passado, de modo a criar um perfil suficiente informativo. Além disso, buscou-se selecionar apenas avaliações relevantes (R2), de modo a limitar a investigação aos efeitos da novidade e da imprevisibilidade. Foram selecionados 2.305 candidatos, dos quais 481 efetivamente participaram do estudo².

²O número de participantes diverge daquele publicado no estudo porque novas respostas foram submetidas por candidatos depois da data de corte considerada pelos autores, conforme nota constante da documentação do conjunto de dados.

3.2 Amostragem de itens para composição do levantamento

O levantamento de dados foi conduzido junto aos participantes por meio de um questionário online. O questionário é composto pelas nove questões listadas na Tabela 1, que também registra a escala de resposta e sua associação com os fatores postulados como determinantes da serendipidade (coluna “Indicador para”). Dentre os filmes identificados para cada participante no processo de seleção de candidatos (filmes cujas avaliações atendem aos requisitos R1, R2 e R3), os cinco menos populares foram selecionados (R4) para composição do levantamento, de modo que cada participante foi solicitado a preencher um formulário para cada um desses cinco filmes. Esse critério foi aplicado com o intuito de selecionar itens potencialmente desconhecidos pelo usuário, o que limita ainda mais o foco da investigação nos efeitos da novidade e da imprevisibilidade sobre a serendipidade. Os participantes responderam 2.150 dos 2.405 questionários submetidos, totalizando 19.350 questões respondidas.

3.3 Sumarização dos dados

Como ilustrado na Figura 2a, uma fração significativa das respostas coletadas é nula (i.e., itens preenchidos pelo participante com NA, compreendendo 1.075 das 19.350 respostas coletadas). Como estratégia para dados ausentes, as respostas nulas foram substituídas pela mediana da distribuição de cada indicador, o que se mostrou uma abordagem eficaz³ para todas as questões, com exceção da questão “s2”. Como complemento da estratégia, foram excluídas todas as respostas para as quais a questão “s2” recebeu resposta nula (E1, 740 respostas). Além disso, com o intuito de aumentar a consistência das respostas, foram excluídas (E2) 11 respostas em que a avaliação dada ao item é menor do que 3,5 estrelas; (E3) 29 respostas em que a questão “q” recebeu resposta nula; (E4) 451 respostas em que o participante afirma ter assistido ao filme há mais de um ano; e (E5) 313 respostas em que o participante demonstra baixa assertividade ao responder à questão “s1”. Os critérios de exclusão E3 e E4 buscam remover as respostas que potencialmente foram afetadas por vieses de recuperação (memória). O critério E5 foi introduzido porque respostas não polarizadas (i.e., 2 a 4) dadas à resposta “s1” podem ser interpretadas como sinal de baixa assertividade do participante. O conjunto de respostas remanescentes ao final do pré-processamento inclui 606 respostas, que representa um tamanho de amostra suficientemente adequado para condução de análise fatorial segundo critérios discutidos em [12]. Vale notar que a distribuição de respostas por item na amostra pré-processada é similar à distribuição original, como ilustra a Figura 2b.

4 Exploração da estrutura de fatores latentes

Na literatura sobre interação humano-computador, há diversos estudos em que a análise fatorial exploratória (*exploratory factor analysis*, em Inglês) é empregada tanto para identificar o número de fatores latentes do construto de usabilidade quanto para avaliar as propriedades psicométricas de instrumentos desenvolvidos para sua mensuração [17, 22]. A análise fatorial exploratória é fundamentada em um modelo estatístico multivariado e se presta à sumarização de relacionamentos entre variáveis de forma a satisfazer um balanceamento racional entre concisão e precisão para benefício da tarefa de interpretação de dados [7, 1, 12]. Por vezes confundida com a análise de componentes principais (*principal component analysis*, em Inglês), diverge desta em pelo menos três aspectos importantes:

- A análise fatorial exploratória assume que a variância observada em uma variável pode ter causas externas ao sistema analisado (um sistema aberto, por analogia à taxonomia de sistemas físicos), ao passo que a análise de componentes principais assume que a variância observada em uma variável necessariamente é compartilhada com alguma outra variável que é interna ao sistema monitorado (um sistema fechado, sem variáveis exógenas).
- De modo formal, ambas as técnicas buscam apoiar a resolução de um problema de otimização dado por [12]:

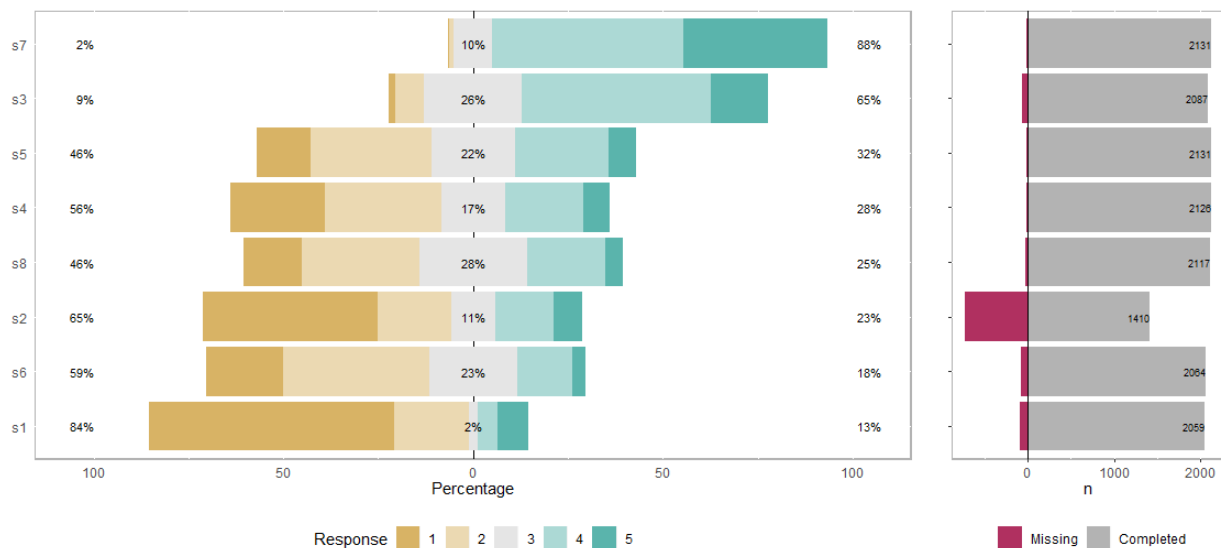
$$\arg \min_Q ||\mathbf{X} - (\mathbf{\Lambda}\mathbf{\Theta} + \mathbf{E})|| + r(\mathbf{\Lambda}, \mathbf{\Theta}). \quad (1)$$

³Eficaz no sentido de que a intervenção não modificou de forma expressiva a distribuição original de respostas.

