MeDML: Med-Dynamic Meta Learning - A multi-layered representation to identify provider fraud in healthcare

Nitish Kumar, Deepak Chaurasiya, Alok Singh, Siddhartha Asthana, Kushagra Agarwal, Ankur Arora
{Nitish.Srivasatava, Deepak.Chaurasiya, Alok.Singh2, Siddhartha.Asthana,
Kushagra.Agarwal, Ankur.Arora}@mastercard.com

Abstract

Every year, health insurance fraud costs taxpayers billions of dollars and puts patient's health and welfare at risk. Existing solutions to detect fraudulent providers (hospitals, physicians, etc.) aim to find unusual pattern at claim level features but fail to harness provider-provider and provider-patient interaction information. We propose a novel framework, Med-Dynamic meta learning (MeDML), that extends the capability of traditional fraud detection by learning patterns from 1) patient-provider interaction using temporal and geospatial characteristics 2) provider's treatment using encounter data (e.g. medical codes, mix of attended patients) and 3) referral using underlying provider-provider relationships based on common patient visits within 30 days. To the best of our knowledge, MeDML is first framework that can model fraud using multi-aspect representation of provider. MeDML also encapsulates provider's phantom billing index, which identifies excessive and unnecessary services provided to patients, by segmenting frequently co-occurring diagnosis and procedures in non-fraudulent provider's claims. It uses a novel framework to aggregate the learned representations capturing their task-specific relative importance via attention mechanism. We test the dynamically generated meta embedding using various downstream models and show that it outperforms all baseline algorithms for provider fraud prediction task.

Introduction

The National Health Care Anti-Fraud Association estimates a loss of $68 billion annually (i.e. 3% of total healthcare spending) due to fraud [1] most of which are due to provider's (hospitals, physicians, etc.) malpractices. Growing use of Electronic Health Records (EHR) has helped many organizations to build systems for detecting provider fraud [2] [3]. Existing fraud detection methods capture a lot of information about providers and patients available in EHR but do not consider modeling the relations among providers and between provider and patients. By modeling these relations fraud detecting system can uncover more sophisticated frauds such as collusion among fraudulent providers, inconsistent or unnecessary treatment claimed for a patient, etc.

In this work, we focus on learning multi-faceted information available in EHR data about providers, patients, and interaction among them. From EHR data, we can learn three faceted information – 1) provider level aggregated claim attributes, 2) their treatment profile and 3) their interaction and referral pattern. Towards this, we propose MeDML, Med-Dynamic Meta-learning, an end-to-end framework for provider fraud prediction which learns multi-faceted provider representations and aggregates them using their task-specific relative importance calculated via attention mechanism to generate meta-embedding incorporating the aforementioned information for each provider.

MeDML: MeDML can be broadly classified into three components – provider representation learning, embedding aggregation and classification component. The first component in MeDML learns multi-faceted provider representation capturing treatment, interaction and referral characteristics. The aggregation component generates a meta-embedding of provider by combining the embeddings learnt in first component dynamically and in a supervised manner, detailed in later sections. Further, the downstream classifier is trained using these meta-embeddings as inputs and provider fraud indicator as labels.

In this section, we elaborate on the three facets of EHR data used by MeDML:

- **Derived claim attributes**: Derived claim attributes are engineered features aggregated at the provider level. The features are broadly categorized into cost and utilization features, derived features from medical codes, and patient diagnosis and demographic features.
- **Treatment profile**: Treatment profile of a provider learns treatment behaviour based on its encounter with patients. It covers two aspects- "treatment pattern of provider" referred as provider's speciality representation and "mix of patients visiting the provider" referred as patient diversity representation. We leverage seq2seq models [4] to generate provider and patient representation in a way that captures their treatment pattern based on medical codes.

  Generated provider embeddings are directly used as provider's speciality embedding whereas patient diversity embedding of a provider is generated by taking mean of visiting patients embedding weighted by number of claims. The above two embeddings together capture complete treatment profile of a provider. In the process we also learn representation of diagnosis and procedure codes which are clustered to get segment of cooccurring diagnosis and procedure codes. This information is later used to create phantom billing index to identify the excessive/unnecessary services provided by fraudulent providers (Details explained in later sections).

