Geração de Receitas Culinárias para Usuários com Restrições Alimentares pela Substituição Automática de Ingredientes

Luciano D. S. Pacifico[1], Larissa F. S. Britto[2], Teresa B. Ludermir[2]

[1]Departamento de Computação (DC) – Universidade Federal Rural de Pernambuco (UFRPE) – Recife – PE – Brazil

[2]Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE) – Recife - PE - Brazil

luciano.pacifico@ufrpe.br,{lfsb,tbl}@cin.ufpe.br

Abstract. *In this work, a data-driven approach for restriction-free cooking recipe generation is presented. The proposed system is composed by a filtering process, where a recipe containing a forbidden ingredient, for a group of users, is automatically adapted to a food restriction domain by single ingredient substitution, helping to improve the number of recipes available for those users. The proposed filtering process is evaluated by means of a qualitative analysis, showing promising results.*

Resumo. *Neste trabalho, uma abordagem voltada aos dados para a recomendação de receitas culinárias para usuários com restrições alimentares é apresentada. O sistema proposto é composto por um processo de filtragem, no qual uma receita contendo ingredientes proibidos, para um grupo de usuários, é automaticamente adaptada a um domínio de restrição alimentar pela substituição de um único ingrediente, auxiliando no aumento do número de receitas disponíveis para tais usuários. O sistema de filtragem proposto é avaliado através de uma análise qualitativa, apresentando resultados promissores.*

1. Introdução

A alimentação desempenha um papel indispensável para os seres humanos, tanto por questões biológicas, quanto por questões socioculturais. A escolha do alimento a ser consumido em um dado momento do dia é um processo decisório complexo, que leva vários fatores em consideração, tais como a quantidade de tempo disponível, local, disponibilidade dos alimentos, gostos e restrições pessoais, etc.

A popularização de programas televisivos e canais na internet relacionados à gastronomia e à culinária tornou possível um rápido compartilhamento dos conhecimentos relacionados ao preparo dos mais diversos tipos de pratos e refeições, sendo grande parte desta popularização decorrente do avanço de tecnologias como a internet e as redes sociais. Receitas culinárias que anteriormente eram compartilhadas através de livros e revistas, atualmente são distribuídos livremente através de websites e blogs especializados no assunto.

A enorme quantidade de receitas disponíveis atualmente, através de repositórios web, tem atraído cada vez mais estudos relacionados à tentativa de geração de

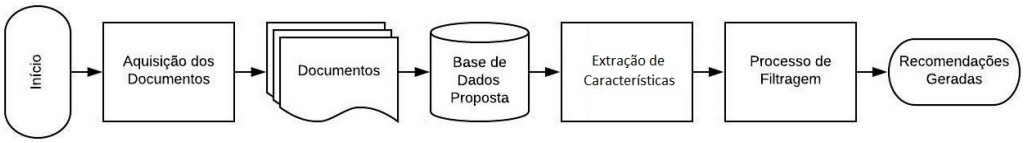

Figura 1. Etapas do sistema proposto.

recomendações que auxiliem usuários a encontrar receitas que sejam de seus interesses, e que também sejam saudáveis e contemplem às necessidades nutricionais desses usuários [Jiang et al. 2019, Majumder et al. 2019, Pecune et al. 2020, Twomey et al. 2020].

Embora as páginas web especializadas tenham trazido a vantagem da enorme quantidade de receitas disponíveis, usuários que apresentam algum tipo de restrição alimentar, como alérgicos, pessoas com problemas de saúde (ex.: diabéticos, hipertensos, etc.), veganos, vegetarianos, e alguns grupos religiosos, ainda são pouco contemplados por esses repositórios [Britto et al. 2020]. Um número bastante reduzido de receitas estão disponíveis para esses grupos, o que torna bastante restrita a variação no cardápio diário desses usuários. Muitas das receitas existentes nos sites de gastronomia poderiam ser facilmente adaptadas, pela substituição automática dos ingredientes causadores dessas restrições, ao consumo de usuários com restrições, porém, poucos trabalhos na literatura têm explorado esse tópico [Ooi et al. 2015, Oliveira et al. 2019, Britto et al. 2020].

