

Proposta de Arquitetura de Software para um Chatbot com aprendizagem

Jéferson do Nascimento Soares, José E. B. Maia

¹MACC - Mestrado Acadêmico em Ciência da Computação
Universidade Estadual do Ceará - UECE
60714-903 - Fortaleza - Ceará - Brasil.

Abstract. *A weakness of the most commonly used chatbot agents today is that they are not designed to learn during the conversation process. Their knowledge is fixed during the project and they do not take advantage of the interaction with users to evolve their knowledge and their ability to dialogue. In this work, the design of a Chatbot with lifelong learning is inserted in the Reinforcement Learning framework. Although reinforcement learning has been used before, the literature review presented shows that this proposal is innovative.*

Keywords: *Conversational agent, Continuous learning, Dialogue man-machine, Chatbot, Virtual assistant.*

Resumo. *Uma fraqueza dos agentes de conversação (chatbot) mais utilizados hoje é que eles não são projetados para aprender durante o processo de conversação. O seu conhecimento é fixado durante o projeto e eles não tiram vantagem da interação com os usuários para evoluir o seu conhecimento e a sua capacidade de diálogo. Neste trabalho, o projeto de um Chatbot com aprendizagem ao longo da vida é inserido no arcabouço de Aprendizagem por Reforço. Embora aprendizagem por reforço já tenha sido utilizada antes, a revisão da literatura apresentada mostra que essa proposta é inovadora.*

Palavras-Chave: *Agente de conversação, Aprendizagem contínua, Diálogo homem-máquina, Chatbot, Assistente virtual.*

1. Introdução

Cada vez mais as pessoas buscam a tecnologia como aliada para resolver tarefas do dia-a-dia tais como fazer uma compra, pagar uma conta ou pedir uma orientação médica. Dentre essas tecnologias que facilitam o cotidiano das pessoas está o chatbot ou assistente virtual, que é um programa de computador que utiliza inteligência artificial para promover a interação entre máquina e ser humano por meio de uma linguagem natural típica de uma conversa face a face, mais geralmente, um sistema de perguntas e respostas.

Uma fraqueza dos agentes de conversação (chatbot) mais utilizados hoje é que eles não são projetados para aprender durante o processo de conversação. O seu conhecimento é fixado durante o projeto e eles não tiram vantagem da interação com os usuários para evoluir o seu conhecimento e a sua capacidade de diálogo.

Já existe na literatura propostas para incorporar aprendizagem ao longo da vida em chatbots baseada nos conceitos de aprendizagem por reforço, a qual é uma técnica de aprendizagem por tentativa e erro bem estabelecida [Ricciardelli and Biswas 2019, Liu et al. 2020, Yin et al. 2018, Nishimoto and Costa 2019, Li et al. 2016] . Os algoritmos de aprendizagem por reforço podem ser divididos em duas categorias: baseados em

modelo e livres de modelo [Sutton and Barto 1998]. A dificuldade de modelar ambientes da vida real fez com que os algoritmos livres de modelo tenham adquirido ampla popularidade, sendo *Q-Learning* aquele mais conhecido e utilizado. Este trabalho também aborda o problema da aprendizagem ao longo da vida em chatbots da perspectiva de aprendizagem por reforço. Porém, ao contrário dos trabalhos revisados, a abordagem aqui proposta é baseada em modelo e utiliza Programação Dinâmica Adaptativa (PDA).

O restante deste trabalho está assim organizado. A Seção 2 apresenta uma revisão conceitual do projeto tradicional de chatbots (2.1) e de aprendizagem por reforço (2.3). A seção é concluída com a descrição da arquitetura de chatbot proposta (2.3). A Seção 3 é uma breve revisão de trabalhos relacionados procurando contrastar com a proposta deste trabalho. O artigo é concluído na seção 4.