- **Interaction profile**: Provider's interaction profile learns three representations - patient-provider temporal interactions representation, location representation of the attended patient and provider-provider referral relationship representation. MeDML learns the temporal representation by using a TGAT layer [5] on a provider-patient interaction graph. Location representation is generated using seq2seq model on sequences of the attending patient's location. Provider-provider relationship representation (henceforth referred to as referral embedding) is learned using GraphSAGE on a homogeneous provider-provider graph with an edge existing between providers when there is a visit by common patient within 30 days.

Combining multiple embeddings belonging to different embedding spaces is an active area of research [6] [7]. Traditionally, aggregation techniques like concatenation [8] and weighted averaging [9] have been used but they have limited capability to capture the relative task-specific importance of individual embeddings. MeDML not only learns treatment and interaction embeddings but also aggregates them dynamically using their task-specific relative importance to generate a single representation of a provider referred as provider meta-embedding. Furthermore, MeDML concatenates provider's meta embedding with their respective phantom billing indexes to train the downstream fraud predictive model. We show that provider's meta embedding capture various fraud aspects from the EHR data and that these meta embeddings, when used to train a supervised classifier, outperforms other baseline algorithms on metrics such as AUC-PR and F1 score.

Related Works

In a preliminary provider fraud detection study, Chandola et. al. [10] used Medicare claims and provider's matriculate data to detect fraud. The authors employed various techniques to mine information such as social network analysis [11], text-mining [12] and temporal analysis and used extracted features in logistic regression model to classify fraud using labelled data from the Texas office of Inspector General's exclusion database only.

Johnshon et. al. [13] compared six deep learning methods designed to address high class imbalance in healthcare fraud labels employing certain data sampling techniques (ROS, RUS and hybrid ROS-RUS) and a cost sensitive lost function – Focal loss using the CMS PUF data. This comparative study focusses on optimizing the sampling technique and ratio but doesn't mention the problem of effective representation of data for optimal learning. In [14], Zhang et. al. attempt to improve the existing fraud model by quantifying the disease–prescription correlation score and use it along with a few hand-crafted features for multilabel decision tree (ML-DT), rank SVM and NN learners. Focussing on feature engineering for data expression, [15] provided a comprehensive study leveraging supervised machine learning methods to detect fraudulent Medicare providers. Bauder et. al. again, generated utilization features along with provider speciality type and performed a comparative study using SVM, LR and C4.5 as provider fraud classifiers.

[14] and [15] partially address the problem of comprehensive representation of data by generating the aforementioned utilization variables but fail to capture multiple aspects of fraud pattern, like temporal, geo-spatial and referral trend of provider visits, in the EHR data. These features also do not highlight the excessive and unnecessary use of resources usually implying high risk of FWA.

Using structured longitudinal visit records of patients, Choi et al. [16], and Choi et al. [17] learned the representation of medical codes by using a seq2seq model. However, naïve aggregation of learned representations doesn't result in optimal representation as it ignores the latent relationship existing between them. MeDML aggregates the representations by learning their provider specific importance in a supervised manner for fraud prediction task.

In the field of multi-modal learning, Kiela et al. [18] examined combining multiple embeddings of words to better represent sentences and predict its category and using a BiLSTM layer [19]. However, this concept isn't transferable to healthcare domain as EHR data is tabular and nonsequential. Our framework uses dense layers [20] to retain the semanticity of features and combine their multiple embeddings in a supervised fashion.

Dataset

For this experiment we use insurance claim data available on Kaggle [1]. Along with the insurance claim data we also use patient data provided which contains their demographic and medical information. The insurance data contains Inpatient and Outpatient claim data along with provider level fraud labels. The training data comprises of claims from $5,410$ providers, out of which 504 are fraud, $138,557$ patients during a period of Nov 2008 to Dec 2009, $517,738$ outpatients claims and $40,475$ inpatient claims.

Design and architecture of MeDML

This section describes the complete architecture and working of MeDML. Firstly the details of the representation learning component (generation of treatment profile, interaction profile, and creating derived claim attributes) are discussed. Secondly, the aggregation component of MeDML which aggregates all the representations to generate provider meta embedding is elaborated. Finally, the downstream model for identifying fraudulent providers is discussed. Figure 2 shows the complete architecture of MeDML.

Treatment Profile Generator

Treatment profile generator generates 2 types of provider representation - Provider specialty representation and Provider patient diversity representation shown in Figure 2.

