

Chatbots Generativos como Ferramentas de Apoio ao Ensino em Cursos na Área de Ferrovias

Eduardo dos Santos Lopes¹, Hilário Tomaz Alves de Oliveira¹, Kelly Assis de Souza Gazolli¹

¹Programa de Pós-graduação em Computação Aplicada (PPComp),
Instituto Federal do Espírito Santo (IFES), Serra, Brasil

lopes.lopes.edu@gmail.com, {hilario.oliveira, kasouza}@ifes.edu.br

Abstract. *Chatbots are conversational systems capable of simulating interactions using natural language. This technology allows quick interaction with users, and its use as a teaching support tool offers students a new way of accessing content. This work presents a domain-specific database and its use in constructing a generative chatbot to aid students in the Railways area. The BiLSTM and GRU neural networks were adopted in an encoder-decoder architecture with an attention mechanism. The experiments showed that the GRU architecture achieved the best results based on the BLEU and ROUGE-L evaluation measures.*

Resumo. *Chatbots são sistemas de conversação capazes de simular interações utilizando linguagem natural. Essa tecnologia permite a interação com os usuários de forma rápida, e seu uso como ferramenta de apoio ao ensino oferece aos estudantes uma nova forma de acesso ao conteúdo. Este trabalho apresenta uma base de dados de domínio específico, bem como sua utilização na construção de um chatbot generativo para auxiliar alunos na área de Ferrovias. Para isso, foram utilizadas as redes neurais BiLSTM e GRU, ambas em uma arquitetura do tipo codificador-decodificador com mecanismo de atenção. Os experimentos realizados demonstraram que a arquitetura usando GRUs obteve melhor desempenho com base nas medidas de avaliação do BLEU e ROUGE-L.*

1. Introdução

Uma das consequências do isolamento social provocado pela pandemia de Covid-19 foi a intensificação do uso de tecnologias no ambiente escolar [Vieira and Silva 2020]. Diversas ferramentas passaram a fazer parte do cotidiano dos estudantes, exercendo um papel fundamental no apoio ao ensino, dentre elas, pode-se destacar o uso de Ambientes Virtuais de Aprendizagem (AVA). Nesse tipo de recurso, a construção do conhecimento é direcionada por meio dos instrumentos disponibilizados (tarefas, vídeos, lições, entre outras), proporcionando uma maior autonomia ao estudante. Contudo, para solucionar as dúvidas ou curiosidades que possam surgir durante o processo de aprendizado é necessária a interação com o professor, o que pode não acontecer de forma imediata.

Uma alternativa para preencher essa lacuna é a utilização de *chatbots*, também conhecidos como sistemas de conversação ou diálogo. Esse tipo de ferramenta possibilita a comunicação com humanos por meio da linguagem natural, seja ela falada ou escrita [Jerald 2015, McTear 2020]. A utilização de um sistema de diálogo de conhecimento específico como ferramenta de apoio educacional oferece ao aluno acesso ao conhecimento de forma rápida e individualizada. Além disso, com o uso de *chatbots* é possível oferecer

responsividade, que é a capacidade de reagir de maneira rápida e positiva em um determinado momento [Mondal et al. 2018], e disponibilidade, já que o sistema pode ficar disponível sempre que o aluno precisar.

O uso de sistemas de conversação tem se tornado cada vez mais popular devido ao seu potencial em simular diálogos humanos, possibilitando assim automatizar tarefas e reduzir custos [Kuhail et al. 2022]. Existem diversos exemplos de aplicações de *chatbots* em vários domínios, incluindo saúde [Chang et al. 2022], serviços ao consumidor [Mischia et al. 2022], educação [Oliveira et al. 2010, Mondal et al. 2018], entre outros. Recentemente, modelos generativos baseados em redes neurais profundas têm se tornado o estado da arte no desenvolvimento de *chatbots*, pois não se limitam a respostas pré-definidas. Tais modelos geram novas respostas a partir de uma base de dados de treinamento de conversação utilizando uma variedade de abordagens da área de aprendizado profundo, do inglês *Deep Learning* [Sawant et al. 2021].