2. Métodos

2.1. O Projeto de Chatbots

As modalidades mais comuns de chatbots segundo [Io and Lee 2017] são baseados em regras e os baseados em inteligência artificial. Os baseados em regras são os mais simples de serem desenvolvidos, no qual o bot vai identificando o que o usuário quer dizer, com base de dados já previamente implementada no sistema, em que cada comando é escrito em forma de expressão regular. Isso, de certa forma, limita a comunicação do assistente virtual pois caso as mensagens fornecidas saiam do script do bot, este terá dificuldade de manter a coerência da conversa, uma vez que o contexto das perguntas dos usuários estão fora dos comandos já processados no chatbot. Já os baseados em inteligência artificial (IA) são mais complexos, pois se baseiam em modelos de IA para se comunicar com os usuários.

No começo, para se desenvolver um chatbot era necessário conhecimento avançado de programação e profissionais para atingir exatidão no protótipo. Contudo, com a popularização desta tecnologia, há diversas plataformas que facilitam a criação de chatbot [Nuruzzaman and Hussain 2018]. O dialogflow, desenvolvido pela empresa Google, é uma plataforma que dispensa ao seu usuário a necessidade de programação avançada para desenvolver seu chatbot gratuitamente. Este modelo já vem com diversos algoritmos de aprendizagem de máquina previamente treinados com uma grande base de dados do próprio google e que está disponível para integração ao chatbot em diversas plataformas como facebook, telegram, skype, twitter entre outros.

Outra plataforma de desenvolvimento de chatbots bastante famosa é a IBM Watson. Essa plataforma é paga e vem com um período de teste gratuito. Ela dá opções de suporte em diversos nichos da tecnologia de processamento de linguagem natural, nos quais já oferece modelos de PLN prontos para aplicar em negócios como reconhecimento de entidades nomeadas, análise de sentimentos, extração de palavras-chave e outros recursos de processamento de linguagem natural.

As desvantagens destas APIs são inúmeras. Primeiro, o gerenciamento do modelo é um tanto trabalhoso, pois o usuário deve ir adicionando os roteiros de perguntas e respostas na plataforma manualmente. Devido serem plataformas de código fechado, isto é, o usuário não tem acesso ao código fonte, o chatbot não pode ser personalizado, portanto, não poderá adicionar ou alterar algo no seu modelo de PNL (Processamento de

Linguagem Natural), não podendo ser aprimorado por seu desenvolvedor, ficando limitado apenas aos modelos da própria API.

Contudo há também a possibilidade de desenvolver chatbots por meio de linguagem de programação de propósito geral. Praticamente toda linguagem de programação já possui um cliente para se conectar e manipular os seus próprios modelos proveniente das plataformas de mensagens (facebook messenger, telegram, bot web, whatsapp). A maioria delas fornecem, de maneira simples, uma API em alto nível para receber, processar e responder cada mensagem ou evento enviado pelas plataformas.

Embora fáceis de programar, nenhuma destas tecnologias oferece um modelo em que o chatbot pode aprender ao longo da sua vida a partir de diálogos futuros. Sempre que o usuário decidir realizar novas atualizações deve fazê-las de forma manual.

2.2. Aprendizagem por Reforço

As técnicas de aprendizagem de máquina pode ser divididas em quatro tipos [Norvig and Russell 2014]. Aprendizagem supervisionada, na qual o modelo é treinado com exemplos completos e entrada e resposta; aprendizagem não supervisionada, na qual a tarefa é encontrar relações ou agrupamentos entre exemplos não rotulados; aprendizagem semi-supervisionada, na qual procura-se treinar um modelo com uma mistura de exemplos rotulados e não rotulados; e aprendizagem por reforço, na qual um agente pode iniciar sem qualquer exemplos inicial rotulado e adquirir conhecimento do ambiente por tentativa e erro.

A figura 1 [Sutton and Barto 1998] é uma representação simbólica do cenário de aprendizagem por reforço: a cada ação executada pelo agente sobre o ambiente, este emite uma resposta que deve ser interpretada e mapeada pelo agente em uma recompensa e em uma representação do estado. O objetivo de um agente de aprendizagem por reforço é aprender uma política de atuação no ambiente a partir da interação com ele.