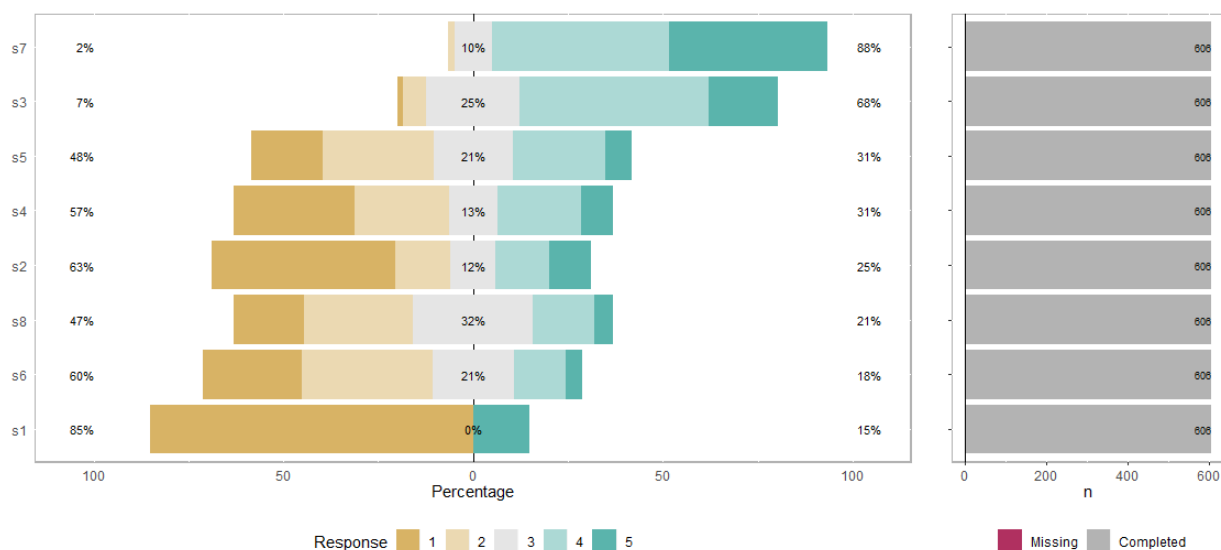
Tabela 1: Questões componentes do questionário submetido aos participantes. A questão “q” apresenta uma pergunta direta ao participante e aceita respostas em uma escala Likert de 7 níveis (1 - na semana passada, 2 - no mês passado, 3 - entre 1 e 6 meses, 4 - entre 6 e 12 meses, 5 - de 1 a 3 anos, 6 - há mais de 3 anos, 7 - não me lembro). Já as questões “s1” a “s8” apresentam proposições apresentadas ao participante, que expressa seu julgamento por meio de uma escala Likert de 5 níveis (1 - discordo totalmente, 2 - discordo, 3 - não concordo nem discordo, 4 - concordo, 5 - concordo totalmente, NA - não me lembro). De acordo com os autores [14], as questões “s1” e “s2” foram introduzidos no questionário para mensurar fator “novidade”, e as questões de “s2” a “s6” estimam o fator “imprevisibilidade”. As questões “s7” e “s8” estimam, respectivamente, os fatores de satisfação e de ampliação da preferência; esses fatores não são postulados como determinantes da serendipidade, e sim influenciados por ela.

Item	Descrição Original	Tradução	Escala	Indicador para
q	When did you watch this movie for the first time?	Quando você assistiu esse filme pela primeira vez?	7 níveis	Não se aplica
s1	The first time I heard of this movie was when MovieLens suggested it to me.	A primeira vez que ouvi sobre este filme foi quando o MovieLens o sugeriu para mim.	5 níveis	Novidade
s2	MovieLens influenced my decision to watch this movie.	O MovieLens influenciou minha decisão de assistir a esse filme.	5 níveis	Novidade
s3	I expected to enjoy this movie before watching it for the first time.	Eu esperava gostar desse filme antes de assisti-lo pela primeira vez.	5 níveis	Imprevisibilidade
s4	This is the type of movie I would not normally discover on my own; I need a recommender system like MovieLens to find movies like this one.	Este é o tipo de filme que eu normalmente não descobriria sozinho; Preciso de um sistema de recomendação como o MovieLens para encontrar filmes como este.	5 níveis	Imprevisibilidade
s5	This movie is different (e.g., in style, genre, topic) from the movies I usually watch.	Este filme é diferente (e.g., em estilo, gênero, tópico) dos filmes que normalmente assisto.	5 níveis	Imprevisibilidade
s6	I was (or, would have been) surprised that MovieLens picked this movie to recommend to me.	Eu fiquei (teria ficado) surpreso com (se) o MovieLens ter escolhido (tivesse escolhido) este filme para recomendar a mim.	5 níveis	Imprevisibilidade
s7	I am glad I watched this movie.	Estou feliz por ter assistido a este filme.	5 níveis	Satisfação
s8	Watching this movie broadened my preferences. Now I am interested in a wider selection of movies.	Assistir a esse filme ampliou minhas preferências. Agora estou interessado em uma seleção mais ampla de filmes.	5 níveis	Ampliação da Preferência

Figura 2: Distribuição de respostas por item antes e depois do pré-processamento. Em cada subfigura, o painel à esquerda ilustra a proporção de respostas por nível e o painel à direita apresenta a proporção de respostas nulas (“Missing”) e não nulas (“Completed”). No painel à esquerda, cada barra representa uma questão e está anotada com três valores que correspondem, da esquerda para a direita, à proporção de respostas negativas, neutras e positivas. No painel à direita, cada barra está anotada com o número de respostas não nulas. Note que o empilhamento das barras segue a ordem decrescente da proporção de respostas positivas e, como consequência do pré-processamento, ocorre uma inversão no ranqueamento das questões “s2” e “s8”.



(a) Distribuição de respostas antes do pré-processamento



(b) Distribuição de respostas depois do pré-processamento.

Na Equação 1, $\mathbf{X}_{(I \times n)}$ é a matriz que registra respostas a um questionário contendo I questões submetido a n participantes, $\mathbf{\Lambda}_{(I \times Q)}$ (*loading matrix*, em Inglês) é uma matriz que mapeia o grau de afinidade (carregamento) de cada uma das I questões em relação aos Q fatores latentes, $\mathbf{\Theta}_{(Q \times n)}$ é uma matriz que mapeia o grau de afinidade de cada um dos n participantes com cada um dos Q fatores latentes, e $\mathbf{E}_{(I \times n)}$ é a matriz de erro, onde se expressam as variâncias associadas a erros de medição. A função r é uma função de regularização que privilegia soluções de maior parsimônia (i.e., com menor número de fatores latentes). Em análise de componentes principais, a matrix E é assumida como nula, diferentemente do que ocorre na análise fatorial, que assume uma matriz não-nula e, adicionalmente, assume uma distribuição normal para os erros de medição.

- Algoritmo de otimização: na resolução da tarefa de otimização, algoritmos distintos costumam ser recomendados para análise fatorial e análise de componentes [12].

A seguir são apresentados os resultados obtidos da aplicação dessa técnica ao conjunto de dados *Serendipity 2018*: os resultados da análise exploratória são reportados na Seção 4.1 e a análise interpretativa é desenvolvida na Seção 4.2.

4.1 Análise Fatorial Exploratória

Com o intuito de reduzir vieses nos resultados, a análise reportada nesta seção leva em consideração o fato de que as escalas empregadas no questionário são ordinais. Seguindo recomendações em [12], a matriz de correlação entre os indicadores (respostas dos questionários) reflete correlações policóricas e, além disso, uma variação robusta do tradicional método dos mínimos quadrados ponderados (*Weighted Least Squares* ou WLS, em Inglês) foi selecionada⁴. Como o fator “Satisfação” conta com um único indicador (questão “s7” na Tabela 1), a variável “rating” foi introduzida na análise para satisfazer a recomendação de que cada fator deve possuir pelo menos dois indicadores⁵. Em relação à qualidade dos dados, a versão pré-processada e estendida do conjunto de dados atende aos requisitos recomendados [4]: (1) a aplicação do teste de esfericidade de Bartlett aponta diferença significativa entre a matriz de correlação e a matriz identidade ($\chi^2(36) = 11.471$, $p\text{-value} < 0,001$) e (2) a aplicação do teste de Keiser-Meyer-Olkin aponta amostra de qualidade mediana para análise fatorial ($kmo = 0,70$ na amostra e os indicadores obtêm individualmente $kmo > 0,50$).

