

USO DE IA PARA MELHORIA EM BASE DE DADOS ELASTIC

THOMAS DE AGUIAR DE OLIVEIRA¹; GUSTAVO LIMA DE MAGALHÃES²;
LEANDRO WEIGE DIAS³

¹Universidade Federal de Pelotas – thomas.aguiar@ufpel.edu.br

²Universidade Federal de Pelotas – gustavo.magalhaes@ufpel.edu.br

³Universidade Federal de Pelotas – leandro.dias@ufpel.edu.br

1. INTRODUÇÃO

Nos últimos anos inúmeros sistemas vem sendo desenvolvido com a finalidade de obtenção de dados, sendo o principal processo de obtenção dessas informações o meio conhecido como mineração de dados, e ou, o envio espontâneo por meio de sites de pesquisas ou redes sociais. Este processo vem tornando a área conhecida como big data cada vez mais popular e presente no dia-a-dia de quem utiliza os meios de comunicação.

O avanço no acúmulo de informações permitiu a criação de sistemas robustos, especializados em sintetizar as informações para utilizações futuras, que vão desde a sugestão de sites que o usuário deseja abrir ao iniciar uma pesquisa no navegador, até o de enviar propagandas à vários sites referentes a produtos que o usuário possui interesse em comprar.

No entanto, transformar esta base de dados em algo útil requer um custo considerável de processamento, pois na medida que essas bases de dados se expandem é exigido cada vez mais capacidade no processamento destas informações, no entanto o avanço nos hardwares de processamento não tem acompanhado o avanço das tecnologias de expansão da memória, causando um grande problema para estas empresas que precisam trabalhar com um grande volume de dados com uma certa urgência.

Uma vez explorado e mantido estas informações chega a hora de atualizá-las e processá-las para que possam atingir o melhor resultado possível e é neste momento que entra a solução que será apresentada neste trabalho. Utilizaremos para isto, duas ferramentas que se convergem. Se por um lado temos o poder de uma base de dados voltada a documentos e um motor de pesquisa de informações potente conhecido como elasticsearch, por outro lado, temos o poder e a capacidade de entender e aprender o que está sendo pesquisado através da utilização de redes neurais artificiais, que neste trabalho, será analisado pela ferramenta Tensor Flow.

O objetivo deste trabalho é desenvolver a hipótese de que dados pré processados e desenvolvidos por uma rede neural artificial pode impactar positivamente os resultados do elasticsearch quando utilizados de maneira adequada.

2. METODOLOGIA

Será utilizado o elasticsearch como mecanismo de busca e indexação de documentos, sua escolha se deu por ser um sistema de código aberto e o buscador empresarial mais popular atualmente no mercado, além de possuir uma API que a torna extremamente compatível com diversas outras soluções.

Para o desenvolvimento da rede neural artificial, será utilizado o Tensor Flow, pois além de ser o sistema mais utilizado para esta finalidade devido a sua

facilidade de uso e instalação, este também possui uma API, o que facilitará esta integração ao elasticsearch.

Como amostra, foi utilizado os dados do portal institucional da UFPel, que trata de um conjunto de dados organizados em cinco grandes grupos, tais como: cursos, disciplinas, projetos, servidores e unidades.

O trabalho consiste em fazer a rede neural artificial entender e aprender, que tipo de resultado (dos cinco grandes grupos descritos acima) o usuário deseja obter quando digitar um determinado termo no campo de busca e passar ao ElasticSearch esta informação.

O processo de aprendizagem profunda, portanto, se dará através da comparação e análise entre dados pesquisados e dados acessados depois da pesquisa em um determinado período. Fazendo com que os scores obtidos pelos resultados no ElasticSearch sejam incrementados com os scores obtidos pelo Tensor Flow.

Através de logs será registrado as buscas bem-sucedidas no portal institucional.ufpel.edu.br, na qual será registrado a posição que este item estava na pesquisa naquele momento com e sem o incremento do Tensor Flow.

