Intent2Vec: Representation learning of cardholder and merchant intent from temporal interaction sequences for Fraud detection

Nitish Kumar , Shiny John Shaju , Kanishka Kayathwal , Kushagra Agarwal , Alok Singh , Deepak Chaurasia , Siddhartha Asthana , Ankur Arora
[AI Garage, Mastercard, India]

Abstract

Fraud detection has been a challenging problem for financial institutions as it causes a loss of $24.2 billion per annum globally. This paper focuses on transaction fraud, the most prevalent type of fraud in the payment industry. The ability to detect and decline potential fraudulent transactions in real-time is crucial to guarantee a robust and secure environment for both, cardholders and merchants. Conventional fraud detection techniques predominantly use rule-based methods or extensive manual feature engineering for machine learning models. These fraud models rely on detecting anomalies in the attributes of a transaction. However, they fail to capture any type of interaction between cardholder and merchant involved in a transaction. The proposed approach, Intent2Vec, extends the capability of traditional fraud models by learning representation of payment entities using approaches of NLP to semantically capture the intent of doing a transaction. The modelled intent enables us to predict the next set of plausible merchants for a card and vice versa. Any deviation from the predicted and observed card or merchant can point towards a potential fraud. We test the relevance of intent based semantics on the downstream task of fraud detection wherein classifiers utilizing the entities' learnt intent outperform other baseline algorithms on metrics such as AUC-PR and F1 score.

Fraud, Payment Network, representation learning.

1 Introduction

According to a McKinsey report, the average number of fraudulent transactions attempted per merchant has been increasing at a rate of 34% per annum [Hasham *et al.*,]. Fraud detection frameworks were traditionally designed as expert systems which relied on a large set of rules stored in a knowledge base and have served the industry with high precision and business interpretability. However, due to frequent rule updates and unmanageable growth in the number of rules required to capture sophisticated fraud patterns, the payment industry has moved to data-driven methods such as statistical and machine learning models.

Primarily, existing solutions use transaction attributes as features and model fraud detection as a binary classification problem. As the fraud event rate is very low, binary classification models suffer due to severe class imbalance. Anomaly detection based models overcome this problem but have their limitations in terms of high false-positive rate. Furthermore, the sophistication of emerging fraud behavior is generally too complex to be described using only transaction attributes.

Deep learning methods have shown promising results in capturing complex transaction attributes' interactions, resulting in more accurate and precise payment fraud detection models. However, one consistent concern with most existing methodologies is that the feature set describing a card's transaction does not consider the historical purchase and selling pattern of the involved entities (merchant and cardholder).

The transaction history of a card can be leveraged to capture the information about merchants frequented by that card. Similarly, for merchants, their historical transactions can be used to learn the information of cards frequenting that merchant. Any mismatch in the learned interaction pattern of involved entities could be a behavioral anomaly and point towards a potential fraud. For example, transaction of a card at a gambling merchant, given that its recent transactions were at a domestic ATM, fuel and grocery merchants, could be a behavioural anomaly as the intent of this cardholder, defined by this recent history, is more semantically related to that of an everyday outing merchant like a retail shop than an infrequent leisure merchant like a casino. This interaction pattern of card and merchant involved in a transaction is referred to as the intent of that transaction, which can be leveraged to extend the capability of existing models to detect fraud.

In this paper, we propose Intent2Vec - a representation learning [Zhou *et al.*, 2018; Goyal and Ferrara, 2018] method that semantically captures the intent of a transaction by learning from prior sequence of transactions, enabling us to discern plausible merchants for a card. In case, the representation of the next transaction doesn't match the collective intent of involved entities, that transaction could be a fraud.

Representation learning models like word2vec [Mikolov *et al.*, 2013b], Glove [Pennington *et al.*, 2014] have been extensively used in Natural language processing problems to learn word representations. In Intent2Vec, we adopt an approach using skip-gram model on sentences of cards and merchants, based on their interaction with each other, to learn embeddings . These learned semantic representations, when used as features in modeling fraud detection on transactions, outperform all baseline fraud detection models.

To guarantee responses for real-time fraud prediction requests, a low latency system is crucial. Due to the drift in behavior of cards and merchants over time, there's also a need for using new transaction data to update the embeddings. Hence, we propose a 2 part deployable solution – periodic offline training and online real-time prediction. Periodic offline training recreates static card and merchant embeddings using new transactions, and then uploaded online, which along with transaction attributes, are used for real-time fraud prediction.

