EECE 5642 HW3

Abstract

This is a report of paper [1] Slow Learning and Fast Inference: Efficient Graph Similarity Computation via Knowledge Distillation, wich analyzes the research direction and perspective of the paper and analyzes the superiority of this paper in data set visualization along with result set visualization.

1. Task/Motivation

The starting point of the paper is to address or optimize the NP-hard algorithm for computing Graph Edit Distance (GED) in graph data mining systems, since this traditional algorithm cannot produce industrial output at a sizable economic gain, which requires expensive economic costs. Also the paper gives an example of drug screening in the chemical field to help understanding, e.g. to efficiently perform queries on target drug structures in large molecular compound databases.

2. Relevant Data Visualization

1. Dataset

The paper uses four datasets, namely, AIDS700nef, LINUX, IMDB-MULTI and ALKANE, which are used to evaluate the performance effectiveness of efficient graph similarity computation based on knowledge distillation.

Table 1: Quantitative GED results of baselines and our method over AIDS, LINUX, IMDB and ALKANE.

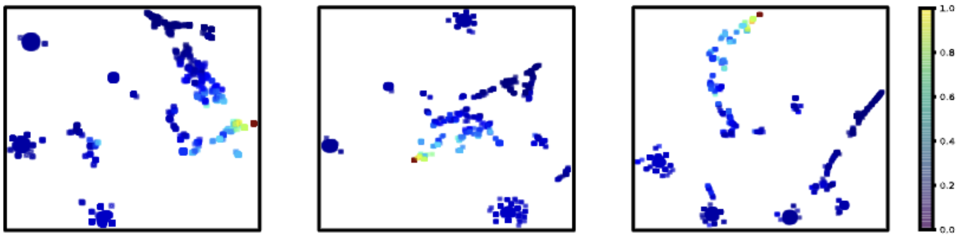
Methods	AIDS					LINUX				
	MSE ↓	ρ ↑	τ ↑	p@10 ↑	p@20 ↑	MSE ↓	ρ ↑	τ ↑	p@10 ↑	p@20 ↑
Beam	12.09	0.609	0.463	0.481	0.493	9.268	0.827	0.714	0.973	0.924
Hungarian	25.30	0.510	0.378	0.360	0.392	29.81	0.638	0.517	0.913	0.836
VJ	29.16	0.517	0.383	0.310	0.345	63.86	0.581	0.450	0.287	0.251
GENN-A*	**0.635**	**0.959**	-	**0.871**	-	0.324	**0.991**	-	0.962	-
SimGNN	1.189	0.843	0.690	0.421	0.514	1.509	0.939	0.830	0.942	0.933
E-SimGNN	2.096	0.869	0.699	0.534	0.641	0.469	0.982	0.892	0.971	0.968
GMN	1.886	0.751	-	0.401	-	1.027	0.933	-	0.833	-
GraphSim	0.787	0.874	-	0.534	-	**0.058**	0.981	-	0.992	-
EGSC-T	1.601	0.901	**0.739**	0.658	**0.729**	0.163	0.988	**0.908**	**0.994**	**0.998**
EGSC-S	1.546	0.898	0.736	0.649	0.724	0.293	0.984	0.898	0.978	0.983

Methods	IMDB					ALKANE				
	MSE ↓	ρ ↑	τ ↑	p@10 ↑	p@20 ↑	MSE ↓	ρ ↑	τ ↑	p@10 ↑	p@20 ↑
SimGNN	1.264	0.878	0.770	0.759	0.777	2.446	0.859	0.686	0.87	0.782
E-SimGNN	1.148	0.864	0.75	0.806	0.807	1.622	0.886	0.722	0.982	0.955
GMN	4.422	0.725	-	0.604	-	-	-	-	-	-
GraphSim	0.743	0.926	-	0.828	-	-	-	-	-	-
EGSC-T	**0.553**	**0.938**	**0.829**	**0.872**	**0.878**	**0.533**	**0.930**	**0.787**	**0.998**	**0.991**
EGSC-S	0.581	0.935	0.826	0.857	0.869	1.198	0.899	0.741	0.993	0.978

A brief description of the relevant parameters of the four datasets such as the number of included charts, the maximum number of nodes, node labels and edge labels is first given using textual descriptions, and then the table is used to present the quantitative GED results for the four datasets so that the reader has a clear understanding of the datasets used. The horizontal coordinates have several values such as MSE, p@10, p@20, etc., and are supported by up and down arrows indicating the trend to the right of the value to make the change in value more intuitive. The vertical axis of the table has Beam, Hungarin, etc. The key data crossed by the horizontal and vertical axes are bolded, which allows the reader to focus more on the key values.