Provider speciality representation

In our work, seq2seq models have been used to generate provider's representation based on their "treatment pattern"

[1] https://bit.ly/3dkaesm

shown in Figure 1. Sequences of diagnosis codes and procedure codes are created at the claim level. Since there is no inherent order in the sequence of procedure and diagnosis in a claim, procedure and diagnosis codes are randomly permuted to generate multiple sequences. Finally, provider ID is sandwiched in between every two codes to form the final sequences. The provider ID is sandwiched in such a way that a seq2seq model could capture provider and the corresponding medical codes in the same context window and learn through cross entity interaction. Thus the sequences created for provider P_i will be $\langle DC_1^i, DC_2^i, P^i, \dots, PC_m^i \rangle$, where DC_j^i , and PC_k^i are all diagnosis and procedure codes in a claim and P_i is the provider ID of the i^{th} claim. These sequences are passed through the word2vec model to get provider's specialty representation (S_i, where $S_i \in \mathbb{R}^{128}$). Seq2Seq model also outputs diagnosis and procedure codes representation along with provider representation in the same embedding space.

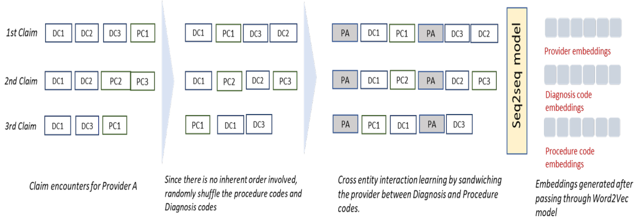

1st Claim
DC1
DC2
DC3
PC1
2nd Claim
DC1
DC2
PC2
PC3
3rd Claim
DC1
DC3
PC1
PA
Seq2seq model
Provider embeddings
Diagnosis code embeddings
Procedure code embeddings
Claim encounters for Provider A
Since there is no inherent order involved, randomly shuffle the procedure codes and Diagnosis codes
Cross entity interaction learning by sandwiching the provider between Diagnosis and Procedure codes.
Embeddings generated after passing through Word2Vec model

Figure 1: Creating provider embedding using medical codes

Provider's patient diversity representation

Provider's patient diversity embedding represents the mix of patients attending a provider. It uses patient embedding generated in a way similar to provider speciality embedding (S_i). We generate patient embedding (J_i, where $J_i \in \mathbb{R}^{128}$) by creating sequences of procedure and diagnosis codes with Patient ID sandwiched in between in all the claims for a patient. The sequences $\langle B_1^i, PC_1^i, DC_1^i, B_1^i \dots, B_1^i \rangle$ (B_i is the i^{th} patient ID) are passed through word2vec that generates patients embeddings. Finally, Provider's patient diversity embedding (PD_i, where $PD_i \in \mathbb{R}^{128}$) is learned by taking the mean of visiting patients embedding weighted by the number of claims.

Interaction Profile Generator

Interaction profile generator learns temporal and geo-spatial characteristics of patient-provider interaction and also learns referral patterns of underlying provider relationships based on common patient visits within 30 days.

Provider temporal representation

Temporal pattern of patients visiting a provider can be a key indicator to identify fraudulent providers. A sudden increase in claim density or an unusual time of filing claims can indicate fraudulent activities. Provider's temporal representation (T_i, where $T_i \in \mathbb{R}^{128}$) is learnt using TGAT [5] on provider-patient interaction graph where nodes are providers ($P_i \in$ P) and patients ($B_i \in$ B) and an edge $\langle e_{ij} \rangle$ exist between two nodes P_i and B_j if patient B_j has a visit with provider P_i. The edge of the temporal graph carries the timestamp of the provider-patient encounter. TGAT generates embeddings of each nodes using neighboring nodes features and aggregates them via attention mechanism.

Provider spatial representation

Provider spatial representation is learnt from historical location data of visiting patients. A provider filing claims for patients visiting from out of pattern location should trigger a suspicion. To capture this information, we generate sequences of attending patients' county with Provider ID sandwiched in between. The final sequences generated are $\langle P_1, C_1, P_1, C_1, P_1, C_2 \rangle$, where C_1 and C_2 are the county location of the patients visited at provider P_1. The generated sequences are passed to a skip-gram model to generate provider spatial embeddings (G_i, where $G_i \in \mathbb{R}^{128}$).