Key takeaways from this work are summarized as follows:

- Intent2Vec improves existing fraud detection models by learning the semantic intent of transaction via their historical sequence in real-time.
- It shows that the combined embeddings of involved merchant and card when passed along with transaction-level features as an input to the downstream supervised model for fraud detection, leads to improved performance on all classification metrics.

Outline: Section 2 mentions related work done in the field of fraud detection, particularly in payment industry domain. Section 3 details proposed approach of detecting transaction fraud, followed by section 4 on synthetic data generation process. Section 5 describes the experimental setup for our research, section 6 compares various methodologies and explains why the proposed method outperforms all baseline algorithms and section 7 concludes this study.

2 Related Works

This section introduces related literature around machine learning algorithms for detecting transaction fraud in payment domain.

2.1 Traditional supervised methods

Supervised machine learning solutions model fraud detection as a binary classification problem. [Awoyemi *et al.*, 2017; Shen *et al.*, 2007] compares the performance of multiple machine learning algorithms such as logistic regression, Naïve Bayes, K-nearest Neighbours, decision tree and neural networks on transaction data using various data sampling techniques. Use of Bayesian and neural networks for supervised fraud detection is described in [Hu *et al.*, 2020]. Traditional algorithms consider every entity independent of one another but in a financial transaction, there exists a historical relation between card and merchant. In our work, we leverage the semantic approach used in NLP to capture the interaction between cards and merchants.

2.2 Deep learning Based Fraud detection

Many recent literature include use of deep learning based methods to detect fraudulent transaction for example, [Misra *et al.*, 2020] uses Autoencoders to transform attributes in a lower dimensional space followed by a classifier to identify fraud. Use of denoising autoencoders with a fully connected network on credit card data is demonstrated in [Zou *et al.*, 2019]. To capture the sequential behaviour, models like Long Short-Term Memory (LSTM) are proposed in [Roy *et al.*, 2018; Jurgovsky *et al.*, 2018]. An ensemble of deep learning models for credit card fraud detection is detailed in [Heryadi and Warnars, 2017].

2.3 Anomaly based Fraud Detection

Most traditional approaches are distance based [Whitrow *et al.*, 2019] which compute anomaly scores using KNN or density-based methods. In [Zhou and Paffenroth, 2017; Schlegl *et al.*, 2017], autoencoders or generative adversarial networks are used to learn the unlabelled data's latent representation, which is used to determine class of an unknown entity using reconstruction error. A recent paper on anomaly detection [Pang *et al.*, 2018] propose a deviation network framework to directly learn an anomaly informed neural network by using limited number of labelled anomaly scores in a semi-supervised manner.

2.4 Embedding based Fraud detection

Representation of words in a sentence in the form of a vector capturing its semantic meaning has been an active area of research in Natural Language Processing. [Mikolov *et al.*, 2013a] proposed Word2Vec which constructs sequences from sentences using negative sampling and further uses a skipgram model to generate vector representations of words in the sequence. Various methods involving different sequence generation techniques have been proposed, with each capturing different domain relative meanings into the representation [Hu *et al.*, 2020; Zhou, 2019; Van Belle *et al.*, 2020].

Research in the field of graph representation learning has received a lot of attention recently. Network embeddings aim to map graph into low-dimensional representations which can be used for network analysis problems. To incorporate the profile of node's neighbors in a network, [Perozzi *et al.*, 2014] proposed DeepWalk which generates sequences of nodes in a graph by performing random walks and feeds these sequences to a skipgram model to generate embeddings. Since then, many random walk methods with efficient node sampling techniques like LINE [Tang *et al.*, 2015] or more flexible objective function like node2vec [Grover and Leskovec, 2016] have been proposed. Few research like GraRep [Cao *et al.*, 2015] construct a particular objective matrix and use matrix factorization techniques to generate node representations. [Dai *et al.*, 2020] proposed an efficient and scalable algorithm for data representation by performing a sequence of function mappings analogous to graphical model inference procedures. To improve fraud detection, recent studies have used network analysis to generate embeddings for entities involved in an encounter. [Cao *et al.*, 2019] proposed an end to end real time transaction fraud detection framework wherein DeepWalk [Perozzi *et al.*, 2014] is used to generate representations capturing user–user relationship in a P2P setting.

