12º ENTEC – Encontro de Tecnologia: 16 de outubro a 29 de novembro de 2018

A APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS RECORRENTES NO PROCESSAMENTO DE LINGUAGEM NATURAL.

Germano Renner de Oliveira Araújo[1]*,William de Oliveira Vittorazzi*[2]*, Florisvaldo Cardozo Bomfim Junior*[3]
[1,2] *Universidade de Uberaba*
germanorenner010@gmail.com, wvittorazzi@gmail.com, florisvaldo.bomfim@uniube.br

Resumo

Este trabalho visa abordar a utilização de técnicas de *Machine Learning*. Estas são amplamente utilizadas na área da computação cognitiva e fomentaram os estudos em diversas esferas da tecnologia, como segurança da informação, entretenimento e no setor fabril, ocasionando o episódio que se denominou como a quarta revolução industrial.

Tais circunstâncias inspiraram o surgimento de aplicações que rompem barreiras, antes insuperáveis. Entre os desafios enfrentados pela ciência, podemos destacar o Processamento de Linguagem Natural (PLN), que é a subárea da Inteligência Artificial, a qual, permite que computadores criem redações, entendam contextos e até mesmo emoções extraídas de histórias.

Com a utilização de métodos clássicos e modernos relacionados à Inteligência Artificial, buscamos desenvolver uma aplicação que fosse capaz de interagir com usuários de maneira coerente e racional.

O procedimento consiste em diferentes etapas, como a alimentação de informações ao sistema, o gerenciamento e processamento das mesmas, e por fim a interface gráfica, sendo esta o meio por onde o usuário irá realizar a interação com o software.

Vale destacar que o objeto em estudo não é a interface do aplicativo, mas sim o aprendizado da máquina após o devido treinamento, de modo com que as respostas emitidas tenham coerência, apresentando a capacidade de interpretação e comunicação, desenvolvidas através do longo processo de capacitação da rede.

Palavras-chave: Machine Learning. Inteligência Artificial. Processamento de Linguagem Natural. Deep Learning. Redes Neurais Artificiais.

1 Introdução

O atual cenário tem como peça fundamental, a tecnologia da informação, que utiliza a abundância dos dados providos pelos usuários para a criação de novas estratégias. Devido a isso, estão surgindo novas ferramentas que enfatizam a importância da mineração de dados na elaboração de soluções para problemas computacionais, como é o caso do PLN, que tem como foco principal, converter a linguagem humana para linguagem de máquina.

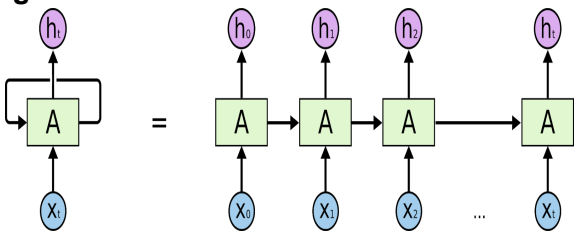
Conforme definido por Elizabeth Liddy, “O processamento de linguagem natural são teorias motivadas por uma série de técnicas computacionais para análise e representação de textos

decorrentes da linguagem natural. Essas técnicas são utilizadas com o objetivo de processar linguagens humanas para diversas aplicações."

Em virtude da complexidade dos processos neste domínio, a evolução das pesquisas no campo de Redes Neurais Artificiais (RNA) alavancaram os estudos e desenvolvimento de metodologias que permitiram superar algumas adversidades impostas pelo PLN.

Este trabalho tem como ênfase o aprendizado da máquina utilizando um modelo de RNA bem popular em aplicações de PLN, que são as Redes Neurais Artificiais Recorrentes (*Recurrents Neural Networks - Rnn*), com o intuito de explorar os desafios proporcionados e destacar o campo de estudo em questão. Tal modelo de rede tem uma grande vantagem em relação aos modelos mais básicos que é a persistência da informação. Em redes neurais tradicionais a informação se propaga pela rede de forma com que ela não seja armazenada, isso acaba dificultando o desenvolvimento de certas aplicações. As RNNs são estruturadas de modo que possuam um loop no qual a informação é passada de uma camada para outra, permitindo assim, que se propague através da rede.

Figura 1: Modelo de Rede Neural Recorrente.



Fonte: colah.github.io (2015)

Além do reconhecimento de padrões em textos, as RNNs tem um histórico com resultados positivos quando aplicadas a outros problemas como: traduções, criação de legendas para imagens, reconhecimento de falas, entre outros.

O objetivo central deste trabalho é o desenvolvimento de uma aplicação que seja capaz de associar entradas de texto de um usuário à saídas condizentes, aprendidas por uma RNA no seu processo didático, de modo que possamos aprofundar o conhecimento e contribuir com os estudos nas áreas de Redes Neurais Artificiais e Processamento de Linguagem Natural, uma vez que as pesquisas nesses campos vem se intensificando nos últimos anos, resultado do crescimento da exploração e uso da inteligência artificial.

