# Natural Language Processing
# A study of Language Models using word2vec

**Isabel B. Amaro**[1], **Adriano Veloso**[2]

[1]Department of Computer Science – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, Brazil

{isabel.amaro,adrianov}@dcc.ufmg.br

***Abstract.***

## Introduction

With the appearing of social networks, a massive amount of data is constantly being generated weekly in the last few years (CAMBRIA, 2014). Natural Language Processing is a Computer Science field that studies the best ways to retrieve, process and generate from representations of human language (CAMBRIA, 2014). To

This work analyses

## Language Models

### word2vec

Word2vec can utilize either of two model architectures to produce a distributed representation of words: continuous bag-of-words (CBOW) or continuous skip-gram. In the continuous bag-of-words architecture, the model predicts the current word from a window of surrounding context words. The order of context words does not influence prediction (bag-of-words assumption). In the continuous skip-gram architecture, the model uses the current word to predict the surrounding window of context words. The skip-gram architecture weighs nearby context words more heavily than more distant context words.[1][4] According to the authors' note,[5] CBOW is faster while skip-gram is slower but does a better job for infrequent words.

### Continuous Bag-Of-Words (CBOW)

### Skip-Gram

### Result and analysis

### Conclusion