Université
franco-allemande
Deutsch-Französische
Hochschule

Activity report

Yvan Lucas

Credit card fraud detection using machine learning with integration of contextual knowledge

Credit card fraud detection present several characteristics that makes it a challenging task. First, the feature set describing a credit card transaction usually ignores detailed sequential information which was proven to be very relevant for the detection of credit card fraudulent transactions. Second, purchase behaviours and fraudster strategies may change over time, making a learnt fraud detection decision function irrelevant if not updated. This phenomenon named dataset shift (change in the distribution p(x,y)) may hinder fraud detection systems to obtain good performances. We conducted an exploratory analysis in order to quantify the day by day dataset shift and identified calendar related time periods that show different properties. Third, credit card transactions data suffer from a strong imbalance regarding the class labels which needs to be considered either from the classifier perspective or from the data perspective (less than 1% of the transactions are fraudulent transactions).
We proposed a multi-perspective HMM-based automated feature engineering strategy in order to incorporate a broad spectrum of sequential information in the transactions feature sets. In fact, we model the genuine and fraudulent behaviours of the merchants and the card-holders according to two univariate features: the timing and the amount of the transactions. Moreover, the HMM-based features are created in a supervised way and therefore lower the need of expert knowledge for the creation of the fraud detection system. In the end, our multiple perspectives HMM-based approach offers automated feature engineering to model temporal correlations so as to complement and possibly supplement the use of transaction aggregation strategies in order to improve the effectiveness of the classification task.
Experiments conducted on a large real world credit card transaction dataset (46 million transactions from belgium card-holders between March and May 2015) have shown that the proposed HMM-based feature engineering allows for an increase in the detection of fraudulent transactions when combined with the state of the art expert based feature engineering strategy for credit card fraud detection.
To conclude, this work leads to a better understanding of what can be considered contextual knowledge for a credit card fraud detection task and how to include it in the classification task in order to get an increase in fraud detection. The method proposed can be extended to any supervised task with sequential datasets.

The funds gratefully provided by the French-German university have been of a great help in achieving the aforementioned contributions. They were used in order to participate to conferences where thesis contributions are published and to cover some travel expenses between Germany and France.

Overall, completing a joint PhD thesis within the IRIXYS center in the franco-german context, has been a great scientific and human experience. These three years wouldn't have quite as enjoyable without the help of the UFA and the franco-bavarian cotutuelle.