2. Features

   A dimensionality-reducing feature visualization method, t-SNE, was introduced in this article with the aim of obtaining two-dimensional visualization of graphical data using a high-dimensional to low-dimensional mapping. Of course, to better show the effectiveness of this improved method, t-SNE using two baseline methods, namely SimGNN and Extended-SimGNN, is also taken out for comparison. As shown in the figure, the authors placed each of the three baseline methods within a figure, and after training, each point represented a graph, and color coded these graphs according to the true similarity between the graphs. It can be seen that the blue and ruddy colors with rich contrast are used, and finally a temperature scale from blue to ruddy colors is placed on the right side of this figure, which allows the reader to capture the basic similarity of the graphs more clearly and intuitively. And the clustering of points in the three figures demonstrates that figures with similar feature similarity are clustered together. Data-ink are used appropriately.



(a) (b) (c)

3. Key Insights

The paper presents a novel solution that employs a more robust graph isomorphic network that incorporates multiscale features from different GIN layers and uses more easily deployable MLP behavioral feature learning. The paper makes a good trade-off between speed and accuracy, as to overcome accuracy degradation, it proposes an early feature fusion network as a teacher model to train the student model. They also propose an original GSC Knowledge Distillation (KD) method to achieve efficient inference and offline embedding collection. They use GED benchmarks to test their model and gain 10 times faster performance compared with the best competitor model on AIDS dataset.

4. Model/Results Visualization

1. Table

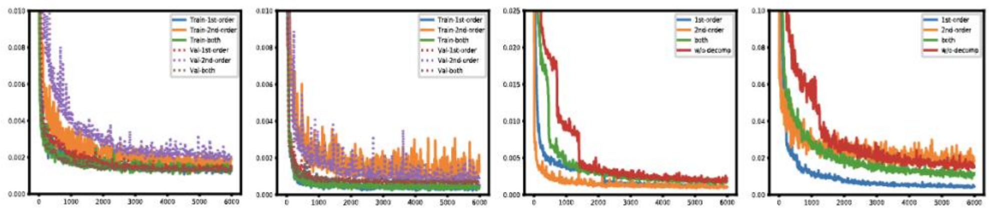
   The paper generally uses tables for values that are more representative of model efficiency and accuracy. For example, the Inference time to solve GED computations on AIDS and other AIDS-like models.

Model	GENN-A*	SimGNN	E-SimGNN	E-SimGNN-F	Teacher	Student-R	Student-F
Time	290.1h	11.139s	9.672s	3.464s	11.139s	10.149s	**0.148s**

   Student-R means the student model with raw input graphs. Student-F denotes that the embeddings are stored offline, which can be online loaded for inference. Important data is bolded.

2. Line Chart

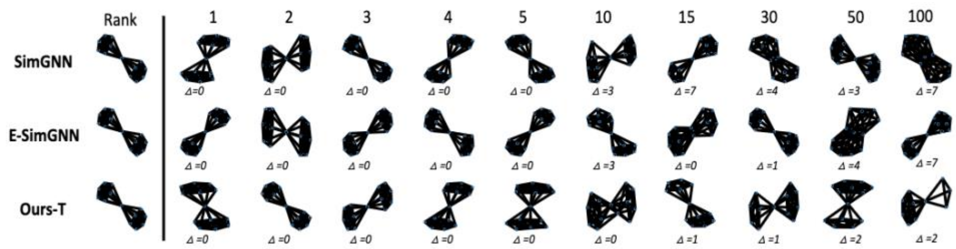
   The paper uses line graphs to represent quantitative Results. As shown in the figure below, this is the figure that represents the knowledge distillation loss results, where the horizontal coordinate indicates the execution time and the vertical coordinate indicates the training loss, and different colors and different line falsities indicate different models. The color match is good making the contrast clear but not making it visually blurred and uncomfortable. As time progresses, the curves gradually converge, indicating the specific progress of the student model fitting the teacher model.



(a) (b) (c) (d)

3. Drawing Chart

   The paper uses figurative graphical examples to show the ranking results of the algorithm. There are no clear differences and errors in the top 5 ranking results, and this corresponds to the change in rank of various GNNs. In the subsequent iteration order, the original baseline method gradually starts to weaken compared to their method. This graph eliminates some of the wireframes and appears clear and tidy, while the graphical representation of rank clearly shows the accuracy results.



5. Summary

This paper has beautiful illustrations, proper details, proper data-ink, accurate values, and conforms to scientific norms. I also learned about the domain knowledge related to graph similarity computation, graph deep learning, graph data mining.

6. Answers to Questions

1. What is the novelty of paper [1]?

   The paper presents a novel solution that employs a more robust graph isomorphic network that incorporates multiscale features from different GIN layers and uses more easily deployable MLP behavioral feature learning.

2. In paper [3], what is the difference of using strong augmentation and weak augmentation?

   Weak augmentation cannot acquire more complex changes because it uses minimal transformation and can only change a limited amount of the original image. Strong augmentation, on the other hand, allows for more extensive and diverse image transformations, such as adding sketches and lines to an image.