Notes on:
DeepWalk - Online Learning of Social Representations

August 27, 2024

1 Abstract

DeepWalk was presentes as a novel approach for learning representation of vertices in networks. This is done by random walks to learn representations and converting them to sentences (embeddings). The increase in F1 scores was up to 10% compared to other methods for social networks like Youtube.

2 Introduction

Machine learning for networks must be able to cope with the sparsity of network classification, anomaly detection and missing link prediction. DeepWalk generalises neural language models to process a special language composed of a set of randomly-generated walks. It takes a Graph as an input and produces a latent representation as output. To demonstrate potential of this method, it was utilised in challenging multi-label network classification in large heterogeneous graphs.

Contributions:

1. Introduce deep learning as a tool to analyse graphs. DeepWalk learns structural features.
2. Evaluated results with a 5-10% improvement compared to toher methods.
3. Demonstrated the scalability of the algorithm by using Youtube's big scale data/graphs.

3 Problem definition

Let $G = (V, E)$, where V are the memebers of the network (nodes or vertices) and E its edges, $E \subseteq (V \times V)$ (A subset of the cartesian product of V) Given a partially labeled social network $G_L = (V, E, X, Y)$, with attributes $X \in \mathbb{R}^{|V| \times S}$

where S is the size of the feature space for each attribute vector, and $Y \in \mathbb{R}^{|V| \times |\mathcal{Y}|}$, $\mathcal{Y}$ is the set of labels.

In traditional machine learning, we aim to learn H that maps elements of X to the set of labels $\mathcal{Y}$. DeepWalk uses the embedded information in the structure G to achieve superior performance.

This method instead of mixing the label space as part of the feature space, an unsupervised method which learns the graph structure independent to the labels.