[Rao *et al.*, 2020] presented a fraud predictive framework with an explainer block, to reason why a transaction was marked fraud. They construct a heterogeneous cardholder-merchant graph and employ a graph neural network with self-attention to build representations of transactions. These representation are then fed to a DNN downstream model for fraud classification. [Zhong *et al.*, 2020] proposed a framework for default card prediction employing meta-paths over various views and devising an encoder to capture local structural patterns on nodes and links. To learn the importance of attributes and data view, attention mechanisms are adopted on node, link and meta-path levels.

Graph representation methods [Perozzi *et al.*, 2014; Cao *et al.*, 2015] embed neighboring node information but fail to capture the temporal relation of node interaction. Algorithms proposed in [Xu *et al.*, 2020] and [Dong *et al.*, 2017] successfully capture this information but are unscalable due to high computational requirement. [Yeşilkanat *et al.*, 2020] show the use of embeddings in credit card fraud detection by training character level embedding model using fastText library on a corpus of merchant names, representing each upcoming transaction as a sum of involved merchant name's vectors. This approach captures the profile of the merchant visited historically using character level representation but doesn't consider the time-ordered sequential nature of card and merchant interaction. Our proposed representation generation technique, Intent2Vec, gives embedding of each card and merchant incorporating information about the profile of card or merchant encounter and its time-ordered sequential relation, along with the card–merchant relationship.

3 Proposed Approach

Here, we describe the complete architecture (Fig. 2) and working of Intent2Vec. The architecture can be grouped into three components: 1) Entity representation learning - cardholder and merchant embedding generation, 2) Autoencoder - optimal transaction embedding generation and 3) Downstream classification model - fraud detection.

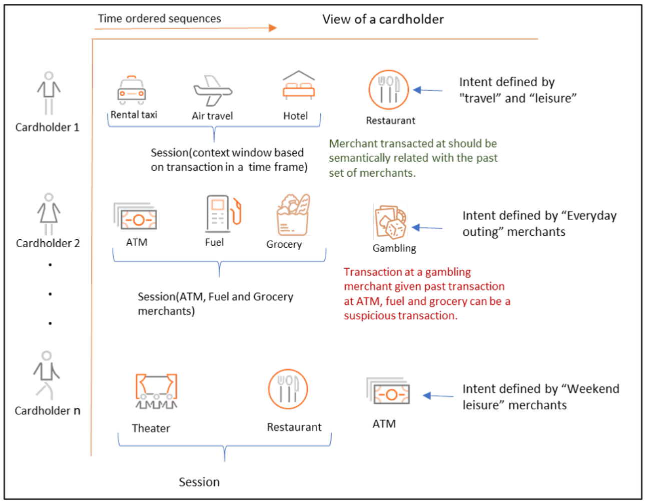

Time ordered sequences
View of a cardholder
Cardholder 1
Rental taxi
Air travel
Hotel
Restaurant
Intent defined by "travel" and "leisure"
Session(context window based on transaction in a time frame)
Merchant transacted at should be semantically related with the past set of merchants.
Cardholder 2
ATM
Fuel
Grocery
Gambling
Intent defined by "Everyday outing" merchants
Session(ATM, Fuel and Grocery merchants)
Transaction at a gambling merchant given past transaction at ATM, fuel and grocery can be a suspicious transaction.
Cardholder n
Theater
Restaurant
ATM
Intent defined by "Weekend leisure" merchants
Session

Figure 1: Cardholder view showing time ordered merchant sequences. Given a session, created based on merchants transacted within a defined time frame, the next merchant should be semantically related by the intent of the cardholder. A similar view can be generated for merchants also.

3.1 Entity representation learning

The intuition behind using time-ordered sequences of entities is to capture the general semantic intent of the corresponding entities for a particular set of transactions. For example, if a card is transacting in JFK, New York, all the merchants transacted on are related by this intent of this card (taxi, lounge access, airport shops). Entity representation learning in intent2vec is analogous to word2vec model where embeddings are generated using historical co-occurring sequences which capture the aforementioned intent of card and merchant encounters by extracting the semantics behind these interactions. Word2vec using skipgram approach takes sentences as input and considers each distinct word as an entity. The key principle behind Skipgram is that, given an entity, the model should be able to predict its neighbouring entities in the corpus. The objective of skipgram is to maximize the following average log probability.

