Statistical Analysis of Hospital Data using MPC

Aditi Kumari
Department of Computer Engineering
San Jose State University
San Jose, USA
aditi.kumari@sjsu.edu

Darshil Kapadia
Department of Computer Engineering
San Jose State University
San Jose, USA
darshilpareshbhai.kapadia@sjsu.edu

Aditya Doshatti
Department of Computer Engineering
San Jose State University
San Jose, USA
aditya.doshatti@sjsu.edu

Devashish Nyati
Department of Computer Engineering
San Jose State University
San Jose, USA
devashish.nyati@sjsu.edu

Tech Elves
Team 12

***Abstract*— This paper describes the implementation of PatientAlyze, which is our implementation of Statistical Analysis of hospital's private data using MPC. PatientAlyze uses machine learning algorithms like time series forecasting using ARIMA model to predict the future of hospitals and compare the results of each hospital using Multi Party Computation to compare the results with other hospitals.**

***Keywords*— Multi Party Computation, Machine Learning, Time Series Forecasting, Hospital Data, Data Privacy**

I. Introduction

A lot of hospitals are competing with each other in order to be better than others. With the amount of data each hospital generates and maintains, there is a huge opportunity to analyse these data to bring forward some sensible information which can help in future. But for the analysis, data is most the important part. When we talk about hospital data, there are lots of information which should be kept private viz. patient details, revenue, # of patients they get every month due to which no hospital would ever participate in any analysis and would never share the private data with other hospitals. To unlock the potential of analysing different patterns and information b combining private data from different hospitals, we are thinking of developing a solution where multiple hospitals can share their data without compromising on the privacy for the analysis like which hospital is going to attract more patients in the upcoming year.

II. Multi Party Computation

A. Theory

MPC is a general idea that can be implemented by utilizing diverse conventions, for example, secret sharing, in which private information from each party, participating in the computation, is separated and divided as encoded "shares" amongst the parties, that when eventually consolidated, give the coveted measurable outcome. The hidden information of any of the party, whenever intercepted, would prove to be futile.
MPC could have an extensive effect in various zones, from strengthening the web security of individual information, to opening the possibility of healthcare institution's data to be accessible that are at presently out of reach for analysis purposes.

B. Example

Three hospitals (say X, Y and Z) want to compute who is wealthiest of all. But they are reluctant to share their incomes with each other. In such case, they would provide their income information to the third party. Let's say, X shares 42 million, Y shares 50 million and Z shares 10M. The third party will divide these data in a way like, 42M (as 10M, 30M and 2M) and 50M (as 20M,14M and 16M) and 10M (as 3M, 3M and 4M). X will receive his shares (10M, 20M and 3M), Y will receive his shares(30M,14M,3M), Z will receive his shares (2M,16M, and 4M). X will compute his result, here summation of 3 shares i.e. 33M, and will send the computation to party Y, encrypted with the public key of Y. Y will do the same and will share the result with Z. Z will do the final computation and will send the result to the Third Party. Third Party will take an average of the summation and will tell the parties the average result and their own income. The third party will also inform each party about their own standing.

C. Our Use-Case

In our case, we are going to ask hospitals to share their private data with our application. A sample data set of a hospital is shown in Fig1. Considering 3 hospitals participate in the computation, our product is going to analyse each of the data privately and will predict the future business of the hospitals, as shown in Fig 3. After that, the predictions will be shared

among 3 hospitals securely, which will do the necessary computation and will let each hospital know their predicted business and average predicted business. A sample output is shown in Fig 2.

1	Date	Number of Patients	Disease
2	1/6/14	392	Heart Attack
3	1/6/14	756	Diabetes
4	1/9/14	31	Lung Cancer
5	1/13/14	647	Fever
6	1/15/14	150	Heart Attack
7	1/16/14	124	Diabetes

Fig. 1: A sample data set of hospital.