Neste contexto, o objetivo deste trabalho é desenvolver um sistema de diálogo para auxiliar na monitoria de alunos de disciplinas da área de Ferrovias. Para isso, foi adotada uma abordagem generativa, utilizando os modelos *Bidirectional Long-Short Term Memory* (BiLSTM) e *Gated Recurrent Unit* (GRU), em conjunto com mecanismos de atenção. Além disso, foi construída uma base de domínio específico contendo perguntas e respostas escritas em inglês da área de Ferrovias com base em dúvidas frequentes dos alunos. Atualmente, a base desenvolvida contém aproximadamente 10 mil pares de perguntas e respostas genéricas, comumente usadas em diálogos, e 1 mil e 100 pares específicas do domínio. Experimentos foram realizados para avaliar o desempenho do *chatbot* desenvolvido comparando a arquitetura sendo desenvolvida utilizando as redes neurais do tipo BiLSTM e GRU com base nas medidas de avaliação do BLEU e ROUGE-L. As principais contribuições deste artigo são: (i) a construção de uma base de dados com perguntas e respostas relacionadas à área de ferrovias; (ii) o desenvolvimento de um *chatbot* utilizando uma abordagem generativa; e (iii) uma análise comparativa do desempenho de duas arquiteturas utilizando as redes neurais do tipo BiLSTM e GRU.

O restante deste artigo está organizado como segue: na Seção 2 é apresentada uma breve revisão da literatura sobre o uso *chatbots*. As principais etapas realizadas para a construção e o pré-processamento da base de dados proposta são abordadas na Seção 3.1. Na Seção 3.3 é apresentada a arquitetura e os modelos usados no desenvolvimento do sistema de *chatbot* proposto. Na Seção 4 são apresentados e discutidos os experimentos realizados. Por fim, na Seção 5 são apresentadas as considerações finais e delineadas linhas de pesquisas para trabalhos futuros.

2. Trabalhos Relacionados

A ideia de criar sistemas computacionais capazes de interação com seres humanos utilizando linguagem natural sempre foi muito atrativa e faz parte da história da área de Inteligência Artificial (IA). Em 1950, Alan Turing foi um dos primeiros pesquisadores a questionar se um programa de computador poderia ser capaz de conversar com um grupo de pessoas sem que eles percebessem que seu interlocutor era artificial [Turing 1950]. Posteriormente, diversos sistemas de *chatbots* famosos foram criados, dentre eles se destacam a ELIZA [Weizenbaum 1966] e a ALICE [AbuShawar and Atwell 2015].

O sistema ELIZA foi desenvolvido em 1966, sendo considerado o primeiro *chat-*

bot do mundo. ELIZA foi projetado para emular a interação de um psicoterapeuta com seus pacientes e, para isso, o sistema tinha uma base de conhecimento neste domínio [Weizenbaum 1966]. Apesar de sua capacidade de interação via texto ser limitada, o trabalho foi fonte de inspiração para diversas pesquisas posteriores no desenvolvimento de novos sistemas. Mais recentemente, nos anos noventa, outro *chatbot* famoso foi a *Artificial Linguistic Internet Computer Entity* (ALICE). Esse sistema utilizava técnicas de Processamento de Linguagem Natural (PLN) em conjunto com uma linguagem de marcação específica chamada de *Artificial Intelligence Markup Language* (AIML) [AbuShawar and Atwell 2015]. A base de conhecimento do sistema ALICE foi desenvolvida utilizando a linguagem AIML para descrever padrões que possibilitam mapear uma pergunta do usuário com uma resposta presente na sua base de conhecimento.

Diversos *chatbots* foram desenvolvidos utilizando a linguagem AIML ou algumas de suas extensões como, por exemplo, iAIML [Neves et al. 2006] ou Persona-AIML [Galvão et al. 2004]. A ideia dessas extensões era permitir descrever padrões mais ricos de interação entre os sistemas e os usuários [Oliveira et al. 2010]. Contudo, esses *chatbots* são limitados a perguntas e respostas que estejam presentes em suas bases de conhecimento. A aplicação de *chatbots*, via voz ou texto, expandiu consideravelmente com o surgimento de algoritmos de Aprendizado de Profundo, especialmente utilizando modelos capazes de geração de linguagem natural, chamados de generativos [Caldarini et al. 2022]. Dessa forma, os sistemas não são limitados a perguntas e respostas existentes em uma base pré-definida.

