IBM
Research

# D’Avatar – Reincarnation of Personal Data Entities in Unstructured Text Datasets

IIM
Ahmedabad

Balaji Ganesan[1], Vaidyanathan Ramachandran[2], Riddhiman Dasgupta[1], Kalapriya Kannan[1], Kranti P Athalye[1], Arnab K. Laha[2]

[1]IBM, [2]IIM Ahmedabad Contact Email: bganesa1@in.ibm.com, x07vaidya@iima.ac.in

## Motivation

With the advent of GDPR in the EU, and the Personal Data Protection Bill, 2018 in India, there is increasing regulatory backing for our privacy, and the protection of our personal data shared with the govt and commercial organisations.

Because of recent advances in Deep Learning, it has become possible to detect personal data entities in documents with high precision and recall. However this research requires large amounts of annotated data.

**Aim**: To produce a dataset for research in privacy and protection of personal data, especially in India.

## Hackathon – Jan 29 – Mar 12, 2019

**Problem Statement:** A method to automatically impute values for the redacted portions in a text, which are known to have contained Personal Data Entities.

*“My credit card number is xxxx and I wish to raise a compliant ….”.*

The entity masked with xxxx is the redacted portion. We might be able to guess that a 16 digit credit card number was originally present in this text.

An acceptable output is to replace xxxx with some variant of a 16 digit number.

*“My credit card number is 1234-5678-9012-3456 and I wish to raise a compliant ….”*

However, a better output will be credit card number which is not completely random, but obeys the Luhn algorithm.

## Research

Solutions to this personal data entities imputation problem can include the following:

- **Rule Based Systems**: Using patterns and dictionaries, a number of rules can be written to impute personal data entities in the redacted spaces. This process can then be done at scale using Data Programming (ex: Snorkel).
- **Word Embeddings and Language Models**: Word embeddings can predict a word given other words in the sentence. But restricting the predicted words to personal data entities and types will require smarter methods.
- **Natural Language Generation:** Auto-encoders and other models can be trained to generate the natural language text. Here the words generated will be in context and not just based on the proximity in vector space.

We encouraged participants to try these approaches.

## Participation, Evaluation and Results

- 166 teams from various colleges and organizations in India registered for the hackathon.
- The dataset contained of 30916 sentences from a Bank Credit Card complaints dataset. Each sentence has one personal data entity already redacted from it.
- We evaluated the results by comparing the imputed values with manually generated ground truth of 100 randomly chosen sentences.
- We used F1 score to evaluate the solutions. We reported the Strict F1 along with Loose Macro and Loose Micro F1.
- The winning team achieved the following F1.
  - Team Members: Piyush Sharma (TCS), Manoj Tiwari, Raghu Ramakrishnan (TCS)
  - strict f1: 0.42
    loose macro f1: {'p': 0.49, 'r': 0.45, 'f1': 0.47}
    loose micro f1: {'p': 0.49, 'r': 0.41, 'f1': 0.45}

## Related Work

**Fine Grained Classification:** A corollary to this hackathon problem of imputing values is the task of detecting and classifying personal data entities in unstructured text.

Riddhiman Dasgupta, Balaji Ganesan, Aswin Kannan, Berthold Reinwald, and Arun Kumar. "Fine Grained Classification of Personal Data Entities." *arXiv preprint arXiv:1811.09368* (2018).

**Bias Detection in Datasets:** While imputing values to create a dataset, it is desirable that the dataset draws personal data entities without bias in any of the classes (known as protected variables).

Bellamy, Rachel KE, Kuntal Dey, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia et al. "Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias." *arXiv preprint arXiv:1810.01943* (2018).

## Resources

- Hackathon Page: https://sites.google.com/iimahd.ernet.in/hackathon-icadabai2019/home
- Github: https://github.com/hackathoniima/ICADABAI2019
- ICADABAI 2019 Conference: https://conference.iima.ac.in/icadabai/2019/