

Recipe Recommendation and Generation Based on Ingredient Substitution

Emilia G. Oliveira¹, Larissa F. S. Britto¹, Luciano D. S. Pacifico¹, Teresa B. Ludermit²

¹Departamento de Informática (DC) – Universidade Federal do Rural de Pernambuco (UFRPE)
Recife – PE – Brazil

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

{emilia.galdino,larissa.feliciana, luciano.pacifico}@ufrpe.br, tbl@cin.ufpe.br

Abstract. *Nowadays, even with the increasing number of recipe sharing websites and systems, users may have difficult to search specific dishes through the massive amount of data contained in such repositories. Also, finding recipes which best fit a handy set of ingredients, while at the same time contemplate some user wishes and restrictions, may become a very time consuming or even impossible task. In this work, we propose a new recipe recommendation and generation system, based on the substitution of recipe ingredients and a data-driven approach, in an attempt to help users finding a recipe that contemplates both their desires and food restrictions, avoiding food wastes.*

Resumo. *Atualmente, mesmo com o aumento no número de páginas web e sistemas de compartilhamento de receitas, usuários podem ter dificuldade na busca por pratos específicos através da enorme quantidade de dados contidas nesses repositórios. Encontrar receitas que se adequem a um conjunto de ingredientes em mãos, contemplando as vontades e restrições desses usuários, pode ser uma tarefa demorada ou mesmo impossível. Neste trabalho, um sistema de recomendação e geração de receitas é proposto, baseado na substituição de ingredientes das receitas e em uma abordagem focada nos dados, em uma tentativa de ajudar os usuários a encontrarem receitas que contemplem tanto seus desejos, quanto suas restrições alimentares, evitando desperdícios.*

1. Introdução

A presença de conteúdo relacionado à alimentação na Internet tem aumentado consideravelmente em decorrência do avanço das Tecnologias da Informação (como mídias sociais, blogs, páginas especializadas, etc.). A culinária tornou-se bastante popular nos últimos anos, e qualquer pessoa pode aprender a preparar pratos que lhe satisfaçam sem sair de sua casa. Com o processo de acelerado de globalização, também tornou-se possível à aquisição de ingredientes relacionados às mais diversas culturas, povos, estilos culinários, dentre outros [Mokdara et al. 2018].

A escolha do prato que se deseja comer em um determinado momento do dia é um processo decisório complexo, que envolve vários fatores, como gostos pessoais, valor nutricional dos alimentos, restrições de dieta, alergias, etc. Outros fatores que também influenciam nessa decisão são a cor, textura (crocância) e temperatura da refeição [Nirmal et al. 2018].

Se por um lado a popularização das páginas web especializadas em receitas ajudaram a difundir a cultura do compartilhamento de receitas, por outro, a enorme quantidade de dados armazenados nesses repositórios torna a busca por uma receita que se adeque aos interesses dos usuários uma tarefa árdua. Nesse contexto de dados massivos, técnicas de Mineração de Dados, como os Sistemas de Recomendação, têm sido cada vez mais empregadas como ferramentas de suporte aos usuários na busca por informações úteis em seus contextos de aplicação.

Os Sistemas de Recomendação [Isinkaye et al. 2015] são técnicas de filtragem automática que reduzem a quantidade de dados recuperada por uma busca realizada por um usuário, tendo tais técnicas sido adaptadas em várias aplicações no contexto de culinária [Mokdara et al. 2018, Nirmal et al. 2018, Nezis et al. 2018, Gorbonos et al. 2018]. Os Sistemas de Recuperação de Receitas (SRRs) visam auxiliar seus usuários a encontrarem receitas e dietas que sejam balanceadas nutricionalmente e que atendam seus gostos pessoais.

Porém, mesmo com o auxílio dos SRRs, às vezes é difícil encontrar todos os ingredientes contidos na lista da receita recomendada, o que torna o preparo de um prato impossível. Existe ainda o problema de que na maioria dos casos, os repositórios existentes possuem poucas receitas que atendam públicos que possuam algum tipo de restrição alimentar, como alérgicos, veganos, vegetarianos, grupos religiosos que tenham restrições a certos tipos de alimentos (como bebidas alcoólicas), dentre outros [Ooi et al. 2015]. Uma forma de evitar esses problemas é a geração de receitas novas pela substituição de ingredientes da lista de ingredientes de receitas já existentes [Shino et al. 2016], porém o processo de substituição de ingredientes é bastante complexo, requerendo conhecimento especializado para que a receita não perca seus sabores ou valores nutricionais.