O primeiro passo da análise consiste em selecionar o número de fatores latentes a considerar. Para este fim, foram empregados diagramas de variância por componente (*scree plots*, em Inglês) e heurísticas de Kaiser e Jolliffe. Devido ao fato de que divergências foram introduzidas na distribuição dos dados pelo pré-processamento, como ilustrado na Figura 2, esta etapa da análise foi aplicada nas duas versões dos dados. Como mostra a Figura 3, independentemente da versão dos dados empregada, a aplicação da heurística de Kaiser sugere a retenção de três fatores latentes, mas a heurística de Jolliffe é ambígua com relação à retenção de três ou quatro fatores. Além disso, a heurística de inflexão da curva aplicada aos dados originais (Figura 2a) sugere a retenção de dois fatores também.

O segundo passo da análise consiste em induzir modelos com número de fatores no intervalo estabelecido na etapa anterior. A exemplo de uma análise fatorial conduzida para usabilidade [17], o presente estudo considerou soluções com o número de fatores sugerido por todas as heurísticas: com dois, três e quatro fatores latentes⁶, como ilustra a Tabela 2. Um problema observado na indução dos modelos por meio do método de mínimos quadrados ponderados é que os modelos resultantes falham em atender (1) os critérios de Kaiser para comunalidade mínima entre variáveis, e (2) os critérios que sugerem níveis aceitáveis de resíduos produzidos na reconstituição da matriz de correlação a partir do modelo induzido [4]. Como alternativa, o método de resíduos mínimos foi empregado, o que produziu modelos aceitáveis, segundo os critérios mencionados, para as soluções com três e quatro fatores.

⁴A autora oferece um volume extenso de recomendações, mas focamos naquelas que são específicas ao uso do modelo “fa-poly-wlsmv”, que é o modelo que mais se adequa aos dados analisados e também ao tipo de análise conduzida neste estudo.

⁵O indicador “rating”, que mede a relevância do item para o usuário, também reflete a satisfação do usuário com o item e, assim, serve como indicador para o fator “Satisfação”. Note que o fator “Ampliação da Preferência” também conta apenas com um indicador, mas não há outros indicadores disponíveis no conjunto de dados que possam ser introduzidos na análise para remediar essa situação.

⁶Respeitada a recomendação de que cada fator latente deve contar ao menos com dois indicadores, então uma solução com cinco fatores exigiria ao menos dez indicadores, que é um número maior do que o conjunto de dados em mãos oferece.

Figura 3: Diagramas de variância explicada por componente. Nos diagramas, a abscissa representa os componentes (eigenvectors) e a ordenada representa a quantidade relativa de variância associada a cada componente (eigenvalues). Há duas anotações nos diagramas: a primeira reflete o critério de Kaiser que sugere a retenção de componentes com eigenvalues maiores do que um; a segunda reflete o critério de Jolliffe, que estabelece esse mesmo limite a partir de 0.7. Ainda, pelo critério de inflexão da curva, a retenção de dois fatores também deveria ser considerada.

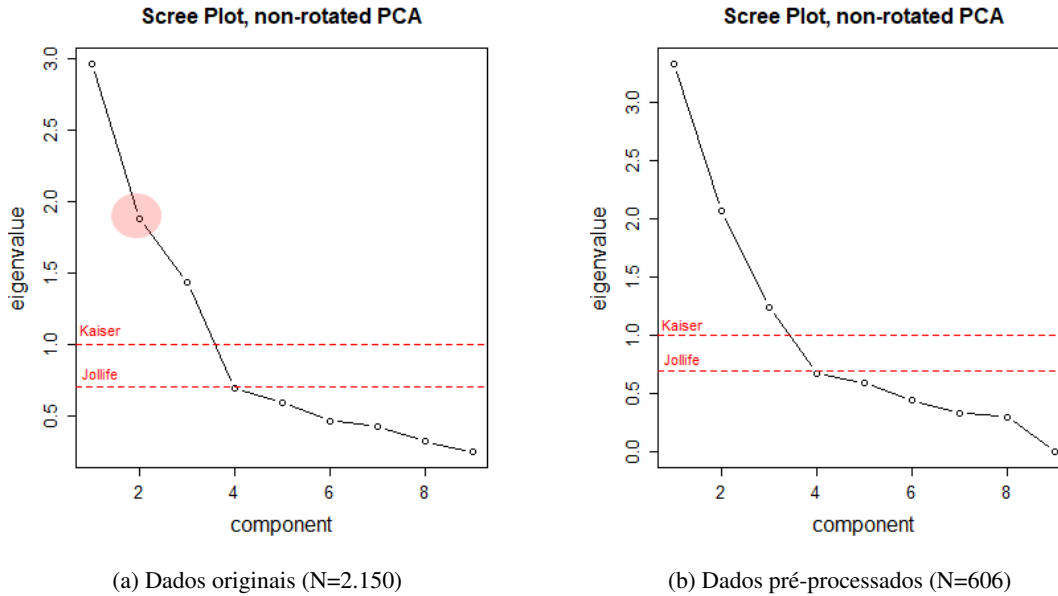


Tabela 2: Avaliação dos modelos induzidos. Foram considerados modelos com 2, 3 e 4 fatores (coluna NF). Dois métodos foram empregados para indução dos modelos: wls - mínimos quadrados ponderados, e minres - resíduos mínimos. Os critérios de Kaiser para comunalidade estão expressos nas colunas (a) e (b), ao passo que os critérios para resíduos estão nas colunas (c), (d) e (e). O critério (a) avalia se todas as variáveis possuem comunalidade acima de 0,7 (e assume que o número de indicadores é menor do que 30); o critério (b) testa se a comunalidade média é superior a 0.6 (e assume que o número de respostas é maior do que 250); o critério (c) demanda que o fit de reconstrução seja superior a 0,9; o critério (d) avalia se a proporção de resíduos com valor maior do que 0,05 é inferior a 50%, e o critério (e) verifica que se há uma diferença estatisticamente significativa entre a distribuição dos resíduos e uma distribuição normal. Os valores que atendem a cada critério estão grafados **com ênfase**.