Provider referral representation

Graph techniques can capture information such as nexus between providers referring to each other. Often these nexus are responsible for committing institutional large scale fraud. [11]. In provider referral representation, we aim to capture the provider-provider relationship in a homogeneous provider referral graph with nodes as $P_i \in$ P and edge $\langle e_{ij} \rangle$ between two providers P_i and P_j exist if the same patient visits the two within 30 days interval. These edges act as a proxy for provider-provider referral due to unavailability of referral information. Derived claim attributes are used as node feature and graphSAGE [21], an inductive representation learning method, is applied on this network to learn the provider referral embeddings (CV_i, where $CV_i \in \mathbb{R}^{128}$).

Derived claim attributes

The last part of representation learning component is generating claim attributes which are aggregated at provider level. Features generated are categorized into three categories - cost and utilisation features, derived features from medical codes, and patient's diagnosis and demographic features. Cost and utilization features are claim count per patient, billed amount per claim, service unit per claim, average length of stay, readmission rate, % of planned visits, etc. Derived features from medical codes are created by classifying medical codes into broader category using the ICD-9 categorization and co-morbidity indices such as Elixhauser and Charlson co-morbidity index. Aggregated provider features are generated as the distribution of patients in each class e.g. % of patient attended with Cardiovascular disease, etc. Patients with chronic conditions are aggregated at provider level to get distribution of patients for each chronic condition. Patient demographic details such as patient age, gender, and location are used to finally get a set of base features.

Phantom billing Index

Phantom billing index points to excessive/unnecessary services provided by fraudulent providers. In the process of generating speciality embedding using word2vec model, we also generate embeddings for the diagnosis codes (DC_i where $DC_i \in \mathbb{R}^{128}$) and procedures codes ($PC_i$, where PC_i $\in \mathbb{R}^{128}$). These medical codes are clustered using K- means to get segments of co-occurring diagnosis and procedures. Phantom billing index is computed at the claim level and aggregated at the provider level as the average number of clusters of diagnosis and procedures in a claim. High variability in diagnosis and procedure segment at claim level may indicate out-of-norm diagnosis/procedure filed by provider.

