

# Uma proposta de método adaptativo para seleção automática de preenchimento de texto

Francisco de Faria

Laboratório de Tomada de Decisão  
Escola Politécnica da Universidade de São Paulo  
São Paulo, Brasil

João Pereira

Escola Politécnica da Universidade de São Paulo  
São Paulo, Brasil  
rajoma@usp.br

**Resumo**—*Esta pesquisa busca propor um método probabilístico para preenchimento automático de texto, com potencial aplicação em dispositivos computacionais como telefones celulares e computadores pessoais. O substrato formal utilizado como base na proposição do método é o autômato adaptativo, ao qual é associado o conceito frequentista de probabilidade, de forma a orientar o preenchimento de textos segundo uma dependência local de contexto. Esta dependência de contexto se traduz na probabilidade de reincidência de bigramas. Caso o contexto local não seja reconhecido no texto em curso de digitação, passa-se a uma análise livre de contexto baseada apenas na reincidência de palavras.*

**Palavras-chave**—*Preenchimento de texto; autômato adaptativo; probabilidade; bigrama; dependência do contexto.*

## I. INTRODUÇÃO

O conceito de adaptabilidade aplicado a autômatos finitos lhes confere a capacidade de se auto-modificar conforme as necessidades encontradas [1]. Essa metodologia adaptativa compreende-se de maneira suplementar à teoria dos autômatos [5].

O autômato adaptativo é uma máquina de estados e pode ser representado por um grafo direcionado, que inicialmente se apresenta como um autômato finito, porém, à medida que lhe são incutidas ações adaptativas durante o fluxo operacional, o grafo original pode ser modificado [3]. Estas ações adaptativas são geradas pela própria movimentação do autômato e resultam da aplicação de suas regras de transição. Se houver uma única regra aplicável à situação corrente, a próxima situação do dispositivo é determinada pela regra em questão, e caso haja mais de uma regra aderente à situação corrente do estado, as diversas situações seguintes são tratadas em paralelo. No caso de dispositivos voltados à sugestão de preenchimento de texto, baseados em análise léxico-morfológica, por exemplo, pode haver mais de uma regra a ser seguida, cada regra levando a uma cadeia diferente [2].

Os mecanismos adaptativos são formados por três tipos de ações adaptativas básicas: consulta (cheça um conjunto de regras), exclusão de regras e inclusão de regras. As ações básicas de inclusão e exclusão são baseadas em padrões pré-definidos [4]. A execução de uma transição adaptativa acarreta mudanças na topologia do autômato. Isto faz com que uma nova máquina de estados substitua a anterior, caracterizando

um novo passo do autômato no espaço das máquinas de estado [6].

O caminho em um espaço de máquinas de estado é gerado devido à máquina original não estar mais presente. Dessa forma, o autômato começa seu trabalho em uma máquina peculiar e vai continuá-lo em outra máquina, descrevendo uma trajetória em um espaço de trabalho que contempla todas as possibilidades de autômatos [4].

Devido a esta possibilidade de adaptação das cadeias às regras dentro do autômato, inúmeras pesquisas confirmam a viabilidade prática da utilização de autômatos adaptativos para preenchimento de texto não-estruturados [6].

### A. Visão do problema

Os editores de texto atuais requerem o desenvolvimento de programas que sejam capazes de determinar uma estrutura de sequência lógica, a fim de completar palavras de maneira simples, sem qualquer análise léxico-morfológica ou pragmática. O histórico de ocorrência de uma palavra em uma frase ou texto pode ser um indicador da probabilidade de sua reutilização. Na linguagem coloquial é comum a reutilização de termos, palavras, bigramas e até trigramas.

Atualmente é muito comum que dispositivos computacionais, sejam celulares ou computadores, se utilizem de algum tipo de software para efetuar o trabalho de complemento de texto ou indicação de sugestões para complementar o texto. Há softwares que sugerem o complemento de frases ou palavras, porém, não seguem uma regra lógica ou estatística, de forma que as sugestões não são ótimas e muitas vezes não aderem ao texto.

A motivação deste trabalho é tornar precisas as recomendações de complemento de texto, não apenas na área tecnológica, mas também dentro de outras áreas que exijam este tipo de comportamento. Com essa finalidade é proposto um método adaptativo para reconhecimento e seleção automática de preenchimento de palavras à medida que os caracteres do texto são inseridos.

O arcabouço sugerido é baseado em um autômato adaptativo que pode ser visto como uma extensão do formalismo do autômato de pilha estruturado usado para reconhecimento de linguagens. O dinamismo ou adaptabilidade neste caso está na capacidade do autômato modificar sua própria topologia a fim de reconhecer palavras, bigramas ou frases reincidentes. Neste contexto, o autômato pode ser

utilizado para sugerir complementos para palavras em curso de digitação em um celular ou computador pessoal.