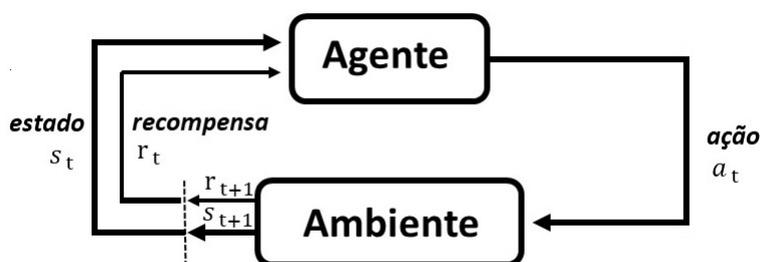


Figura 1. Cenário de Aprendizagem por Reforço [Sutton and Barto 1998]

No paradigma de Aprendizagem por Reforço (do inglês, RL) há duas classes de algoritmos: baseados em modelo ou livres de modelo (ativa). Quando um modelo do ambiente é conhecido são utilizados algoritmos baseados em modelo. Um representante desta classe é o algoritmo de Programação Dinâmica Adaptativa (PDA). Entretanto, frequentemente, o ambiente no qual o agente deve atuar é desconhecido. Para esses cenários foram desenvolvidos os algoritmos livres de modelo. O algoritmo *Q-Learning* sendo o principal representante desta família.

Seguindo [Sutton and Barto 1998], um agente de aprendizagem por reforço é definido por quatro componentes básicos: uma política, uma função de recompensa, uma função de valor e um modelo de ambiente.

- **Política:** é a função de tomada de decisão do agente correspondendo ao conjunto de estímulos que levam o agente a determinar o próximo comportamento correspondente;
- **Função recompensa:** é o reforço recebido pelo agente a cada uma ou mais ações executadas (decisão);
- **Função de valor:** é uma função que representa um estímulo ou atratividade para cada estado do agente;
- **Modelo ambiente:** é um modelo do mundo externo onde o agente atua. Um modelo do ambiente possibilita o agente prever os resultados das suas ações.

2.3. Arquitetura Proposta

Segundo [Adamopoulou and Moussiades 2020], os principais requisitos para projetar um chatbot tradicional são: representação do conhecimento (i.e. um conjunto de dados sobre o domínio para o qual se deseja aplicar), uma estratégia de geração de respostas e um conjunto de respostas neutras predefinidas para responder quando a expressão do usuário não é entendida. Considerando esses elementos, partiu-se da arquitetura geral apresentada em [Adamopoulou and Moussiades 2020] Para para elaborar a arquitetura aqui proposta.

A Figura 2 apresenta a arquitetura proposta, que é um mapeamento da arquitetura em [Adamopoulou and Moussiades 2020] para os elementos de aprendizagem por reforço. Os dois principais elementos que suportam a aprendizagem por reforço são o bloco de intenção do usuário e o bloco bases de dados.

Em um chatbot tradicional, o bloco de intenção do usuário tem a função de juntar a pergunta recebida com o contexto da conversa no sentido de guiar a escolha da resposta a ser dada ao usuário. No modelo com RL, o chatbot vai criar e manter um modelo do ambiente para guiar o diálogo. Esse modelo será alimentado por feedback de relevância implícito e explícito (solicitado pelo chatbot), quando o modelo indicar. A ideia aqui é utilizar os mesmos tipos de modelos já utilizados para modelar usuários (estudantes, p.e.) em sistemas tutoriais inteligentes. Portanto, ao contrário do modelos encontrado na revisão da literatura, esta proposta inclui um modelo do ambiente e, portanto, algoritmos baseados em modelo poderão ser utilizados.

O outro componente de destaque são as bases de conhecimento. Em um chatbot tradicional, existe apenas uma base de conhecimento que é construída em tempo de projeto. Na arquitetura de chatbot com RL aqui proposta existem duas bases de dados: uma permanente, dependente do domínio, construída com conhecimento de especialista em tempo de projeto e uma segunda base de conhecimento que evolui ao longo da vida do chatbot sendo alimentada pelos diálogos e pelos feedback de relevância. Os demais elementos da arquitetura se modificam para atender as necessidades deste dois componentes, além de cumprir suas próprias funções.