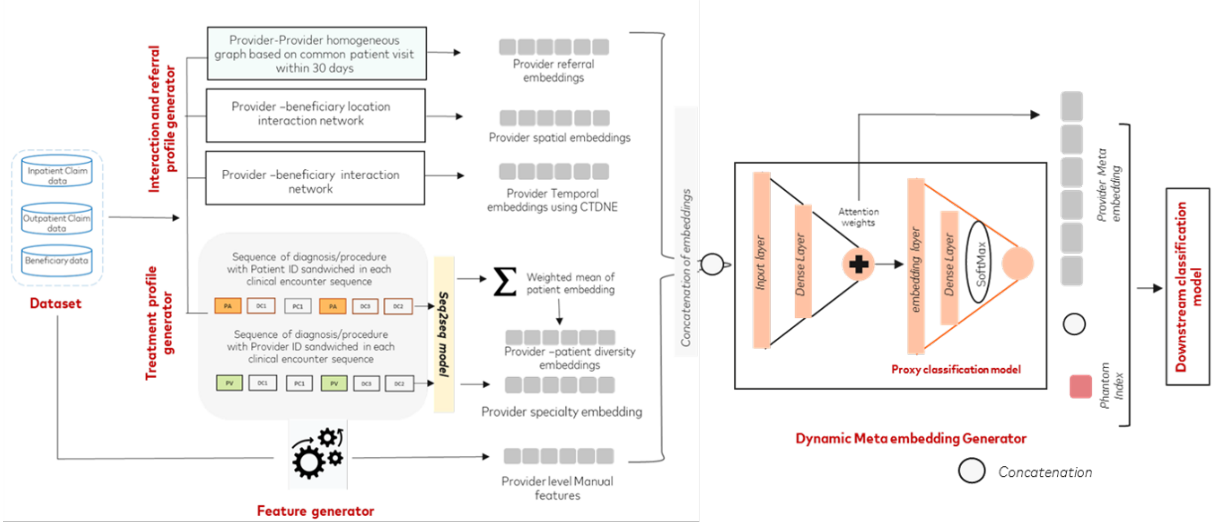

Interaction and referral profile generator
Provider-Provider homogeneous graph based on common patient visit within 30 days
Provider –beneficiary location interaction network
Provider –beneficiary interaction network
Provider referral embeddings
Provider spatial embeddings
Provider Temporal embeddings using CTDNE
Inpatient Claim data
Outpatient Claim data
Beneficiary data
Dataset
Treatment profile generator
Sequence of diagnosis/procedure with Patient ID sandwiched in each clinical encounter sequence
PA
DC1
PC1
PA
DC3
DC2
Sequence of diagnosis/procedure with Provider ID sandwiched in each clinical encounter sequence
PV
DC1
PC1
PV
DC3
DC2
Seq2seq model
Σ
Weighted mean of patient embedding
Provider –patient diversity embeddings
Provider specialty embedding
Provider level Manual features
Feature generator
Concatenation of embeddings
Input layer
Dense Layer
Attention weights
embedding layer
Dense Layer
SoftMax
Proxy classification model
Provider Meta embedding
Phantom Index
Downstream classification model
Dynamic Meta embedding Generator
Concatenation

Figure 2: An overview of the complete approach of training a predictive model for provider fraud classification

Dynamic Meta Learner

Dynamic Meta learner is the aggregation component of MeDML where treatment representation, interaction representation, and derived claim attributes generated are aggregated together to generate a single meta embedding for each provider. The aggregation layer uses concatenated embeddings of S_i, PD_i, T_i, G_i, CV_i and derived claim attributes and combines them dynamically i.e weights of these representations are learned for each provider by propagating the gradient of a proxy classification task backward to the aggregation task to generate a $\mathbb{R}^{128}$ representation. Dynamic meta learner learns task relative importance of individual representations using attention mechanism [22] tied to a proxy classification model to finally generate provider's meta embedding (P_i, where $P_i \in \mathbb{R}^{128}$).

$$P_j^{meta} = \sum_{i=1}^{n} \alpha_{i,j} w'_{i,j} \text{ where, } \alpha_{i,j} = g(w'_{i,j}) \text{ are scalar attention weights, } w = S||PD||T||G||CV$$

$$\alpha_{i,j} = \phi(a.w'_{i,j} + b) \text{ where, } \phi \text{ is the softmax function}$$

$$\alpha_{i,j} := \alpha_{i,j} - \frac{d[\mathcal{L}(\hat{y}, y)]^m}{d\theta} \text{ where, } [\mathcal{L}(\hat{y}, y)]^m \text{ is the error at the } m^{th} \text{ layer of the dynamic meta learner, } \theta \text{ are the learnable params}$$

Downstream classification model

The final part of MeDML is the downstream classification task of identifying a fraudulent provider. Provider meta embeddings (P_i, where $P_i \in \mathbb{R}^{128}$) generated after aggregation task is used as feature vector as input to a downstream provider fraud prediction task. The meta embedding is concatenated with the phantom billing index and passed through a classifier with provider fraud labels are ground truth.

Experiments and Results

In this section, we evaluate our method on a publicly available claim level data. We demonstrate MeDML's evaluation by testing the efficacy of generated embeddings. We further provide architecture details and compare MeDML performance with standard baselines. We also compare performance of different aggregation techniques with our aggregations method and finally show an ablation study of MeDML to show importance of each representation.