Método	NF	Número de variáveis com comunalidade acima de 0.7 (a)	Comunalidade média por variável (b)	Fit da reconstrução (c)	Proporção de resíduos maiores que 0,05 (d)	Distribuição dos resíduos (e)	Resultado
wls	2	0	0,378	0,868	58%	não normal	Rejeitar
wls	3	0	0,405	0,896	50%	não normal	Rejeitar
wls	4	0	0,405	0,896	50%	não normal	Rejeitar
minres	2	2	0,490	0,915	61%	normal	Rejeitar
minres	3	4	0,630	0,994	8%	normal	Aceitar
minres	4	5	0,714	0,997	0%	normal	Aceitar

O passo seguinte da análise consiste em avaliar os modelos com base nos agrupamentos de indicadores sugeridos por cada modelo. A Figura 4 apresenta os agrupamentos para soluções com dois, três e quatro fatores, empregando dois tipos de rotação (promax e oblimin) para ajustar os pesos que representam a associação entre indicadores e fatores latentes. As soluções com dois fatores foram incluídas, apesar da recomendação de rejeição pelos critérios descritos na Tabela 2, com o intuito de mostrar a evolução da associação entre indicadores e fatores. Apesar da variação em número de fatores e métodos de rotação, alguns agrupamentos permanecem estáveis: (1) os indicadores “s4”, “s5”, “s6” e “s8” permanecem associados a um fator comum em todas as variações, (2) o mesmo pode ser dito a respeito dos indicadores “s1” e “s2” e (3) dos indicadores “s7” e “rating”. O indicador “s3” é o único indicador que não se mantém em um agrupamento estável: na solução com 2 fatores, ele aparece agrupado aos indicadores “s7” e “rating”; na solução com 3 fatores, ele aparece agrupado com outro conjunto de indicadores, mas está negativamente correlacionado ao fator comum; já na solução com 4 fatores, o indicador aparece isolado. Dessa forma, exceto pelo indicador “s3”, os agrupamentos produzidos pelas soluções com três e quatro fatores são similares. Essa conclusão se mantém mesmo após a inspeção dos fatores de carregamento de cada indicador, que são ilustrados na Figura 5. Por esta razão, o indicador “s3” foi excluído da análise interpretativa desenvolvida a seguir.

4.2 Análise Interpretativa

Uma análise do conteúdo das questões que aparecem agrupadas em modelos distintos é apresentada na Figura 6. As questões foram mantidas no idioma original com o intuito de preservar a fidelidade semântica ao longo da análise.

4.2.1 Agrupamento MR1

Pela inspeção dos diagramas de fator de carregamento (Figuras 5b e 5c), é possível determinar que o indicador “s5” é o indicador predominante do fator latente MR1, seguido dos fatores “s8”, “s4” e “s6” (em ordem de predominância). O respondimento das questões agrupadas sob o fator MR1 parece requerer que o participante expresse seu julgamento sobre a (dis)similaridade entre o item (filme) referido no questionário e os itens dos quais o participante se recorda. Na Figura 6a, a primeira sentença da questão “s4” é associada à questão “s5” por meio de sentenças mediadoras. A semelhança entre essas sentenças é postulada pela proximidade semântica das sentenças desenvolvidas a seguir:

- (s4) *“This is the type of movie I would not normally discover on my own.”*
- (m1) *This is the type of movie I would **hardly find** on my own.*
- (m2) *This is the type of movie I **rarely find**.*
- (m3) *This _ type of movie **is not like the ones I usually find** .*
- (m4) *This type of movie **is different from** the ones I usually find.*
- (m5) *This type of movie is different from the ones I usually **watch**.*
- (m6) *This **movie (e.g., in style, genre, topic)** is different from the ones I usually watch.*
- (s5) *“This movie is different (e.g., **in style, genre, topic**) from the ones I usually watch.”*

Por outro lado, a segunda sentença da questão “s4” parece evocar o julgamento do participante a respeito do impacto que o sistema Movielens exerceu na sua experiência, que é a interpretação desenvolvida mais adiante para o fator latente MR3. Por ora, vale ressaltar que os fatores MR1 e MR3 possuem uma correlação relativamente alta entre si (entre 0,3 e 0,5, conforme respectivamente ilustram as Figuras 4b e 4c). É possível notar que o indicador “s4” é o indicador que possui o maior carregamento positivo comum agregado nos dois fatores, MR1 e MR3, como ilustram as Figuras 5b e 5c. Postulamos que a primeira sentença da questão “s4” explica seu carregamento relativamente alto no fator MR1, e a segunda sentença explica o carregamento no fator MR3.

Na Figura 6a, a primeira sentença da questão “s8” também é associada à questão “s5” por meio de sentenças mediadoras:

- (s8) *“Watching this movie broadened my preferences.”*

Figura 4: Diagramas de agrupamento de indicadores e fatores latentes obtidos a partir de modelos induzidos com o método de resíduos mínimos (minres). Os indicadores são representados como retângulos e os fatores latentes são representados como círculos. As setas que partem de um fator latente para um indicador denotam um modelo reflexivo (i.e., o fator latente causa a manifestação mensurada pelos indicadores). Neste caso, a anotação na seta representa o fator de carregamento do indicador sobre o fator. As setas que conectam dois fatores latentes representam a correlação entre esses fatores; a magnitude da correlação é expressa na anotação e apenas correlações de magnitude superior a 0,3 estão representadas. O critério de agrupamento adotado associa cada indicador ao fator latente para o qual apresenta o maior carregamento absoluto. Nos diagramas na primeira linha, os pesos foram ajustados por meio de rotação promax, e na segunda linha por oblimin. Indicadores que permanecem sob um mesmo agrupamento em todas as variações aparecem ilustrados com uma mesma cor; o indicador “s3” é a única exceção observada. Vale ressaltar que, exceto por pequenas diferenças nos fatores de carregamento e nas correlações entre fatores, não há variação na composição dos agrupamentos produzidos pelas duas rotações.

