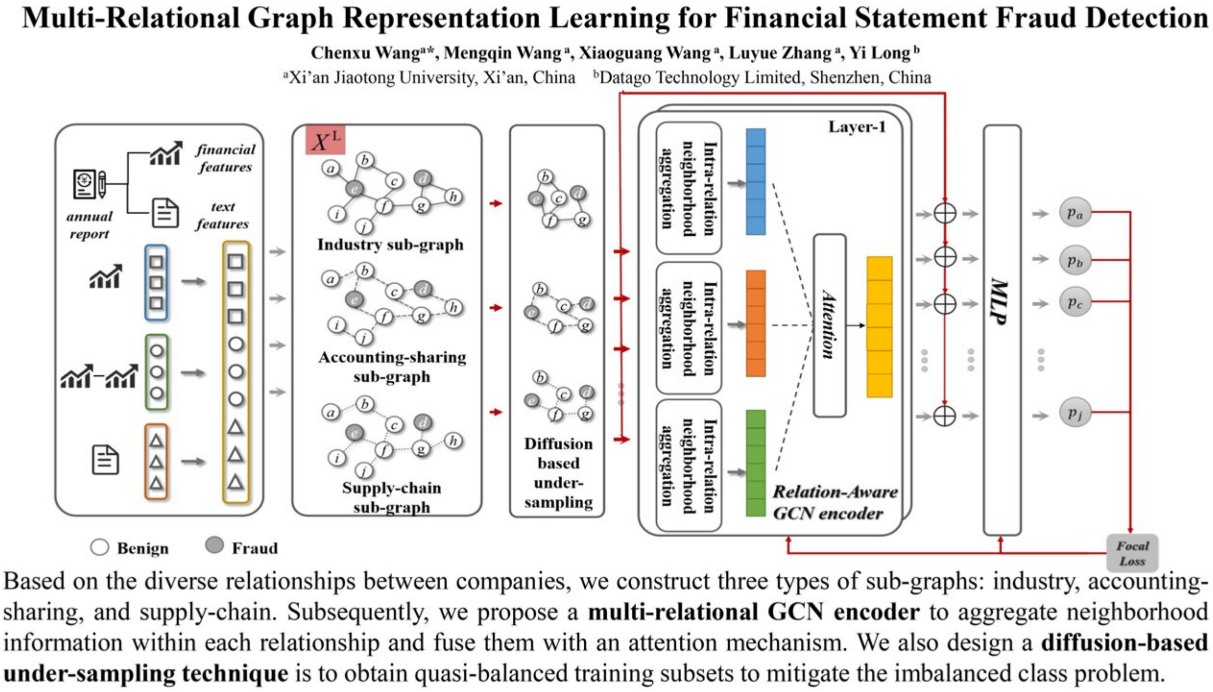
Machine learning technique predicts likely accounting fraud across supply chains

September 3 2024



Multi-Relational Graph Representation Learning for Financial Statement Fraud Detection

Chenxu Wang[a*], Mengqin Wang[a], Xiaoguang Wang[a], Luyue Zhang[a], Yi Long[b]

[a]Xi'an Jiaotong University, Xi'an, China [b]Datago Technology Limited, Shenzhen, China

financial features
text features
annual report
X^L
Industry sub-graph
Accounting-sharing sub-graph
Supply-chain sub-graph
Diffusion based under-sampling
Layer-1
Intra-relation neighborhood aggregation
Intra-relation neighborhood aggregation
Intra-relation neighborhood aggregation
Attention
Relation-Aware GCN encoder
MLP
p_a
p_b
p_c
p_j
Focal Loss
○ Benign
● Fraud

Based on the diverse relationships between companies, we construct three types of sub-graphs: industry, accounting-sharing, and supply-chain. Subsequently, we propose a **multi-relational GCN encoder** to aggregate neighborhood information within each relationship and fuse them with an attention mechanism. We also design a **diffusion-based under-sampling technique** is to obtain quasi-balanced training subsets to mitigate the imbalanced class problem.

Overview of the FraudGCN approach. The researchers constructed three types of 'sub-graphs' depending on the type of relationships between companies: with accounting firms; along supply chains; and throughout an industry. The training direction of the machine learning model is depicted by red arrows. Grey circles ('nodes') represent fraudulent firms and white circles represent normal firms. Credit: Big Data Mining and Analytics, Tsinghua University Press

As the perpetrators of accounting fraud become ever more sophisticated in their techniques, fraud detection needs to step up its game. Thankfully, a group of researchers have devised a new machine learning 'detective' that is able to analyze not just fraud at a single firm, but predict likely fraud across whole supply chains and industries.

A paper describing the team's approach was published in the journal *Big Data Mining and Analytics* on August 28.