Os *chatbots* generativos podem ser construídos utilizando vários tipos de redes neurais, mantendo uma mesma estrutura, no qual o sistema recebe uma entrada, faz representações vetoriais e a partir de um modelo pré-treinado produz uma resposta a ser apresentada ao usuário como saída [Adamopoulou and Moussiades 2020]. Em [Aleedy et al. 2019], os autores compararam três modelos para a construção de um *chatbot* para a área de atendimento ao cliente, baseados nos seguintes tipos de redes neurais: *Long Short-Term Memory* (LSTM), *Gated Recurrent Units* (GRU) e *Convolutional Neural Network* (CNN). Os resultados obtidos demonstraram que as redes do tipo LSTM e GRU são promissoras para o desenvolvimento de *chatbots* generativos. Aproveitando a capacidade das redes do tipo LSTM, alguns trabalhos [Zalake and Naik 2019, Dhyani and Kumar 2021] implementaram *chatbots* utilizando o formato bidirecional, isto é, duas redes LSTM separadas, recebendo a entrada tanto do início para o fim, como do fim para o início, técnica conhecida como BiLSTM, gerando, desse modo, diálogos mais apropriados.

O uso de mecanismos de atenção se tornaram parte integrante de diversos modelos neurais na área de PLN, especialmente em tarefas envolvendo geração de linguagem, pois conseguem oferecer um melhor desempenho e se adaptam a uma grande variedade de arquiteturas [Galassi et al. 2020]. Em [Kim et al. 2019], os autores propuseram um *chatbot* generativo utilizando mecanismos de atenção com a capacidade de lidar com o contexto da conversa. Além disso, para geração do diálogo o sistema desenvolvido utiliza uma rede do tipo *Wasserstein Generative Adversarial Network* (WGAN). A abordagem proposta obteve resultados do estado da arte quando aplicado ao conjunto de dados do *DailyDialog*.

3. Materiais e Métodos

Nesta seção são descritos o processo de construção da base de dados e as etapas de pré-processamento realizadas. Em seguida, a arquitetura utilizada no desenvolvimento do *chatbot* proposto, bem como as técnicas envolvidas são apresentadas.

3.1. Base de dados

Para viabilizar o desenvolvimento do sistema de *chatbot* proposto neste trabalho, foi inicialmente construída uma base de dados contendo perguntas e respostas comumente usadas em diálogos no idioma inglês.

Para isso, foram usadas perguntas e respostas de domínio aberto e de conhecimento específico da área de ferrovias que é área de interesse deste trabalho. Os diálogos de domínio aberto são aqueles utilizados em conversas rotineiras e não estão associados a um assunto específico. A intenção de incluir esse tipo de diálogo na base de dados é possibilitar que o sistema possa interagir com o aluno de uma maneira mais receptiva e acolhedora. Dessa forma, o sistema pode ser capaz de responder a perguntas simples do dia a dia, por exemplo, “Bom dia, como você está?”, tornando a comunicação com o usuário mais natural. Na construção da base de dados proposta, os diálogos de domínio aberto foram extraídos da base de perguntas e respostas do DialoGPT [Zhang et al. 2020], disponível em¹. Essa base possui 10 mil pares de perguntas e respostas escritas em inglês.

Tendo em vista a pouca disponibilidade de diálogos a respeito da área de sistemas ferroviários, foi necessária a construção de um conjunto de dados próprio de perguntas e respostas sobre esse domínio. Para isso, foram utilizados textos disponibilizadas no site *American-Rails*² que tratam não só sobre a história das ferrovias americanas, mas também sobre o maquinário e as tecnologias. Além disso, também foi utilizado o conteúdo presente no site *Railway Technical*³, bem como materiais produzidos em disciplinas relacionadas ao tema no Instituto Federal do Espírito Santo (IFES), instituto esse que oferta o curso Técnico em Manutenção de Sistemas Metroferroviários. Outra fonte adotada foi o site *a2z-online*⁴ que conta com centenas de perguntas e repostas a respeito do tema de manutenção de vagões. Atualmente, a base de dados possui cerca de 1 mil e 100 pares de perguntas e respostas sobre conteúdo específico que ao serem reunidos com os diálogos de domínio aberto, geraram uma base de dados com 11 mil e 100 perguntas e respostas. Na Tabela 1 são apresentados alguns exemplos de perguntas e respostas relativas ao domínio específico presentes na base de dados.

3.2. Pré-Processamento da Base de Dados

A realização de etapas de pré-processamento visando normalizar e simplificar o vocabulário dos textos presentes na base de dados antes da fase de treinamento é uma tarefa muito importante para aplicações de PLN. A presença de caracteres maiúsculos, minúsculos, símbolos de pontuação, erros ortográficos, dentre outros, afetam o desempenho dos modelos durante a etapa de treinamento. Por exemplo, as palavras “railroad”, “Railroad” e “railRoad!” são considerados diferentes apesar de terem o mesmo significado.

¹<https://github.com/microsoft/DialoGPT>.