Neste trabalho, um sistema de substituição automática de ingredientes é proposto, baseado em uma abordagem puramente voltada aos dados, visando a recomendação e geração de receitas livres de ingredientes que podem causar restrições alimentares aos usuários. A abordagem adotada será a de substituição de um único ingrediente por vez [Shino et al. 2016, Nirmal et al. 2018, Oliveira et al. 2019, Britto et al. 2020]. O presente trabalho estende as propostas apresentadas em [Oliveira et al. 2019] e [Britto et al. 2020] das seguintes formas:

1. O sistema de filtragem da proposta atual melhora a abordagem desenvolvida em [Oliveira et al. 2019] por garantir que ingredientes que não se adequem às restrições do usuário sejam automaticamente desconsiderados da lista final de recomendações.
2. Como o sistema de ranqueamento dos ingredientes substitutos apresentado não necessita do treinamento de algoritmos de Aprendizagem de Máquina, sua execução é mais rápida que a abordagem apresentada em [Britto et al. 2020].

O trabalho está organizado da seguinte forma: a metodologia proposta é apresentada na Seção 2. Em seguida, uma abordagem qualitativa é empregada para a análise do sistema proposto (Seção 3), seguida pela apresentação das conclusões e linhas para pesquisas futuras (Seção 4).

2. Metodologia

Nesta seção, a metodologia de substituição de ingredientes para a geração de receitas sem ingredientes causadores de restrições é apresentada. A metodologia adotada é ilustrada na Fig. 1. O sistema proposto é dividido em três etapas principais: a aquisição da base de dados (Seção 2.1), a etapa de extração de características (Seção 2.2) e o sistema de filtragem proposto (Seção 2.3). Cada etapa será descrita em detalhes na sequência.

Tabela 1. Informações gerais referentes à base de dados adotada.

Categoria	Sem Restrições (Domínio Original)	Sem Lactose (Domínio Alvo)	Base Combinada
Número de Receitas	20000	4105	24105
Total de Ingredientes	8130	3544	8633
Média de Ingredientes por Receita	9.8808	9.2285	9.7697
Máximo de Ingredientes por Receita	40	26	40
Mínimo de Ingredientes por Receita	2	1	1

2.1. Base de Dados

A base de dados utilizada neste trabalho foi proposta em [Majumder et al. 2019], extraída da página Food.com[1]. A base de dados já apresenta um pré-processamento dos documentos, contendo originalmente mais de 230 mil receitas (em inglês). Para cada receita estão disponíveis informações como lista de ingredientes, modos de preparo, valores nutricionais por categoria, comentários e avaliações dos usuários, além de uma categorização feita por seus usuários através de um sistema de *tags* (uma receita pode conter nenhuma, uma ou múltiplas *tags*).

Como o intuito deste trabalho é a recomendação automática de ingredientes substitutos para a adequação de uma receita às restrições do usuário, apenas as informações relacionadas à **lista de ingredientes** e às ***tags*** das receitas serão levadas em consideração. A partir da base de dados original, um conjunto de receitas foi selecionado no intuito de demonstrar a capacidade do sistema de filtragem proposto. As receitas utilizadas foram selecionadas a partir de dois domínios: Receitas sem Restrições (considerado o **Domínio Original** - D_o), formado por receitas nas quais os usuários não realizaram a classificação por *tags*; Receitas sem Lactose (**Domínio Alvo** - D_t), que são receitas nas quais as *tags* fornecidas pelos usuários (*lactose-free* e *dairy-free*) indicam a restrição do usuário no consumo desses alimentos. Como forma de preservar o desbalanceamento natural das bases de dados reais de websites especializados em receitas culinárias, a base de dados proposta apresentam um desbalanceamento entre o número de receitas no D_o e no D_t. As estatísticas da base de dados proposta são apresentadas na Tabela 1.

