Université
franco-allemande
Deutsch-Französische
Hochschule

Context-Aware Credit Card Fraud Detection

La fraude par carte de crédit est devenue un problème majeur dans le secteur des paiements électroniques. Dans cette thèse, nous étudions la détection de fraude basée sur les données transactionnelles et abordons plusieurs de ces défis complexes en utilisant des méthodes d'apprentissage automatique visant à identifier les transactions frauduleuses qui ont été émises illégitimement au nom du titulaire légitime de la carte. En particulier, nous explorons plusieurs moyens d'exploiter les informations contextuelles au-delà des attributs de base d'une transaction, notamment au niveau de la transaction, au niveau de la séquence et au niveau de l'utilisateur.

Au niveau des transactions, nous cherchons à identifier les transactions frauduleuses qui présentent des caractéristiques distinctes des transactions authentiques. Nous avons mené une étude empirique de l'influence du déséquilibre des classes et des horizons de prévision sur
la performance d d'un classifieur de type random forest. Nous augmentons les transactions avec des attributs supplémentaires extraits de sources de connaissances externes et montrons que des informations sur les pays et les événements du calendrier améliorent les performances de classification, particulièrement pour les transactions ayant lieu sur le Web.

Au niveau de la séquence, nous cherchons à détecter les fraudes qui sont difficiles à identifier en elles-mêmes, mais particulières en ce qui concerne la séquence à court terme dans laquelle elles apparaissent. Nous utilisons un réseau de neurone récurrent (LSTM) pour modéliser la séquence de transactions. Nos résultats suggèrent que la modélisation basée sur des LSTM est une stratégie prometteuse pour caractériser des séquences de transactions ayant lieu en face à face, mais elle n'est pas adéquate pour les transactions ayant lieu sur le Web.

Au niveau de l'utilisateur, nous travaillons sur une stratégie existante d'agrégation d'attributs et proposons un concept flexible
nous permettant de calculer de nombreux attributs au moyen d'une syntaxe simple. Nous fournissons une implémentation basée sur CUDA pour pour accélerer le temps de calcul de deux ordres de grandeur. Notre étude de sélection des attributs révèle que les agrégats extraits de séquences de transactions des utilisateurs sont plus utiles que ceux extraits des séquences de marchands. De plus, nous découvrons plusieurs ensembles d'attributs candidats avec des performances équivalentes à celles des agrégats fabriqués manuellement tout en étant très différents en termes de structure.

En ce qui concerne les travaux futurs, nous évoquons des méthodes d'apprentissage artificiel simples et transparentes pour la détection des fraudes par carte de crédit et nous esquissons une modélisation simple axée sur l'utilisateur.