

Sistema de Recomendação de Frases Para Nomes de Bebês

Lara Esquivel de Brito Santos, João B. Rocha Junior

¹Departamento de Ciências Exatas – Universidade Estadual de Feira de Santana (UEFS)
Caixa Postal 44036-900 – Feira de Santana – BA – Brasil

laraesquivelsantos@gmail.com, joao@uefs.br

Abstract. *Part of the studies on modern recommender systems emphasized system performance rather than user experiences with recommender interfaces. In this way, it was interesting to analyze a new form of recommendation to offering a new experience for the users interested in construction of new products. We propose a new interface for recommended systems based on frases. We evaluate this new interface in a system to recommend baby names.*

Resumo. *Parte dos estudos sobre sistemas de recomendação modernos enfatizou no desempenho do sistema ao invés das experiências dos usuários com as interfaces do recomendador. Dessa forma foi interessante analisar uma nova forma de recomendação para oferecer uma nova experiência para o usuário interessados na construção de novos produtos. Propomos uma nova interface para sistemas recomendados baseado em frases. Avaliamos esta nova interface em um sistema para recomendar nomes de bebês.*

1. Introdução

Nos últimos anos, a explosão de dados digitais e a crescente presença da Internet na vida cotidiana transformaram drasticamente a maneira como consumimos informações e produtos. Em meio a essa evolução tecnológica, surgiram os sistemas de recomendação, que desempenham um papel fundamental na simplificação e personalização das escolhas dos usuários. Devido a capacidade de avaliação, o usuário pode escolher o que gosta ou não gosta, como ocorre com a Netflix, Amazon e o Google News, por exemplo [Aggarwal et al. 2016].

Basicamente os Sistemas de Recomendação vão analisar as interações *usuário-item* providas pelos próprios usuários, a quais a partir dessas o sistema irá aprender as preferências e as características dos itens que o usuário provavelmente gosta e recomendar os *Top-N* itens que mais correspondem aos gostos de cada usuário [Fang et al. 2020].

Todavia parte dos estudos sobre sistemas de recomendação modernos focam no desempenho do sistema ao invés das experiências dos usuários com as interfaces do recomendador [Chen and Pu 2014]. Dessa maneira é interessante analisar novas formas de recomendação, que além das clássicas listas de recomendação para propor uma nova experiência para o usuário. Este artigo propõe um nova interface para sistemas de recomendação, utilizando frases (tipo chat).

Ao invés de um sistema de recomendação clássico aonde produtos são recomendados com base na interação do usuário, a ideia é recomendar frases personalizadas que

possam auxiliar o usuário na tomada de decisão. A medida que mais informações vão sendo obtidas do usuário, frases mais específicas podem ser recomendadas. Por exemplo, para um usuário que acaba de entrar um *e-commerce*, frases genéricas do tipo “Deseja conhecer os itens mais vendidos da loja?” ou “Gostaria de saber quais itens está em promoção?”. Entretanto, a partir da interação do usuário com os produtos do *e-commerce*, frases mais específicas podem ser recomendadas tipo “*Gostaria de saber quais são os vestidos vermelhos mais vendidos?*”.

Acreditamos que esse novo método de interação com o usuário através de frases, possa ser interessante para construção de novos produtos. Para testar a recomendação de frases, nós criamos um Sistema que Recomenda nomes para bebês. A partir de um nome fornecido pelo usuário, outros nomes são recomendados. Por exemplo, caso o usuário digite “*João*”, nomes de bebês associados a *João Pedro, Rafael e José* são recomendados. O foco deste artigo não está na recomendação desses nomes, mas sim na recomendação de frases que serão exibidas para o usuário à medida que ele interage com o sistema. Por exemplo, para um usuário que está na região nordeste e está procurando por nomes masculino, o sistema de recomendação pode sugerir frases como “*Gostaria de conhecer os nomes masculinos mais procurados por usuários que vivem no nordeste?*”

Este artigo apresenta esse problema e apresenta os primeiros resultados obtidos no intuito de construir um Sistema de Recomendação de Frases. O restante deste artigo está organizado da seguinte forma, a Seção 2 apresenta os trabalhos relacionados e suas contribuições para a pesquisa, enquanto a Seção 3 apresenta metodologia utilizada, a Seção 4 descreve o desenvolvimento, enquanto na Seção 5 descreve os resultados obtidos, e por fim a Seção 6 apresenta a conclusão da pesquisa.

