A step ahead for generation of explainable embeddings using feature maximization principle: Application to graph and word embeddings

Jean-Charles Lamirel[1, 2] : *mail: lamirel@loria.fr; phone: +33624365491*

[1] Infocom Dept., University of Strasbourg, France
[2] WISELAB, University of Technology of Dalian, China .

Keynote abstract: On the one side, graph embedding goal is to learn low-dimensional representations of nodes enclosing the graph topology. On the other side, word embedding aims at learning word vectors that encode semantic properties of the vocabulary. Graph embedding has numerous applications on complex networks related tasks such as link prediction and node classification, whilst the range of application of word embedding is very large in natural language processing tasks. Graph and word embeddings are usually considered as dissimilar tasks. However, word co-occurrence matrices, widely exploited to perform word embeddings have a graph-based representation. Furthermore, the majority of network embedding techniques rely either on methodologies like random walk or matrix factorization that are also used for word embedding and take as a prior to obtain low-dimensional representations. These are usually computationally expensive methods that depend on several parameters. In all cases, the dimensions of the obtained embedding space are not interpretable. To cope with these issues, we introduce a new framework that is based on the already validated hypothesis that all graphs can be represented as bipartite graphs. This underlying bipartite structure may be explicit, as it is meet with co-author networks based on paper-author relationships. However, with our framework we focus on discovering latent bipartite structures, lying for instance in social networks or even in word co-occurrence networks. We also put the focus on structures providing condensed and interpretable representations of the considered graph. We experimentally show on usual test benchmarks that our approach provides as efficient embeddings as the concurrent methods, whilst preserving the sparsity, which represents a necessary condition for providing explainable representation in an embedding space. Additionally, we also show that the computational complexity of our approach is far beyond the ones of the concurrent methods.

Keywords: *node metric, word metric, feature maximization, graph embedding, word embedding, sparsity, explicable AI.*

* Short Bio *

Prof. Jean-Charles Lamirel
University of Strasbourg – France
University of Dalian – China
LORIA - France

Email : lamirel@loria.fr
GSM : +33624365491

Prof. Jean-Charles Lamirel is both invited Sea-Sky full professor at DUT - University of Dalian China since 2016 and lecturer at the University of Strasbourg since 1997. He is also associated researcher in SAMM – Paris 1 Panthéon-Sorbonne since 2021. He got his PhD in 1995 and his Research Accreditation in 2010. He is currently teaching Artificial Intelligence at the University of Dalian and Information Science (including Library Science) and Computer Science at the University of Strasbourg and achieving his research at the WISELab laboratory of DUT-Dalian and at the INRIA-LORIA laboratory of Nancy (France). He was a research member of the INRIA-CORTEX project whose scope is Neural Networks and Biological Systems. He has then integrated the INRIA-TALARIS project (recently becoming the LORIA SYNALP Project) whose main concern is automatic language and text processing. His main domains of research are Textual Data Mining based on Neural Networks, Scientometrics and Social Networks analysis. He has interests in both theoretical models for Data Mining and Data Mining applications. He is more specifically specialized in unsupervised learning methods. He is the creator of the paradigms of Data Analysis based on Multiple Viewpoints (MVDA) and Metric based on Feature Maximization (F-Max). The related models for which it has been proven that they outperform classical models begin to be used in many challenging Data Mining applications. His other main topics of research concern Visualization Methods for Data Analysis, Quality Issues in Data Analysis, Novelty Detection models, Social Networks Analysis Techniques, Geographical Data Analysis Methods and Deep Learning Methods for Text Analytics. He and his tools have been currently involved in several European projects on Webometrics and Data Analysis, like the recent EISCTES or QUAERO projects. He is a board member of international Webometrics journals and organizer and board member of international conferences in Machine Learning, like the Kohonen's WSOM+ conference. He is also editor of special issue of international journals in Neural Network and Machine Learning, like NCAA. He committee member of the main conferences in data mining, like ICDM or ICTAI. He is the unique foreign member of the CAASP (Speciality Committee of Science of Science and Discipline Construction) - China. From 2021 he is a member of the the Intemational Scientific Advisory Board for the Priority Research Area (in Polish: Priorytetowy Obszar Badawczy - POB). His research work and direction led to the successful presentation of more than 12 different PhDs (4 PhDs are still ongoing). It also generated an important scientific production: more than 58 invited conferences, organizer of 4 international conferences and of more than 21 special sessions in international conferences, main author of more than 168 publications in international conferences and journals. This work also was worthy for him the recognition of many prestigious foreign institutional partners like NIEHS (USA), NSC (Taiwan), KU Leuven (Belgium), UTS (Australia), UQAM (Montréal, Canada), University of Santiago (Chile), University of Tartu (Estonia), University of Gliwice (Poland). Prof. Lamirel is one of the winners of the recent IDA Frontier Prize 2021.