2 Materiais e Métodos

Inicialmente foi feita uma busca para conseguir uma base de dados com diálogos humanos, de maneira que pudesse ser realizado o treinamento da rede neural. Após a aquisição, deu se início à fase de tratamento e a formatação da mesma. Neste estágio, realizamos o pré-processamento do conjunto de informações, eliminando palavras irrelevantes e tornando o vocabulário mais objetivo.

Quando se fala em manipulação e análise de dados não pode se esquecer de citar três ferramentas que auxiliaram no desenvolvimento, são elas: *Anaconda, Spyder e TensorFlow.* Anaconda é uma ferramenta que permite a criação de ambientes e módulos de trabalho para análise de dados baseada em Python. Já o Spyder é uma plataforma de desenvolvimento utilizada no Anaconda que possibilitou escrever, debugar e executar todo o código. Por último e não menos importante devemos destacar a importância do Framework em Python TensorFlow, que possui funções para manipulação de matrizes e vetores

multidimensionais facilitando a criação de redes neurais artificiais.

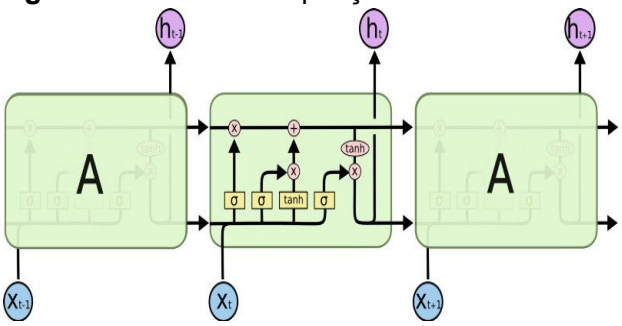
O tratamento das informações contidas na base original permitiu que a mesma ficasse organizada, de forma que a rede neural pudesse acessar e utilizá-la como entrada para o treinamento a fim de começar o reconhecimento dos padrões. Para realizar a manipulação da base contamos com a biblioteca *Pandas,* que permitiu analisar e manipular os dados de forma mais simplificada e ágil, através das funções integradas.

Com os recursos oferecidos pelas ferramentas citadas acima, a fase de formatação dos dados não foi problema, e então se iniciou a próxima etapa que foi a definição dos parâmetros que seriam utilizados na rede, como a taxa de aprendizado em **0.01**, a taxa de *dropout* em **0.5**, e o número de épocas em **100.** Depois disso, teve início o desenvolvimento da Rede Neural Recorrente. Como dito anteriormente, neste trabalho foi utilizado um modelo específico de rede capaz de armazenar a informação através das camadas. Esta é uma derivação das RNNs e é chamada de "*Long Short Term Memory Networks - LSTM".*

As LSTMs foram introduzidas por Hochreiter & Schmidhuber (1997) quando surgiram alguns problemas em redes recorrentes e o grande desafio era guardar informações requisitadas em um futuro não tão próximo do processamento atual. A solução nesse caso seria uma rede que respondesse com saídas plausíveis, à entradas não tão recentes, e com esse intuito as LSTMs foram desenvolvidas.

Originadas a partir das RNNs podemos observar que a estrutura das LSTMs não difere tanto dos tipos tradicionais de redes recorrentes, fazendo com que a diferença maior se encontre no número de camadas das redes neurais nos módulos de repetição, o qual em LSTMs são quatro camadas e nas redes tradicionais são apenas uma.

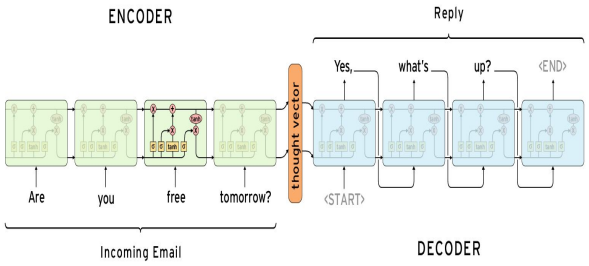
Figura 2: Módulo de repetição de uma LSTM.



Fonte: complx.me (2016).

Baseando-se na arquitetura chamada de "*Sequence to Sequence - Seq2Seq*" iniciamos mais uma etapa do desenvolvimento, que era a rede em questão. Esta arquitetura conta com duas RNNs uma é camada de codificação (*encoder layer*) dos valores da entrada e a outra é camada de decodificação (*decoder layer*) dos valores recebidos da saída da camada de entrada. Conforme explica o pesquisador Ramamoorthy (2016), " O codificador lê a seqüência de entrada, palavra por palavra e emite um contexto (uma função do estado oculto final do codificador), que idealmente capturaria a essência (resumo semântico) da seqüência de entrada. Com base nesse contexto, o decodificador gera a sequência de saída, uma palavra por vez, enquanto observa o contexto e a palavra anterior durante cada timestep".