2.2. Extração de Características

As listas de ingredientes das receitas originais estão apresentadas em formato **textual**, sendo necessária a padronização das mesmas para um formato que seja mais adequado à manipulação desses dados pelos mecanismos de filtragem propostos. Neste trabalho, duas abordagens de extração de características são empregadas:

- ***Bag-of-Words* Binária**: Cada ingrediente é representado pelo seu nome em inglês (que pode ser formado por uma ou mais palavras). Na representação por *Bag-of-Words* Binária, o conjunto de documentos será convertido em uma **Matriz de Termos do Documento** (*Document-Term Matrix*, ou DTM), na qual cada linha representará um documento (ou seja, uma **receita**), e cada coluna representa um termo (ou seja, um **ingrediente**). Cada posição na DTM indicará se um dado ingrediente faz parte (indicado por '1') ou não (indicado por '0') da lista de ingredientes da receita em questão. Esta abordagem se mostra bastante adequada à

[1] `https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions`

proposta atual, uma vez que desejamos avaliar apenas se um ingrediente faz parte ou não à lista de uma dada receita.

- **TF-IDF (*Term Frequency - Inverse Document Frequency*)**: O TF-IDF avalia o quão importante um termo é para um documento na base de dados, como uma forma de diminuir a influência de termos que ocorrem com frequências muito altas, mas que possuem pouca relevância em cada documento. Para a proposta atual, este método pode ser bastante útil para a identificação automática de ingredientes muito comuns a diversas receitas (como *água*, *trigo*, etc.), mas que para o contexto de uma receita específica não sejam tão relevantes.

2.3. Sistema de Filtragem Proposto

Após a extração de características, tem início a etapa de filtragem dos ingredientes. O sistema de filtragem proposto tem por objetivo encontrar os ingredientes candidatos x_c mais adequados à substituição do ingrediente alvo x_t da receita alvo R_t, sendo R_t uma receita no domínio original D_o que deseja-se adequar ao domínio alvo D_t. Nesse intuito, três filtros são propostos, sendo cada filtro detalhado na sequência.

O primeiro filtro do sistema proposto tem por objetivo analisar o comportamento de cada ingrediente candidato $x_c \in I_t$ (onde I_t se refere ao conjunto total de ingredientes existentes em D_t) à substituição do ingrediente alvo x_t em relação aos demais ingredientes x_i da receita alvo R_t. O Valor de Suporte Médio (VSM), adaptado de [Nirmal et al. 2018, Oliveira et al. 2019] é então adotado. Primeiramente, o valor da ocorrência pareada $C_v(x_c, x_i)$ entre cada ingrediente $\{x_c \in I_t | x_c \notin R_t\}$ e todos os ingredientes $x_i \in R_t (i \neq t)$ é calculado, considerando todas as receitas no domínio alvo D_t. Em seguida, o VSM de cada ingrediente candidato x_c é obtido pela Eq. (1):

$$VSM(x_c) = \sum_{x_i \in R_t (i \neq t)} C_v(x_c, x_i)/(|R_t| - 1) \tag{1}$$

onde $|R_t|$ indica o número de ingredientes na receita alvo R_t.

A ocorrência pareada dos ingredientes utilizada nesse filtro leva em consideração apenas as receitas do domínio alvo D_t, representadas pela DTM binária desse domínio.

O segundo filtro que compõe o sistema proposto é obtido pela análise da relevância do ingrediente candidato x_c no domínio alvo D_t. Neste trabalho, uma abordagem automática de extração de características em textos é utilizada para a avaliação da relevância de cada x_c em D_t, o TF-IDF. Desta forma, o Valor de Relevância do Ingrediente (VRI) é dado pela Eq. (2):

$$VRI(x_c) = \sum_{\forall R_j \in D_t} \textit{TF-IDF}(x_c, R_j) / \sum_{\forall x_i \in I_t} \sum_{\forall R_j \in D_t} \textit{TF-IDF}(x_i, R_j) \tag{2}$$

O terceiro e último filtro do sistema proposto consiste na exclusão dos ingredientes do Domínio Original D_o que não existam no Domínio Alvo D_t. Esta operação tem por objetivo impedir que ingredientes causadores de alergias de D_o sejam considerados como candidatos à substituição do ingrediente alvo x_t, evitando, desta forma, sugestões prejudiciais ao usuário. A Pertinência ao Domínio Alvo (PDA) é obtida pela conjunção

Tabela 2. Receitas selecionadas.