Neste trabalho, um sistema de recomendação e geração de receitas é proposto, baseado na substituição de um único ingrediente da lista de ingredientes de uma receita já existente. O sistema proposto faz uso de uma abordagem focada nos dados (*data-driven approach*), empregada através de um sistema de filtragem que acessa as relações inerentes entre receitas e seus ingredientes, de modo a oferecer melhores recomendações de ingredientes candidatos à substituição de um ingrediente escolhido. O sistema proposto avalia tanto a similaridade entre ingredientes e receitas, quanto às relações de coocorrência entre pares de ingredientes nas receitas. Com o intuito de popularizar o uso dessas ferramentas no Brasil, uma base de dados nova também é proposta, obtida através do emprego de técnicas de *web scraping* em páginas web nacionais, e pelo emprego de modelos de análise de textos em Português Brasileiro. O trabalho está organizado da seguinte forma: uma breve discussão dos trabalhos relacionados é apresentado na Seção 2, seguida pela apresentação da metodologia proposta (Seção 3). Em seguida, uma abordagem qualitativa é empregada para a análise do sistema proposto (Seção 4), seguida pela apresentação das conclusões e linhas para pesquisas futuras (Seção 5).

2. Trabalhos Relacionados

Uma das primeiras tentativas no que diz respeito ao desenvolvimento de sistemas computacionais de auxílio ao planejamento de dietas foi proposta por [Hammond 1986]. O sistema proposto (CHEF) simulava o processo cognitivo e decisório humano no preparo de receitas da culinária Szechwan.

Atualmente, existem três principais abordagens para a recomendação de receitas: a aplicação de técnicas tradicionais de recuperação da informação ao contexto de recomendação de alimentos; a recomendação de receitas baseada na análise ontológica; e a análise contextual das correlações entre ingredientes e suas informações nutricionais.

No contexto de modelos tradicionais de recomendação, as técnicas de Filtragem Baseada em Conteúdo e Filtragem Colaborativa foram adotadas por [Freyne and Berkovsky 2010], onde os autores analisaram ambas as técnicas e uma versão híbrida das mesmas para a recomendação de alimentos. A Navegação Social [Svensson et al. 2005] e a Análise de Sentimentos [Trattner and Elsweiler 2017], que são técnicas populares de análise de textos, também foram empregadas na recomendação de receitas. Em [Trattner and Elsweiler 2017], a Análise de Sentimentos é empregada para avaliar o quão saudáveis as receitas recomendadas pelos usuários eram. O trabalho revelou que os usuários tendem a recomendar receitas principalmente com base nos sabores finais apresentados, o que geralmente acarreta em recomendação de receitas pouco saudáveis. Em [Mokdara et al. 2018], um modelo temporal, que faz uso de uma Rede Neural Profunda, foi empregado para a recomendação de comida tailandesa baseada no comportamento individual dos usuários, em suas preferências e em seu histórico de consumo de alimentos. Em [Gorbonos et al. 2018], um sistema baseado em Redes Neurais Artificiais e Fatorização Matricial é usado para ajudar usuários à encontrarem receitas que satisfaçam um determinado conjunto de ingredientes já possuídos.

A categorização dos alimentos e dos ingredientes, da mesma forma que suas relações, têm sido adotadas em vários trabalhos na recomendação e geração de receitas. Em [Shino et al. 2016], um sistema de recomendação para a substituição de um único ingrediente em receitas é proposto, baseado na análise da frequência de coocorrência dos ingredientes na base de receitas. Os autores também empregaram um método de categorização dos ingredientes pelo uso de uma ontologia culinária. Em [Ooi et al. 2015], um método de substituição de ingredientes alergênicos é desenvolvido pela análise do contexto dos alimentos, obtido através da similaridade entre receitas. A medida de avaliação de similaridade empregada examina três grupos de metadados sobre as receitas: metadados relacionados aos ingredientes, metadados relacionados ao equipamento usado no preparo da receita e metadados relacionados ao contexto do alimento (como por exemplo, a cor e textura do preparo, o papel do ingrediente alergênico no prato, etc.). Em [Nirmal et al. 2018], um sistema de recomendação é proposto para a geração de receitas pela substituição de um único ingrediente, empregando tal sistema técnicas de otimização tanto do sabor, quanto do valor nutricional dos ingredientes nas recomendações geradas. O sistema proposto faz uso de uma abordagem de filtragem em três etapas para a otimização da similaridade dos sabores entre os ingredientes pertencentes à lista de candidatos à substituição e o ingrediente alvo, assim como também emprega uma otimização dos valores nutricionais da receita a ser gerada, levando-se em consideração a substituição do ingrediente alvo por um ingrediente da lista de candidatos.