Model Evaluation

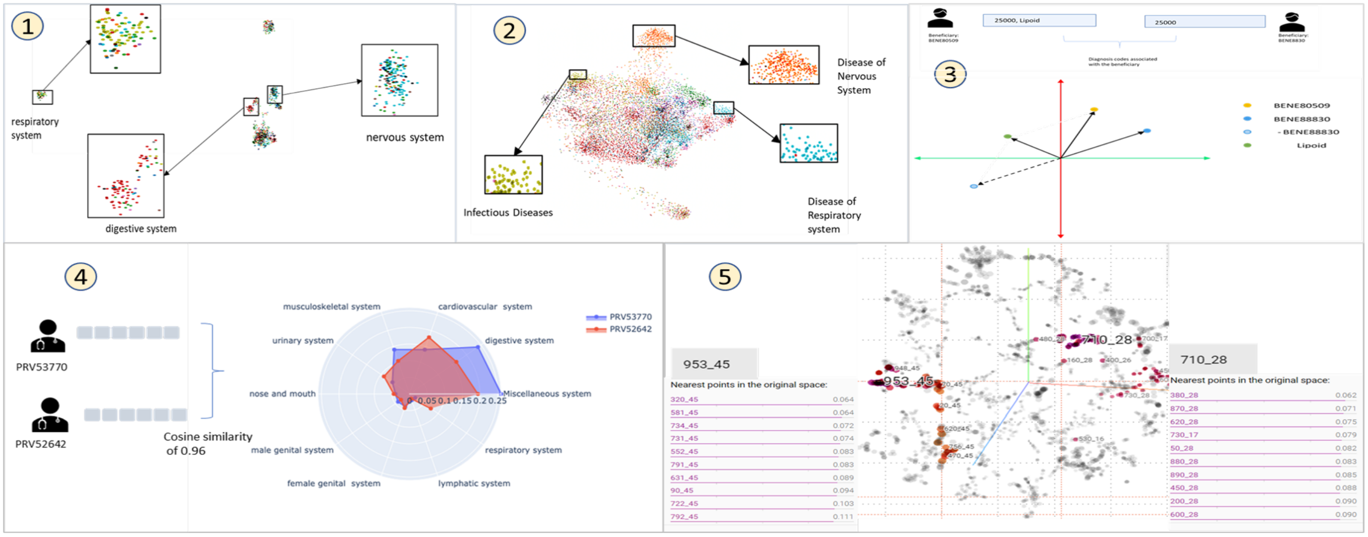
MeDML generates multi-faceted provider embeddings and aggregates them to create provider meta-embedding to be used as an input in fraud classification task. The quality of embeddings fed to the aggregation layer plays a huge role in determining our classification score. In this section, we start with discussing the efficacy of generated embeddings.

Efficacy of generated embeddings

Procedure and diagnosis code: Figure 3, shows 2D t-SNE representations of procedure and diagnosis codes. It is observed that the embeddings of closely related diagnosis and procedures are in close proximity in the embedding space. For example, in Figure 3 diagnosis codes related to the nervous system are clustered at the top of the t-SNE plot.

Patient embeddings: Mikolov et al. [4] showed that word vectors generated using skip-grams obey semantically meaningful linear operations like 'King' – 'Man' + 'Woman' being very close to 'Queen' word vector. Similar semantic relation holds for patient embedding as well. For example, Patient BENE80509 has diagnosis codes 'V7791' and '37636' whereas patient BENE8330 has only 'V7791' diagnosis code in their treatment profile. We observed a similarity score of 0.99 between BENE8830 + '37636' and BENE80509 embeddings as shown in Figure 3.

Provider speciality embeddings: To compare the efficacy of provider speciality embeddings, we state that two providers having similar distribution of diagnosis and procedure code categories will be in closer proximity in embedding space. Radar plot in Figure 3 shows distribution of medical code categories of two providers namely, PRV53770 and PRV52642. The plot has 12 categories of medical codes with claim % for that category denoted by radial distance . We observe, the intersection area for these two providers is



953_45

Nearest points in the original space:

320_45	0.064
581_45	0.064
734_45	0.072
731_45	0.074
552_45	0.083
791_45	0.083
631_45	0.089
90_45	0.094
722_45	0.103
792_45	0.111

710_28

Nearest points in the original space:

380_28	0.062
870_28	0.071
620_28	0.075
730_17	0.079
50_28	0.082
880_28	0.083
890_28	0.085
450_28	0.088
200_28	0.090
600_28	0.090

Figure 3: 1. t-SNE representation of the diagnosis code embedding categorized by their classes. 2. 2D t-SNE representation of the procedure code embedding categorized by their classes. 3. Vector representation of patient embedding showing semantically meaningful linear operations. 4. Radar plot showing 2 providers having high co-sine similarity in the embedding space have high intersection area on categories of diagnosis codes prescribed. 5. t-SNE plot of county embedding obtained from word2vec.

very high, which indicates that they have similar treatment pattern and should be very close in the embedding space. We observe their cosine similarity to be 0.96 which indicates that the speciality embedding is effectively capturing the provider treatment behaviour.

Spatial embedding: While generating provider spatial embedding, we also generate county representation. To check the efficacy of provider's spatial embeddings, we check how close the counties are in the embedding space. Figure 3 is a t-SNE representation showing that counties from the same state are closer to each other in the embedding space. For example, in the embedding space, counties closer to county#953 in state#45 belong to the same state.