Receita	Número de Ingredientes	Ingredientes	Ingrediente Alvo
Quick Pound Cake	5	*flour*, *eggs*, *sugar*, *butter*, *almond extract*	***butter***
Whole Wheat Corn Bread	8	*baking powder*, *cornmeal*, *eggs*, *milk*, *salt*, *sugar*, *vegetable oil*, *whole wheat flour*	***milk***

simples da lista total de ingredientes das receitas em D_t, e pela exclusão dos ingredientes tais que $\{x_c \in D_o, x_c \notin D_t\}$ (Eq. (3)):

$$PDA(x_c) = \begin{cases} \textbf{Verdadeiro}, \text{ se } x_c \in I_t \\ \textbf{Falso}, \text{ se } x_c \notin I_t \end{cases} \tag{3}$$

Os filtros VRI e PDA independem da receita alvo R_t, enquanto o VSM varia de acordo com a R_t escolhida. Para o cálculo do Escore Final do Ingrediente (EFI), os valores do VSM e do VRI são normalizados na escala $[0, 1]$, evitando que os resultados de um dos filtros seja dominante em relação ao outro. Desta forma, temos que o EFI é obtido como (Eq. (4)):

$$EFI(x_c) = \begin{cases} (VSM(x_c) + VRI(x_c))/2, \text{ se } PDA(x_c) \text{ é } \textbf{Verdadeiro} \\ 0, \text{ caso contrário} \end{cases} \tag{4}$$

3. Análise dos Resultados

Nesta seção, o sistema de substituição de ingredientes proposto será avaliado através de uma análise qualitativa. Neste intuito, três questões de pesquisa principais são elaboradas:

1. O sistema é capaz de encontrar bons ingredientes candidatos à substituição automaticamente, levando em consideração a receita alvo escolhida?
2. Os ingredientes candidatos sugeridos infringem às restrições alimentares dos usuários?
3. Os filtros propostos e sua combinação se mostraram adequados?

A avaliação do sistema proposto levará em consideração duas receitas exemplo, selecionadas do conjunto de receitas sem restrições (Domínio Original) da base de dados adotada. De modo semelhante, um ingrediente alvo de cada uma das receitas será escolhido para exemplificar a execução do sistema proposto, sendo esse ingrediente um ingrediente restrito no Domínio Alvo. A lista de ingredientes de cada uma das receitas selecionadas é apresentada na Tabela 2.

A análise qualitativa proposta levará em consideração a avaliação dos dez melhores ingredientes candidatos à substituição do ingrediente alvo (análise ***Top-10***), obtidos pelo filtro VSM, assim como o resultado *Top-10* para o sistema de filtragem final, uma vez que os resultados para o filtro VRI independem da receita alvo, e que o filtro PDA não ranqueia os ingredientes candidatos.

Os resultados experimentais para o filtro VSM são apresentados na Tabela 3. Podemos observar que, embora as listas geradas sejam bastante diferentes entre si, e que os

Tabela 3. Resultados experimentais *Top-10* para o Valor de Suporte Médio (VSM) para cada uma das receitas alvo. Ingredientes considerados bons substitutos em relação à receita alvo estão em negrito.

Receita (Ingrediente Alvo)	Ingrediente Candidato	VSM	Receita (Ingrediente Alvo)	Ingrediente Candidato	VSM
Quick Pound Cake (butter)	*baking powder*	0.1288	*Whole Wheat Corn Bread (milk)*	*baking soda*	0.1398
	baking soda	0.1238		*cinnamon*	0.0903
	salt	0.0997		*flour*	0.0894
	vanilla	0.0896		***water***	**0.0868**
	cinnamon	0.0763		*vanilla*	0.0804
	vanilla extract	0.0694		*vanilla extract*	0.0705
	water	0.0675		*all-purpose flour*	0.0694
	oil	**0.0616**		***soymilk***	**0.0567**
	vegetable oil	**0.0579**		*oil*	0.0552
	brown sugar	0.0459		*egg*	0.0543

Tabela 4. Resultados experimentais *Top-10* para o Valor de Relevância do Ingrediente (VRI) na categoria alvo.

