Multi-View Data Mining

Kick-off meeting

27 October 2014

max planck institut
informatik

Today's Agenda

- Check attendance
- Goals of the seminar
- Organization of the seminar
- Grading & guidelines
- A brief introduction to the topic

First Things First

- A block seminar
  - Preliminary work + one (or two) day(s) of presentations
- 7 ECTS credits
- Meeting all DLs and attending all talks is **mandatory** for passing the seminar
- Attending this kick-off meeting is mandatory

Head Count

Goals

- To learn how to read and understand recent research literature

- To learn how to write a concise report of a research article

- To learn how to present research

- To boldly read what no one (at this seminar) has read before

- To keep young people out of streets

Workflow

1. You read the paper (+ other papers)
2. You write a draft report and send it to me
3. I comment your report
4. You improve your report and prepare your presentation, which you also send to me
5. I comment your presentation
6. You improve your presentation and send me the final report
7. I distribute the reports to everybody
8. You read others' reports
9. You present your work and follow and discuss others' presentations

Schedule (1)

Day	Topic
27 October	Kick-off
December	Report draft DL
January	Slides draft DL
1 February	Report DL
February/March	Seminar

Schedule (2)

Day	Topic
27 October	Kick-off
late November	Report draft DL
mid December	Slides draft DL
late December	Report DL
early January	Seminar

December and January's Midterms

Klausurenkalender/ Examination Schedule - Saarland University - Department of Computer Science

December 2014 — Month view

Sun	Mon	Tue	Wed	Thu	Fri	Sat
30	Dec 1	2	3	4	5	6
7	8	9	10 2pm Grundzüg	11	12	13 9am Programn
14	15	16	17 2pm Algorithm	18 4pm Grundzüg	19	20
21	22	23	24 Heiligabend	25 1. Weihnacht	26 2. Weihnacht	27
28	29	30	31	Jan 1 Neujahr	2	3

Events shown in time zone: Berlin

Klausurenkalender/ Examination Schedule - Saarland University - Department of Computer Science

January 2015 — Month view

Sun	Mon	Tue	Wed	Thu	Fri	Sat
28	29	30	31	Jan 1 Neujahr	2	3
4	5	6	7	8	9	10 9am Lineare Al
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Events shown in time zone: Berlin

February and March's Exams

Klausurenkalender/ Examination Schedule - Saarland University - Department of Computer Science

February 2015 — Month view

Sun	Mon	Tue	Wed	Thu	Fri	Sat
Feb 1	2	3	4 12pm Grundlag	5	6	7
8	9 10am Automat 4pm Ideen der	10	11	12 4pm Security -	13 12pm Convex 4pm An Introd	14 9am Mathemat
15	16 Rosenmontag	17 Fastnachtsdie 10am Advance	18 9am Lineare A 2pm Differenti	19 9am Höhere M 10am Future M 2pm Image Cc 3pm Grundzüg	20 9am Algorithm 2pm Modelling	21 9am Programn
22	23 9am Analysis : 9am Grundlag	24 9am Mathema	25	26	27	28

Events shown in time zone: Berlin

Klausurenkalender/ Examination Schedule - Saarland University - Department of Computer Science

March 2015 — Month view

Sun	Mon	Tue	Wed	Thu	Fri	Sat
Mar 1	2 9am Grundzüg	3	4	5	6	7
8	9	10	11	12 10am Unixkurs	13	14
15	16	17 10am Convex	18 2pm Differenti	19	20	21
22	23	24 10am Advance	25 12pm Grundlag	26 9am Grundlag 10am Databas	27	28
29	30 MNU Bundes	31	Apr 1	2 9am Programm	3 Karfreitag	4

Events shown in time zone: Berlin

Grading Overview

- Report (2–4 pages):

  - Correctness, connections, criticism, style

- Slides & presentation (20 min):

  - Delivery, clearness, presentation skills

- Discussion (5–10 min):