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{-c\leq j\leq c;c\neq 0}\log p(m_{n+j}|m_n) \tag{1}$$

where c is the context window size, m_n is the target entity, $\{m_{n+j} \mid |j| \leq c\}$ is the set of context entities and the conditional probability $p(m_{n+j}|m_n)$ is defined by:

$$p(m_0|m_i) = \frac{\exp\left(v_{m_i}\cdot v_{m_0}^T\right)}{\sum_{j=1}^{N}\exp\left(v_{m_i}\cdot v_{m_j}^T\right)} \tag{2}$$

where v_{m_i} is the vector representation of m_i.
Intent2vec defines a sentence in the corpus as the time-ordered sequence of merchants encountered by a card, for card representation generation, and cards transacted at a merchant, for merchant representation generation, in a defined time frame (referred as a session). To generate merchant embedding, we construct merchant sessions, "$C_1, C_2, C_3, C_4, \ldots$" ($C_i$: i^{th} card), as shown in Fig.1 and feed them to a skipgram model. This approach captures the co-occuring card relationship and incorporates them into the embeddings. Similarly, for card embeddings, we construct card sessions and use it to train another skipgram model. Drawing analogy from NLP, similar merchants, for example, all grocery stores, will be in close proximity in embedding space because of their nature of being transacted at interchangeably by cards.

3.2 Autoencoder

Autoencoder is a viable approach for capturing latent semantic features, where a bottleneck layer forces our network to summarize, thus enhancing the model's ability to capture non linear relations existing among various transaction entities.

Each record in transaction data consists of 1) Real time transaction attributes such as amount, time, etc. , 2) Merchant ID and 3) Card ID. Instead of representing merchant and card IDs as one-hot encoded vectors, we use learnt entity embeddings. Further, these representations are passed through an autoencoder to generate a latent representation referred as transaction embedding.

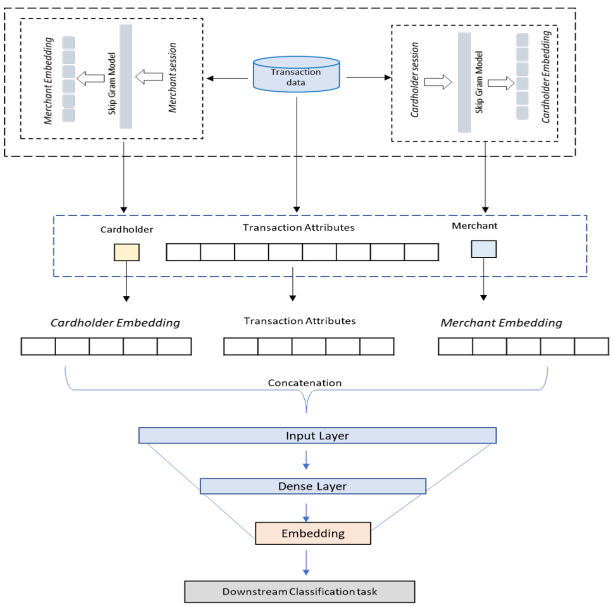


Figure 2: Card and Merchant embedding is passed through an autoencoder along with transaction features. Encoder layer is used as a feature vector for downstream fraud classification task.

3.3 Classification Model

In previous sub-section, we generated a generic latent representation of the transaction data. We use these generic embeddings to model fraud, using a multi-layer feed forward network, as shown in Fig. 2.

4 Dataset

Pursuant to internal controls to protect data, confidentiality and privacy, use of real data is avoided unless otherwise necessary. As this is an exploratory research, a synthetic dataset is created in lieu of leveraging actual data as shown in figure 3. For this, SMOTE-NC (Synthetic Minority Oversampling Technique), which is a data augmentation technique for generating synthetic data, is used. SMOTE-NC is a variant of SMOTE wherein generated data has continuous as well as non-continuous features. SMOTE-NC traditionally is used to oversample data from minority class to handle class imbalance. For our experiment, we create synthetic data for 2 classes – minority class with 1100 data points and majority class with 537k data points. The synthetic data points are generated using data with transactions labelled as fraud (minority class) and non-fraud (majority class). We demonstrate all our experiments and results on the synthetically generated data only as shown in Fig. 3.