Ingrediente Candidato	*water*	*olive oil*	*salt*	*sugar*	*garlic cloves*	*honey*	*onion*	*lemon juice*	*cinnamon*	*soymilk*
VRI	0.0111	0.0102	0.0097	0.0079	0.0072	0.0068	0.0064	0.0063	0.0057	0.0056

ingredientes alvo não sejam os mesmos, existem ingredientes candidatos em comum em ambas as listas. Tal fato pode ser facilmente explicado em decorrência desses ingredientes candidatos (como *oil*, *vanilla*, e *baking soda*) ocorrerem com alta frequência em outras receitas do Domínio Alvo D_t. Em uma análise dos *Top-10* ingredientes candidatos, podemos verificar que o filtro VSM foi capaz de identificar dois bons ingredientes substitutos aos ingredientes alvo, de acordo com o contexto e tipos de preparos das receitas alvo.

Como forma de melhorar as recomendações geradas no Domínio Alvo, o filtro VRI é aplicado ao D_t, gerando os resultados apresentados na Tabela 4. Uma única lista de valores é gerada, uma vez que os resultados para este filtro independem da receita alvo escolhida. Cada ingrediente que ocorreu ao menos em uma receita em D_t é considerado como fazendo parte deste domínio, e, desta forma, todos os 3544 ingredientes neste domínio (vide Tabela 1) retornarão **Verdadeiro** em relação ao filtro PDA.

Os resultados finais obtidos pelos sistema de filtragem proposto (ou seja, os valores do EFI) são apresentados na Tabela 5. Como podemos observar, ao menos um ingrediente candidato adequado é apresentado para cada uma das receitas alvo. Pode-se observar ainda que, após a execução de cada um dos filtros propostos, nenhum ingrediente que infringiria a restrição do domínio alvo (receitas livres de lactose) é apresentado, mostrando a efetividade da proposta. Com a combinação dos filtros realizada, uma nova substituição adequada foi incluída à lista *Top-10* da receita *Quick Pound Cake*, resultando em três sugestões aceitáveis, o que pode ser considerável bastante promissor, uma vez que todo o processo de recomendação e filtragem empregado pelo sistema proposto ocorre de forma automática.

Em uma análise geral, o sistema proposto mostrou-se capaz de encontrar ao menos um ingrediente candidato adequado para as receitas alvo, respondendo de forma satisfatória à questão de pesquisa 1. De modo semelhante, o processo de filtragem proposto

Tabela 5. Resultados experimentais *Top-10* para o Escore Final do Ingrediente (EFI) em relação às receitas alvo. Ingredientes considerados bons substitutos em relação à receita alvo estão em negrito.

Receita (Ingrediente Alvo)	Ingrediente Candidato	EFI	Receita (Ingrediente Alvo)	Ingrediente Candidato	EFI
Quick Pound Cake (butter)	*baking powder*	0.0139	*Whole Wheat Corn Bread (milk)*	*baking soda*	0.0136
	salt	0.0138		***water***	**0.0127**
	baking soda	0.0132		*cinnamon*	0.0103
	water	0.0116		*flour*	0.0093
	vanilla	0.0100		*olive oil*	0.0089
	cinnamon	0.0097		*vanilla*	0.0085
	vanilla extract	0.0081		*vanilla extract*	0.0077
	oil	**0.0080**		***soymilk***	**0.0074**
	olive oil	**0.0078**		*oil*	0.0070
	vegetable oil	**0.0074**		*honey*	0.0069

tende a fazer recomendações de ingredientes existentes apenas no domínio alvo, o que dificulta a ocorrência de ingredientes que não se adequariam às restrições dos usuários, cumprindo, desta forma, os requisitos da questão de pesquisa 2. Por fim, a combinação dos filtros propostos para a composição do sistema final resultou em um sistema ainda mais preciso do que o uso dos filtros individuais, gerando recomendações mais adequadas no contexto de cada uma das receitas avaliadas (questão de pesquisa 3).