Implementation details

Here, we mention the parameters used while learning multiple representations and training the aggregation and classification model. Seq2Seq models trained for generating provider's speciality, patient diversity and geo-location representation use a window size of 3, negative sampling rate of 10 and a learning rate of 0.03. TGAT layer, which learns the temporal aspect of provider-patient interaction, takes a learning rate of .0003, batch size of 30 and no. of sampled neighbours as 15. It uses 2 hidden layers and a dropout layer with 0.1 probability. To generate referral embedding of providers, graphSAGE uses a learning rate of 0.001 and batch size of 50 trained for 50 epochs using Adam optimizer. To generate the phantom billing index, MeDML clusters the diagnosis and procedure representations into 27 clusters as calculated using the SSE v/s number of clusters plot. Aggregation layer has two components - concatenation and attention. Concatenation component uses non-trainable embeddings layers which are further projected to 100 dimensions using a dense layer. The attention layer uses two cascading dense layers (32-1) to learn the weights. For fraud prediction, the provider level data is split in train, validation and test set in 0.7 : 0.1 : 0.2 ratio. The proxy classification layer attached to the aggregation layer of MeDML consists of cascading dense blocks with dropouts, Dense(32)-Dense(8)-Dropout(.25)-Dense(2) and is trained only on the training dataset in a supervised fashion. The attention weights are updated dynamically using the derivative of loss function percolated back from the proxy classification layer. The meta embedding is further used to train various downstream provider fraud predictive learners like logistic regression, SVM and xgboost with default parameters, performance compared in Table 1. We evaluate our models on Precision, Recall, F1 score and AUC-PR.

Feature	Model	Prec.[1]	Recall	F1	PR[3]
Base	LR	0.24	0.49	0.36	0.20
Base	XGB	0.45	0.64	0.61	0.58
Base	SVM	0.50	0.43	0.52	0.50
Base	MLP	0.58	0.47	0.55	0.50
Autoencoder	LR	0.30	0.79	0.52	0.40
Autoencoder	XGB	0.45	0.70	0.55	0.53
Autoencoder	SVM	0.71	0.26	0.55	0.53
Autoencoder	MLP	0.49	0.45	0.52	0.50
MeDML	LR	0.39	**0.93**	0.66	0.55
MeDML	XGB	**0.62**	0.79	0.71	**0.74**
MeDML	SVM	0.60	0.80	**0.72**	0.71
MeDML	MLP	0.57	0.64	0.64	0.55

Table 1: MeDML performance compared with baseline

Baseline results: We benchmark the performance of MeDML against existing provider fraud detection solutions which use aggregated claim level features (base model) [13][14][15] or their representation generated using autoencoder [23]. We performed experiments using Logistic Regression (LR), XGBoost (XGB), Support Vector Machine (SVM) and MLP (Multi layer Perceptron). Table 1 details the performance results for learners across all the performance metrics. We can see that the learners using MeDML as input out-performs those using Base feature and autoencoder embedding.

MeDML aggregation vs standard baselines: Table 2 presents the comparison of various standard aggregation techniques - Concatenation, averaging, F1-weighing and stacking with MeDML's dynamic aggregation method using vanilla logistic regression classifier as the downstream model. Results show that MeDML outperforms all the other aggregation strategies on different classification metrics.

Model	Prec.[1]	Recall	F1	AUC[2]	PR[3]
Concatenate	0.09	0.29	0.53	0.037	0.52
Average	0.16	0.51	0.54	0.48	0.52
F1-weighted	0.23	0.55	0.56	0.5	0.53
Stacking	0.35	0.63	0.58	0.57	0.55
MeDML	**0.39**	**0.93**	**0.66**	**0.76**	**0.55**

Table 2: MeDML aggregation vs standard baselines

Model	Prec.[1]	Recall	F1	AUC[2]	PR[3]
BaseNet	0.23	0.50	0.30	0.66	0.18
BSNet	0.10	0.71	0.48	0.56	0.44
BSGNet	0.09	0.73	0.51	0.49	0.5
BSGCNet	0.09	0.73	0.52	0.52	0.5
MeDML	**0.39**	**0.93**	**0.66**	**0.76**	**0.55**

Table 3: Ablation study of MeDML

Ablation study of MeDML

In this section we show the importance of each component of MeDML, with Table 3 showing the incremental results in the performance. We show the impact of each type of representation upon addition to the model. It is clearly evident from the Table 3 that each component is critical and improves the model performance.