Figura 3: Arquitetura Sequence to Sequence.



Fonte: complx.me (2016)

Com a nossa Seq2Seq finalizada partimos para uma nova fase do processo, que é o treinamento da rede. Esta é a fase mais crítica e demorada, pois é a fase onde serão recalculados todos dos pesos da rede que se responsabilizarão pelas saídas geradas.

Após alguns dias de treinamento, observamos que já tínhamos uma rede neural capaz de receber entradas de textos e gerar respostas com as informações recebidas. Diante disso, pensando em proporcionar uma melhor experiência para os usuários, a fase seguinte e final se fez com o desenvolvimento de uma aplicação *Front-end* que é composta por uma interface gráfica desenvolvida com o framework Javascript *Angular 6.* A interface, foi responsável por realizar a interação do usuário com o chatbot, tornando mais agradável a experiência.

3 Resultados

Devido a limitações de tempo e baixo poder de processamento (comparado ao processamento de dados a nível de aplicações científicas), a fase de treinamento da rede teve de ser interrompida, significando que a rede não alcançou as 100 épocas como programado.

Tal interrupção expôs o projeto a falhas e resultados não tão desejáveis. Como a rede não foi capaz de realizar a associação dos dados por mais tempo, o estímulo responsivo a certas entradas foi de palavras fora do contexto das frases. Isso não quer dizer que a rede ou a lógica não esteja funcionando, mas sim que o treinamento só precisa de mais tempo para praticar ou um maior poder de processamento e ter uma melhor compreensão de seu propósito na aplicação. Tal empecilho enfrentado é na verdade o mais comum no estudo científico da atualidade: o poder de processamento alcançando os limites físicos das unidades de processamento (CPUs).

4 Discussão

Levando em consideração os trabalhos de estudiosos da área, como Hochreiter e Schmidhuber, é possível notar a diferença no método de funcionamento das redes quando são expostas a cenários com grande volume de dados e grande requisição dos mesmos.

As RNAs simples, que funcionam de maneira excelente quando se tratam de problemas com soluções lineares, podem ser a base para a solução de grandes problemas, mas com certeza não são a resposta. Principalmente quando o sistema é composto de características como as tais da área de Processamento de Linguagem Natural.

Em um artigo divulgado pela rede de publicação inglesa *The Economist* em 2013, revela que um adulto fluente na língua inglesa tem um vocabulário médio variando de 20.000 a 35.000 palavras. Comparando se com este estudo e sabendo que o banco de informações utilizado tem como base a língua inglesa, conseguimos estimar que um Chatbot,

para conseguir alcançar um estado de simulação comparável com um humano, deve realizar um grande número de cálculos associativos, aprendendo as palavras e em que contexto utilizá-las.

5 Conclusão

Reparando os passos tomados e inferindo os resultados obtidos, é possível compreender que redes neurais artificiais são um dos principais métodos para manipulação de dados e tratamento de informações, o que automaticamente transforma grande parte dos estudos de Inteligência Artificial, dependentes de RNAs. Ao aprofundar na matéria de manipulação de dados, rapidamente percebe-se que são muitas as variações dos tipos de redes neurais artificiais e métodos existentes para criação das mesmas. Podendo de tal forma, tomar como exemplo as próprias RNNs, que não se satisfazem com uma simples estrutura e exigem uma formação personalizada para o Processamento e Aprendizagem de linguagem Natural.

Tudo isso fica muito claro ao realizar o desenvolvimento de uma aplicação que interage diretamente com os usuários utilizando linguagem natural.

Referências

BENGIO, Y.; SIMARD P.; FRASCONI P. **Learning Long-Term Dependencies with Gradient Descent is Difficult.** 1994.

BRITZ, Denny. Deep Learning for Chatbots, Part 1 - Introduction. 2016. Disponível em: <**http://www.wildml.com/2016/04/deep-learning-for-chatbots-part-1-introduction/**>. Acesso em: **26 Out**. 2018.

OLAH, Christopher. **Understanding LSTM Networks**. 2015. Disponível em: <**http://colah.github.io/posts/2015-08-Understanding-LSTMs/**>. Acesso em: **26 Out**. 2018.

RAMAMOORTHY, Suriyadeepan. **Revisiting sequence to sequence learning, with focus on implementation details**. 2011. Disponível em: <http://complx.me/2016-12-31-practical-seq2seq//>. Acesso em: 25 Out. 2018.