  - Participation, correctness, connections

Report

- 2–4 pages
- **In your own words**
  - No verbatim copy
- Explain the main ideas of the paper
  - Research questions
  - Proposed solutions and their evaluation
- Write to somebody who hasn't read the paper
- Provide also extra connections & criticism towards the approach

Presentation

- 20 minutes (+ 5–10 minutes of discussion)
- Explain the main ideas of the paper
  - What, why, and how
- Build connections and provide criticism where appropriate
- Target to audience that knows CS and basics of DM and ML
  - Some audience knows more on your topic, others less; cater for both groups

Discussion

- After you've followed a presentation, discuss
  - You know the topic a bit: you've read the report
- Ask if something was left unclear
- Tell if you know more on something
- After technical part, give constructive feedback on the presentation itself
- Actually discussing is **mandatory**

Introduction to Topics
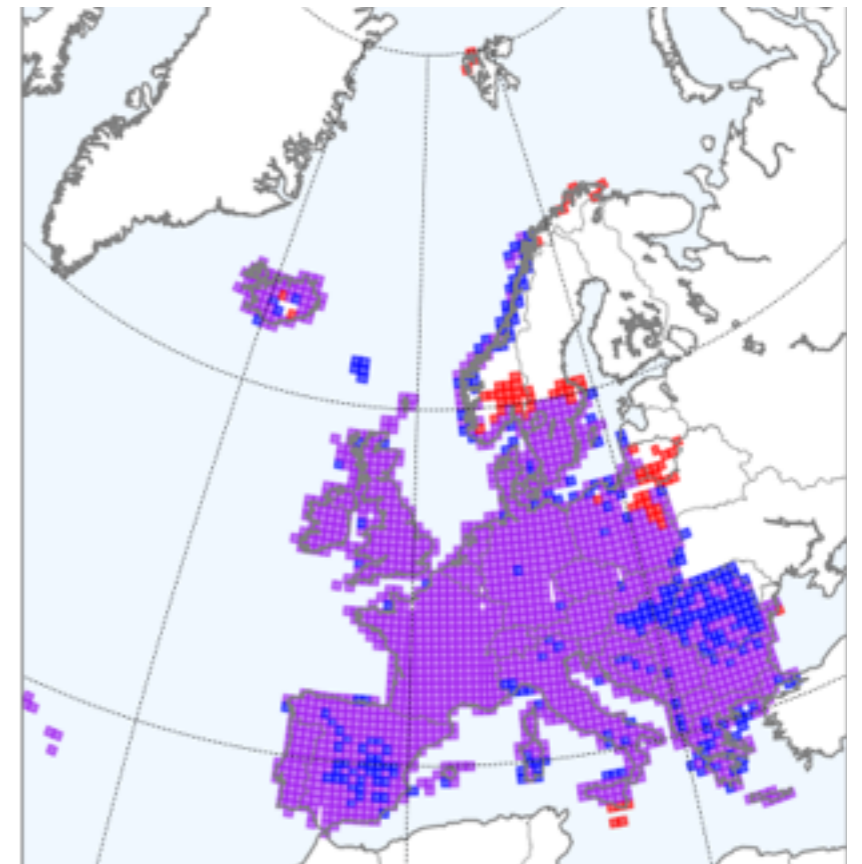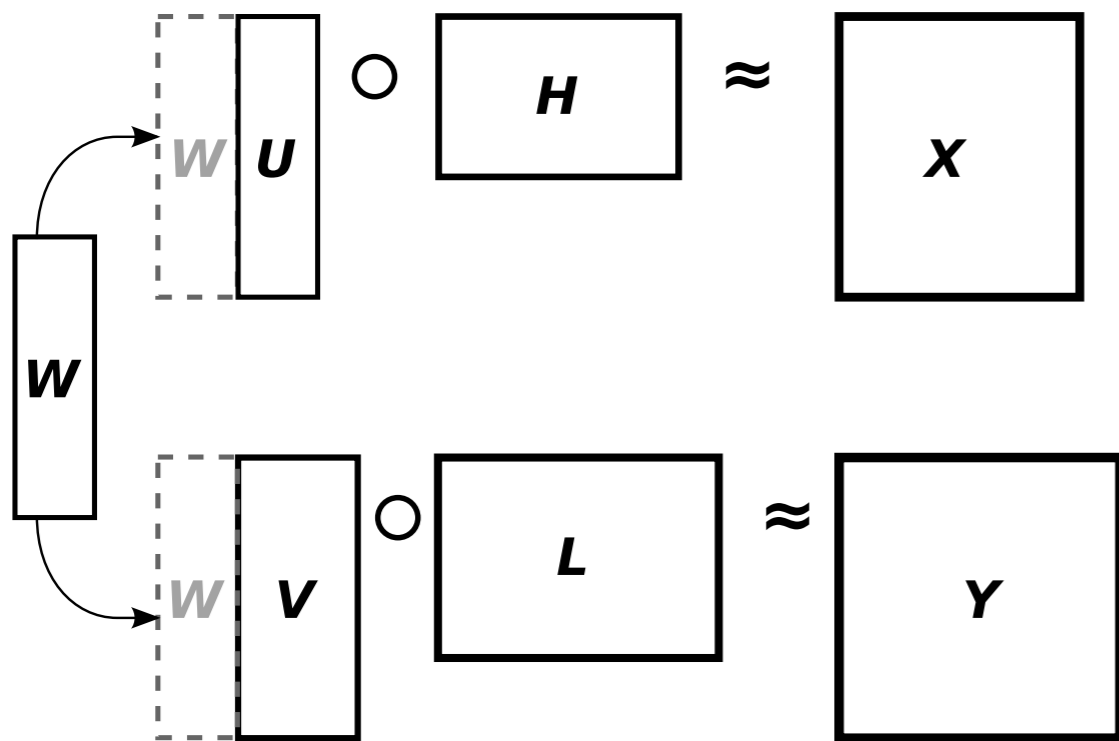
Wood mouse ∨ Azores Noctule ∨ Harp Seal