BaseNet: BaseNet is the model with all the derived base features (f_i) of a provider. This gives an F1 score of 0.30 and AUCPR of 0.18.

BSNet: BSNet is the model with base features (f_i) combined with speciality embeddings (S_i). This improves F1 and AUCPR to 0.48 and 0.44 respectively.

BSGNet: Upon adding the spatial embedding to the BSNet, a further jump in F1 score and AUCPR is observed to 0.51 and 0.49 respectively.

BSGCNet: BSGCNet is obtained by adding referral embedding to the BSGNet. This gives further boost in F1 score and AUCPR to 0.53 and 0.52.

Conclusion

In this study, the primary objective was to design and train a provider fraud prediction model by looking at data beyond claim level and capture such intrinsic relationships. In this work we present a new deep learning architecture, MeDML-Med Dynamic Meta Learning, an end to end framework to capture fraud providers. We focus on capturing provider's interactional and treatment profile derived from provider's encounter with other providers, patients and diagnosis/procedures. We propose an aggregation framework to combine multiple representation that captures the interrelationship between each other via attention mechanism. Finally, these embeddings are used for down-stream task of provider fraud prediction. We show baseline results and present ablation study of the architecture to show the criticality of individual components of MeDML. In future, this work can be extended by using advanced semantic approaches in NLP like BERT, GPT, etc.

References

[1] *Kernel Description*. URL: https://bit.ly/3px6J4s. (accessed: 02.11.2020).

[2] Muhammad Suleiman et al. "A Generic Data Driven Approach for Medicaid Fraud Detection". In:

[3] V. Snorovikhina and A. Zaytsev. *Unsupervised anomaly detection for discrete sequence healthcare data*.

[4] T. Mikolov et al. *Efficient Estimation of Word Representations in Vector Space*. 2013.

[5] Da Xu et al. *Inductive Representation Learning on Temporal Graphs*. 2020.

[6] Wenpeng Yin and Hinrich Schütze. *Learning Meta-Embeddings by Using Ensembles of Embedding Sets*. 2015.

[7] Akira Fukui et al. *Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding*. 2016.

[8] Brian Lester et al. *Multiple Word Embeddings for Increased Diversity of Representation*. 2020.

[9] Joshua Coates and Danushka Bollegala. *Frustratingly Easy Meta-Embedding – Computing Meta-Embeddings by Averaging Source Word Embeddings*. 2018.

[10] Varun Chandola, Sreenivas Rangan Sukumar, and Jack Schryver. "Knowledge discovery from massive healthcare claims data". In: 2013.

[11] L. K. Branting et al. "Graph analytics for healthcare fraud risk estimation". In: 2016.

[12] Fred Popowich. "Using Text Mining and Natural Language Processing for Health Care Claims Processing". In: (2005).

[13] Justin Johnson and Taghi Khoshgoftaar. *Medicare fraud detection using neural networks*.

[14] Xiao X Zhang C and Wu C. "Medical Fraud and Abuse Detection System Based on Machine Learning". In: (2020).

[15] Richard A. Bauder and T. Khoshgoftaar. "The Detection of Medicare Fraud Using Machine Learning Methods with Excluded Provider Labels". In: *FLAIRS Conference*. 2018.

[16] E. Choi et al. *Medical Concept Representation Learning from Electronic Health Records and its Application on Heart Failure Prediction*. 2017.

[17] Y. Choi, Chill Yi-I Chiu, and D. Sontag. *Learning Low-Dimensional Representations of Medical Concepts*.

[18] D. Kiela, C. Wang, and K. Cho. *Dynamic Meta-Embeddings for Improved Sentence Representations*. 2018.

[19] M. Schuster et. al. *Bidirectional recurrent neural networks*.

[20] G. Huang et al. *Densely Connected Convolutional Networks*.

[21] William L. Hamilton, Rex Ying, and Jure Leskovec. *Inductive Representation Learning on Large Graphs*. 2018.

[22] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. *Effective Approaches to Attention-based Neural Machine Translation*. 2015.

[23] *HCF using auto-encoder*. URL: https://bit.ly/2NfqsZ3. (accessed: 16.02.2021).

[1] Precision [2] AUC-ROC [3] AUC-PR