Financial statement fraud, or, more commonly, accounting fraud may be a less frequent form of corporate fraud, but it is by far the costliest crime in the world. Perhaps the most famous cases of white-collar crime can be considered accounting fraud, when an enterprise manipulates the figures on its financial statements or other valuation data in order to make it appear more profitable than it is.

The collapse of US energy firm Enron, the largest bankruptcy in US history, came from their cooking of the books in collusion with their accounting firm. In 2008, Lehman Brothers declared bankruptcy due to insolvency, having concealed approximately $50 billion in debt through balance sheet fraud. In the late 2010s, American investment advisor Bernie Madoff managed to cheat clients out of a whopping $65 billion.

It is not only investors who are hurt by financial statement fraud. Hundreds of thousands of jobs can be lost, communities devastated, and, in the most extreme cases, through knock-on effects, it can threaten the stability of national economies.

Despite the threat that such fraud poses, it remains very hard for authorities to catch. Red flags such as a sudden surge in a company's performance just before the end of a reporting period, or soaring sales growth while competing firms' sales remain sluggish could turn out to be just the result of good luck or a superior product. And so for decades, forensic auditors have used statistical analysis to spot manipulation.

But such efforts are enormously labor intensive and require trawling through huge volumes of data. As a result, authorities tend to depend upon random audits, but this means that most firms most of the time go unchecked.

"Making matters even worse, in recent years, fraudsters have become increasingly sophisticated in the techniques they deploy," said Chenxu Wang, lead author of the paper and an associate professor with the School of Software Engineering and the Key Lab of Intelligent Networks and Network Security at Xi'an Jiaotong University. "It's an unending, mathematical arms race between the authorities and the fraudsters."

"What is needed is an effective and accurate algorithm to automatically identify accounting fraud, and leave the days of random auditing behind," said Mengqin Wang, also of Xi'an Jiaotong University.

A number of mathematicians and computer scientists specializing in the topic have achieved some success in this regard by the use of machine learning. But up to now, this approach has only been applied to individual firms.

"This overlooks the often-intricate relationships between different firms that may also offer up indicators of fraud," said Yi Long, another team member, but from Shenzhen Finance Institute, at the Chinese University of Hong Kong, Shenzhen. "An accounting firm that colludes in financial statement fraud with one company has an increased likelihood of engaging in fraudulent activities with other companies."

And it is not just between accounting firms and their clients where the fraudulent relationships are propagated. Accounting fraud practices can spread up and down supply chains, or, be perpetuated horizontally across industries.

But to incorporate data beyond a single firm means a commensurate increase in the computational expense. Moreover, existing machine-learning approaches suffer from a severe imbalance in the samples used to train the computer model how to classify something as fraudulent because normal, non-fraudulent samples significantly outnumber actual fraud cases. This imbalance can lead to biased computer models that prioritize the majority class, the non-fraudulent cases, making it difficult to accurately detect fraudulent activities.

To overcome all of these challenges, the research team developed a machine-learning technique combined with mathematical methods taken from the realm of graph theory.

The cutting-edge artificial intelligence financial-fraud detective they devised involves a graph, a structure that mathematically represents the connections or relations (described as edges) between different companies, individuals and products (described as nodes). And multi-relational graphs allow for multiple types of edges, allowing the representation of diverse relationships between nodes, and offer a more comprehensive representation of the complexity of connections among them.

And the detective itself, called FraudGCN, is a graph convolutional network, or GCN, a type of neural network designed to operate on graph-structured data. Unlike traditional neural networks that operate on grid-like data such as images, GCNs can operate on data represented as graphs.

FraudGCN itself constructs a multi-relational graph representing various industry connections, supply chain links, and shared accounting firm auditing practices, and by doing so, capture rich information arising from these relationships, in particular details uncovered in particular 'neighborhoods' of nodes in the graphs. By aggregating such information, FraudGCN not only enhances the ability to identify patterns indicative of existing likely fraudulent activities, but also predict where they are likely to arise.

Finally, unlike previous efforts at machine-learning assisted fraud detection, FraudGCN is able to handle the addition of new nodes without the need for the model to be retrained, enhancing its adaptability and scalability.

The team trialed FraudGCN on a real-world dataset from Chinese listed companies to assess its performance, and found that it beat state-of-the-art approaches by between 3.15% and 3.86%.

Moving forward, the team hope to develop their approach to be able to deal with medium-sized enterprises, not just larger ones.

More information: Chenxu Wang et al, Multi-Relational Graph Representation Learning for Financial Statement Fraud Detection, *Big Data Mining and Analytics* (2024). DOI: 10.26599/BDMA.2024.9020013

Provided by Tsinghua University Press

Citation: Machine learning technique predicts likely accounting fraud across supply chains (2024, September 3) retrieved 5 September 2024 from https://techxplore.com/news/2024-09-machine-technique-accounting-fraud-chains.html