[−0.8 ≤ t2+] ∧ [−0.141 ≤ t10~ ≤ 19.6] ∧ [26.6 ≤ p4]

Multi-View Data

- More and more multi-view (or multi-relational) data
  - Phenotypes and genotypes of patients
  - User preferences and genres of music and movies
  - Bioclimatic conditions and fauna or flora of geographical areas
  - …

Mining Multi-View Data

- The goal of mining multi-view data is to find the connections between the data *beyond just (simple) correlations*

- The connections (hopefully) tell us something any single view cannot

  - These species require these kinds of conditions, this is how genre effects the song's popularity vs. personal preferences, etc.

Redescription Mining: An Informal Definition

- A **redescription** provides two ways of describing the same set of entities

  - Descriptions are statements over entities' attributes

    - Tells us something about interesting attributes

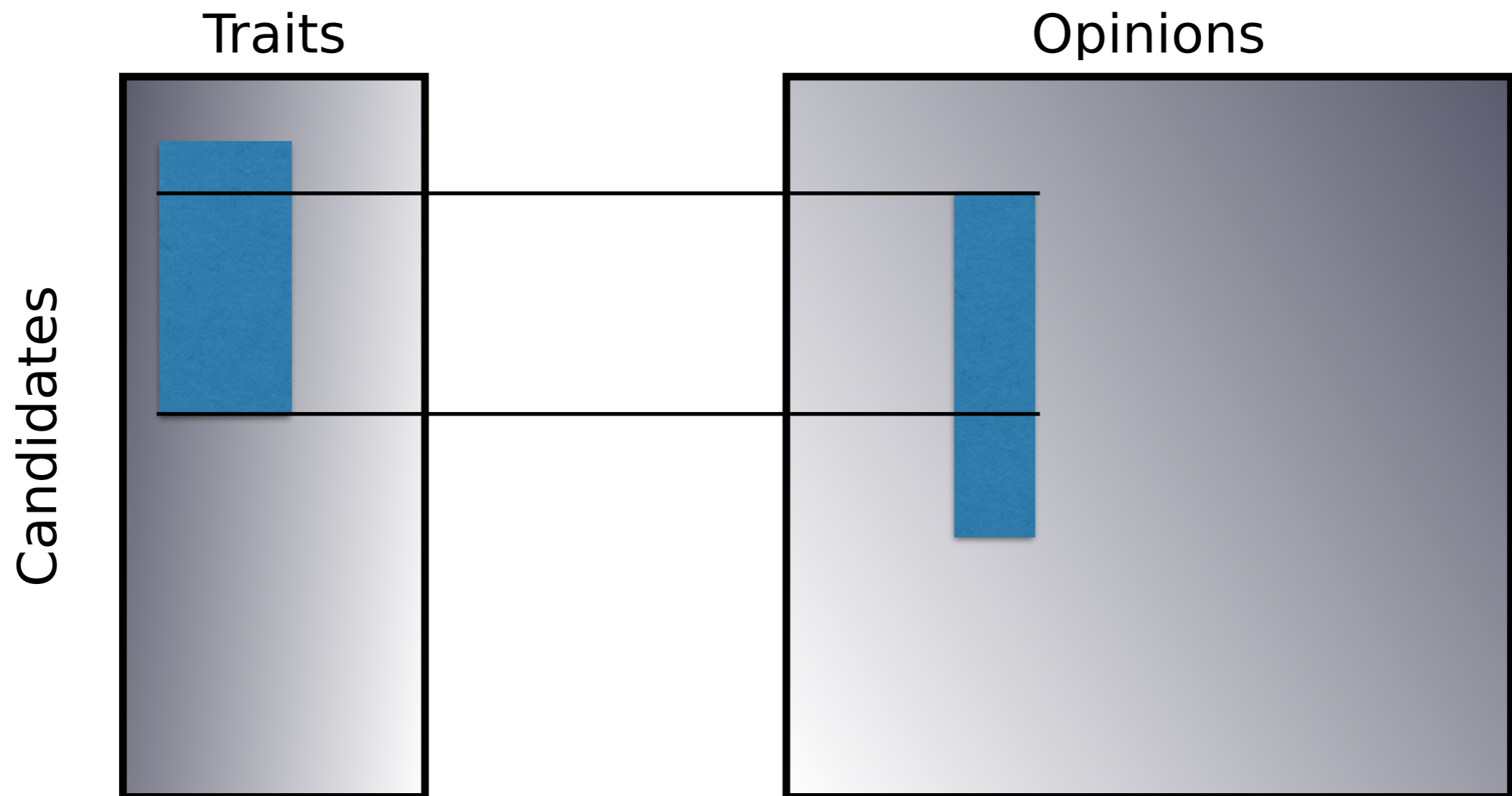 - Also the set of entities is interesting

Example

EDUSKUNTA
RIKSDAGEN

In the 2011 parliamentary elections in Finland, the candidates who were female
or were at most 39 years old
were (approximately) the candidates who
supported gay families' right to adopt outside the family

