

Destaques da leitura de **A Neural Probabilistic Language Model**, Bengio 2003

Nome: Caio Petrucci dos Santos Rosa

RA: 248245

Os destaques da leitura do artigo foram os seguintes:

- A abordagem proposta para resolver o problema da **curse of dimensionality** baseado da redução da dimensão do espaço através da aprendizagem de representações semânticas das palavras, de forma que o modelo aprenda em conjunto uma probabilidade de representação das palavras e a probabilidade de sequências de palavras representadas nesse espaço semântico. Isso é interessante pois, na predição de sequências de palavras, há uma gigantesca quantidade de possibilidades para a próxima palavra, que é o que o autor trata como a **curse of dimensionality**;
- A apresentação das técnicas consideradas *state-of-the-art* anteriores a esse trabalho: era basicamente a utilização de *n-grams* que modelavam a probabilidade de uma sequência de tamanho *n* aparecer para predizer a próxima palavra. Nesse sentido, a abordagem proposta é um pouco similar, dado que o modelo neural elaborado precisa de uma janela de contexto de *n* palavras para funcionar, sendo a maior mudança o aprendizado de uma representação semântica das palavras, o que possibilita aproximar sinônimos, enquanto os *n-grams* não são capazes de fazer isso.
- A forma computacional como o treinamento da rede neural proposta é realizado, com 40 CPUs treinando de forma paralela, o que mostrou a necessidade de lidar com problemas de sincronização entre os processadores. Isso mostra a diferença entre como estes modelos eram treinados naquela época se comparado a como são treinados hoje em dia com a existência de GPUs;
- Achei bastante interessante as comparações entre diferentes implementações da arquitetura proposta e os modelos já existentes (assim como combinações entre eles) que foram realizadas e cujo os resultados foram mostrados na Tabela 1. Essas comparações, por exemplo, mostraram que o uso de *skip-connections*, representadas como *direct* na Tabela 1, não foram eficientes para a tarefa de predição nessa arquitetura;
- Os autores também propõem possíveis melhorias para a arquitetura elaborada, entre elas a alteração da camada de saída para o *output* ser o *embedding* de uma palavra em vez de seu índice. Essa abordagem é interessante pois facilitaria a predição de sinônimos de palavras fora do vocabulário do modelo.