Intro

Class imbalance is prevalent across all problems.

→ How to do anomaly detection.

- Medical
- Net. Security
-

Carleton Data Science

Focus on data science

↳ for Masters Dgr.

2 New Programs of D.S

→ w- PHD program

Data Day on Mar 31st!

Benjamin Fung Presentation

Data-mine & ML for (on Authorship & Malware Analysis).
funded by CFI.

Today focus on National Defense Sector

Authorship Analysis

feature-engineering w-embedding

→ can be used for different problems

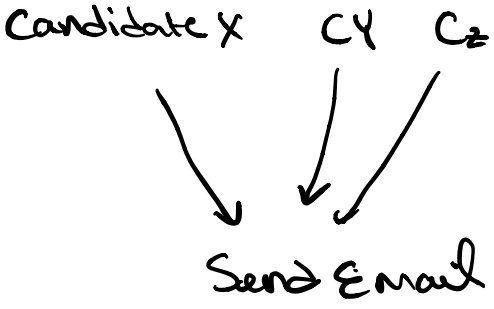
- deep learning not so reusable.

Cheque Scam

1. Send cheque to someone
2. Ask for return on some money
3. Bounce the cheque

How to prove the scammer is the author of the email?

Authorship Identification



There is networking proof but no link between keyboard & typer

Handcrafted Stylometric Features

- lexical (caps/no caps)
- syntactic (spaces, commas)
- structural
- content-specific
- idiosyncratic features

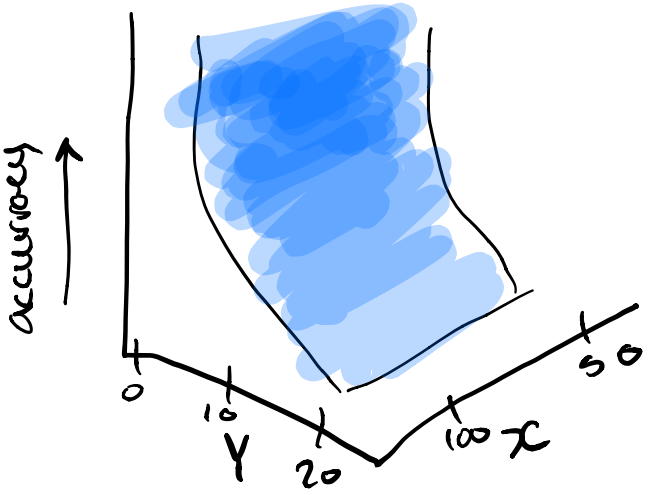
why Models don't work:

Naive Bayes
- low classification accuracy.

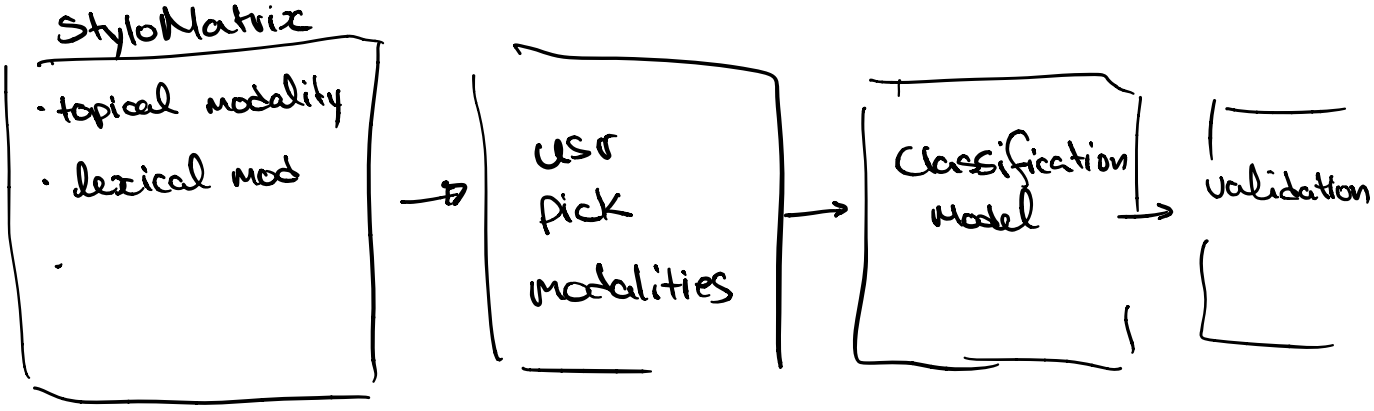
Decision Trees
- accuracy ≈ SVM

Support Vector Machine

- small data



X: # writing samples
Y: number of candidates



topic-lexic-2vec

1. find topic
2. find context
3. find bias

} find in emails.

Also is Authorship Verification, Auth. Characterization.

↓ age ↓ political orientation → sex.