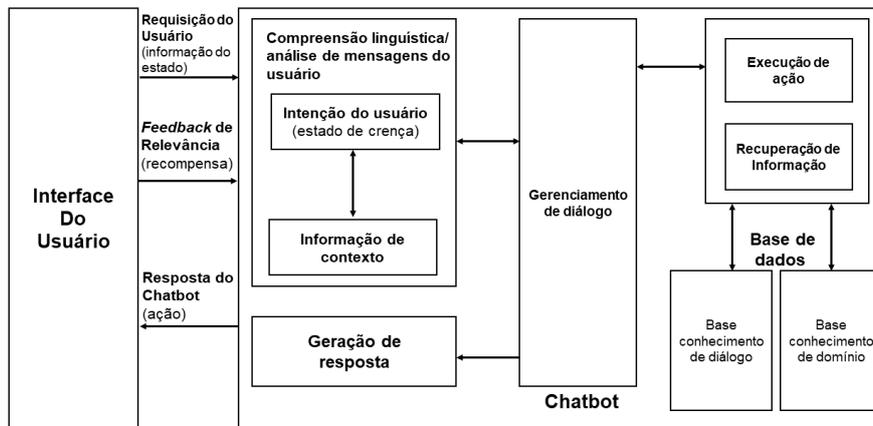


Figura 2. Arquitetura de software do chatbot com aprendizagem por reforço modificada de [Adamopoulou and Moussiades 2020].

3. Trabalhos Relacionados

Esta seção faz uma revisão breve de alguns trabalhos com chatbots baseados em aprendizagem por reforço relacionados a esta pesquisa [Serban et al. 2017, Ricciardelli and Biswas 2019, Liu et al. 2020, Yin et al. 2018, Nishimoto and Costa 2019].

[Liu et al. 2020] apresenta uma estrutura de chatbot com aprendizado de reforço hierárquico orientada a objetivos chamado GoChat. A proposta do autor é treinar o agente de diálogo com o algoritmo de aprendizagem por reforço, Advantage Actor-Critic (A2C), sem a necessidade de uma estrutura específica de diálogo, pois este utiliza um conjunto de diálogos coletados para treinar o modelo com um foco específico, onde a política de alto nível orienta a conversa em direção ao objetivo final. O autor utilizou um conjunto de dados de diálogo a partir de dados financeiros que visa a detecção de fraudes durante o diálogo, adotou-se, entre outras, inúmeras métricas de avaliação, como por exemplo, a avaliação humana para avaliar a qualidade e a diversidade das respostas geradas, respectivamente. As experiências realizadas pelos autores aprovou a eficácia do modelo.

[Ricciardelli and Biswas 2019] introduz um modelo de RL para chatbot visando chatbots do tipo FAQ, ou seja, chatbots com perguntas e respostas frequentes, geralmente utilizados por empresas para responder perguntas constantes de seus cliente. O modelo visa aproveitar dados de conversas com o usuário para melhorar o desempenho do assistente virtual (offline). A abordagem considera uma aplicação de pontuação dos enunciados do chatbot por meio de um feedback do usuário. Para o aprendizado de políticas é implementado um agente usando Deep Q-Network (DQN) com exploração *epsilon-greedy*. As pontuações previstas por este modelo são usadas como recompensa para o agente de RL. Segundo o autor o potencial da abordagem é mostrado em um caso extraído de um chatbot corporativo, melhorando o desempenho inicial de 50% de sucesso para 75% em 20 a 30 épocas de treinamento.

Yin et al. [Yin et al. 2018] apresenta um modelo denominado *parameterized auxiliary asynchronous advantage actor-critic* (PA4C). Ele é um modelo de chatbot baseado

em RL o qual utiliza o modelo PA4C para otimizar os modelos convencionais de RL com parametrização de ações e tarefas auxiliares para o treinamento de chatbot. Resultados experimentais mostram que o modelo PA4C apresentado pelo autor supera os modelos da literatura nas métricas de taxa de sucesso, duração do diálogo e recompensa do episódio.

O trabalho [Nishimoto and Costa 2019] propõe um sistema Dialog Managers (DM), ou seja, um sistema responsável pelo fluxo de conversa de um agente. Utiliza-se algoritmo Deep Reinforcement Learning (DRL) para treinamento dos dados de diálogos e usa uma estratégia softmax combinada com o aprendizado de transferência do conhecimento do designer para equilibrar a exploração do modelo. O grande diferencial é a inclusão de uma função softmax com transferência de aprendizado durante a fase de aprendizado. Os resultados experimentais indicam que a proposta é bastante promissora.

