

Pattern Recognition

Exercise Session 10

Last Group Projects + Evaluation Infos

Paul Märgner (paul.maergner@unifr.ch)

- You have to register for the exam!
 - Deadline: May 18th, 2018
 - Withdrawal without reason until June 4th, 2018
- Exam is on June 18th, 2018 10:00-11:00AM

Task 3 – Keyword Spotting

Don't forget the current group task

- Keyword Spotting
- Deadline: **April 30, 2018 (end of day)**
- **Group Presentations on May 7, 2018**

Data and Info on Github:

https://github.com/lunactic/PatRec17_KWS_Data

Deadline: Tuesday, May 22nd, 23:59

Signature Verification

Can be solved with DTW

Molecules

Use approximate Graph Edit Distance

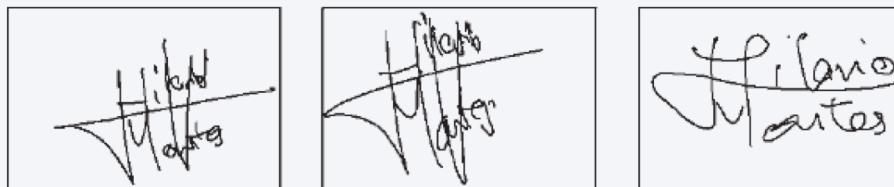
You need to solve one!

30 writers

Enrollment: 5 genuine signatures each

Verification: 45 signatures each (20 genuine, 25 forgeries)

Ground truth in verification-gt.txt



Task

Compute dissimilarity for each verification signature wrt
the 5 genuine ones

Input

t x y pressure penup azimuth inclination

Penup 1 if change between pen-up and pen-down

Azimuth / inclination angles of the pen

Recommendation: DTW

Features: x , y , v_x , v_y , pressure

v_x , v_y velocity in x and y with respect to Δt

Normalize for each signature individually

Sakoe-Chiba band can be helpful

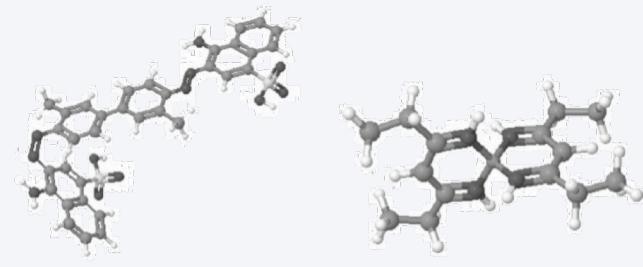
Evaluation: mean average-precision

AIDS Antiviral Screen Database of Active Compounds

250 training, 250 validation molecules

Two classes active 'a' and inactive 'i'

Annotated in train.txt and valid.txt



Task

Classify the molecules of the validation set using KNN

Distance: approximate GED

Input: Graph xml

Nodes labeled with their chemical symbol

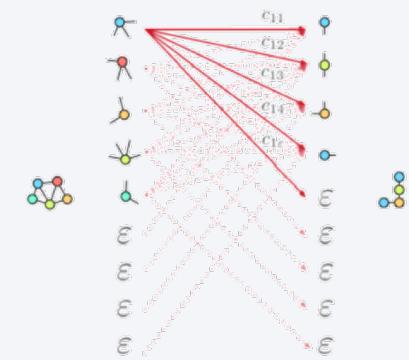
Unlabeled undirected edges

Compute approximate GED between pairs of molecules with bipartite graph matching

(lecture 10, slide 21)

Build cost matrix (*Dirac*)

$$\mathbf{C} = \left[\begin{array}{cccc|cccc|c} c_{11} & c_{12} & \cdots & c_{1m} & c_{1e} & \infty & \cdots & \infty \\ c_{21} & c_{22} & \cdots & c_{2m} & \infty & c_{2e} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \infty \\ c_{n1} & c_{n2} & \cdots & c_{nm} & \infty & \cdots & \infty & c_{ne} \\ \hline c_{e1} & \infty & \cdots & \infty & 0 & 0 & \cdots & 0 \\ \infty & c_{e2} & \ddots & \vdots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \infty & \vdots & \ddots & \ddots & 0 \\ \infty & \cdots & \infty & c_{em} & 0 & \cdots & 0 & 0 \end{array} \right]$$



Hungarian Algorithm
To find optimal assignment

Derive Edit Path costs from the result
(distance for classification)

KNN for classification (optimize for K)

Recommendation

Use *Dirac cost function* for GED (optimize C_n and C_e)
(lecture 9, slide 36)

Node substitution: $2 * C_n$ if symbols \neq , 0 otherwise

Node deletion/insertion: C_n

Edge deletion/insertion: C_e

Use an existing framework for the Hungarian algorithm

Evaluation

30/04/2018

You have to run evaluations on all group tasks

Dates and times are still tentative

Test Data available:

Tuesday, May 22nd, 10:00

Results/Software/Readme on github:

Thursday, May 24th, 23:59

Your results will be crosschecked

Input:

Images of numbers

Expected Output:

ASCII plaintext file

One line per test sample

test_ID1, predicted_classX
test_ID2, predicted_classY

Input:

As keywords.txt

Expected Output:

ASCII plaintext file

One line per keyword

Keyword1, testword_ID1, dissimilarity1, testword_ID2, dissimilarity2, ...

Keyword2, testword_ID1, dissimilarity1, testword_ID2, dissimilarity2, ...

Input

Same folder and input structures (enrollment/verification)

Expected Output:

ASCII plaintext file

One line per user (can be sorted by dissimilarity)

```
user1, signature_ID11, dissimilarity11, signature_ID12, dissimilarity12, ...
user2, signature_ID21, dissimilarity21, signature_ID22, dissimilarity22, ...
```

Final Group Presentations on **Monday, May 28th**

Part 1:

Signatures / Molecules results

Part 2:

Team and Software organization

What makes your solutions special

Lessons learned

Deadline for slides upload: **Monday, May 28, 14:00**

Questions?