ملخص أطروحة الدكتوراه بعنوان

تطوير خوارزمية لمعالجة الاضطرابات في الشبكات العصبونية الممثلة بالبيان

اسم الطالب

ضياء حسن هرموش

المشرف

أ.د. سمير كرمان

المشرف المشارك

د.هيام خدام

القسم والاختصاص

قسم هندسة الحواسيب والأتمتة

اختصاص هندسة التحكم والأتمتة

الملخـــــص

حققت الشبكات العصبونية الممثلة بالبيان GNN (Graph Neural Networks) نجاحا ملحوظا في العديد من التطبيقات الخاصة بتحليل الرسوم البيانية ونمذجتها.

ويعود سر النجاح الكبير الذي حققته GNN في العديد من التطبيقات المتعلقة بالرسوم البيانية الى مخطط تمرير الرسائل الذي تعتمده أثناء التعلم حيث تقوم بتجميع رسائل الجوار لكل عقدة في كل طبقة من طبقاتها أثناء التدريب مما يسمح للنموذج في الطبقة النهائية من معرفة البيان بشكل كامل وفقا للرسائل المجمعة من كل عقدة وجوارها.

وعلى الرغم من قوة هذا المبدأ في مهام تصنيف العقد الخاصة بالبيان الا أن اعتماد GNN على بنية البيان بشكل كبير أثناء تبادل الرسائل يجعلها عرضة للاضطرابات الناجمة عن الهجمات العدائية والتي تحدث على طوبولوجيا البيان وتؤثر سلبا على متانة هذه الشبكات واستقرارها وبالتالي انخفاض كبير في الأداء ونتائج غير دقيقة ينجم عنها اعطاء العقد تسمية مختلفة عن تسميانها الحقيقية.

تم في هذا البحث دراسة الأثر السلبي للهجمات العدائية على GNN واقتراح نموذج هجوم على البيان المعروف بشبكة الاقتباساتCitation Network لقاعدة البيانات المعروفة بـ CORA-DATSET بعد تحويل هذا البيان الى بيان موزون واقتراح خوارزمية لمعالجة الاضطرابات الناجمة عن هذه الهجمات وذلك وقت اختبارGNN, بعد ذلك تم اجراء دراسة تحليلية لمعاملات النموذج للكشف عن الهجمات العدائية والاستفادة من المعاملات لتطوير الخوارزمية المقترحة ومعالجة الاضطرابات الناجمة عن الهجمات العدائية التي تحدث أثناء تدريب النموذج, كما تم اجراء نوعين من الهجمات (Random attack, metattack) على كل من قاعدتي البيانات (CORA Polblogs) لتظهر النتائج التحسن الملحوظ في أداء GNN وامكانية الخوارزمية من تقليل الآثار السلبية للهجمات.

PhD dissertation summary

Developing an Algorithm to Handle the Perturbations of Graph Neural Networks

Student Name

Diaa Hasan Harmosh

Co-Supervisor

Dr. Hiyam Khaddam

Supervisor

Dr. Samir karman

Department

Computer and Automation Engineering Department

Summary

Graph neural networks (GNNs) have achieved remarkable success in many applications for graph analysis and modeling. The secret of the great success achieved by GNN in many applications related to graphs is due to the message passing scheme that it adopts during learning, as it collects neighbor messages for each node in each of its layers during training, which allows the model in the final layer to know the statement completely according to the collected messages. From each node and its neighbors. Despite the strength of this principle in node classification tasks, the GNN's heavy reliance on the graph structure during message passing makes it vulnerable to perturbation caused by adversarial attacked Which occurs on the manifest topology and negatively affects the robustness and stability of these networks and thus a significant decrease in performance and inaccurate results that result in giving the nodes a label different from their real label. in this research, we studied the negative impact of adversarial attacks on GNN and proposed an attack model on the graph known as the Citation Network for the database known as CORA-DATSET after converting this graph into a weighted graph and proposing an algorithm to address the perturbation resulting from this attack at the time of testing the GNN. After that, we conducted a study an analysis of the model's parameters to detect adversarial attacks and take advantage of the parameters to develop the proposed algorithm and address the perturbation resulting from adversarial attacks that occur during training the model .Two types of attacks (random attack, mettattack) were also conducted on both databases (CORA Polblogs) so that the results show a noticeable improvement in GNN performance and the algorithm's ability to reduce the negative effects of attacks.