3. RESULTADOS E DISCUSSÃO

O trabalho está em etapa de obtenção e análise de amostragem, esperamos baseado na bibliografia utilizada, que aconteça uma melhoria de no mínimo dez por cento (10%) nas soluções apresentadas, ou seja, com a utilização de uma rede neural artificial, mesmo que de maneira básica como a observação de grandes grupos, tenhamos uma sensível diferença entre as pesquisas com e sem o Tensor Flow.

Acredita-se também, que quanto mais especializada for a pesquisa e a tipificação dos dados enviados para a rede neural artificial, mais impacto causará no resultado final da pesquisa com o elasticsearch, portanto, treinar uma rede neural artificial para entender o tipo de pesquisa e atribuir score de acordo com o tipo de conteúdo que se deseja encontrar, apesar de parecer um passo a mais no momento de se desenvolver, parece ter um resultado positivo para os usuários finais no resultado das pesquisas.

4. CONCLUSÕES

Como o sistema de Rede Neural Artificial foi implementado após a construção da base e dos índices do elasticsearch, e naquele momento não se tinha previsão para a utilização deste quando o sistema foi desenvolvido, então alguns erros poderiam ter sido observados e corrigidos no início, tal como a construção de índices próprios para a rede neural diretamente nos itens dos objetos cadastrados, evitando a necessidade da inclusão de um score externo.

Este trabalho se focou na utilização do elasticsearch em seu modo de uso tradicional, evitamos a ativação de sistemas como o do slowlog que tenta simular o uso de uma rede neural artificial interna do elasticsearch, este modo não é aconselhável ainda para uso em ambientes de produção devido a lentidão que causa nos sistemas de pesquisa. Em um trabalho futuro será realizado a implementação e comparação do slowlog com a rede neural artificial do tensorflow.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Arnold, S. (2014). Rumble in the Search Jungle: The ELK Invasion. *Online Searcher*, 38(4), 16-20.

Awate, G., Bangare, S., Pradeepini, G., & Patil, S. (2018). Detection of Alzheimers Disease from MRI using Convolutional Neural Network with Tensorflow. *IEEE Xplore* 2018.

Chen, Dequan, Chen, Yi, Brownlow, Brian N., Kanjamala, Pradip P., Arredondo, Carlos A. Garcia, Radspinner, Bryan L., & Raveling, Matthew A. (2017). Real-Time or Near Real-Time Persisting Daily Healthcare Data Into HDFS and ElasticSearch Index Inside a Big Data Platform. *Industrial Informatics, IEEE Transactions on*, 13(2), 595-606.

Neubig, G., Dyer, C., Goldberg, Y., Matthews, A., Ammar, W., Anastasopoulos, A., . . . Yin, P. (2017). *DyNet: The Dynamic Neural Network Toolkit*.

Shanshan Han, Fu Ren, Chao Wu, Ying Chen, Qingyun Du, & Xinyue Ye. (2018). Using the TensorFlow Deep Neural Network to Classify Mainland China Visitor Behaviours in Hong Kong from Check-in Data. *ISPRS International Journal of Geo-Information*.

Takase, W., Nakamura, T., Watase, Y., & Sasaki, T. (2017). A solution for secure use of Kibana and Elasticsearch in multi-user environment.

Mu, C., Zhao, J., Yang, G., Zhang, J., & Yan, Z. (2018). Towards Practical Visual Search Engine within Elasticsearch.

Taylor, Ali, & Varley. (2018). Automating the processing of data in research. A proof of concept using elasticsearch. *International Journal of Surgery*, 55, S41.

Thomas Johnson. (2013). Indexing Linked Bibliographic Data with JSON-LD, BibJSON and Elasticsearch. *Code4Lib Journal*, (19), *Code4Lib Journal*, 01 January 2013, Issue 19.