MILABOT [Serban et al. 2017] é um framework de diálogo baseado em *ensemble* desenvolvido pelo Montreal Institute for Learning Algorithms (MILA) for the Amazon Alexa Prize competition. A arquitetura é baseada em aprendizagem por reforço profundo e consiste de um *ensemble* de modelos de recuperação de informação e geração de linguagem natural incluindo modelos baseados em template, modelos bag-of-words, modelos de redes neurais sequencia-a-sequencia e de variáveis latentes. O sistema é treinado para selecionar a resposta apropriada usando dados de interações de usuários reais. Diversos modelos de redes neurais são avaliados para aproximar a função ação-valor, entre elas LSTM, Q-learning AMC e *off-policy* REINFORCE. Para os dados testados, os autores registram desempenho competitivo.

Comum a todos os modelos descritos nesta seção está o fato de aplicarem algoritmos de RL livres de modelo. Este trabalho diferencia-se destes por propor uma arquitetura para utilizar algoritmos de RL baseados em modelo.

4. Conclusão

A arquitetura de um sistema software refere-se a definição dos componentes de software e seus relacionamentos entre si e com outros softwares, do ponto de vista das suas propriedades nas interfaces. A arquitetura do software é determinante para o sucesso de um projeto de software.

Neste trabalho, investigou-se, via revisão da literatura, algumas arquiteturas propostas para o projeto de chatbots com o intuito de modificá-las para acomodar aprendizagem ao longo da vida. Especificamente, a arquitetura geral apresentada em [Adamopoulou and Moussiades 2020] foi modificada para acomodar o paradigma de aprendizagem por reforço baseada em modelo. O que distingue este trabalho dos demais revisados na literatura é a aplicação de RL baseada em modelo na qual o chatbot vai criar e manter um modelo do usuário no estilo daqueles utilizados em sistemas tutoriais inteligentes. Isso possibilitará o projeto utilizar algoritmos de aprendizagem por reforço baseados em modelo cuja convergência é sabido ser muito mais rápida. Isso é importante pois um chatbot poderá ter de aprender a partir de alguns poucos diálogos.

Este é um trabalho em andamento. Um protótipo e experimentos prova-de-conceito está sendo desenvolvido para um chatbot de perguntas e respostas em um domínio restrito com interface textual. O chatbot é baseado em Recuperação de Informação e utiliza feedback de relevância implícita e explícita, via comportamento ativo

do chatbot. Estes experimentos serão utilizados também para avaliar comparativamente os algoritmos de RL e os modelos de usuário.

Referências

- Adamopoulou, E. and Moussiades, L. (2020). An overview of chatbot technology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 373–383. Springer.
- Io, H. and Lee, C. (2017). Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 215–219. IEEE.
- Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., and Jurafsky, D. (2016). Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.
- Liu, J., Pan, F., and Luo, L. (2020). Gochat: Goal-oriented chatbots with hierarchical reinforcement learning. *arXiv preprint arXiv:2005.11729*.
- Nishimoto, B. and Costa, A. H. R. (2019). Dialogue management with deep reinforcement learning: Balancing exploration and exploitation. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 449–454. IEEE.
- Norvig, P. and Russell, S. (2014). *Inteligência artificial: Tradução da 3a Edição*. Elsevier Brasil.
- Nuruzzaman, M. and Hussain, O. K. (2018). A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, pages 54–61. IEEE.
- Ricciardelli, E. and Biswas, D. (2019). Self-improving chatbots based on reinforcement learning. In *4th Multidisciplinary Conference on Reinforcement Learning and Decision Making*.
- Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., Ke, N. R., et al. (2017). A deep reinforcement learning chatbot. *arXiv preprint arXiv:1709.02349*.
- Sutton, R. S. and Barto, A. G. (1998). Reinforcement learning i: Introduction.
- Yin, C., Zhang, R., Qi, J., Sun, Y., and Tan, T. (2018). Context-uncertainty-aware chatbot action selection via parameterized auxiliary reinforcement learning. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 500–512. Springer.