2. Trabalhos Relacionados

Alguns autores apresentam como a interface afeta as experiências dos usuários, Chen and Pu (2014) desenvolveram a *ORG (Interface de Organização*, que diferentemente da clássica lista de recomendação, agrupam os itens em categorias anotadas. As categorias são geradas pelo algoritmo de *Método de Organização Baseado em Preferências*, o qual é utilizado para descobrir propriedades de compensação semelhantes entres os produtos, mas não necessariamente um novo método de recomendação.

Enquanto Habib and et al(2020) propôs um sistema de recomendação baseada em *chatbot*, que o autor denomina como *IAI MovieBot*, que faz recomendação de filmes por meio da plataforma *Telegram*, o seu trabalho mostra uma nova forma de recomendação, que difere completamente das clássicas listas, pois mostra as recomendações por meio de uma conversação, e que auxiliar o usuário a compreender as próprias preferências. Diferente do nosso trabalho, esse artigo assume que o usuário está usando o chatbot, enquanto no nosso, frases são recomendadas em uma plataforma tradicional como um browser.

Diferente dos trabalhos acima, Reviriego et al(2022) estuda a similaridade, em especial analisando as assinaturas de bit, embora não tenhamos utilizado o *Bitwise Signature Comparison (BSC)* proposto pelo autor, o seu trabalho foi essencial para resolver um problema de desempenho relacionado a recomendação das frases porque foi por esse trabalho que veio a ideia de usar assinaturas para a recomendação de frases, como está descrito na Seção de desenvolvimento.

3. Metodologia

Inicialmente desenvolvemos um banco de nomes de pessoas, o qual foi extraído diretamente do site do IBGE, e filtrado de acordo com as necessidades do projeto. A qual foi complementado com um *web crawler*, que é um programa que faz pesquisa na *Word Wide Web* de maneira sistemática e autônoma [Kausar et al. 2013], no contexto da aplicação sobre a origem dos nomes e o significado de cada um deles. O significado e a origem dos nomes extraídos da base de dados do IBGE foram encontrados no site *Dicionário de Nomes Próprios*.

Para as recomendações dos nomes de bebês foi utilizado um algoritmo baseado em popularidade de forma não personalizada, o que isso significa que as recomendações serão as mesmas para todos os usuários [Poriya et al. 2014], ou seja todos os usuários que pesquisam pelo nome "Sandra" recebem os dez nomes mais associados para Sandra em cima da popularidade desses nomes em relação ao nome Sandra, por exemplo.

Em seguida, foi feito um Banco de Frases, pensando na comunicação com o usuário, perguntando se ele gostaria de nomes com determinadas características. Como por exemplo: os mais pesquisados de hoje. Assim cada frase possui uma *query* associada, a qual é uma consulta enviada ao banco de dados com o objetivo de recuperar informações específicas [Elmasri et al. 2005], no caso os nomes com determinadas características associadas à expressão. Enquanto as frases são recomendações personalizadas para cada usuário, ou seja, cada usuário baseado no conhecimento que se tem dele vai receber uma frase.

Por fim, é desenvolvida a interface do recomendador que gera a clássica lista de recomendações e o chat para recomendar as frases com nomes associados. O sistema de recomendação foi desenvolvido por meio da plataforma *Google Scripts App* para que o seu desenvolvimento fosse acelerado e permitisse cumprir os objetivos em um curto prazo e com baixo custo, permitindo guardar os dados em planilhas.

4. Desenvolvimento

O Algoritmo 1 produz uma lista de recomendação de nomes, a partir de um nome ofertado como parâmetro 1. O algoritmo recebe como parâmetro esse nome n e retorna a lista de nomes recomendados para este nome. Inicialmente, o algoritmo pega a lista de nomes mais frequentes (linha 1) associadas a um nome n . Essa lista é obtida a partir das consultas feitas pelos usuários para o nome n . A cada vez que o nome é pesquisado, os nomes retornados para esta consulta recebem um incremento de 1. Por exemplo, caso o nome "João" tenha sido pesquisado 3 vezes e o nome "Paulo" tenha sido recomendado 2 vezes, a frequência de "Paulo" para esta consulta é 2.