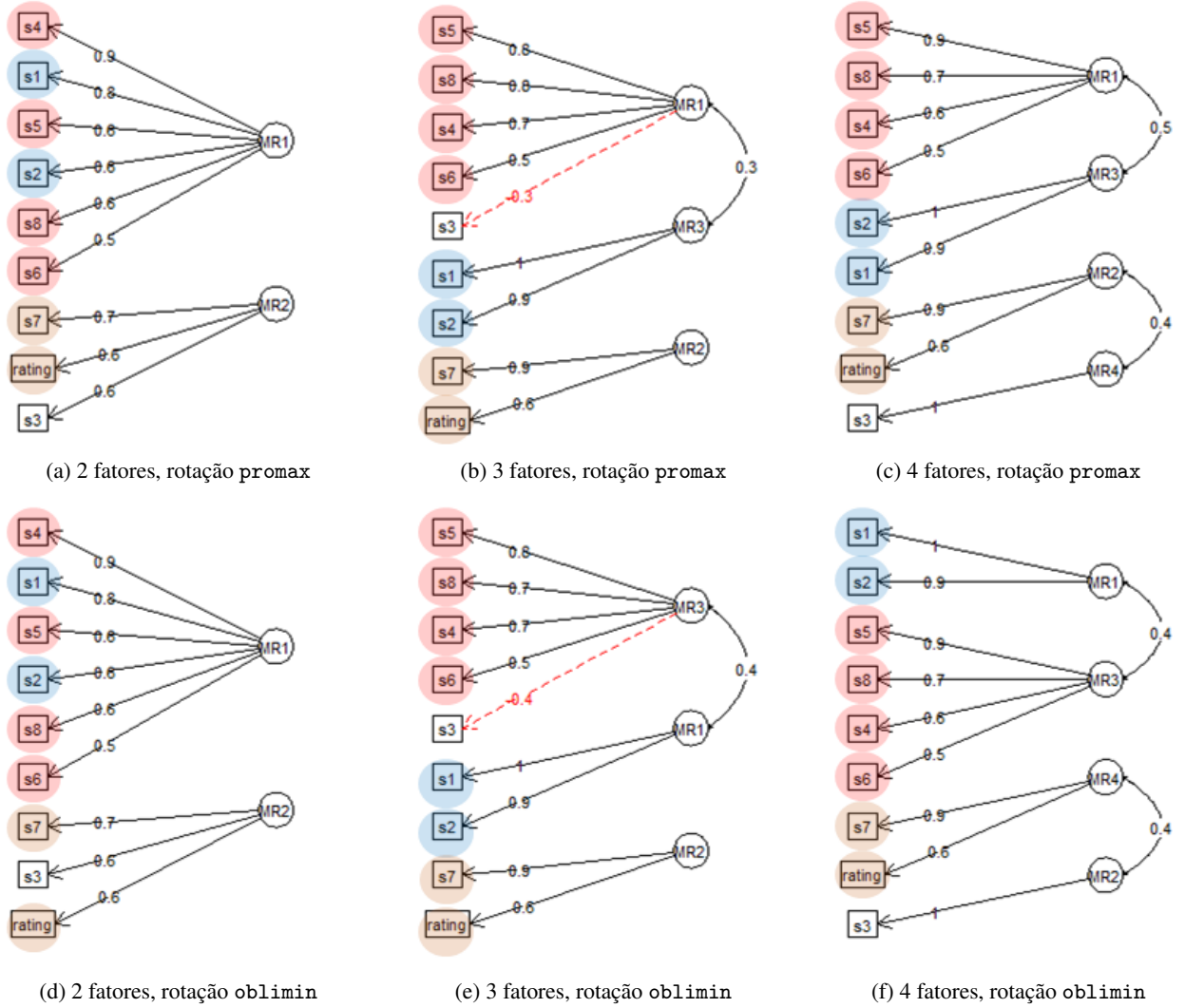
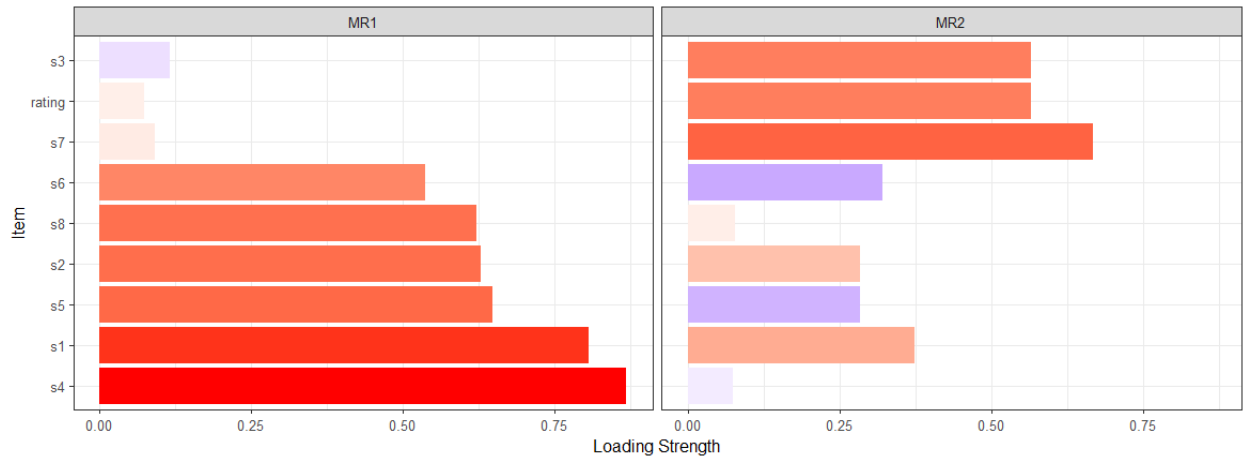
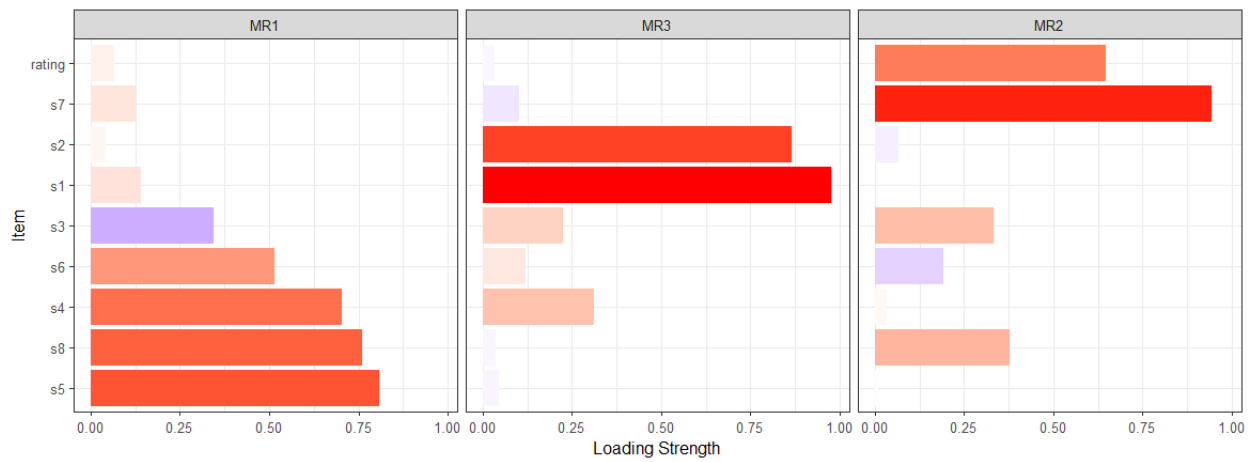


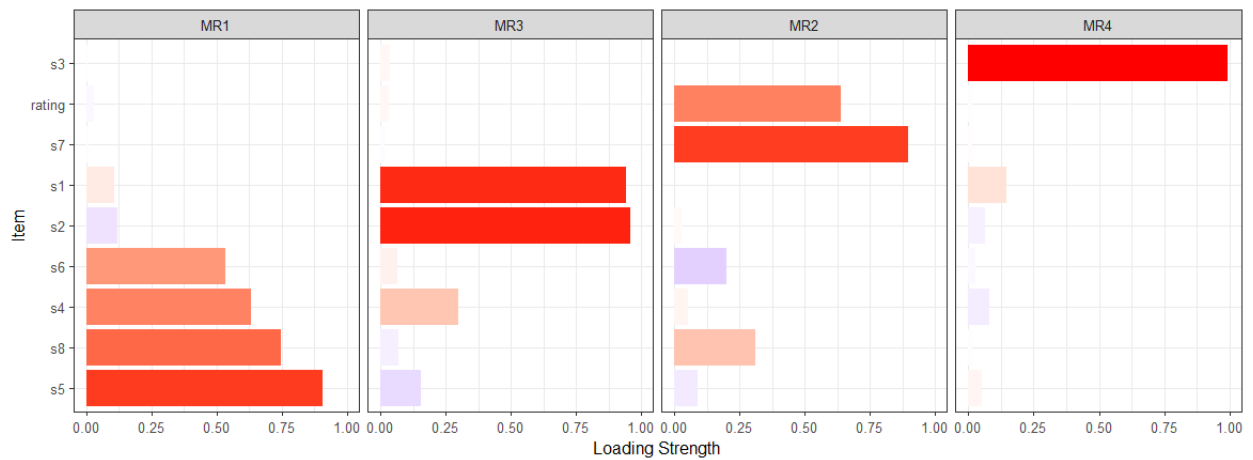
Figura 5: Diagramas de fatores de carregamento por indicador. Nos diagramas, os indicadores estão representados na ordenada e a abscissa apresenta a graduação da carga do indicador sobre o fator MR1, MR2, MR3 ou MR4. As barras em tons de vermelho representam fatores positivos de carregamento e as barras em tons de azul representam fatores negativos de carregamento.