²<https://www.american-rails.com>

³<http://www.railway-technical.com/>

⁴<https://a2z-online.org/carriage-wagon-objectivequestions-with-answers/>

Tabela 1. Exemplos de perguntas e respostas presentes na base de domínio específico.

Perguntas	Respostas
<i>What is the definition of AEI?</i>	<i>Short for Automatic Equipament Identification.</i>
<i>What is Recrew?</i>	<i>When a train receives a new crew.</i>
<i>What is the definition of Adhesion factor?</i>	<i>Describes how well locomotive can grip the rails.</i>
<i>What it the definition of Affilliate?</i>	<i>A company owned or controlled by another.</i>

Por isso, os textos dos diálogos da base de dados foram normalizados, sendo substituídas as letras maiúsculas por minúsculas. Além disso, visando diminuir o tamanho do vocabulário criado, palavras com frequência menor ou igual a cinco foram substituídas pela marcação < *OUT* >. Tal ação foi realizada para otimizar o treinamento do modelo sem impactar no funcionamento do *chatbot*, tendo em vista que essas palavras são pouco utilizadas.

3.3. Arquitetura do *Chatbot*

Para a construção do sistema de *chatbot* proposto, foi utilizada a arquitetura do tipo Codificador-Decodificador, do inglês *Encoder-Decoder*. Esse tipo de arquitetura é comumente usada em tarefas de PLN que envolvem geração de linguagem, por exemplo, sumarização automática de texto [Hou et al. 2021], tradução automática [Dabre et al. 2020], entre outras. A escolha por utilizar esse tipo de arquitetura é por ela possibilitar o desenvolvimento de um sistema generativo que possa aprender automaticamente padrões de alinhamento entre as perguntas dadas como entrada e as respostas a serem geradas como saída. Além disso, por ser capaz de gerar respostas, palavra por palavra, dependendo da pergunta de entrada [Arora et al. 2020], não existe a limitação de ter que gerar uma resposta de saída com a mesma quantidade de palavras da entrada [Aleedy et al. 2019].

As redes neurais recorrentes são comumente usadas para desenvolver os módulos de codificador e de decodificador [Cho et al. 2014]. O módulo codificador (*Encoder*) recebe como entrada uma sequência de tamanho variado e tem o objeto de converter em um vetor de representação contextual de tamanho fixo. Por outro lado, o módulo Decodificador (*Decoder*) recebe como entrada o vetor gerado pelo codificador e a mapeia em uma sequência de tamanho variado que deve ser gerado como saída [Cho et al. 2014].

Neste trabalho, a implementação dos módulos codificador e decodificador foi realizada utilizando duas variações de redes recorrentes, a BiLSTM e GRU. Além disso, em ambas as implementações foi utilizado um componente de mecanismo de atenção proposto em [Bahdanau et al. 2014], que possibilita com que a rede neural identifique e foque em pontos de atenção específicos nas entradas [Sutskever et al. 2014, Bahdanau et al. 2014]. O módulo Decodificador foi desenvolvido usando um método de decodificação guloso, do inglês *Greedy Search Decoding*, para realizar o processo de escolha das palavras a serem geradas como saída.

Na Figura 1 é apresentada a arquitetura do *chatbot* proposto utilizando redes BiLSTMs para implementar os módulos codificador e decodificador. As redes do

tipo LSTM possuem a capacidade de aprender padrões em dados que possuem dependências temporais, isto é, que dependem do contexto nos quais estão inseridos [Sutskever et al. 2014]. Por isso, esse tipo de rede tem sido muito usado em tarefas que envolvem dados sequenciais, por exemplo, tarefas de PLN e séries temporais.