O algoritmo continua obtendo uma lista de nomes N (linha 2), a partir dos nomes mais frequentemente pesquisados, por exemplo caso João tenha sido pesquisado 3 vezes, o retorno para essa frequência é 3. Depois para cada nome recomendado, é incrementado com as frequências dos nomes pesquisados. Caso o nome não exista, cria-se e acrescenta 1 para cada nova aparição.

Em seguida é chamado a função *sortByKeys* (linha 9) passando como parâmetro a frequência de nomes, ela ordenada os nomes baseado na frequência e retorna uma lista de nomes ordenados, caso a lista possua mais de 10 nomes, retorna apenas os 10 primeiros,

caso contrário retorna a lista.

Algorithm 1: namesFiltering

Input: nome a ser pesquisado n
Output: Array contendo 10 nomes para o nome pesquisado

```
1  $F \leftarrow \text{getMostFrequent}(n)$ 
2  $N \leftarrow \text{getMostSearched}(F)$ 
3 for  $n \in N$  do
4   | if  $F[n] == \emptyset$  then
5   |   |  $F[n] \leftarrow 0$ ;
6   | end
7   |  $F[n] \leftarrow F[n] + 1$ ;
8 end
9  $V \leftarrow \text{sortByKeys}(F)$ 
10 if  $|V| > 10$  then
11 |    $\text{return } V[:, 10]$ ;
12 end
13  $\text{return } V$  ;
```

O Algoritmo 2 recebe como parâmetro a assinatura de um usuário, identificando as características dos nomes de suas preferências e produz uma frase recomendada para esse. Como por exemplo a assinatura *10100001000001000*, a qual é um identificador das preferências de um determinado usuário, cada campo corresponde a um padrão de comportamento referentes a origem dos nomes, o gênero dos nomes que o usuário gosta e a região que o usuário mora. Como temos 5 regiões no país, temos 2 gêneros e o *crawler*, no momento, encontrou 10 origens para os nomes, temos 17 campos.

Supondo que tenhamos um usuário que gosta de nomes de origem grega femininos, o campo origem grega e nomes femininos possuirá 1 e o de nomes masculinos e lusitanos terá 0, e o conjunto desses campos formam uma assinatura.

Além da assinatura do usuário, o Algoritmo 2 também recebe F um conjunto de frases com as assinaturas correspondentes às frases. Assim, um conjunto de frases a serem recomendadas é declarado (linha 1), em seguida verifica-se se o usuário é conhecido, ou seja se a assinatura é diferente de vazio (linha 2).

Caso seja igual a vazio, ou seja, não se tem informação sobre o usuário, um número aleatório é escolhido para representar uma posição do conjunto de uma frase, assim uma frase aleatória do conjunto. Caso o contrário, o conjunto de frases é percorrido e as frases que coincidem com a assinatura do usuário são adicionada a outro conjunto (linha 5). Ao final, um número aleatório é escolhido e a frase que está na posição correspondente a ao número sorteado é retornada.

Dessa maneira supondo que tenhamos um usuário denominado Afonso, o qual é nordestino e ele e sua esposa esperam por um menino, a frase mais indicada para Afonso é a terceira frase mostrada na Tabela 1, que contém a frase e os campos correspondente ao perfil de usuário ideal para aquela frase, onde 1 indica verdadeiro e 0 falso, porque esta possui a assinatura *11000*, isso quer dizer que essa frase é indicada para pessoas que