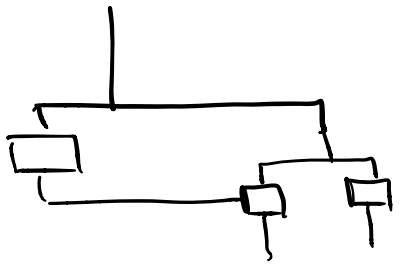
Assembly — Reverse Engineering

Given binary code - how to understand its behaviour.

push eax
. . . .
. . .
test ...
jz ...

⟹

Control Flow Graph



PROBLEM — how to find subgraph clone

Approach

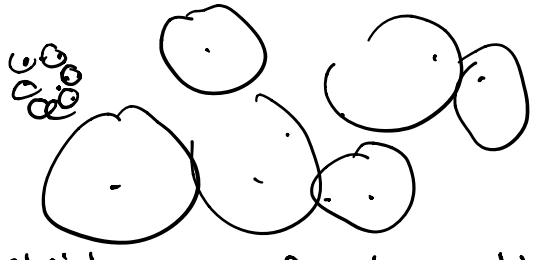
① when saving Hash code & compare

② diff operators
normalize & hash

③ almost same - idiosyncracies
can't hash → do subgraph clone w- search engine.

Adaptive Locality Sens Hashing

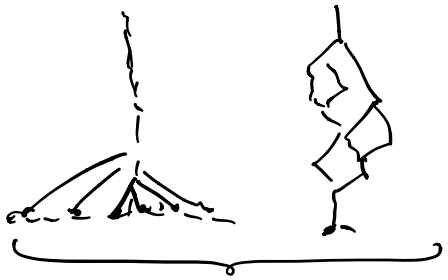
locality sensing hashing



shrink scope for dense distrib.

- find circular reference
- expand circ. ref as much possible.

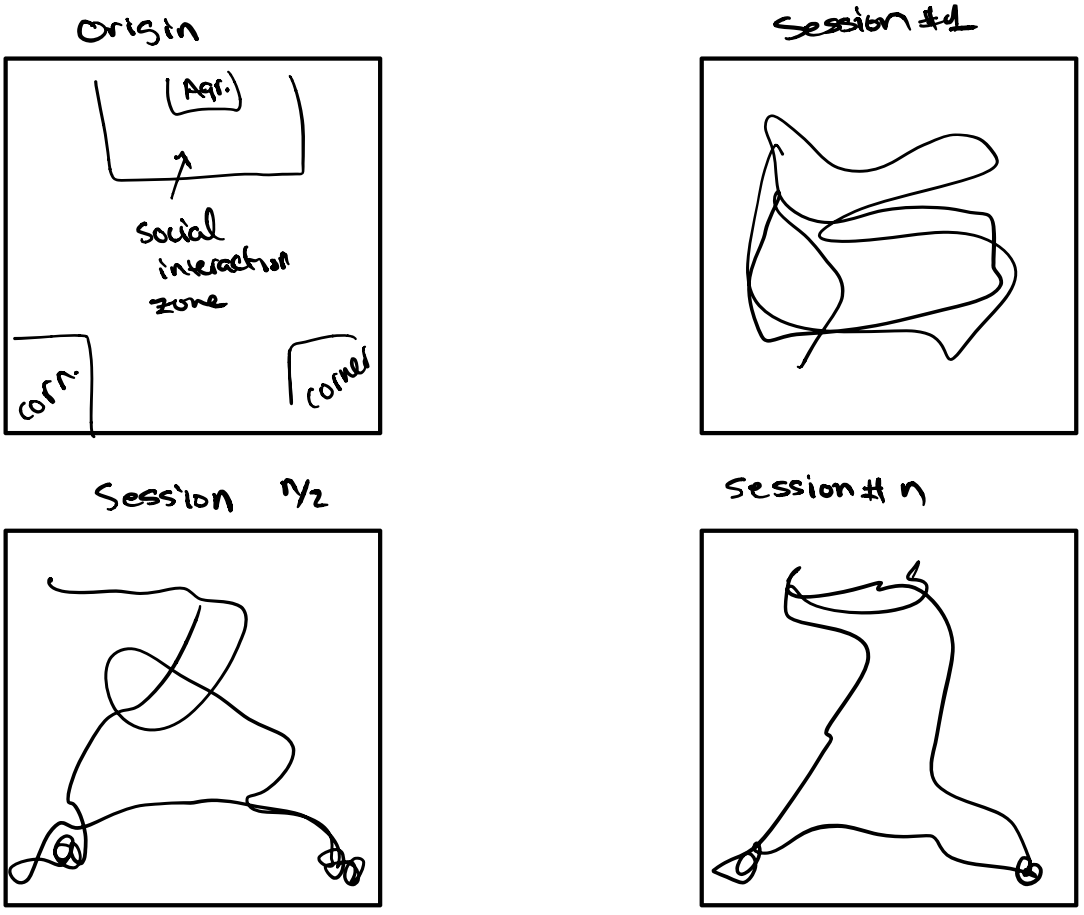
Obfuscation & Optimization



Same code – how detect?

↳ use ML/AI (KamInO).

Fighting Rats



(*) Could classify political sentiment using 2013 twitter data.

→ StyloMatrix. App on GitHub