(a) 2 fatores, rotação promax



(b) 3 fatores, rotação promax



(c) 4 fatores, rotação promax

- (m7) *Watching this movie broadened my **experience**.*
- (m8) *Watching this movie broadened my experience **because it is different from the ones I watched**.*
- (m9) *This movie is different from the ones I **usually watch**.*
- (s5) *“This movie is different (e.g., **in style, genre, topic**) from the ones I usually watch.”*

Já a segunda sentença da questão “s8” parece evocar o julgamento do participante a respeito do impacto que o item em questão (filme) exerceu na sua experiência, que é a interpretação desenvolvida mais adiante para o fator latente MR2. É possível notar que o indicador “s8” é o indicador que possui o maior carregamento positivo comum agregado nos dois fatores, MR1 e MR2, como ilustram as Figuras 5b e 5c. Postulamos que a primeira sentença da questão “s8” explica seu carregamento relativamente alto no fator MR1, e a segunda sentença justifica o carregamento no fator MR2.

Por fim, a Figura 6a mostra que a questão “s6” também pode ser associada à questão “s5” por meio de sentenças mediadoras:

- (s6) *“I was surprised that Movielens picked this movie to recommend to me.”*
- (m10) *I was surprised that _ this movie **was recommended** to me.*
- (m11) *I was surprised that this movie was recommended to me **because it is different from the ones I watch**.*
- (m5) *_ This **type of movie** is different from the ones I usually watch.*
- (s5) *“This movie is different (e.g., **in style, genre, topic**) from the ones I usually watch.”*

Vale ressaltar que o indicador “s6” apresenta carregamento positivo em MR3 e negativo em MR2, como ilustram as Figuras 5b e 5c. Nossa interpretação é que o participante tende a atribuir ao Movielens, que é explicitamente referido como um agente em “s6”, o impacto sobre sua experiência (o que justifica algum carregamento positivo em MR3); por exclusão, o impacto não pode ter sido exercido por um outro agente (o que justifica o carregamento negativo em MR2).

4.2.2 Agrupamento MR2

Pela inspeção dos diagramas de fator de carregamento (Figuras 5b e 5c), é possível estabelecer que “s7” é o indicador predominante do fator latente MR2, seguido do fator “rating”. O respondimento das questões agrupadas sob esse fator parece exigir que o participante expresse seu julgamento a respeito do impacto que a exposição ao item (i.e., assistir ao filme) exerceu sobre sua experiência. Postulamos que o indicador “s7” mede a atitude do participante em relação à sua exposição ao item (i.e., assistir ao filme foi uma experiência positiva), ao passo que o indicador “rating” mede a mesma atitude, mas em uma escala com graduação mais fina (com 10 pontos).

Como mencionado, a segunda sentença da questão “s8” parece demandar do participante o mesmo tipo de julgamento, conforme ilustra a sequência de sentenças mediadoras:

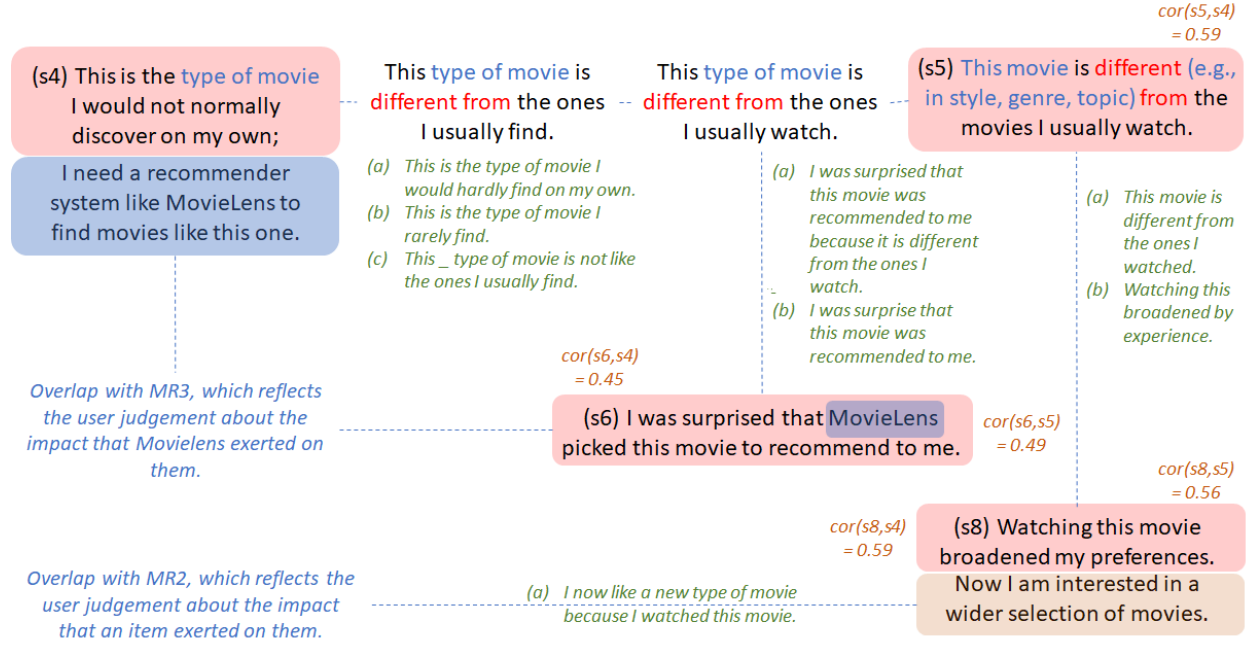
- (s8) *“Now I am interested in a wider selection of movies.”*
- (m12) ***Watching this movie made me** interested in a wider selection of movies.*
- (m13) *Watching this movie made me interested in a wider selection of movies **because I liked the movie**.*
- (m14) *I liked **watching** this movie.*
- (s7) *“I am glad I watched this movie.”*

Apesar da afinidade do indicador “s8” com o fator MR2 (impacto no participante da exposição ao item), seu carregamento é significativamente maior no fator MR1 (dissimilaridade entre o item e a experiência do usuário), o que justifica seu agrupamento sob este fator.

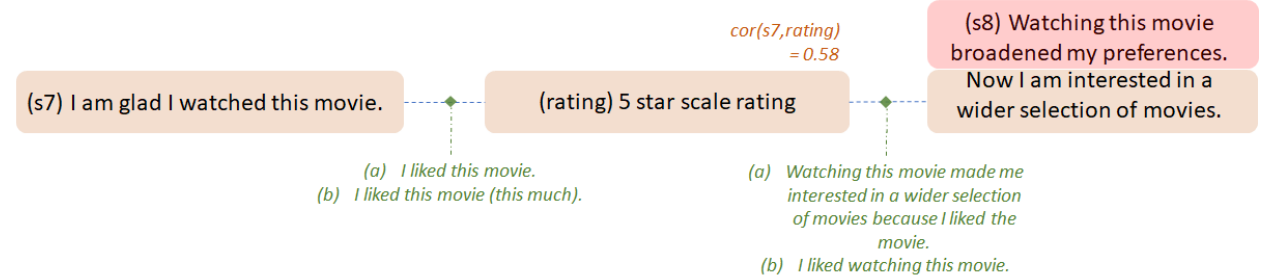
4.2.3 Agrupamento MR3

Os diagramas de fator de carregamento (Figuras 5b e 5c) mostram que “s1” é o indicador predominante do fator latente MR3 na solução com três fatores, seguido do fator “s2”, mas essa condição se inverte na solução com quatro fatores. O