3. Metodologia

A metodologia adotada é apresentada na Fig. 1. O sistema proposto é dividido em duas etapas principais: a aquisição da base de dados (Seção 3.1) e a geração de receitas pela substituição de um único ingrediente (Seção 3.2), em uma abordagem inspirada nos trabalhos desenvolvidos por [Ooi et al. 2015, Shino et al. 2016, Nirmal et al. 2018]. Cada

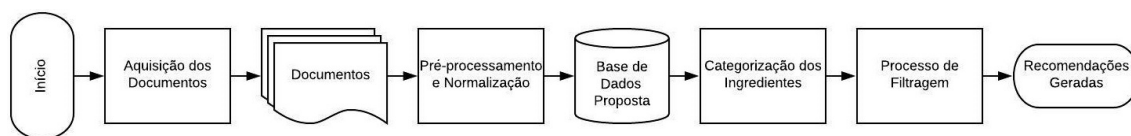


Figura 1. O Sistema de Geração de Receitas Proposto.

etapa será descrita em detalhes na sequência.

3.1. Base de Dados Proposta

Neste trabalho, uma nova base de dados é proposta. Como uma receita é um documento textual [Su et al. 2014, Jayaraman et al. 2017, Kalajdziski et al. 2018, Nirmal et al. 2018], técnicas de análise de documentos escritos em Português Brasileiro foram empregadas na obtenção da base de dados.

Através da aplicação da técnica de *web scraping* a páginas web brasileiras especializadas em receitas culinárias, um total de 4448 documentos foram obtidos. Após a etapa de aquisição dos documentos, os processos de pré-processamento e extração de características foram executados.

O pré-processamento foi iniciado pela remoção de redundâncias nos documentos adquiridos. A remoção foi realizada pela comparação simples dos links para as páginas dos documentos, seguida da remoção de ocorrências redundantes. Como o interesse do presente estudo está focado na elaboração de um sistema de recomendação em Português, documentos escritos em outros idiomas foram removidos. De modo semelhante, optou-se pela exclusão de documentos que continham mais de uma receita, o que resultou na elaboração de uma base de dados contendo 3106 documentos (*receitas*).

Dado que existem poucas ferramentas para a análise automática de textos em Português Brasileiro (como o spaCy¹ e o NLTK²), e que as ferramentas existente são bastante limitadas e passíveis a erros [de Souza et al. 2018], optamos pelo desenvolvimento de uma ferramenta própria para a análise das receitas obtidas. Como a abordagem atual foca apenas na análise da lista de ingredientes de cada receita, dois *lexicons* foram desenvolvidos, o primeiro para a remoção de adjetivos (como *picado*, *quente*, etc.), e o segundo para a remoção de quantidades (como *xícara*, *kg*, valores numéricos, etc.) associadas aos ingredientes. Desta forma, cada documento (*receita*) foi reduzido a uma lista de seus ingredientes, e, de modo semelhante, uma lista contendo todos os ingredientes da base de dados foi formada.

Após a obtenção da lista composta por todos os ingredientes da base de dados, uma etapa de normalização dos dados foi realizada. A normalização faz-se necessária, uma vez que as receitas originais podem conter inúmeros problemas, como palavras escritas incorretamente, erros de digitação, abreviações, ingredientes que possuem várias nomenclaturas, dentre outros problemas [Yasukawa and Scholer 2017]. Após a normalização, a lista final obtida contém 1312 ingredientes diferentes. Cada receita da base de dados é então convertida para uma representação vetorial binária, que indica a presença de um determinado ingrediente na receita (1) ou sua ausência (0), obtendo-se a matrix binária

¹<https://spacy.io/models/pt>

²<https://www.nltk.org/>

de termos da base de dados. Em seguida, a técnica de TF-IDF [Manning et al. 2010] é aplicada à matrix de termos para a obtenção da relevância de cada ingrediente em cada receita [Jayaraman et al. 2017].

Com a obtenção da base de dados, a categorização dos ingredientes é realizada. Na abordagem proposta, os ingredientes foram distribuídos em 11 (onze) categorias: 0-molhos e temperos, 1-leite e derivados, 2-carnes, 3-peixes e frutos do mar, 4-vegetais e cogumelos, 5-frutas, 6-cereais, oleagenos e raízes, 7-alimentos processados e industrializados, 8-bebidas alcoólicas, 9-outras bebidas, 10-outros. Essa categorização visa agrupar ingredientes que podem vir a desrespeitar alguma restrição alimentar dos usuários (como alérgenos, carnes, bebidas alcoólicas, entre outros).

3.2. Geração de Receitas pela Substituição de Um Único Ingrediente

Com a base de dados pronta e os ingredientes categorizados, o sistema proposto é desenvolvido pela aplicação de um processo de filtragem em busca da obtenção da melhor lista de substitutos para um determinado ingrediente x_r em uma receita alvo R_c . O processo de filtragem adotado é descrito na sequência (da Seção 3.2.1 à Seção 3.2.4).