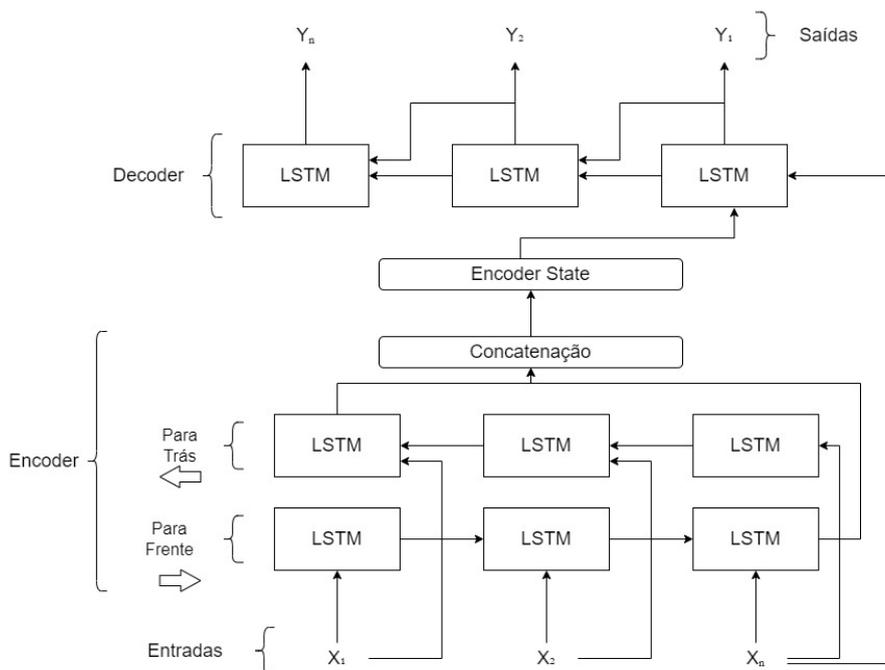


Figura 1. Arquitetura do *chatbot* proposto utilizando as redes BiLSTMs nos módulos codificador e decodificador.

Na Figura 2 é apresentada a arquitetura do *chatbot* utilizando as redes GRU. Segundo [Li et al. 2021] as redes GRU buscam otimizar a estrutura das redes LSTM sem perder o desempenho. Esse tipo de rede simplifica o funcionamento das LSTMs, o que as tornaram populares em várias áreas de aplicação [Rana 2016].

4. Experimentos

Nesta seção são apresentados e discutidos os experimentos realizados para avaliar o desempenho do *chatbot* desenvolvido. Para isso, experimentos foram realizados para analisar as seguintes questões: **(i)** uma comparação entre o uso das redes do tipo BiLSTM e GRU; e **(ii)** uma análise do impacto da variação do número de épocas usada no treinamento do modelo e no número de células usadas nos modelos. Antes de apresentar e discutir os resultados obtidos, uma breve apresentação do ambiente experimental e alguns detalhes de implementação são apresentados na próxima seção.

4.1. Configurações dos Experimentos

Treinamento dos Modelos. Os modelos utilizados para o desenvolvimento do *chatbot* proposto neste trabalho são compostos de diversos parâmetros, tais como o número de células das redes neurais. Além disso, durante o treinamento dos modelos é possível configurar diferentes parâmetros, por exemplo, o número de épocas usado para treinar os modelos. Para realização dos experimentos a base de dados foi dividida usando a

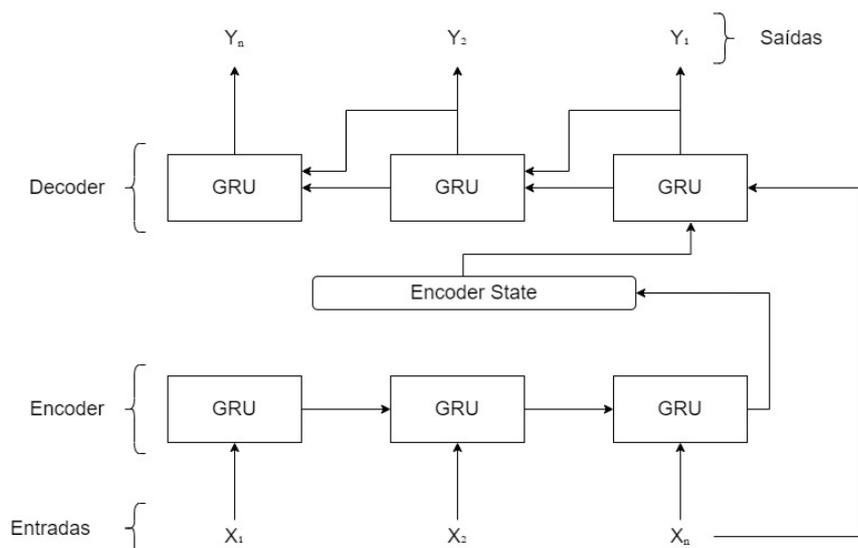


Figura 2. Arquitetura do *chatbot* utilizando GRU nos módulos codificador e decodificador.

proporção de 90% para treino e 10% para testes. Essa divisão foi realizada dessa maneira porque a quantidade de perguntas e respostas presente na base de dados é considerada pequena para treinar modelos de redes neurais profundas. Na Tabela 2 são apresentadas as configurações utilizadas nos experimentos.

