Kernel Node Embeddings

Abdulkadir Çelikkanat and Fragkiskos D. Malliaros
CentraleSupélec, Inria, Université Paris-Saclay

Introduction

• Network representation learning (NRL) aims to encode the structure of a network into low-dimensional vectors
• Applications in network analysis: visualization, classification, community detection and link prediction
• In this work:
  ○ We propose a novel approach for learning node embeddings by incorporating kernel functions with models relying on weighted matrix factorization
  ○ We perform extensive performance evaluation of the proposed method in two downstream tasks

Proposed Approach

• We define the general objective function of our problem as a weighted matrix factorization:
  \[
  \arg \min_{A,B} \frac{1}{2} \| W \odot (M - AB^T) \|^2_F \tag{1}
  \]
  
  • By setting each term \( W_{uv} \) as the square root of the number of occurrences of \( u \) in the contexts of \( v \), the objective in (1) becomes:
    \[
    \arg \min_{A,B} \frac{1}{2} \| \sqrt{F} \odot (M - AB^T) \|^2_F = \arg \min_{A,B} \frac{1}{2} \sum_{u,v} F_{u,v} (M_{u,v} - (A_u, B_v))^2 \tag{2}
    \]

  \( M_{u,v} \) represents if \( u \) appears in the context of \( v \) in any walk \( \langle v_1, v_3, v_5, v_4, v_7 \rangle \)
  \( F_{u,v} \) is the number of occurrences of \( u \) in the contexts of \( v \) \( \langle v_1, v_3, v_5, v_4, v_7, v_6 \rangle \)

  Figure: Schematic representation of node embeddings

• The inner product in Eq. (2) can be expressed in the feature space as follows:
  \[
  \arg \min_{A,B} \frac{1}{2} \sum_{u,v} \sum_{w \in W} \sum_{l \in u,v} (M_{u,v} - (\Phi(A_u), \Phi(B_v)))^2 \]
  \[
  \approx \arg \min_{A,B} \frac{1}{2} \sum_{u,v} \sum_{w \in W} \sum_{l \in u,v} (M_{u,v} - (A_u, B_v))^2 \]

  We use the following universal kernels [1, 2] in our evaluation:
  \[
  \kappa_C(x, y) = \exp \left( -\frac{\| x - y \|^2}{\sigma^2} \right) \quad \kappa_S(x, y) = \frac{1}{1 + \| x - y \|^2}
  \]

Experimental Setup

• For optimization, we employ Stochastic Gradient Descent (SGD)
  • We apply negative sampling strategy: \( k \) negative instances \( u\'s \) are sampled from the noise distribution \( \rho \) for each context node \( u \):
    \[
    \left( 1 - \kappa(A_u, B_u) \right)^2 + \sum_{u\' \neq u} \left( \kappa(A_u, B_u) \right)^2.
    \]
  • In experiments, we use logistic regression with \( L_2 \) regularization.

Table: Statistics of networks used in the experiments. \( |V| \): number of nodes, \( |E| \): number of edges, \( |K| \): number of labels and \( |C| \): number of connected components.

Parameter Sensitivity

Numerical Tests

Table: Area Under Curve (AUC) scores for the link prediction task

References