Algorithm 2: RecommendPhrase

Input: assinatura do usuário u , F : conjunto de frases e suas assinaturas

Output: Frase para o usuário

```
1  $R \leftarrow []$ ;  
2 if  $u \neq \emptyset$  then  
3   for  $phrase(f), signature(f) \in F$  do  
4     if  $(signature(f) \text{ AND } u) == u$  then  
5        $R \leftarrow R \cup phrase(f)$ ;  
6     end  
7   end  
8    $s \leftarrow \text{randomNumber}()$ ;  
9   return  $R[s]$ ;  
10 end  
11  $s \leftarrow \text{randomNumber}()$ ;  
12 return  $F[s]$ ;
```

| Frases | Assinatura | | | | |
|--|------------|------|-----|------|-----|
| | NE | Masc | Fem | Greg | Rom |
| Você gostaria de conhecer os nomes mais populares da região Nordeste? | 1 | 0 | 0 | 0 | 0 |
| Você gostaria de conhecer os nomes masculinos mais populares? | 0 | 1 | 0 | 0 | 0 |
| Você gostaria de conhecer os nomes masculinos mais populares da região Nordeste? | 1 | 1 | 0 | 0 | 0 |

Table 1. Tabela Ilustrativa Exemplar representando as frases com as suas assinaturas

buscam nomes masculinos e estão curiosos sobre os nomes mais pesquisados no Nordeste, que provavelmente é a sua região habitacional, o que corresponde ao perfil de Afonso.

O Algoritmo 2 seleciona as frases que atendem a assinatura, neste caso, apenas 1 frase atendeu a assinatura do usuário que foi a frase 3. Entretanto, a assinatura da frase coincidiu com a assinatura de Afonso porque foi gerada baseada no seu histórico: local onde acessou o site e gênero dos nomes pesquisados e votados. Enquanto Sônia que está acessando pela primeira vez, não se tem conhecimento sobre os seus gostos, logo essa não possui assinatura alguma. Assim uma frase qualquer da tabela irá ser sorteada é recomendada para Sônia.

5. Resultados

Nesta seção está descrito os resultados da pesquisa, na subseção Interface apresentamos a interface e o seu funcionamento, enquanto na subseção Coleta de Dados e Recomendações os produtos consequentemente gerados pelo novo método.

5.1. Interface

A interface do recomendador, que pode ser vista na Figura 1, foi pensando para pais que procuram um nome ideal para os seus filhos. Basta procurar um nome na barra de pesquisa centralizada na imagem, se quiser pode clicar no oráculo, ícone de vidente com bola de cristal no canto inferior direito das telas, para receber uma recomendação de frase, como pode ser visualizado na Figura 3.

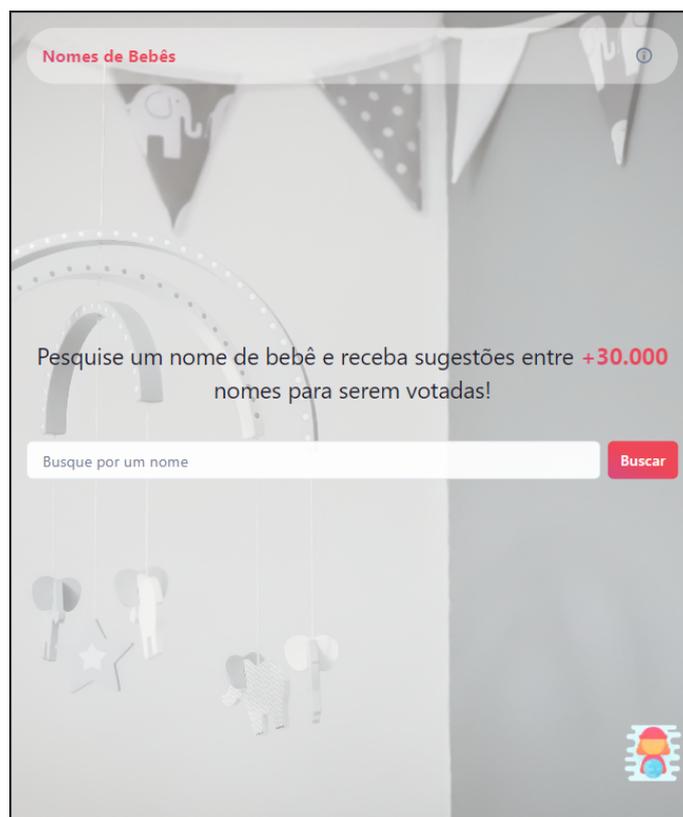


Figure 1. Interface do Recomendador

Supondo que o usuário buscou o nome *Helena*, o sistema retorna os 10 nomes mais recomendados para esse nome, baseado nos nomes mais populares para *Helena*. Conforme está na Figura 2, podendo atribuir "like" ou "deslike". Assim, mais informações sobre o indivíduo são detectadas e sobre os próprios nomes.