3.2.1. Filtragem por Similaridade Acumulada do Ingrediente (SAI)

O filtro de Similaridade Acumulada do Ingrediente visa encontrar os ingredientes que estão mais proximamente relacionados à receita R_c , através de um processo inspirado no sistema de escores de ingredientes proposto em [Ooi et al. 2015].

Considere a base de dados S que contém todas as receitas R_j ($j = 1, 2, \dots, n$, onde n é o número de receitas em S). A similaridade entre R_c e cada uma das demais receitas em S é calculada usando a Medida do Cosseno (Eq. (1)).

$$sim(R_c, R_j) = \frac{R_c \bullet R_j}{|R_c||R_j|} \quad (1)$$

Uma porcentagem das receitas mais similares à R_c (neste trabalho, 50%) é selecionada para compor um conjunto S_{best} . Para cada ingrediente x_i , seu escore em relação à receita $R_j \in S_{best}$ é determinado pela Eq. (2).

$$score(x_i) = \begin{cases} sim(R_c, R_j), & \text{se } x_i \in R_j \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

A Similaridade Acumulada do Ingrediente x_i (SAI) é dada por (Eq. (3)):

$$SAI(x_i) = \sum_{\forall R_j \in S_{best}} score(x_i) \quad (3)$$

3.2.2. Filtragem por Valor de Suporte Médio (*VSM*) e por Lista de Ingredientes Proibidos (*LIP*)

Para a avaliação da probabilidade de coocorrência pareada entre ingredientes candidatos à substituição de x_r e os demais ingredientes na receita alvo R_c , o Valor de Suporte Médio (*Average Support Value*) proposto por [Nirmal et al. 2018] é adaptado da seguinte forma: primeiramente, o valor da coocorrência pareada $C_v(x_i, x_l)$ entre cada ingrediente $x_i \notin R_c$ da base de dados e todos os ingredientes $x_l \in R_c (l \neq r)$ é calculado considerando toda a base de dados S ; em seguida, o *VSM* de cada ingrediente candidato x_i é obtido pela Eq. (4):

$$VSM(x_i) = \sum_{x_l \in R_c (l \neq r)} C_v(x_i, x_l) / (|R_c| - 1) \quad (4)$$

Como discutido em [Nirmal et al. 2018], a ocorrência de $C_v(x_i, x_l) = 0$ para qualquer ingrediente candidato x_i pode ser um indicativo de que a combinação entre x_i e x_l não ocorre com muita frequência, o que pode indicar que tal combinação pode ser tóxica ou não ser agradável ao paladar. Desta forma, uma Lista de Ingredientes Proibidos (*LIP*) é formada com essa informação, sendo tal lista aplicada como um novo filtro. Quando dois ingredientes possuem alto valor de coocorrência em uma base de receitas, há um forte indicativo de que sua combinação pode ser gostosa ou nutricionalmente balanceada.

3.2.3. Filtragem pela Categoria do Ingrediente

Os últimos filtros são obtidos pela análise da categoria dos ingredientes. O primeiro filtro é aplicado pela atribuição de um valor de escore aos ingredientes candidatos x_i que pertencem à categoria c_r do ingrediente a ser substituído x_r (Eq. (5)).

$$c_{score}(x_i) = \begin{cases} 1, & \text{se } x_i \in c_r \\ 0, & \text{caso contrário} \end{cases} \quad (5)$$

Um segundo filtro é obtido pelo cálculo da probabilidade de ocorrência de um ingrediente $x_i \in c_r$ considerando todas as receitas de S (Eq. (6)).

$$p_{score}(x_i) = \begin{cases} \sum_{x_i \in R_j} \text{TF-IDF}_{score}(x_i) / |S|, & \text{se } x_i \in c_r \\ 0, & \text{caso contrário} \end{cases} \quad (6)$$

Ambos os filtros são implementados baseados na ideia de que ingredientes em uma mesma categoria são mais prováveis de serem bons substitutos uns aos outros do que ingredientes pertencentes a categorias diferentes [Shino et al. 2016].

3.2.4. Escore Total do Ingrediente (*ETI*)

Os valores obtidos para *SAI*, *VSM*, c_{score} e p_{score} são normalizados na faixa $[0, 1]$, para evitar que algum filtro domine os demais [Nirmal et al. 2018]. O Escore Total do

Tabela 1. Receitas Seleccionadas e seus Ingredientes.