4. Conclusões

Neste trabalho, um sistema de recomendação de ingredientes substituitos é proposto para a geração de receitas para usuários com restrições alimentares. Através de um processo automático de filtragem e pela análise da lista de ingredientes das receitas, o sistema proposto é capaz de oferecer ingredientes que sejam bons candidados à substituição de um ingrediente causador de uma restrição em uma receita (ex.: alérgenos, ou ingredientes que não sejam saudáveis ao usuário por algum motivo) por um ingrediente que não cause essa restrição, expandindo o leque de opções de usuários que possuem algum tipo de dieta restritiva, sem comprometer a saúde desses usuários.

O sistema proposto é avaliado de acordo com uma análise qualitativa. Para isso, três questões de pesquisa são formuladas, e uma análise da adequabilidade ao contexto das receitas testadas dos *Top-10* melhores ingredientes retornados por cada etapa do sistema de filtragem proposto, assim como do sistema completo, é realizada. Os resultados experimentais indicam que o sistema proposto é capaz de gerar boas recomendações de ingredientes substitutos, adequados às receitas originais, e que não infrinjam as restrições dos usuários.

Como tendências para trabalhos futuros, pretende-se melhorar a análise da relevância dos ingredientes em cada um dos domínios de receitas estudados, de modo a evitar que ingredientes muito frequentes, porém pouco relevantes no domínio, acabem tendo grande impacto nas recomendações, gerando, desta forma, recomendações mais precisas e adequadas ao contexto das receitas nesses domínios. Também pretendemos expandir a análise proposta pelo acréscimo de mais receitas exemplo, assim como de receitas em diferentes domínios alvo. Uma avaliação mais detalhada da ponderação dos filtros do sistema proposto será realizada automaticamente (por métodos como as meta-heurísticas

de busca), visando a elaboração de sistemas mais otimizados e precisos. Por fim, pretendemos analisar a elaboração de abordagens que façam a substituição de múltiplos ingredientes simultaneamente.

Agradecimentos

Os autores gostariam de agradecer à FACEPE, ao CNPq e a CAPES pelo suporte financeiro.

Referências

Britto, L. F. S., Pacifico, L. D. S., and Ludermir, T. B. (2020). Geração automática de receitas culinárias para pessoas com restrições alimentares. *Revista Eletrônica de Iniciação Científica em Computação*, 18(3).

Jiang, H., Wang, W., Liu, M., Nie, L., Duan, L.-Y., and Xu, C. (2019). Market2dish: A health-aware food recommendation system. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2188–2190.

Majumder, B. P., Li, S., Ni, J., and McAuley, J. (2019). Generating personalized recipes from historical user preferences. *arXiv preprint arXiv:1909.00105*.

Nirmal, I., Caldera, A., and Bandara, R. D. (2018). Optimization framework for flavour and nutrition balanced recipe: A data driven approach. In *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pages 1–9. IEEE.

Oliveira, E. G., Britto, L. F. S., Pacifico, L. D. S., and Ludermir, T. B. (2019). Recipe recommendation and generation based on ingredient substitution. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019)*, 1:238–249.

Ooi, A., Iiba, T., and Takano, K. (2015). Ingredient substitute recommendation for allergy-safe cooking based on food context. In *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pages 444–449. IEEE.

Pecune, F., Callebert, L., and Marsella, S. (2020). A socially-aware conversational recommender system for personalized recipe recommendations. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, pages 78–86.

Shino, N., Yamanishi, R., and Fukumoto, J. (2016). Recommendation system for alternative-ingredients based on co-occurrence relation on recipe database and the ingredient category. In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 173–178. IEEE.

Twomey, N., Fain, M., Ponikar, A., and Sarraf, N. (2020). Towards multi-language recipe personalisation and recommendation. In *Fourteenth ACM Conference on Recommender Systems*, pages 708–713.