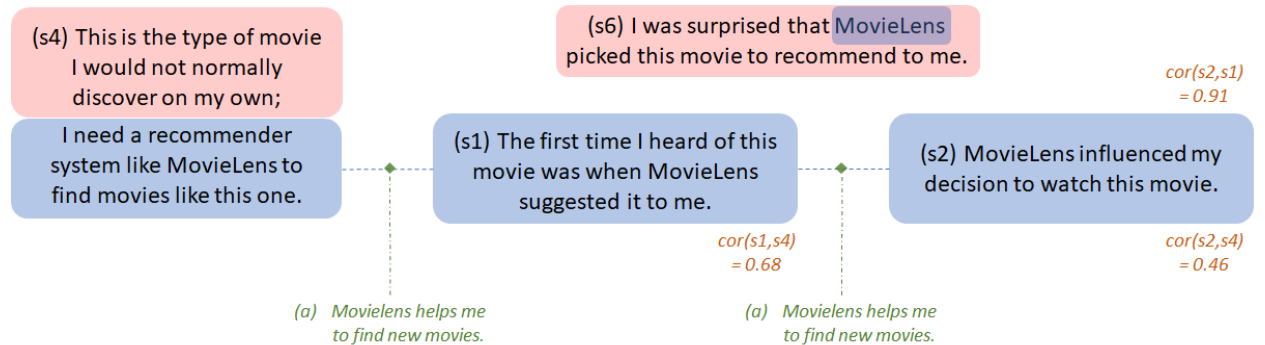
Figura 6: Similaridade semântica entre as questões agrupadas por afinidade aos fatores latentes. Sentenças que compõem questões estão representadas dentro de retângulos, cujo esquema de cor acompanha aquele empregado nos diagramas de agrupamento de indicadores (Figura 4). Se a questão composta de duas sentenças, cada sentença aparece em um retângulo cuja cor indica a afinidade semântica com um fator latente específico. Todas as sentenças que se encontram fora de retângulos são sentenças mediadoras, exceto sentenças em azul, que registram a interpretação dada a um fator latente diferente daquele que é apresentado na sub-figura. Também são anotadas algumas correlações entre indicadores.



(a) Questões agrupadas em torno do fator MR1, que reflete o julgamento do participante sobre a (dis)similaridade entre o item (filme) referido no questionário e os itens dos quais ele se recorda.



(b) Questões agrupadas em torno do fator MR2, que reflete o julgamento do participante a respeito do impacto que a exposição ao item exerceu sobre sua experiência.



(c) Questões agrupadas em torno do fator MR3, que reflete o julgamento do participante a respeito da participação do sistema MovieLens na construção da experiência.

respondimento das questões agrupadas sob esse fator parece exigir que o participante expresse seu julgamento a respeito do impacto que o Movielens exerceu sobre sua experiência. Postulamos que ambos os indicadores, “s1” e “s2”, avaliam a atitude do usuário em relação à participação do (agente) Movielens na construção da experiência (de assistir ao filme):

- (s1) “*The first time I heard of this movie was when Movielens suggested it to me.*”
- (m15) *I watched this movie because Movielens suggested it to me.*
- (m16) *Movielens helped me decide to watch this movie.*
- (s2) “*Movielens influenced my decision to watch this movie.*”

Como mencionado, a segunda sentença da questão “s4” também parece demandar do participante o mesmo tipo de julgamento, conforme ilustra a sequência de sentenças mediadoras:

- (s4) “*I need a recommender system like Movielens to find movies like this one.*”
- (m17) *I need a recommender system like Movielens to find movies like this one that it has suggested to me.*
- (m18) *Movielens suggested this movie to me.*
- (s1) “*The first time I heard of this movie was when Movielens suggested it to me.*”

Apesar da afinidade dos indicadores “s4” e “s6” com o fator MR3 (participação do sistema Movielens na construção da experiência do usuário), seu carregamento é significativamente maior no fator MR1 (dissimilaridade entre o item e a experiência do usuário), o que justifica seu agrupamento sob este fator.

5 Discussão

Como registrado na Tabela 2, os modelos com três e quatro fatores satisfazem os critérios de qualidade tipicamente aplicados em análise fatorial, ao contrário ao modelo com dois fatores, que falhou em ambos os critérios de Kaiser e nos critérios de proporção e distribuição de resíduos que surgem na reconstrução da matriz de correlações. Entretanto, o modelo com quatro fatores, ilustrado nas Figuras 4c e 4f, apresenta um fator com um indicador isolado(s3), o que reduz a qualidade do modelo, de modo que a solução com três fatores parece mais promissora.

Vale ressaltar que o indicador “s3” foi removido da análise subsequente por conta da variação dos fatores aos quais ele aparece associado nas diferentes soluções exploradas. Por exemplo, na solução com dois fatores, este indicador aparece fortemente associado aos indicadores “s7” e “rating”, como ilustra a Figura 5a, o que sugere que ele captura a atitude do usuário com relação ao impacto do item na experiência do participante. Já na solução com três fatores, ele permanece associado aos mesmos indicadores, mas sua afinidade com o fator MR1 é predominante e negativa, como ilustra a Figura 5b, o que sugere que ele captura a percepção de similaridade do item avaliado em relação aos itens conhecidos pelo usuário. Algumas causas podem estar atuando para provocar essa ambiguidade:

- A questão assume que a noção de “expectativa”, evocada na frase “*I expected to enjoy this movie*”, é selecionada pelos participantes de maneira razoavelmente uniforme. Entretanto, o verbo “*to expect*” possui acepções distintas, podendo ser interpretado como “chances de algo acontecer” (noção mais objetiva) ou “o desejo de que algo aconteça” (noção mais subjetiva)⁷. Neste caso, diferenças individuais podem se manifestar como maior variância nas respostas. Vale ressaltar que, uma análise de confiabilidade realizada no modelo ilustrado na Figura 4b indica que a remoção do indicador “s3” aumenta a confiabilidade da medição do fator MR1.
- A questão assume que os participantes registram (no sentido de memorizar) expectativas passadas a respeito de um evento que costuma ser cotidiano e pouco emocional. Nesta situação, a tendência é que o participante não se recorde de ter nutrido uma expectativa específica. Essa hipótese encontra suporte no fato de o número de respostas neutras para “s3” é o segundo maior entre os indicadores (atrás apenas de “s8”).

⁷ Acepções registradas no dicionário Merriam-Webster online: <https://www.merriam-webster.com/dictionary/expect>

- A questão está sujeita ao efeito da busca por consistência caso o participante não se recorde de ter tido uma expectativa em relação ao evento. Neste caso, o fato do participante ter tomado a decisão de assistir ao filme e ter-lhe atribuído uma avaliação positiva podem induzir o participante a responder à questão de forma positiva.