Tabela 2. Configuração dos hiperparâmetros utilizados no treinamento dos modelos.

Parâmetros	Valores
Células	500, 1000
Épocas	2000, 4000
Tamanho do Batch	500
Função de Ativação	SoftMax
<i>Encoder Layers</i>	2
<i>Decoder Layers</i>	2
<i>Dropout</i>	0,1
<i>Decoder</i>	<i>Greedy Search</i>

Medidas de avaliação. A avaliação de sistemas de *chatbots* generativos pode ser realizada comparando a resposta gerada pelo sistema para uma dada pergunta em relação a uma ou mais respostas esperadas (referências). Para isso, medidas de similaridade podem ser utilizadas para avaliar o quão semelhante é a resposta do sistema em comparação com a esperada. Para avaliar o desempenho do *chatbot* desenvolvido neste trabalho foram utilizadas as medidas de precisão, cobertura e medida-F derivadas da *Recall-Oriented Understudy for Gisting Evaluation Longest Common Subsequence* (ROUGE-L) [Lin 2004]. Essa medida compara a maior sequência de caracteres em comum entre um texto candidato gerado automaticamente e uma ou mais referências. A

implementação do ROUGE-L disponível em⁵ foi utilizada. Outra medida adotada foi a *Bilingual Evaluation Understudy* (BLEU) [Papineni et al. 2002]. A medida BLEU foi usada para comparar a sobreposição de n-gramas entre a resposta gerada pelo sistema e a de referência. Foi utilizada a implementação da medida BLEU-4 disponibilizada na ferramenta *Natural Language Toolkit* (NLTK)⁶. A variação BLEU-4 computa a média geométrica da medida de precisão da sobreposição de unigramas, bigramas, trigramas e quadrigramas.

Os experimentos realizados neste trabalho foram executados em um computador equipado com processador AMD Ryzen 7 5700g, 32 GB de memória RAM e placa de vídeo Nvidia RTX 3060 com 12 GB. A placa de vídeo utilizada possui recursos específicos para processamento de alto desempenho, por meio da *Nvidia CUDA Toolkit*, versão 11.3. A implementação do *chatbot* foi feita utilizando a linguagem de programação Python, utilizando especialmente a biblioteca Pytorch versão 1.11.0.

4.2. Resultados e Discussões

O objetivo deste experimento é comparar a utilização das redes BiLSTM e GRU na arquitetura do *chatbot* proposto. Além disso, variações no número de células (500 e 1000) usadas nas redes e na quantidade de épocas (2000 e 4000) de treinamento foram analisadas. Na Tabela 3 são apresentados os resultados computados com base nas medidas de Precisão (P), Cobertura (C) e medida-F (F1) do ROUGE-L, além da medida BLEU. O maior resultado obtido em cada medida está destacado em negrito.

Tabela 3. Resultados (%) da avaliação de diferentes variações da arquitetura do *chatbot* e de parâmetros com base nas medidas do BLEU e ROUGE-L.

Rede	Células	Épocas	BLEU	ROUGE-L		
				P	C	F1
GRU	500	2000	44,96	82,44	59,66	64,49
GRU	500	4000	28,65	85,50	41,46	50,30
GRU	1000	2000	17,91	83,40	45,76	53,94
GRU	1000	4000	11,32	85,90	24,07	34,83
BiLSTM	500	2000	15,80	82,15	30,07	41,25
BiLSTM	500	4000	39,12	84,84	56,59	61,99
BiLSTM	1000	2000	11,25	83,43	32,54	44,11
BiLSTM	1000	4000	10,30	85,00	25,20	36,37

Analisando os resultados obtidos, é possível identificar que o modelo GRU obteve resultados superiores do que a rede BiLSTM em quase todos os cenários avaliados. Em especial, a configuração usando a rede GRU com 500 células e sendo treinada por 2000 épocas obteve o melhor desempenho com base nas medidas de cobertura, medida-F e BLEU. O maior valor obtido na medida de precisão foi alcançado também usando a rede

⁵<https://github.com/Diego999/py-rouge>

⁶https://www.nltk.org/_modules/nltk/translate/bleu_score.html

GRU com 1000 células e treinada por 4000 épocas. Os resultados obtidos indicam que o aumento no número de épocas ou de células não implicou, necessariamente, em um aumento de desempenho.