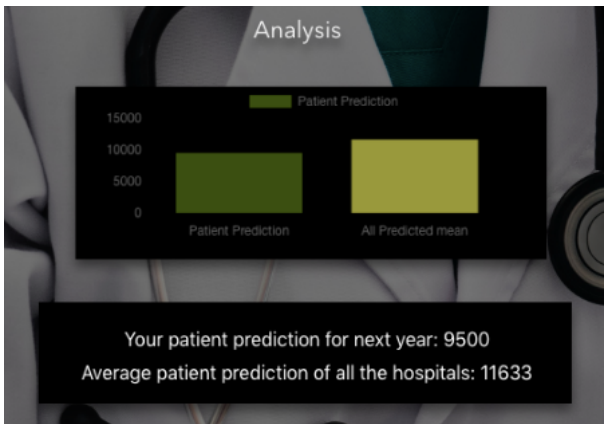

Fig. 2: Analysis after MPC

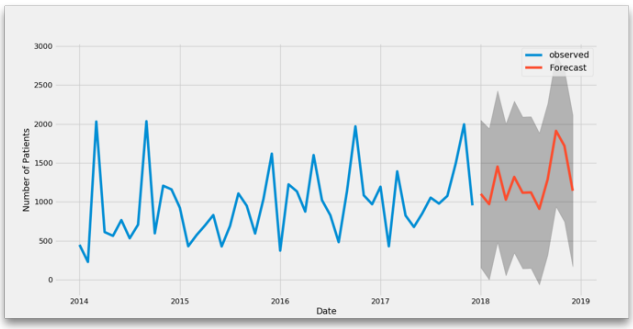

Fig. 3: Visualization of Time Series Forecasting

III. Time Series Forecasting

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

In PatientAlyze, we have analyzed the future prediction for number of patients. To get the predictions, we have done the following things on our data:

- Data Preprocessing
- Indexing with Time Series Data
- Time Series Forecasting with ARIMA
- Producing and Visualizing Results. Fig. 3 shows an example of future prediction.

For Time Series Forecasting, PatientAlyze uses one of the most commonly used method for time-series forecasting, known as ARIMA, which stands for Autoregressive Integrated Moving Average.

ARIMA models are denoted with the notation ARIMA (p, d, q). These three parameters account for seasonality, trend, and noise in data. We calculated the AIC value for different p, d, and q values and selected the one with lowest AIC value. Then we fitted the data with the ARIMA Model. The calculated p, d, and q values were entered in the model. Then we validated the forecasting on data which we already had. After validating, we produced and visualized forecasts. Then after forecasting we use secure multi party computation to compare the results of different hospitals which is explained in part 2.

IV. Architecture Flow

The architecture of PateintAlyze is as follows:

- Hospitals upload their patient's private data on-site with the PatientAlyze importer.
- PatientAlyze importer encrypts the uploaded data and upload the results to the PatientAlyze Application Server.
- PatientAlyze Server decrypts the data, applies Machine Learning algorithms to analyze the data and predicts the future crowd of hospitals.
- PatientAlyze utilizes secure computing technology- Multi Party Computing, to process the predicted data received after ML algorithms without removing the protection.

Fig. 4: Architecture of PatientAlyze, statistical analysis of hospital data using MPC.

- Using MPC, PatientAlyze compares the encrypted hospital data and produces results.
- Hospital get the results without knowing the data of the other hospitals.

As shown in fig. 4, hospital's confidential data is added to PatientAlyze and results are shown to the hospital.