SMOTE-NC on fraud dataset 2 copies of fraud data are made (i.e.original dataset triplicated). The 2 copied datasets (2200 data points) are marked as class 1 and the original dataset of 1100 data points as class 0. Runnig SMOTE-NC on this dataset generates 1100 points of class 0 which are the synthetic data points. All 3300 original data points are removed and only synthetically generated 1100 data points are kept to be used as fraud data points for our experiments.

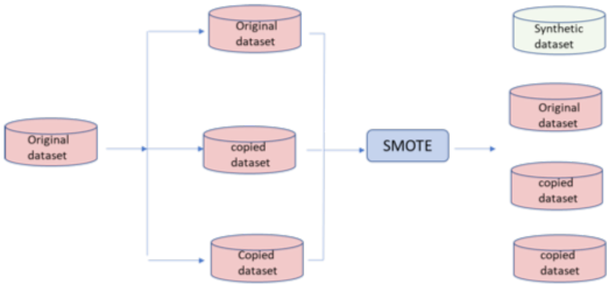


Figure 3: Framework of creating synthetic data from original dataset

SMOTE-NC on non-fraud dataset In a similar manner as explained above, we generate synthetic data points for Non-fraud class with $537k$ data points.

4.1 Data Description

There are $537k$ transactions out of which 1100 belong to the minority fraud class (0.2% event rate). Under the terms of our non-disclosure agreement, we cannot reveal all the details of the database schema nor the contents of the data. So, we only list a few variables which are present in common data schema used by most banks. The features can be grouped into four types: transaction, merchant, card, and issuer features. Transaction features are details such as timestamp, amount, channel, geography, etc.. Card, merchant and issuer features are a combination of static and velocity features.

Velocity is the measure of how fast the variable is changing with respect to time. We create velocities for cards by aggregating total amount and count of transactions on time intervals of $7, 14\&30$ days. For issuers, we create retrospective features such as overall and merchant-wise average approval rates. For merchants, along with static features such as merchant category code, industry and demographic details, dynamic features akin to that of issuers are created, which are a measure of tendency of declining transactions.

4.2 Distribution Similarity

We show how synthetic data is a good representation of real data and the results demonstrated on synthetic data holds on real data as well. Although the two datasets cannot be compared via individual data points, due to privacy protocols, they still can be compared at an aggregate level.

Table 1 shows distribution of total amount spent and average 30 day spends by a card in real dataset. Table 2 shows the same in the synthetic dataset. The distribution of these attributes in the two datasets is observed to be close. We use SDV library, developed by the Data to AI Lab at MIT, which contains a Synthetic Data Evaluation Framework that facilitates the task of evaluating quality of Synthetic Dataset. The Chi-Square test performed on the two datasets gives a value of 0.998, showing that the synthetic dataset generated is a true representation of real dataset. Therefore, the results demonstrated in this paper on synthetic data should hold true for real world data as well. We perform experiments on real data and the results were similar to what we report in this paper.

Statistic	**txn_amt**	**30d_avg**	**14d_avg**
count	538,125	538,125	538,125
mean	46.314	0.867	0.82
std	150.45	0.27	0.34
min	0.01	0.00	0.00
25%	8.97	0.88	0.83
50%	16.98	1.00	1.00
75%	43.96	1.00	1.00
max	44,418.73	1.00	1.00

Table 1: Distribution of transaction amount in real data

Statistic	**txn_amt**	**30d_avg**	**14d_avg**
count	53,8125	53,8125	538,125
mean	46.28	0.88	0.83
std	144.30	0.24	0.30
min	0.01	0.00	0.00
25%	9.08	0.88	0.83
50%	17.02	0.98	0.99
75%	43.95	1.00	1.00
max	36,661.17	1.00	1.00

Table 2: Distribution of transaction amount in synthetic data

5 Experiment Design

In this section, we detail the experimental setup for entity representation learning and fraud classification. Further, we describe few baseline approaches for comparing the performance of our model with existing state-of-art algorithms.

5.1 Training Details

For generating n-dimensional card and merchant representations, Intent2Vec employs a skip-gram model with negative sampling using a window size of 4 and min_count as 5 for pruning the card and merchant data. We use a decaying learning rate initialized as 0.03, minimum value as 0.0007 and $n = 100$. These representations are then fed to an autoencoder to generate transaction embeddings. A 6 layer autoencoder with $1024-512-128-128-512-1024$ units is trained with mean square error loss. Encoder's last layer is used to extract the learned 128-dimensional transaction embedding. For the downstream classification task, class weights are assigned in a $80 : 1$ ratio to address the high class imbalance.