Receita	Ingrediente	Categoria
<i>Cestinha de Carne Seca</i> (ID#1145)	<i>requeijão</i>	#1
	<i>carne seca</i>	#2
	<i>escarola</i>	#4
	<i>alho</i>	#0
	<i>cebola</i>	#4
	<i>massa de pastel</i>	#6
<i>Brownie de Capuccino</i> (ID#1869)	<i>manteiga</i>	#1
	<i>canela em pó</i>	#0
	<i>chocolate</i>	#7
	<i>ovo</i>	#10
	<i>café solúvel</i>	#9
	<i>leite</i>	#1
	<i>chocolate</i>	#7
	<i>açúcar</i>	#10
<i>farinha de trigo</i>	#6	

Ingrediente (ETI) é então calculado como uma combinação dos demais filtros (Eq. (7)).

$$ETI(x_i) = \frac{SAI(x_i) + VSM(x_i) + c_{score}(x_i) + p_{score}(x_i)}{4} \quad (7)$$

Para a lista de ingredientes candidatos formada, um valor alto do ETI indica que o ingrediente x_i é mais adequado para a substituição do ingrediente x_r . Ingredientes pertencentes à LIP (descrita na Seção 3.2.2) são removidos diretamente da lista final de melhores ingredientes candidatos.

4. Experimentos

Como forma de avaliação do sistema proposto, uma análise qualitativa é realizada nesta seção. Os experimentos visam a satisfação de três perguntas principais: 1-os ingredientes substitutos gerados pelo sistema são adequados? 2-ingredientes na mesma categoria do ingrediente a ser substituído representam boas substituições? 3-a combinação de filtros proposta é adequada? A principal suposição do trabalho é a de que ingredientes candidatos na mesma categoria do ingrediente a ser substituído representam uma melhor alternativa para a substituição [Ooi et al. 2015].

A avaliação do sistema proposto levará em consideração duas receitas exemplo seleccionadas da base de dados proposta. De modo semelhante, três ingredientes de cada uma das receitas serão escolhidos para exemplificar a execução do sistema, porém um único ingrediente será substituído por vez na receita original. A lista de ingredientes de cada uma das receitas é apresentada na Tabela 1. Os cinco ingredientes com maior score em relação a cada um dos filtros propostos (exceto para o c_{score} , dado que todos os ingredientes na categoria do ingrediente a ser substituído possuem o mesmo valor) serão apresentados na discussão.

O resultado da Similaridade Acumulada do Ingrediente (SAI) não depende da escolha do ingrediente a ser substituído, sendo calculado para cada receita alvo como um todo. Os valores normalizados para o SAI para as receitas seleccionadas são apresentados

Tabela 2. Resultados da Similaridade Acumulada do Ingrediente (*SAI*) para as Receitas Seleccionadas. Ing.: Ingrediente Candidato; Cat.: Categoria do Ingrediente Candidato.

Receita	Ingr.	Cat.	<i>SAI</i>	Receita	Ingr.	Cat.	<i>SAI</i>
ID#1145	<i>queijo mussarela</i>	#1	1.000	ID#1869	<i>fermento</i>	#10	1.000
	<i>queijo parmesão</i>	#1	0.905		<i>leite condensado</i>	#1	0.836
	<i>tomate</i>	#4	0.892		<i>gema de ovo</i>	#10	0.563
	<i>salsa</i>	#0	0.831		<i>óleo</i>	#10	0.562
	<i>azeite de oliva</i>	#10	0.814		<i>creme de leite</i>	#1	0.531
	<i>cheiro verde</i>	#0	0.716		<i>essência de baunilha</i>	#10	0.461
	<i>bacon</i>	#2	0.693		<i>clara de ovo</i>	#10	0.451
	<i>molho de tomate</i>	#0	0.652		<i>água</i>	#10	0.411
	<i>creme de leite</i>	#1	0.523		<i>doce de leite</i>	#1	0.336
	<i>leite</i>	#1	0.521		<i>chocolate amargo</i>	#7	0.348

na Tabela 2. Dado que as duas receitas são bastantes dissimilares (uma é um *salgado*, e a outra é um *doce*, além de não haver ingrediente presente em ambas as receitas ao mesmo tempo), os resultados obtidos para o *SAI* são, da mesma forma, dissimilares. Dos resultados, podemos observar que a base de dados proposta é bastante desbalanceada em termos de frequência dos ingredientes. Por exemplo, embora não haja ingredientes da Categoria #10 na receita de *Cestinha de Carne Seca*, o *azeite de oliva* atingiu um alto valor de similaridade em relação a essa receita. Mais uma vez, tais resultados também eram esperados, uma vez que alguns ingredientes (como *azeite de oliva*, *sal*, *cebola* e *alho*) são associados a várias categorias diferentes de receitas, tornando suas frequências de ocorrência muito elevadas.