### B. Objetivo

O objetivo deste trabalho é estudar a possibilidade de utilização de autômatos adaptativos para sugerir o preenchimento automático de palavras, segundo um método probabilístico e dependente de contexto baseado no histórico de ocorrências destas em textos anteriores.

Além do objetivo principal, pode-se ressaltar também a utilização da teoria de autômatos, retomando um caminho percorrido anteriormente pelos precursores na área de Inteligência Artificial, utilizando como base para o modelo de solução o autômato adaptativo no lugar do finito, já que propicia um tratamento mais adequado e uniforme às tratativas computacionais.

Um mecanismo baseado no arcabouço é desenvolvido a fim de efetuar testes de sugestão e as probabilidades de ocorrência de cada palavra em qualquer tipo de texto construído na etimologia latim/grega da língua portuguesa. Este mecanismo deve ser simulado em um computador, isto é, deve ser um mecanismo de software.

## II. MÉTODO PROPOSTO

O método proposto baseia-se nas características do autômato adaptativo, sendo que a criação de novas máquinas de estado é regida pela ocorrência prévia de uma palavra ou bigrama. A cada estado do autômato corresponde um conjunto de sugestões de complemento. As sugestões de complemento são associadas à frequência estatística com a qual aquele estado culmina no complemento em questão. Esta abordagem estatística tem por objetivo maximizar as chances de aceitação dos complementos sugeridos.

### A. Definição do método e evolução

O método proposto tem por finalidade:

- A sugestão de complementos para uma palavra em curso de digitação;
- O aprendizado de novas palavras e bigramas digitados.

O autômato inicial corresponde a um único estado de entrada, ao qual é associado um único estado de saída. À medida que textos são inseridos o autômato se adapta de forma a reconhecer palavras ou bigramas reincidentes. Quando um primeiro caractere é recebido, o autômato não o reconhece como entrada válida. Dessa maneira, uma vez que não há transições de estado associadas a este caractere, passa-se à fase de aprendizado, no qual novos estados são criados via ações adaptativas. Quando a primeira palavra é inserida por completo, o autômato conta com uma sequência de estados e transições capazes de reconhecê-la. O autômato já não se encontra no estado de entrada, mas sim em um novo estado que corresponde ao fim desta primeira palavra. A partir deste estado correspondente ao fim da primeira palavra registrada, o procedimento se repete até que a segunda palavra seja concluída e o autômato se torne capaz de reconhecer este primeiro bigrama inserido. Uma ação adaptativa desencadeada

pelo fim desta segunda palavra liga este último estado ao estado inicial. Volta-se então a este estado inicial e repete-se o procedimento para esta segunda palavra também passe a ser reconhecida. Quando uma terceira palavra, diferente das duas primeiras, for inserida o autômato será capaz de reconhecer não apenas três palavras, mas também dois bigramas, o primeiro bigrama formado pela primeira e pela segunda palavra, e o segundo bigrama pela segunda e pela terceira palavra.

A fim de visualizar a dinâmica de funcionamento do método proposto, vejamos como seria processada a seguinte frase quando digitada: Enquanto a pata bota os ovos, os patos observam a pata e os ovos. Vamos supor que o autômato acaba de ser criado, ou seja, ainda não reconhece palavra alguma. Quando as primeiras palavras são inseridas, nenhuma sugestão é apresentada, pois o autômato ainda não as reconhece.

1) Enquanto a pata bota os o\_

- $P(\text{os} | \text{o}_) = 1.0$

Observa-se que quando se digita pela segunda vez a letra "o", esta já é reconhecida como parte de uma palavra conhecida, "os".

2) Enquanto a pata bota os ovos, o\_

- $P(\text{os} | \text{o}_) = 0.5$
- $P(\text{ovos} | \text{o}_) = 0.5$

Quando se digita pela terceira vez a letra o, são sugeridas as palavras "os" e "ovos", ambas associadas a uma frequência de ocorrência igual a 50%, pois cada palavra apareceu apenas uma vez anteriormente.

3) Enquanto a pata bota os ovos, os p\_

- $P(\text{pata} | \text{p}_) = 1.0$

Já após digitar a letra "p" pela segunda vez, é sugerida apenas a palavra "pata", a única inserida anteriormente começando com "p".

4) Enquanto a pata bota os ovos, os pa\_

- $P(\text{pata} | \text{pa}_) = 1.0$

5) Enquanto a pata bota os ovos, os pat\_

- $P(\text{pata} | \text{pat}_) = 1.0$

6) Enquanto a pata bota os ovos, os patos o\_

- $P(\text{os} | \text{o}_) = 0.66$
- $P(\text{ovos} | \text{o}_) = 0.33$

7) Enquanto a pata bota os ovos, os patos observam a p\_

- $P(\text{pata} | \text{a p}_) = 1.0$

Note que aqui, pela primeira vez, ocorre reincidência de um bigrama, "a p".