Como consequência, a lista de nomes associada a um nome vai se modificando, baseado na opinião dos usuários. Esse mesmo comportamento vai gerar uma assinatura de bits para cada usuário, podemos verificar que o usuário gosta de nomes que os usuários que moram no nordeste gostam ou que o usuário gosta de nomes de origem grega, por exemplo. Sendo possível posteriormente recomendar frases mais apropriadas e personalizadas para cada um, conforme está na figura 3.

5.2. Coleta de Dados e Recomendações

Ao final foi possível criar um banco de frases com aproximadamente 116 frases para auxiliar a tomada de decisão do usuário com nomes interativos por meio de *queries*, resultantes das interações do usuário com a interface descrita na Subseção anterior.

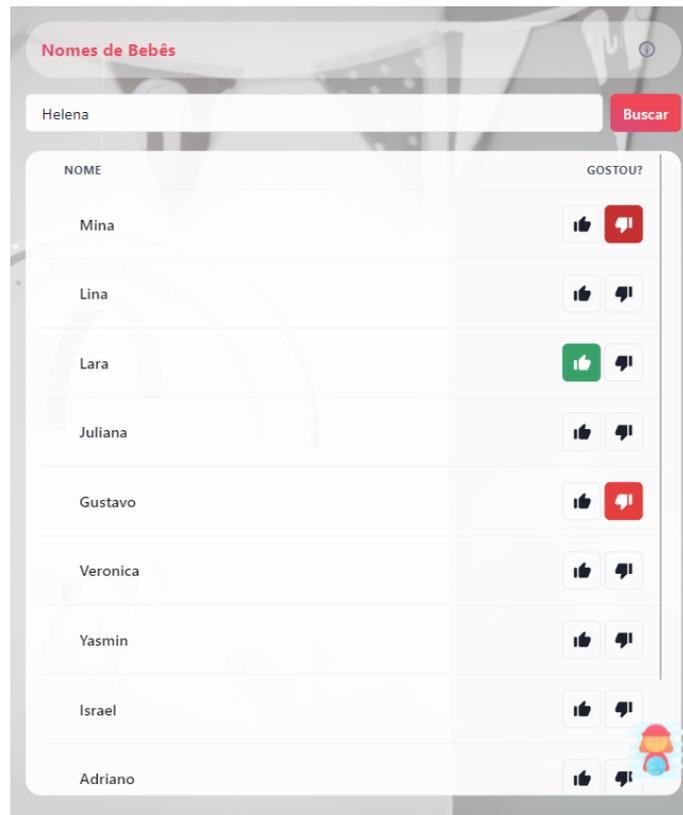


Figure 2. Interface do Recomendador após uma busca por um nome

Como por exemplo, o sistema recebe um usuário novo, o qual não tem nenhuma informação. Inicialmente, as frases recomendadas são aleatórias, a medida que ele pesquisa nomes de interesse e avalia as recomendações, as frases deixam de ser aleatórias e passam a ser um item recomendado para o usuário. Pois as suas buscas e votações são salvas e isso é utilizado no final como gerador de assinatura, que descreve que tipo de usuário é, e o que deve ser recomendado para esse.

Um usuário que mora no nordeste ou busca um nome masculino, como já visto por exemplo, vai receber uma frase associada a região nordeste e a nomes masculinos, com os *Top-N* melhores nomes para essa frase, que no exemplo de Afonso da Seção 4: *Você gostaria de conhecer os nomes masculinos mais populares da região Nordeste?* Os dez nomes masculinos mais populares entre os usuários que moram no nordeste.

Por fim, foi criado um banco de nomes com 50.743 registros, o qual ainda está sendo incrementado.

6. Conclusão

Neste trabalho descrevemos o processo de criação e da aplicação de um novo método de recomendação por meio de frases para nomes de bebês a fim de auxiliar a tomada de decisão do usuário, onde uma frase possui os *Top-N* nomes associados mais recomendados para a frase, dado as características dessa frase.