Os resultados para o *VSM* (Tabela 3) reforçam o desbalanceamento da base em relação aos ingredientes. *Sal*, *água* e *azeite de oliva* estão presentes em muitas receitas, resultando em altos valores do *VSM* em ambas as receitas em análise. Embora as receitas seleccionadas sejam bastante diferentes, seus ingredientes coocorreram pareadamente de forma semelhante com ingredientes de outras receitas, resultado em listas finais semelhantes em termos do *VSM*.

A Tabela 4 apresenta os resultados para o p_{score} . Podemos observar que os ingredientes candidatos com maiores probabilidades nas categorias representam bons substitutos aos ingredientes alvo seleccionados. Porém, alguns ingredientes são de difícil substituição, como o *açúcar*, que está presente em um número elevado de receitas da base, o que faz com que ingredientes que representariam boas opções para sua substituição ocorram com baixa frequência em sua categoria.

Os resultados finais obtidos pelo sistema proposto são apresentados na Tabela 5. Os valores do *ETI* para cada ingrediente exemplo são apresentados para a lista final de ingredientes substitutos (após a eliminação dos ingredientes contidos na *LIP*). Os resultados finais mostram a capacidade do sistema de filtragem proposto de oferecer boas soluções, em termos de ingredientes substitutos, na lista dos cinco melhores ingredientes candidatos. O *ETI* também se mostrou útil na avaliação de quão bom substituto um ingrediente candidato é em relação ao ingrediente a ser substituído, tendo em vista que boas substituições apresentaram altos valores de *ETI*, enquanto substituições menos adequadas resultaram em baixos valores nessa métrica. Por exemplo, *farinha de trigo* obteve

Tabela 3. Resultados do Valor de Suporte Médio (VSM) dos Cinco Melhores Ingredientes Candidatos em Relação à cada Ingrediente a Ser Substituído. Ing. Orig.: Ingrediente Original (a Ser Substituído); Ing. Cand.: Ingrediente Candidato à Substituição; Cat. Cand.: Categoria do Ingrediente Candidato.

<i>Cestinha de Carne Seca (ID#1145)</i>				<i>Brownie de Capuccino (ID#1869)</i>			
Ing. Orig.	Ing. Cand.	Cat. Cand.	VSM	Ing. Orig.	Ing. Cand.	Cat. Cand.	VSM
<i>carne seca</i>	<i>sal</i>	#0	1.000	<i>manteiga</i>	<i>sal</i>	#10	1.000
	<i>azeite de oliva</i>	#10	0.776		<i>água</i>	#10	0.612
	<i>tomate</i>	#4	0.515		<i>óleo</i>	#10	0.605
	<i>salsa</i>	#0	0.460		<i>fermento</i>	#10	0.569
	<i>água</i>	#10	0.402		<i>gema de ovo</i>	#10	0.432
<i>escarola</i>	<i>sal</i>	#0	1.000	<i>café solúvel</i>	<i>sal</i>	#0	1.000
	<i>azeite de oliva</i>	#10	0.7747		<i>cebola</i>	#4	0.587
	<i>tomate</i>	#4	0.518		<i>água</i>	#10	0.571
	<i>salsa</i>	#0	0.463		<i>óleo</i>	#10	0.492
	<i>água</i>	#10	0.403		<i>azeite de oliva</i>	#10	0.470
<i>massa de pastel</i>	<i>sal</i>	#0	1.000	<i>açúcar</i>	<i>sal</i>	#0	1.000
	<i>azeite de oliva</i>	#10	0.776		<i>água</i>	#10	0.516
	<i>tomato</i>	#4	0.519		<i>óleo</i>	#10	0.477
	<i>salsa</i>	#0	0.462		<i>fermento</i>	#10	0.437
	<i>água</i>	#10	0.402		<i>creme de leite</i>	#1	0.389

Tabela 4. Resultados da Probabilidade do Ingrediente (p_{score}) dos Cinco Melhores Ingredientes em Relação à cada Ingrediente a Ser Substituído. Ing. Orig.: Ingrediente Original (a Ser Substituído); Ing. Cand.: Ingrediente Candidato à Substituição.