O resultado essencial dessa análise é que a interpretação dada aos fatores latentes indica uma clara afinidade conceitual com os fatores postulados como determinantes da serendipidade, como definidos na Seção 2: (1) o fator MR2, que representa o julgamento do participante a respeito do impacto que a exposição ao item exerceu sobre sua experiência, encontra consonância com a noção de relevância do item para o usuário; (2) o fator MR1, que expressa o julgamento do participante a respeito da dissimilaridade entre o item referido e os itens dos quais o participante se recorda, se confunde claramente com a noção de surpresa. Entretanto, é necessário destacar que a análise levanta dúvidas a respeito da eficácia dos indicadores “s1” e “s2” em mensurar diferentes noções de novidade que, no estudo original, estão associadas à baixa familiaridade do participante com o item referido no questionário. Esses indicadores capturam, em maior medida, a atitude do participante em relação ao papel de causalidade que o sistema Movielens exerceu na exposição do participante ao item apontado no questionário. Essa hipótese é reforçada pelo fato de que esses indicadores obtêm um baixo carregamento no fator MR1, que está relacionado à dissimilaridade do item em relação aos itens já conhecidos (e recuperados) pelo usuário. Dada a sobreposição conceitual entre a novidade, como definida na Seção 2, e a interpretação dada ao agrupamento MR1 (dissimilaridade entre o item apresentado e os itens recuperados pelo participante), a tendência seria que os indicadores “s1” e “s2” tivessem algum carregamento expressivo no agrupamento MR1, o que não ocorre. Além disso, a noção de imprevisibilidade, que tem as questões “s3” e “s4” como indicadores, não parece estar presente em nenhum dos fatores latentes identificados. Talvez isso deva à ambiguidade na interpretação da questão associada ao indicador “s3” e ao fato de que o indicador “s4” encontrou forte afinidade com o fator MR1, que está alinhado à definição de surpresa. Com base nesses resultados, o modelo formativo com dois fatores, ilustrado na Figura 1a, parece ter mais afinidade com os dados analisados do que o modelo com três fatores.

6 Conclusão, limitações e trabalhos futuros

Neste estudo, traçamos um paralelo entre o estado atual da pesquisa sobre serendipidade em sistemas de recomendação com as pesquisas sobre usabilidade conduzidas na comunidade de interação humano-computador. As semelhanças foram apontadas para motivar a aplicação de análise fatorial no conjunto de dados Serendipity 2018 com o intuito de esclarecer qual o modelo de medição está mais alinhado com os dados. Os resultados apontam que um modelo formativo que combina as noções de relevância e surpresa encontra maior consonância com os indicadores coletados. Um benefício adicional da adoção desse modelo de medição é que ele permite a exploração de resultados da literatura em ciências cognitivas a respeito da experiência de surpresa em agentes humanos [3]. Nessa literatura, os conceitos de surpresa e novidade estão melhor desenvolvidos, no sentido de contar com instrumentos validados, que podem servir de inspiração para a criação de instrumentos que se apliquem ao contexto de sistemas de recomendação.

Uma limitação deste estudo é que, por características das bibliotecas de software utilizadas, não foi possível aplicar a variante robusta do algoritmo de otimização recomendado para análise de dados ordinais, como recomendado em [12]. Ainda, devido à adoção de correlação policórica, os testes de multicolinearidade falharam, mas vale ressaltar que esse problema não se manifesta quando a análise é conduzida com correlação de Pearson, que assume dados intervalares.

Como oportunidade futura, parece promissora a ideia de reproduzir este estudo aplicando modelagem de equações estruturais (*Structural Equation Modelling*, em Inglês), o que permitiria a realização de análise fatorial confirmatória e também avaliar o papel de mediação da serendipidade na satisfação e ampliação da preferência do usuário.

Reprodutibilidade

O conjunto de dados pré-processado e o código fonte empregado neste estudo estão publicamente disponíveis em <https://github.com/andreplima/factor-analysis>. O código fonte se baseia predominantemente em trechos de código em R gentilmente disponibilizados pelos autores em [4] e em funcionalidades implementadas na biblioteca psych [21].

Agradecimentos

Agradeço, sem implicar, ao Pedro Abrahão, da Consultoria estatística do ICMC/NEA, pelo suporte metodológico, e ao Diego Lima, egresso do Programa de Pós-graduação em Psicologia da UFSCar, pelas referências sobre análise fatorial.

Referências

- [1] BOLLEN, K., AND LENNOX, R. Conventional wisdom on measurement: A structural equation perspective. *Psychological bulletin* 110, 2 (1991), 305.
- [2] DE GEMMIS, M., LOPS, P., MUSTO, C., NARDUCCI, F., AND SEMERARO, G. Semantics-aware content-based recommender systems. In *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds., 2nd ed. Springer US, Boston, MA, 2015, ch. 4, pp. 119–160.
- [3] DE LIMA, A. P., AND PERES, S. M. Effect of item representation and item comparison models on metrics for surprise in recommender systems. In *Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 1: ICEIS*, (2019), INSTICC, SciTePress, pp. 513–524.
- [4] FIELD, A., MILES, J., AND FIELD, Z. *Discovering statistics using R*. Sage, 2012.
- [5] GESCHKE, D., LORENZ, J., AND HOLTZ, P. The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology* 58, 1 (2019), 129–149.
- [6] GIERE, R. N. *Scientific perspectivism*. University of Chicago Press, 2006.
- [7] GORSUCH, R. L. *Factor analysis*. Saunders, 1974.
- [8] HERLOCKER, J. L., KONSTAN, J. A., TERVEEN, L. G., AND RIEDL, J. T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 5–53.
- [9] IAQUINTA, L., DE GEMMIS, M., LOPS, P., SEMERARO, G., AND MOLINO, P. Can a recommender system induce serendipitous encounters? In *E-commerce*. IntechOpen, 2010.
- [10] KAMINSKAS, M., AND BRIDGE, D. Measuring surprise in recommender systems. In *Proceedings of the workshop on recommender systems evaluation: dimensions and design (Workshop programme of the 8th ACM conference on recommender systems)* (2014), Citeseer.
- [11] KAMINSKAS, M., AND BRIDGE, D. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiIS)* 7, 1 (2017), 2.
- [12] KAPPENBURG-TEN HOLT, J. *A comparison between factor analysis and item response theory modeling in scale analysis*. PhD thesis, University of Groningen, 2014.
- [13] KOREN, Y., AND BELL, R. Advances in collaborative filtering. In *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds., 2nd ed. Springer US, Boston, MA, 2015, ch. 3, pp. 77–118.
- [14] KOTKOV, D., KONSTAN, J. A., ZHAO, Q., AND VEIJALAINEN, J. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (2018), ACM, pp. 1341–1350.
- [15] KOTKOV, D., WANG, S., AND VEIJALAINEN, J. A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111 (2016), 180–192.
- [16] LEWIS, J. R. Usability: lessons learned... and yet to be learned. *International Journal of Human-Computer Interaction* 30, 9 (2014), 663–684.
- [17] LEWIS, J. R., AND SAURO, J. The factor structure of the system usability scale. In *International conference on human centered design* (2009), Springer, pp. 94–103.

- [18] MCNEE, S. M., RIEDL, J., AND KONSTAN, J. A. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems* (2006), ACM, pp. 1097–1101.
- [19] NGUYEN, T. T., HUI, P.-M., HARPER, F. M., TERVEEN, L., AND KONSTAN, J. A. Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proceedings of the 23rd International Conference on World Wide Web* (New York, NY, USA, 2014), WWW '14, ACM, pp. 677–686.
- [20] PARISER, E. *The filter bubble: What the Internet is hiding from you*. Penguin US, New York, NY, 2011.
- [21] REVELLE, W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2018. R package version 1.8.12.
- [22] TRACTINSKY, N. The usability construct: A dead end? *Human–Computer Interaction* 33, 2 (2018), 131–177.
- [23] ZHANG, Y. C., SÉAGHDHA, D. O., QUERCIA, D., AND JAMBOR, T. Auralist: Introducing serendipity into music recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (New York, NY, USA, 2012), WSDM '12, ACM, pp. 13–22.