V. Implementation

Fig. 5 shows a sample code, which takes care of secretly sharing the information among the parties participating. Here, party X, Y and Z are participating and Shamir Secret Sharing algorithm is being used here to divide the inputs into 3 encrypted parts and are given to each party. After successful computation, it securely opens the connection to send the encrypted data to the receiving party. And the third party takes care of the further computation.

```
x, y, z = rt.shamir_share([1, 2, 3], Zp, input)
sumVal = x + y + z
opened_sum = rt.open(sumVal)
opened_sum.addCallback(got_result)
```

Fig. 5: Multi Party Computation using Shamir Secret Sharing Algorithm

```
# Time series forecasting with ARIMA
mod = sm.tsa.statespace.SARIMAX(y,order=(1, 1, 1),
seasonal_order=(1, 1, 0, 12), enforce_stationarity=False,
enforce_invertibility=False)

results = mod.fit(disp=0)

# Producing and visualizing forecasts
pred_uc = results.get_forecast(steps=12)
pred_ci = pred_uc.conf_int()
```

Fig. 6: Time Series Forecasting using ARIMA

Fig. 6 shows a sample code by which were able to predict the future number of patients of each hospital based on the data they provide. We are using ARIMA model to forecast these behaviors.

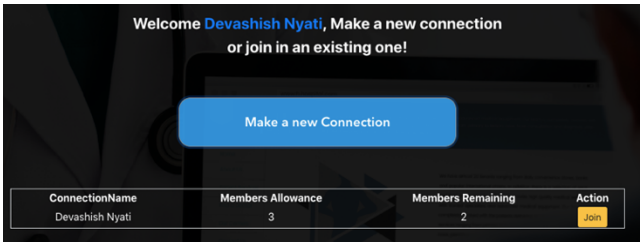

Fig. 7: Connecting with other hospitals

Fig. 7 shows the UI side, where any party can make a new connection (which will prompt user to enter the number of parties that connection will accept), or join an existing connection made by some other party.

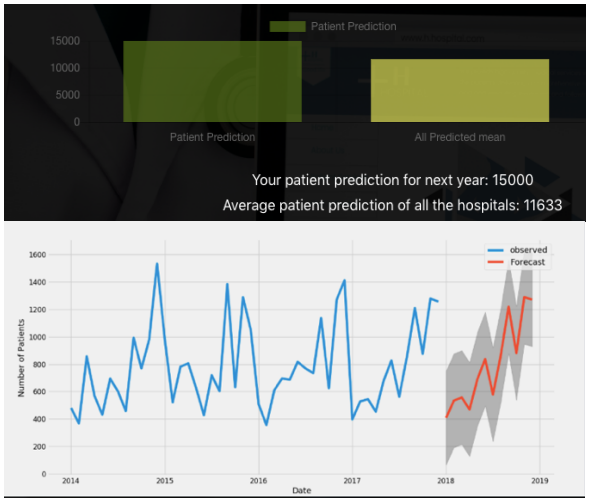

Fig. 8: Results on UI

After the connection gets full, our product does the job in the background and the parties will be able to see their data's interpretation and computation as shown in fig. 8

VI. Conclusions

In this project, we developed a secure system for doing statistical analysis of hospital's data without comprising with their privacy using MPC. This can prove very helpful in healthcare industry where there is need for continuous improvement. Using the data, hospitals can be well prepared with facilities and workforce for the proper treatment of each and every patient in coming years as well as strive to follow best practices.

For the future work, we would like to extend this application for the benefit of many hospitals. Also, we aim to use more secure web MPC capable of handling huge computational data and results. With current expertise and future enhancements in the project, we aim to build more secure and scalable application.

VII. Acknowledgment

We would like to thank Professor Rakesh Ranjan for his exemplary guidance and constant encouragement throughout the course of this project. The constant motivation to design innovative solution for the existing real-time challenges helped us to closely understand the existing problem and succeed in implementation of the project.

VIII. References

[1] VIFF, the Virtual Ideal Functionality Framework. http://viff.dk/
[2] Susan Li, *An end to end project n time series analysis and forecasting with python.* Available: https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b
[3] Dragoş Rotaru, *Awesome MPC.* Available: https://github.com/rdragos/awesome-mpc
[4] Sharemind, https://sharemind.cyber.ee/#
[5] Cybernetica, https://cyber.ee/