Outro aspecto a ser observado é que os resultados alcançados na medida de precisão do ROUGE-L foram maiores que 80,00%, enquanto valores inferiores foram obtidos para medida de cobertura, no qual o melhor resultado foi de 59,66%. Tais resultados indicam que, em geral, as respostas geradas pelo *chatbot* foram mais curtas do que as respostas de referência presentes na base de dados.

Na Figura 3 é apresentado um trecho de diálogo utilizando o *chatbot* usando a rede GRU. A ferramenta foi nomeada de Assistente de Inteligência Técnica, em inglês *Technical Intelligence Assistant* (T.I.A), para melhor interação com os usuários. No trecho de diálogo apresentado, o usuário faz quatro perguntas sobre conceitos distintos e o *chatbot* consegue gerar respostas para cada uma das perguntas. Pela sua abordagem generativa, o sistema pode acabar gerando repetições desnecessárias ou então concluir as respostas de maneira inadequada. Por exemplo, ao responder à pergunta “*Conceptualize Bulletin Order*” a resposta gerada repete a palavra “*railroad*” duas vezes seguidas.

```
Pergunta > : what is Interurban
T.I.A: a termed applied to electrically powered commuter
Pergunta > : what is Lading
T.I.A: refers to the cargo of the truck to the burlington
Pergunta > : conceptualize Bulletin order
T.I.A: notifications given a railroad railroad informing of them to a operational change
Pergunta > : what is the definition of container
T.I.A: standardized foot freight by rail truck or ship by a truck or ship by a ship
```

Figura 3. Exemplo de um trecho de diálogo usando o *chatbot* desenvolvido.

Outra situação que acontece de maneira recorrente em linguagem natural é a formulação de uma mesma pergunta para um determinado assunto de várias formas diferentes. Tal cenário pode fazer com que o *chatbot* gere respostas diferentes, afetando assim o resultado apresentado ao usuário. Na Figura 4 é ilustrada essa situação, no qual o usuário realiza uma mesma pergunta de três maneiras diferentes. Mesmo nesse cenário, o sistema conseguiu identificar que as perguntas possuem semânticas similares e assim gerar a mesma resposta.

```
Pergunta > : container is
T.I.A: standardized foot designed for shipping freight by rail truck or ship car
Pergunta > : what is the definition of container
T.I.A: standardized foot designed for shipping freight by rail truck or ship car
Pergunta > : talk about container
T.I.A: standardized foot designed for shipping freight by rail truck or ship car
```

Figura 4. Exemplo de diálogo com o *chatbot* com uma pergunta sendo feita de maneiras diferentes.

Por fim, ainda é possível que sejam feitas perguntas que estejam fora do domínio para o qual o sistema foi treinado. Nesse cenário, o *chatbot* consegue gerar uma resposta

por conta da sua abordagem generativa, mas observou-se que as saídas geradas não respondiam de fato o que se estava sendo questionado. Como tratar cenários como esse é um desafio para sistemas de conversação.

5. Considerações Finais e Trabalhos Futuros

Neste trabalho foi apresentada a proposta de um sistema de diálogo (*chatbot*) como uma ferramenta de apoio ao ensino de disciplinas na área de ferrovias. Para isso, foi construída uma base de conhecimento contendo perguntas e respostas de domínio geral e específico escritas em inglês. A base construída possui 11108 perguntas e respostas, sendo 10000 de domínio geral e 1108 de domínio específico. A base de dados proposta foi, então, utilizada no desenvolvimento de um *chatbot* generativo utilizando a arquitetura codificador-decodificador com mecanismo de atenção. Duas variações da arquitetura foram analisadas: uma utilizando redes do tipo BiLSTM e outra adotando redes GRUs.

Experimentos foram realizados para comparar as duas variações de arquitetura adotadas no desenvolvimento do *chatbot* proposto com base nas medidas de avaliação do BLEU e ROUGE-L. Além disso, foi realizada uma análise do impacto da variação do número de células usadas nas redes neurais e do número de épocas usadas no treinamento dos modelos. Os resultados experimentais demonstraram que a arquitetura usando redes GRUs obteve o melhor desempenho quando comparado com o modelo baseado na rede BiLSTM, atingindo um valor de 44,96% na medida BLEU e um valor na medida-F igual a 64,69% para o ROUGE-L.

Como trabalhos futuros, pretende-se expandir a base de dados construída, especialmente com a inclusão de novas perguntas e respostas de domínio específico da área de ferrovias. Além disso, uma linha de pesquisa futura que pode ser interessante é explorar a arquitetura *Transformer* [Vaswani et al. 2017], que tem obtido bons resultados em diversas tarefas de PLN que envolvem geração de linguagem.

