Case ID:M18-226P^

Published: 2/26/2020

Inventors

Jundong Li

Huan Liu

Harsh Dani

Xia (ben) Hu

Contact

Shen Yan
shen.yan@skysonginnovations.com

Residual Analysis for Anomaly Detection in Attributed Networks

Background

Attributed networks are networks whose nodes contain rich sets of features or attributes. In social networks, for instance, attributes may be user interests, while for paper citation networks they may be papers associated with an institution or scholar. The pervasiveness of attributed networks across numerous domains has given rise to the development of dedicated analytical tools.

The detection of anomalous network content—that is, finding rare instances that differ singularly from the majority—is an established field of study that holds much practical value for recognizing fraud and spam. Many existing algorithms, however, are based on assumptions about anomaly context, structure, or community. Because real-world anomalies are constantly emerging, evolving, and interacting in unexpected ways, a more dynamic approach is needed. The use of residual analysis (i.e., scoring anomalies based on differences between true and estimated data) has been valuable in generalized settings and its integration into more robust frameworks may greatly advance detection capability.

Invention Description

Researchers at Arizona State University have developed a new learning system for anomaly detection in attributed networks. This technology features two algorithms working in tandem: the first models attribute information while the second models the network itself.

For attribute modeling, a set of representative instances is created using attribute data from the network. A residual analysis process then tracks the degree to which reconstructing network data is possible using these representative instances; the more exact the reconstruction, the lower the anomaly probability. The network modeling algorithm adds a structural dimension to the detection process by recognizing that attribute reconstruction patterns tend to be closely linked. Together, these two modules operate without knowledge of predefined anomaly properties. Experiments conducted using real data sets from Enron and Amazon showed improved AUC (Area Under ROC Curve) values for this framework over baseline methods.

Potential Applications

- Fraud and spam recognition

- System fault diagnosis

- Network intrusion detection

- Data analytics

Benefits and Advantages

- Integrative – Both attribute and network characteristics are combined into one coherent learning process

- Non-restrictive – Method is not constrained by assumed anomaly properties

- Effective – Experiments show improved detection performance over similar methods

Related Publication (PDF)

Homepage of Professor Huan Liu