Example

[Gender = F] ∨ [Age ≤ 39]
⇔
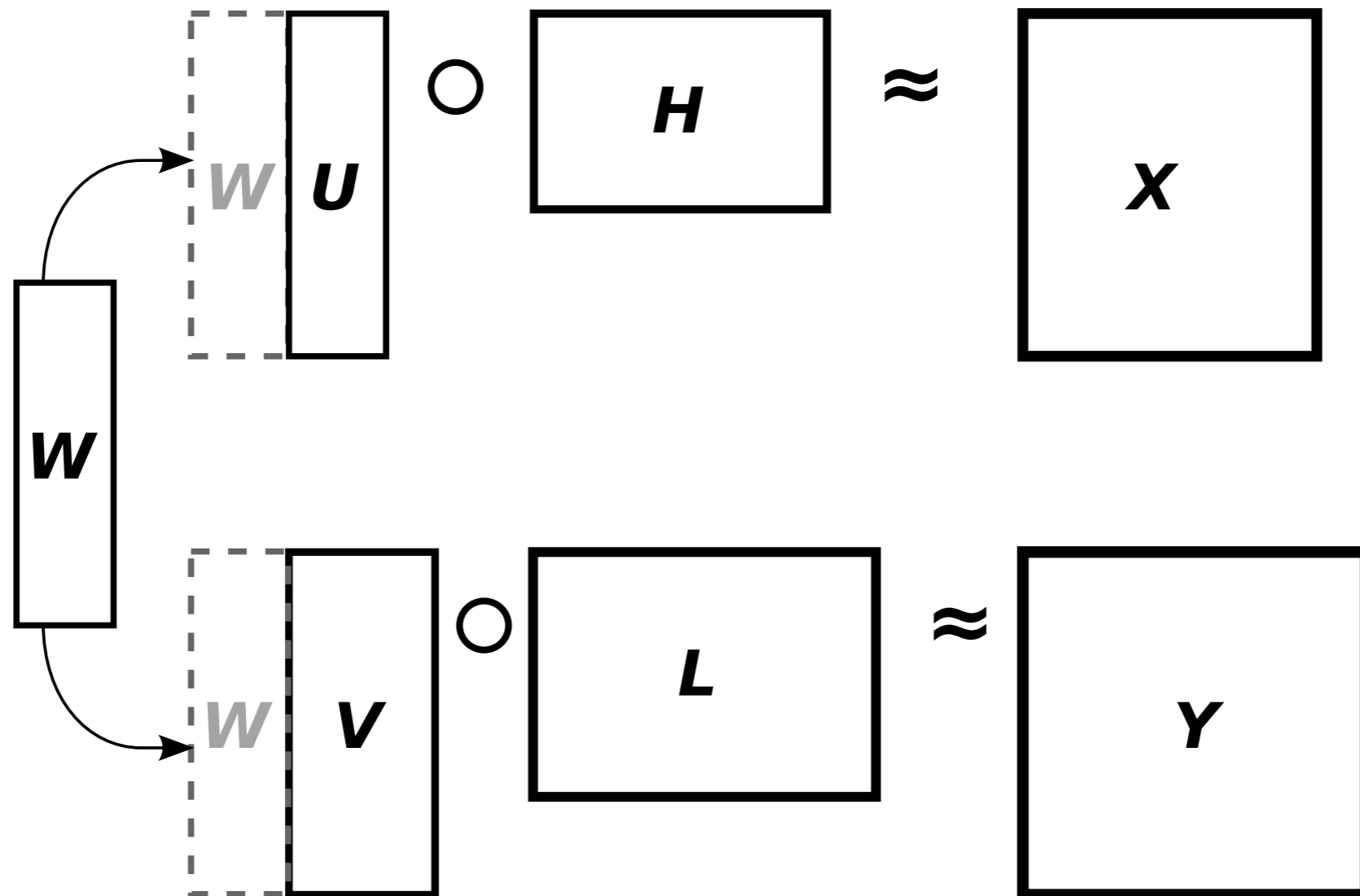[Supports Gay Adoption Rights = True]

Traits

Opinions

Candidates

Shared Subspaces: An Informal Definition

- A **shared-subspace factorization** decomposes two matrices X and Y into five matrices W, U, V, H, and L such that $X \approx [W\ U] \times H$ and $Y \approx [W\ V] \times L$

  - Matrix W contains the shared subspace while U and V contain subspaces specific to X and Y

Image

Example

- **X** is movies-by-actors matrix
- **Y** is movies-by-genres matrix
- **W** maps movies to a subspace that explains the common behaviour between **X** and **Y**
  - RomCom and Hugh Grant; Action and Bruce Willis; …
- **U** and **V** explain the specific behaviour

Caveat Emptor

- These are just broad generalizations
- Different papers take different view on the problems
  - Can be significantly different to the one I presented here
  - Usually the underlying ideas are similar, though