<i>Cestinha de Carne Seca (ID#1145)</i>			<i>Brownie de Capuccino (ID#1869)</i>		
Ing. Orig.	Ing. Cand.	p_{score}	Ing. Orig.	Ing. Cand.	p_{score}
<i>carne seca</i>	<i>bacon</i>	1.000	<i>manteiga</i>	<i>creme de leite</i>	1.000
	<i>carne bovina</i>	0.798		<i>leite condensado</i>	0.868
	<i>peito de frango</i>	0.664		<i>queijo parmesão</i>	0.812
	<i>galinha</i>	0.604		<i>queijo mussarela</i>	0.581
	<i>calabresa</i>	0.601		<i>requeijão</i>	0.464
<i>escarola</i>	<i>tomate</i>	1.000	<i>café solúvel</i>	<i>café</i>	1.000
	<i>cenoura</i>	0.784		<i>suco de maracujá</i>	0.744
	<i>batata</i>	0.677		<i>refrigerante de guaraná</i>	0.733
	<i>pimentão vermelho</i>	0.455		<i>chocolate ao leite</i>	0.593
	<i>cebola roxa</i>	0.443		<i>suco de limão</i>	0.481
<i>massa de pastel</i>	<i>farinha de trigo</i>	1.000	<i>açúcar</i>	<i>água</i>	1.000
	<i>arroz</i>	0.449		<i>azeite de oliva</i>	0.926
	<i>amido de milho</i>	0.439		<i>óleo</i>	0.863
	<i>coco</i>	0.426		<i>fermento</i>	0.704
	<i>farinha de rosca</i>	0.372		<i>gema de ovo</i>	0.699

Tabela 5. Resultados do Escore Total do Ingrediente (ETI) para os Cinco Melhores Ingredientes Candidatos em Relação à cada Ingrediente a Ser Substituído. Os Ingredientes na Lista de Ingredientes Proibidos (LIP) foram Removidos Automaticamente pelo Processo de Filtragem. Ing. Orig.: Ingrediente Original (a Ser Substituído); Ing. Cand.: Ingrediente Candidato à Substituição; Cat. Cand.: Categoria do Ingrediente Candidato.

<i>Cestinha de Carne Seca</i> (ID#1145)				<i>Brownie de Capuccino</i> (ID#1869)			
Ing. Orig.	Ing. Cand.	Cat. Cand.	ETI	Ing. Orig.	Ing. Cand.	Cat. Cand.	ETI
<i>carne seca</i>	<i>carne bovina</i>	#2	0.573	<i>manteiga</i>	<i>leite condensado</i>	#1	0.771
	<i>presunto</i>	#2	0.462		<i>creme de leite</i>	#1	0.739
	<i>azeite de oliva</i>	#10	0.398		<i>margarina</i>	#1	0.446
	<i>sal</i>	#0	0.378		<i>fermento</i>	#10	0.392
	<i>tomate</i>	#4	0.352		<i>sal</i>	#10	0.298
<i>escarola</i>	<i>tomate</i>	#4	0.852	<i>café solúvel</i>	<i>café</i>	#9	0.512
	<i>pimentão vermelho</i>	#4	0.440		<i>fermento</i>	#10	0.367
	<i>pimentão verde</i>	#4	0.405		<i>sal</i>	#0	0.298
	<i>azeite de oliva</i>	#10	0.397		<i>leite condensado</i>	#1	0.290
	<i>sal</i>	#0	0.378		<i>óleo</i>	#10	0.264
<i>massa de pastel</i>	<i>farinha de trigo</i>	#6	0.684	<i>açúcar</i>	<i>fermento</i>	#10	0.785
	<i>arroz</i>	#6	0.491		<i>água</i>	#10	0.732
	<i>amido de milho</i>	#6	0.448		<i>óleo</i>	#10	0.726
	<i>farinha de rosca</i>	#6	0.419		<i>gema de ovo</i>	#10	0.650
	<i>milho</i>	#6	0.409		<i>clara de ovo</i>	#10	0.535

um *ETI* de 0.684 na substituição da *massa de pastel*, enquanto o *arroz* obteve um valor de 0.491 considerando o mesmo ingrediente exemplo, indicando que tratar-se de uma substituição menos adequada na prática. Os resultados finais corroboram a suposição inicial de que ingredientes em uma mesma categoria são mais adequados para a substituição uns dos outros do que ingredientes em categorias diferentes.

5. Conclusões

Neste trabalho, um sistema de recomendação e geração de receitas é proposto. O sistema é capaz de realizar a recomendação de receitas quando uma lista de ingredientes é fornecida como entrada do mesmo, assim como também é capaz de gerar novas receitas automaticamente pela substituição de um único ingrediente de uma receita já existente, seguindo uma abordagem focada nos dados. Um processo de filtragem é proposto baseado na análise da categoria dos ingredientes, assim como de suas probabilidades de coocorrência pareadas nas receitas. Um esquema de categorização dos ingredientes é apresentado, visando prover melhores recomendações de ingredientes candidatos à substituição de um ingrediente pertencente a uma receita e que se deseja substituir. Uma base de dados formada por receitas em Português Brasileiro é proposta, composta por 3106 receitas, que apresentam um total de 1312 ingredientes diferentes em suas composições.

