Bag-Of-Words Format Preserving Encryption

Ciaran McMurtrie
MSc in Information Technology

UNIVERSITY of STIRLING

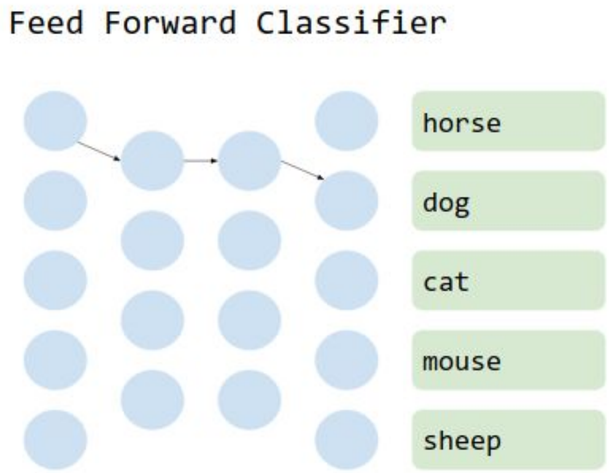
Machine Learning Data Security

Companies are under increasing pressure to share data with third parties, many of which are using machine learning technology to offer predictive services (for example, by predicting a borrower's ability to service a mortgage loan).

Rather than share personally identifiable, plain-text data with third parties, there is a need for encryption methodologies that preserve the data's predictive precision while keeping the user's right to data privacy under general data protection laws.

In this work, I present a simple method of encrypting plain-text data with no loss in terms of recorded accuracy, using standard academic datasets. The technique I propose is a partial encryption method (potentially lossy) that nonetheless adds elements of security to any document classification task.

DigitalGenius



Format-Preserving Encryption

There is not an agreed methodology for data encryption that addresses machine learning problems specifically, but format-preserved encryption presents the most opportunities.

Format-preserved encryption is any encryption methodology that conserves the format of the input data. The meaning of "format" can vary with the specific application.

"the dog is on the table"

are	cat	dog	is	now	on	table	the
0	0	1	1	0	1	1	1

```
Accuracy on test data (no-encryption): 0.634094530005
Encrypted example: m9d YMG tb MF m9d mRsid
Decrypted example: the dog is on the table
Accuracy on test data (with encryption): 0.644184811471
```

Bag-Of-Words FPE

In general, a bag-of-words classifier does not require word-sentence order to be conserved. In the bag-of-words model, the presence of a word in a document vector is enough to train a classification model.

I will offer a few variations on a FPE solution with different degrees of lossiness, and will present results that show that predictive accuracy in bag-of-words classification tasks is conserved.

I will also show that by preserving format, some natural language processing techniques (such as stop-word removal) can port easily into a data encryption cipher, as long as the classifier has access to the encryption keys.

For further information contact: ++447923 165 341

cdm00020@students.stir.ac.uk