5.2 Baseline Approaches

Our proposed model approach is compared with five baseline model approaches, details of which are given below:

- Deep Anomaly detection with Deviation Networks (DADN) [Pang *et al.*, 2019]: We use the default architecture listed in the original paper, network depth of 4 with layers having $1000-250-20-1$ units stacked together with ReLU activation function for first three and linear activation for the last layer. Model is trained for 50 epochs with batch size 16, then 10 runs with known number of outliers as 1100 and contamination rate of 0.002.
- Autoencoder (AE) with downstream classifier : Autoencoder is a deep learning network that learns data encoding efficiently in an unsupervised manner. These set of baselines use Autoencoder to learn transaction embeddings and then employs one of the following downstream models for fraud classification.

  The downstream models includes :

  - Generalized linear model : Logistic Regression (LR) - Default parameters.
  - LightGBM (LGBM) [Machado *et al.*, 2019] - A gradient boosting framework which uses tree based learning algorithm in distributed fashion. 500 trees with regularization.
  - Multilayer perceptron (MLP) - A feedforward artificial neural network for binary classification. Shape of the architecture - $1000 - 500 - 100 - 1$ units. Adam optimizer with batch size as 1024.
- TabNet : TabNet uses sequential attention to choose which features to reason from at each decision step, enabling interpretability and more efficient learning as the learning capacity is used for the most salient features [Arik and Pfister, 2019]. Default parameters were used to train the model.

Baseline	IE[1]	Prec[2]	Rec[3]	F1	AUC-PR
DADN	-	0.09	0.29	0.13	0.03
AE + LR	-	0.09	0.11	0.10	0.03
AE + MLP	-	0.34	0.39	0.36	0.32
AE + RF	-	0.75	0.34	0.47	0.44
AE + LGBM	-	0.68	0.47	0.55	0.53
TabNet	-	0.65	0.59	0.62	0.55
AE + LR	✓	0.21	0.26	0.24	0.16
AE + MLP	✓	0.92	0.66	0.77	0.76
AE + RF	✓	0.77	0.7	0.74	0.72
AE + LGBM	✓	**0.85**	**0.7**	**0.77**	**0.76**

Table 3: Baseline methods & Precision & Recall & F1 score & AUC-PR

6 Results

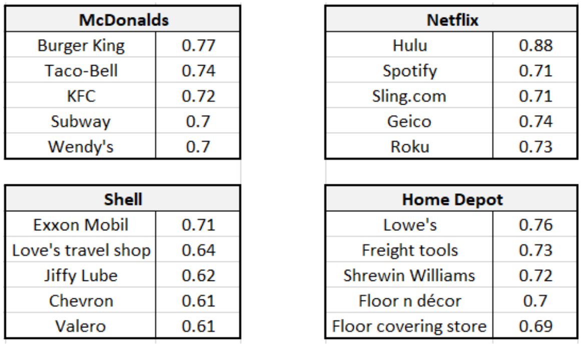
Here, we discuss the performance of our model on the dataset discussed in section 4 and compare it with a few state-of-art baselines. We use a training to testing stratified split of $80 : 20$ and internally, all models are cross-validated to achieve stable performance. Further, we demonstrate the efficacy of learned entity representations in order to verify captured semantics.

6.1 Model Evaluation

The models are evaluated on Precision, Recall, F1 score, and Area Under Precision-Recall Curve. Results in Table 3 shows the performance of our model on test data based on the threshold which gives highest F1 score on validation data.

Results in Table 3 shows that feeding card and merchant embeddings along with transaction features significantly improves the model's performance. It shows the importance of historical information of transactions captured in entities' embedding for detecting fraud. It can be seen that supervised

[1] Intent Embedding [2] Precision [3] Recall



McDonalds	
Burger King	0.77
Taco-Bell	0.74
KFC	0.72
Subway	0.7
Wendy's	0.7

Netflix	
Hulu	0.88
Spotify	0.71
Sling.com	0.71
Geico	0.74
Roku	0.73

Shell	
Exxon Mobil	0.71
Love's travel shop	0.64
Jiffy Lube	0.62
Chevron	0.61
Valero	0.61

Home Depot	
Lowe's	0.76
Freight tools	0.73
Shrewin Williams	0.72
Floor n décor	0.7
Floor covering store	0.69

Figure 4: Merchants with their closest counterpart merchant in the embedding space

methods outperform anomaly based unsupervised method. Also, even using entity representations along with transaction features as input, linear model performs poorly in comparison to non-linear models as they fail to capture the non-linear relationship among entities present in a transaction.