Embora não tenhamos uma avaliação definitiva acerca do novo método, visto que a aplicação ainda não foi disponibilizada para os usuários, o processo gerou um sistema de

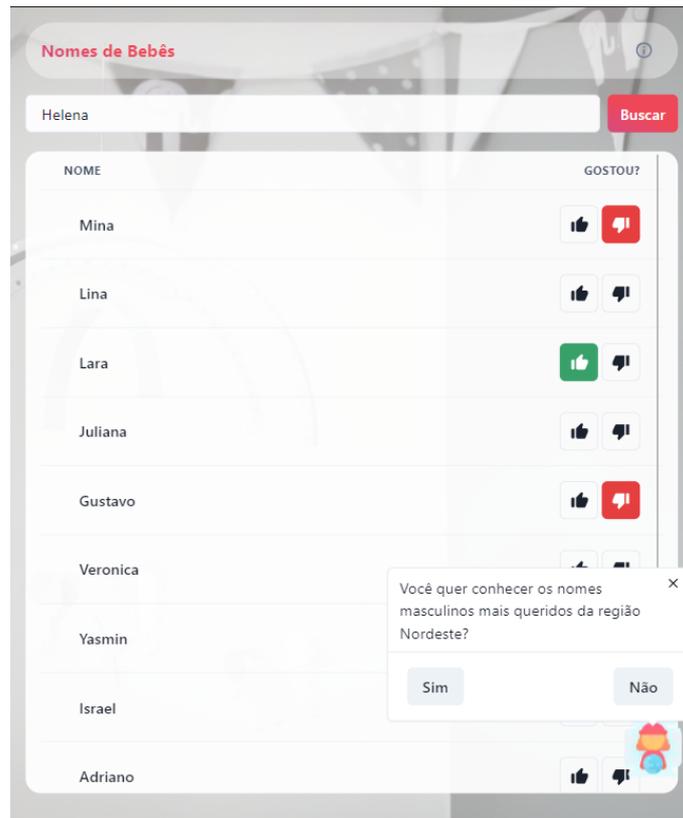


Figure 3. Interface do Recomendador recomendando uma frase

recomendação de nomes de bebês e duas bases de dados: uma de nomes e outra de frases. As quais podem ser utilizadas em trabalhos futuros ou reutilizadas por outras pessoas.

Como uma futura melhoria, é possível criar uma interface mais amigável para o usuário, onde tenha mais formas de interação para diverti-lo e se manter ativo ao sistema. Ao mesmo tempo que se tem mais informações sobre o mesmo, como um *quizzes* para saber o que esse espera em um nome, qual sentimento esperado, o que esse gostaria que um nome fosse associado, para assim criar frases mais complexas e mais personalizadas para cada usuário.

Ao final, tendo mais dados sobre o usuário é possível extrair mais valor dos nomes, por exemplo *Antonella* pode ser recomendado para o usuário que quer que sua filha seja associada a algo valioso, já que este é o significado desse nome ou busca uma versão feminina do nome *Antônio*, que pode ser o nome de alguém querido para o usuário.

Todavia, apesar do Google App Script ter acelerado o desenvolvimento da aplicação, não apresentou bons resultados de desempenho e precisa ser substituído por uma tecnologia que permita processar consultas de forma mais eficiente.

References

- Aggarwal, C. C. et al. (2016). *Recommender systems*, volume 1. Springer.
- Chen, L. and Pu, P. (2014). Experiments on user experiences with recommender interfaces. *Behaviour & Information Technology*, 33(4):372–394.

- Elmasri, R., Navathe, S. B., Pinheiro, M. G., et al. (2005). *Sistemas de banco de dados*.
- Fang, M., Gong, N. Z., and Liu, J. (2020). Influence function based data poisoning attacks to top-n recommender systems. In *Proceedings of The Web Conference 2020*, pages 3019–3025.
- Kausar, M. A., Dhaka, V., and Singh, S. K. (2013). Web crawler: a review. *International Journal of Computer Applications*, 63(2):31–36.
- Poriya, A., Bhagat, T., Patel, N., and Sharma, R. (2014). Non-personalized recommender systems and user-based collaborative recommender systems. *Int. J. Appl. Inf. Syst*, 6(9):22–27.