8) Enquanto a pata bota os ovos, os patos observam a pa\_

- $P(\text{pata} | \text{a pa}_) = 1.0$

9) Enquanto a pata bota os ovos, os patos observam a pat\_

- $P(\text{pata} \mid \text{a pat}_) = 1.0$

10) Enquanto a pata bota os ovos, os patos observam a pata e o\_

- $P(\text{os} \mid \text{o}_) = 0.5$
- $P(\text{ovos} \mid \text{o}_) = 0.25$
- $P(\text{observam} \mid \text{o}_) = 0.25$

11) Enquanto a pata bota os ovos, os patos observam a pata e os o\_

- $P(\text{ovos} \mid \text{os o}_) = 1.0$

12) Enquanto a pata bota os ovos, os patos observam a pata e os ov\_

- $P(\text{ovos} \mid \text{os ov}_) = 1.0$

13) Enquanto a pata bota os ovos, os patos observam a pata e os ovo\_

- $P(\text{ovos} \mid \text{os ovo}_) = 1.0$

14) Enquanto a pata bota os ovos, os patos observam a pata e os ovos.

### III. CONCLUSÃO

O método proposto baseia-se nas características do autômato adaptativo, sendo que a criação de novas máquinas de estado é regida pela ocorrência prévia de uma palavra ou bigrama. A cada estado do autômato corresponde um conjunto de sugestões de complemento. As sugestões de complemento são associadas à frequência estatística com a qual aquele estado culmina no complemento em questão. Esta abordagem estatística tem por objetivo maximizar as chances de aceitação dos complementos sugeridos.

Espera-se ter contribuído com esta pesquisa para a formulação de uma metodologia bem fundamentada do preenchimento automático de textos, que possa ser usada em simulações e também como modelo teórico para criação e execução de programas em diversos ambientes. Espera-se, ainda, ter contribuído com a utilização prática do autômato

adaptativo, através da proposição de um dispositivo de software nele baseado.

Dentro dos objetivos traçados, a forma utilizada para alcançá-los foi interessante e profícua. Isto se deve ao fato de buscar uma visão integradora entre outras áreas como Probabilidade e, dentro dessa diversidade, tentar encontrar a resposta mais adequada às questões colocadas no início desta pesquisa.

O preenchimento automático de texto pode facilitar o trabalho de indivíduos que não dispõem de facilidade para digitação. O trabalho desenvolveu-se dentro das expectativas provando viabilidade do projeto e real vantagem na utilização de um método de sugestão baseado em probabilidade condicional, quando se utiliza linguagem coloquial.

### AGRADECIMENTOS

Os autores gostariam de agradecer à Toshiba Corporation Japan pelo suporte ao primeiro autor. Agradecemos ainda a José Neto pela orientação deste projeto.

### REFERÊNCIAS BIOGRÁFICAS

- [1] Barretto e José Neto. Uma arquitetura adaptativa. Em: Memórias do X Workshop de Tecnologia Adaptativa - WTA 2016. EPUSP, São Paulo. ISBN: 978-85-86686-86-3, pp. 2-3. 28 e 29 de Janeiro, 2016.
- [2] Canovas e Cugnasca. Um mapeamento de modelos adaptativos para dispositivos adaptativos guiados por regras. Em: Memórias do X Workshop de Tecnologia Adaptativa - WTA 2016. EPUSP, São Paulo. ISBN: 978-85-86686-86-3, pp. 86-97. 28 e 29 de Janeiro, 2016.
- [3] Carhuanina e José Neto. Personalização, customização, adaptabilidade e adaptatividade. Em: Memórias do X Workshop de Tecnologia Adaptativa - WTA 2016. EPUSP, São Paulo. ISBN: 978-85-86686-86-3, pp. 52-59. 28 e 29 de Janeiro, 2016.
- [4] Contier, Padovani e José Neto. O reconhecedor gramatical Linguístico: avanços em desambiguação sintática e semântica. Em: Memórias do X Workshop de Tecnologia Adaptativa - WTA 2016. EPUSP, São Paulo. ISBN: 978-85-86686-86-3, pp. 27-36. 28 e 29 de Janeiro, 2016.
- [5] Neto, J. J. Adaptive Automata for Context -Sensitive Languages. SIGPLAN NOTICES, Vol. 29, n. 9, pp. 115-124, September, 1994.
- [6] Rocha, R. L. A. e Neto, J. J. Uma proposta de método adaptativo para a seleção automática de soluções. Proceedings of ICIE Y2K - International Congress on Informatics Engineering, Buenos Aires, 2000.