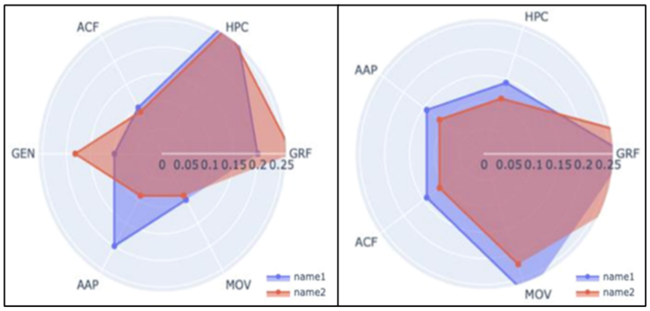


Figure 5: Radar plot shows 2 pairs of cards with high cosine similarity have high intersection area as well.

6.2 Efficacy of Merchant Embedding

Word embeddings hold semantic consistency i.e. words which are semantically similar tend to be in proximity in embedding space. In our work, it is shown that merchants which are similar in nature of their commercial activity are nearby in the embedding space and have a high cosine similarity. Fig.4 shows few Merchants and their nearest neighbours in the embedding space. We observe that most merchants similar to McDonald are food joints and those to Netflix are entertainment based. A similar trend holds true for Shell and Home Depot as well, indicating that the embedding generated are semantically consistent.

6.3 Efficacy of Card Embedding

We hypothesize that two cards having a similar distribution of transactions across industries will be in closer in embedding space. Radar plot in Fig.5 shows transaction distribution across industries for 2 pairs of cards. The plot has 6 industries with % of transaction at an industry denoted by the radial distance. The intersection area for two cards in a pair is very high, indicating their similar transaction patterns. This observation is supported by fact that they are found to be close in the embedding space as their cosine similarity is 0.96 indicating that the card embedding is effectively capturing the card behaviour.

7 Conclusion

In this study, the primary objective was to design and train stable and scalable fraud prediction models for transactions in payment industry. We focus on capturing behavior of cards and merchants to encode semantic relations across transactions. We propose a intent learning model from card's interaction history with the merchants and vice-versa for merchant's history with the cards. The transaction embedding is learnt using card and merchant embeddings with other transaction features, which captures the relationship between features optimally. Finally, these transaction embeddings are used for downstream task of fraud prediction. We conduct experiments on synthetically created transaction data and show that our method significantly outperforms the state-of-the-art systems. We show the importance of the intent based card and merchant embeddings.

References

[Arik and Pfister, 2019] Sercan O. Arik and Tomas Pfister. Tabnet: Attentive interpretable tabular learning. 2019.

[Awoyemi *et al.*, 2017] J. Awoyemi, A. Adetunmbi, and S. Oluwadare. Credit card fraud detection using machine learning techniques: A comparative analysis. In *International Conference on Computing Networking and Informatics*, 2017.

[Cao *et al.*, 2015] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Grarep: Learning graph representations with global structural information. In *Proceedings of ACM International on Conference on Information and Knowledge Management*, CIKM, 2015.

[Cao *et al.*, 2019] Shaosheng Cao, Xinxing Yang, Cen Chen, Jun Zhou, Xiaolong Li, and Yuan Qi. Titant: Online real-time transaction fraud detection in ant financial, 2019.

[Dai *et al.*, 2020] Hanjun Dai, Bo Dai, and Le Song. Discriminative embeddings of latent variable models for structured data, 2020.

[Dong *et al.*, 2017] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*, 2017.

[Goyal and Ferrara, 2018] Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 2018.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks, 2016.

[Hasham *et al.*,] Salim Hasham, Rob Hayden, and Rob Wavra. Using advanced analytics for fraud management—mckinsey. https://mck.co/3laenB4, year = 2018.