A avaliação do sistema proposto foi executada através de uma análise qualitativa, realizada por meio da escolha de duas receitas (*Cestinha de Carne Seca* e *Brownie de Capuccino*) como exemplo, havendo a tentativa de substituição de três ingredientes de cada uma dessas receitas (um ingrediente por vez). Em uma avaliação geral, os resultados finais obtidos pelo sistema podem ser considerados satisfatórios, tendo em vista que ao menos uma boa recomendação foi gerada na lista dos cinco melhores ingredientes

candidatos à substituição para cada ingrediente alvo, na maioria dos casos. Os experimentos mostraram que ingredientes pertencentes à categoria do ingrediente a ser substituído representam melhores candidatos à substituição que ingredientes de outras categorias. Também foi possível avaliar que alguns ingredientes são difíceis de substituir em uma receita (como, por exemplo, o *açúcar*), devido ao fato de tais ingredientes serem bastante representativos em suas categorias e aparecerem em um grande número de receitas, o que diminui a taxa de ocorrência de ingredientes que poderiam representar boas substituições para os mesmos.

Existem diversas linhas para trabalhos futuros relacionados à presente pesquisa. Primeiramente, pretendemos aumentar a base de dados através da busca de mais documentos em páginas web brasileiras. Uma base de dados maior alteraria de forma significativa as taxas de coocorrências pareadas entre ingredientes, da mesma forma que poderia agregar novos ingredientes à lista já existente, evitando a exclusão precipitada de ingredientes candidatos por seu acréscimo à Lista de Ingredientes Proibidos. Também pretende-se melhorar o sistema de filtragem proposto pelo desenvolvimento de uma Ontologia de Ingredientes em Português Brasileiro, de modo a realizar uma categorização e análise das relações entre os ingredientes de forma mais precisa. A ponderação dos filtros propostos também será alvo de uma análise mais aprofundada, de modo a se verificar a relevância e real impacto desses filtros nos resultados finais obtidos. Finalmente, uma análise nutricional e de sabores dos componentes de cada ingrediente em cada receita será realizada, como forma de aprimorar a qualidade das receitas a serem geradas pelo sistema.

Agradecimentos

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

Referências

- de Souza, J. G. R., de Paiva Oliveira, A., and Moreira, A. (2018). Development of a brazilian portuguese hotel’s reviews corpus. In *International Conference on Computational Processing of the Portuguese Language*, pages 353–361. Springer.
- Freyne, J. and Berkovsky, S. (2010). Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 321–324. ACM.
- Gorbonos, E., Liu, Y., and Hoàng, C. T. (2018). Nutrec: Nutrition oriented online recipe recommender. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 25–32. IEEE.
- Hammond, K. J. (1986). Chef: A model of case-based planning. In *AAAI*, pages 267–271.
- Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3):261–273.
- Jayaraman, S., Choudhury, T., and Kumar, P. (2017). Analysis of classification models based on cuisine prediction using machine learning. In *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, pages 1485–1490. IEEE.
- Kalajdziski, S., Radevski, G., Ivanoska, I., Trivodaliev, K., and Stojkoska, B. R. (2018). Cuisine classification using recipe’s ingredients. In *2018 41st International Conven-*

- tion on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pages 1074–1079. IEEE.
- Manning, C., Raghavan, P., and Schütze, H. (2010). Introduction to information retrieval. *Natural Language Engineering*, 16(1):100–103.
- Mokdara, T., Pusawiro, P., and Harnsomburana, J. (2018). Personalized food recommendation using deep neural network. In *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*, pages 1–4. IEEE.
- Nezis, A., Papageorgiou, H., Georgiadis, P., Jiskra, P., Pappas, D., and Pontiki, M. (2018). Towards a fully personalized food recommendation tool. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, page 77. ACM.
- 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.
- 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.
- 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.
- Su, H., Lin, T.-W., Li, C.-T., Shan, M.-K., and Chang, J. (2014). Automatic recipe cuisine classification by ingredients. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication*, pages 565–570. ACM.
- Svensson, M., Höök, K., and Cöster, R. (2005). Designing and evaluating kalas: A social navigation system for food recipes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(3):374–400.
- Trattner, C. and Elswiler, D. (2017). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*, pages 489–498. International World Wide Web Conferences Steering Committee.
- Yasukawa, M. and Scholer, F. (2017). Concurrence of word concepts in cooking recipe search. In *Proceedings of the 9th Workshop on Multimedia for Cooking and Eating Activities in conjunction with The 2017 International Joint Conference on Artificial Intelligence*, pages 25–30. ACM.