Agradecimentos

Os autores agradecem ao Ifes, apoio da FAPES e CAPES (processo 2021-2S6CD, nº FAPES 132/2021) por meio do PDPG (Programa de Desenvolvimento da Pós-Graduação, Parcerias Estratégicas nos Estados).

Referências

- AbuShawar, B. and Atwell, E. (2015). Alice chatbot: Trials and outputs. *Computación y Sistemas*, 19(4):625–632.
- Adamopoulou, E. and Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2:100006.
- Aleedy, M., Shaiba, H., and Bezbradica, M. (2019). Generating and analyzing chatbot responses using natural language processing. *International Journal of Advanced Computer Science and Applications*, 10(9).
- Arora, B., Chaudhary, D. S., Satsangi, M., Yadav, M., Singh, L., and Sudhish, P. S. (2020). Agribot: a natural language generative neural networks engine for agricultural applications. In *2020 International Conference on Contemporary Computing and Applications (IC3A)*, pages 28–33. IEEE.

- Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Caldarini, G., Jaf, S., and McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, 13(1).
- Chang, I.-C., Shih, Y.-S., and Kuo, K.-M. (2022). Why would you use medical chatbots? interview and survey. *International Journal of Medical Informatics*, 165:104827.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.
- Dabre, R., Chu, C., and Kunchukuttan, A. (2020). A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR)*, 53(5):1–38.
- Dhyani, M. and Kumar, R. (2021). An intelligent chatbot using deep learning with bidirectional rnn and attention model. *Materials today: proceedings*, 34:817–824.
- Galassi, A., Lippi, M., and Torroni, P. (2020). Attention in natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10):4291–4308.
- Galvão, A., Barros, F., Neves, A., and Ramalho, G. (2004). Persona-aiml: An architecture developing chatterbots with personality. pages 1266–1267.
- Hou, S.-L., Huang, X.-K., Fei, C.-Q., Zhang, S.-H., Li, Y.-Y., Sun, Q.-L., and Wang, C.-Q. (2021). A survey of text summarization approaches based on deep learning. *Journal of Computer Science and Technology*, 36(3):633–663.
- Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
- Kim, J., Oh, S., Kwon, O.-W., and Kim, H. (2019). Multi-turn chatbot based on query-context attentions and dual wasserstein generative adversarial networks. *Applied Sciences*, 9(18):3908.
- Kuhail, M. A., Alturki, N., Alramlawi, S., and Alhejori, K. (2022). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, pages 1–46.
- Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., and Guo, Y. (2021). Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (gru). *Information Processing in Agriculture*, 8(1):185–193.
- Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.
- McTear, M. (2020). *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Misischia, C. V., Poecze, F., and Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. *Procedia Computer Science*, 201:421–428.

- The 13th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 5th International Conference on Emerging Data and Industry 4.0 (EDI40).
- Mondal, A., Dey, M., Das, D., Nagpal, S., and Garda, K. (2018). Chatbot: An automated conversation system for the educational domain. In *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pages 1–5. IEEE.
- Neves, A., Barros, F., and Hodges, C. (2006). iauml: a mechanism to treat intentionality in aiml chatterbots. In *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*, pages 225–231.
- Oliveira, H. T. A. d., Gadelha, R. N., de Azevedo, R. R., Júnior, J. B. D., Dias, G. A., and Freitas, F. (2010). Dr. pierre: Um chatterbot com intenção e personalidade baseado em ontologias para apoiar o ensino de psiquiatria. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. *ACL '02*, page 311–318, USA. Association for Computational Linguistics.
- Rana, R. (2016). Gated recurrent unit (gru) for emotion classification from noisy speech. *arXiv preprint arXiv:1612.07778*.
- Sawant, S., Vishwakarma, A., Sawant, P., and Bhavathankar, P. (2021). Analytical and sentiment based text generative chatbot. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7.
- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14*, page 3104–3112, Cambridge, MA, USA. MIT Press.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX(236):433–460.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Vieira, M. and Silva, C. (2020). A educação no contexto da pandemia de covid-19: uma revisão sistemática de literatura. *Revista Brasileira de Informática na Educação*, 28(0):1013–1031.
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45.
- Zalake, N. and Naik, G. (2019). Generative chat bot implementation using deep recurrent neural networks and natural language understanding. In *Proceedings 2019: Conference on Technologies for Future Cities (CTFC)*.
- Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., and Dolan, B. (2020). Dialogpt: Large-scale generative pre-training for conversational response generation. In *ACL, system demonstration*.