[Heryadi and Warnars, 2017] Y. Heryadi and H. L. H. S. Warnars. Learning temporal representation of transaction amount for fraudulent transaction recognition using cnn, stacked lstm, and cnn-lstm. In *2017 IEEE International Conference on Cybernetics and Computational Intelligence*, 2017.

[Hu *et al.*, 2020] Jinlong Hu, Tenghui Li, Yi Zhuang, Song Huang, and Shoubin Dong. Gfd: A weighted heterogeneous graph embedding based approach for fraud detection in mobile advertising. *Security and Communication Networks*, 2020.

[Jurgovsky *et al.*, 2018] Johannes Jurgovsky, Michael Granitzer, Konstantin Ziegler, Sylvie Calabretto, Pierre-Edouard Portier, Liyun He, and Olivier Caelen. Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, 2018.

[Machado *et al.*, 2019] M. R. Machado, S. Karray, and I. T. de Sousa. Lightgbm: an effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In *International Conference on Computer Science Education*, 2019.

[Mikolov *et al.*, 2013a] T. Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[Mikolov *et al.*, 2013b] T. Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 2013.

[Misra *et al.*, 2020] Sumit Misra, Soumyadeep Thakur, Manosij Ghosh, and Sanjoy Saha. An autoencoder based model for detecting fraudulent credit card transaction. *Procedia Computer Science*, 2020.

[Pang *et al.*, 2018] Guansong Pang, Longbing Cao, Ling Chen, and Huan Liu. Learning representations of ultrahigh-dimensional data for random distance-based outlier detection. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018.

[Pang *et al.*, 2019] Guansong Pang, Chunhua Shen, and Anton van den Hengel. Deep anomaly detection with deviation networks. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

[Pennington *et al.*, 2014] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *Proceedings of the conference on empirical methods in natural language processing*, 2014.

[Perozzi *et al.*, 2014] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk. *Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

[Rao *et al.*, 2020] Susie Xi Rao, Shuai Zhang, Zhichao Han, Zitao Zhang, Wei Min, Zhiyao Chen, Yinan Shan, Yang Zhao, and Ce Zhang. xfraud: Explainable fraud transaction detection on heterogeneous graphs, 2020.

[Roy *et al.*, 2018] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling. Deep learning detecting fraud in credit card transactions. In *Systems and Information Engineering Design Symposium*, 2018.

[Schlegl *et al.*, 2017] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, 2017.

[Shen *et al.*, 2007] A. Shen, R. Tong, and Y. Deng. Application of classification models on credit card fraud detection. In *International Conference on Service Systems and Service Management*, 2007.

[Tang *et al.*, 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line. *Proceedings of the 24th International Conference on World Wide Web*, 2015.

[Van Belle *et al.*, 2020] Rafael Van Belle, Sandra Mitrovic, and Jochen Weerdt. *Representation Learning in Graphs for Credit Card Fraud Detection*. 2020.

[Whitrow *et al.*, 2019] Christopher Whitrow, David J Hand, Piotr Juszczak, David Weston, and Niall M Adams. Transaction aggregation as a strategy for credit card fraud detection. *Data mining and knowledge discovery*, 2019.

[Xu *et al.*, 2020] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs, 2020.

[Yeşilkanat *et al.*, 2020] Ali Yeşilkanat, Barış Bayram, Bilge Köroğlu, and Seçil Arslan. An adaptive approach on credit card fraud detection using transaction aggregation and word embeddings. In *Artificial Intelligence Applications and Innovations*, 2020.

[Zhong *et al.*, 2020] Qiwei Zhong, Yang Liu, Xiang Ao, Binbin Hu, Jinghua Feng, Jiayu Tang, and Qing He. Financial defaulter detection on online credit payment via multi-view attributed heterogeneous information network. 2020.

[Zhou and Paffenroth, 2017] Chong Zhou and Randy C. Paffenroth. Anomaly detection with robust deep autoencoders. 2017.

[Zhou *et al.*, 2018] Lekui Zhou, Yang Yang, Xiang Ren, Fei Wu, and Yueting Zhuang. Dynamic network embedding by modeling triadic closure process. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Zhou, 2019] Sida Zhou. Empirical effect of graph embeddings on fraud detection/ risk mitigation, 2019.

[Zou *et al.*, 2019] Junyi Zou, Jinliang Zhang, and Ping Jiang. Credit card fraud detection using autoencoder neural network. *arXiv preprint arXiv:1908.11553*, 2019.