Organisation

Katja Markert

PS/HS Bias, Heidelberg 2019

- Tuesday: 14.15 - 15.45
- Bachelor 6 LP, Master 8 LP
- Overview: Which semester? Subject combinations? Prior knowledge?
- Language: English, German possible if ok with everybody
- No Notebooks, smartphones etc
- My office hours: Thursday 11.15 - 12.15

- For Master students: none
- For Bachelor students: Statistical Methods for Computational Linguistics OR Neural Networks: Architectures and Applications for NLP OR Embeddings

Structure

1. Bias operationalisations, legal background (1 session)
2. Bias as Stereotypes: Measuring and Mitigating Bias in Word Embeddings (3 Sessions)
3. Selection Bias: Bias in Wikipedia (1 session)
4. Evaluation Corpora: GBETS (Gender Bias Evaluation Test Sets) for Coreference Resolution (1 Session)
5. Case Studies I: Bias in MT and dialect processing (1 or 2 sessions)
6. Bias as disparate impact of machine learning classification on protected groups (2 Sessions)
7. Case Studies II: Bias in Text Classification (1 Session)
8. Case Studies III: Visual semantic role labeling and/or image search results (1 session)
9. Final discussion (1 session)
10. In holidays (not obligatory): 1 Session to discuss intermediate progress on term papers and projects

See https://www.cl.uni-heidelberg.de/courses/ws19/bias/material

Structure

1. Bias operationalisations, legal background (1 session)
2. Bias as Stereotypes: Measuring and Mitigating Bias in Word Embeddings (3 Sessions)
3. Selection Bias: Bias in Wikipedia (1 session)
4. Evaluation Corpora: GBETS (Gender Bias Evaluation Test Sets) for Coreference Resolution (1 Session)
5. Case Studies I: Bias in MT and dialect processing (1 or 2 sessions)
6. Bias as disparate impact of machine learning classification on protected groups (2 Sessions)
7. Case Studies II: Bias in Text Classification (1 Session)
8. Case Studies III: Visual semantic role labeling and/or image search results (1 session)
9. Final discussion (1 session)
10. In holidays (not obligatory): 1 Session to discuss intermediate progress on term papers and projects

See https:
//www.cl.uni-heidelberg.de/courses/ws19/bias/material

4

Structure

1. Bias operationalisations, legal background (1 session)
2. Bias as Stereotypes: Measuring and Mitigating Bias in Word Embeddings (3 Sessions)
3. Selection Bias: Bias in Wikipedia (1 session)
4. Evaluation Corpora: GBETS (Gender Bias Evaluation Test Sets) for Coreference Resolution (1 Session)
5. Case Studies I: Bias in MT and dialect processing (1 or 2 sessions)
6. Bias as disparate impact of machine learning classification on protected groups (2 Sessions)
7. Case Studies II: Bias in Text Classification (1 Session)
8. Case Studies III: Visual semantic role labeling and/or image search results (1 session)
9. Final discussion (1 session)
10. In holidays (not obligatory): 1 Session to discuss intermediate progress on term papers and projects

See https: //www.cl.uni-heidelberg.de/courses/ws19/bias/material

Structure

1. Bias operationalisations, legal background (1 session)
2. Bias as Stereotypes: Measuring and Mitigating Bias in Word Embeddings (3 Sessions)
3. Selection Bias: Bias in Wikipedia (1 session)
4. Evaluation Corpora: GBETS (Gender Bias Evaluation Test Sets) for Coreference Resolution (1 Session)
5. Case Studies I: Bias in MT and dialect processing (1 or 2 sessions)
6. Bias as disparate impact of machine learning classification on protected groups (2 Sessions)
7. Case Studies II: Bias in Text Classification (1 Session)
8. Case Studies III: Visual semantic role labeling and/or image search results (1 session)
9. Final discussion (1 session)
10. In holidays (not obligatory): 1 Session to discuss intermediate progress on term papers and projects

See https: //www.cl.uni-heidelberg.de/courses/ws19/bias/material

4

Structure

1. Bias operationalisations, legal background (1 session)
2. Bias as Stereotypes: Measuring and Mitigating Bias in Word Embeddings (3 Sessions)
3. Selection Bias: Bias in Wikipedia (1 session)
4. Evaluation Corpora: GBETS (Gender Bias Evaluation Test Sets) for Coreference Resolution (1 Session)
5. Case Studies I: Bias in MT and dialect processing (1 or 2 sessions)
6. Bias as disparate impact of machine learning classification on protected groups (2 Sessions)
7. Case Studies II: Bias in Text Classification (1 Session)
8. Case Studies III: Visual semantic role labeling and/or image search results (1 session)
9. Final discussion (1 session)
10. In holidays (not obligatory): 1 Session to discuss intermediate progress on term papers and projects

See https:
//www.cl.uni-heidelberg.de/courses/ws19/bias/material

4

Structure

1. Bias operationalisations, legal background (1 session)
2. Bias as Stereotypes: Measuring and Mitigating Bias in Word Embeddings (3 Sessions)
3. Selection Bias: Bias in Wikipedia (1 session)
4. Evaluation Corpora: GBETS (Gender Bias Evaluation Test Sets) for Coreference Resolution (1 Session)
5. Case Studies I: Bias in MT and dialect processing (1 or 2 sessions)
6. Bias as disparate impact of machine learning classification on protected groups (2 Sessions)
7. Case Studies II: Bias in Text Classification (1 Session)
8. Case Studies III: Visual semantic role labeling and/or image search results (1 session)
9. Final discussion (1 session)
10. In holidays (not obligatory): 1 Session to discuss intermediate progress on term papers and projects

See https:

//www.cl.uni-heidelberg.de/courses/ws19/bias/material

- Active Participation: 33%
- Presentation: 33%
- Term Paper or Project or Second Presentation: 33%
- All components have to be passed **individually**

How do I pass the module: Active Participation

- Attendance: maximum twice unexcused missing of seminar
- Participating during seminar discussions
- **Written exercises/questions:** max 3 to be submitted to me (markert@cl.uni-heidelberg.de) per email by Monday 12 o'clock before the seminar with the subject line BIAS: *Date of session*
- **Reviewing:** You will be reviewer for two sessions that you do not orally present with responsibility for discussion session and marking feedback (peer review of talk):
  - Ask questions on paper and evaluate paper
  - Steer and get involved in discussion after talk
  - Give additional background on paper or problem
  - Give feedback to module leader on talk

- Assessment criteria: Research plan, Implementation, Demo and/or Documentation, Evaluation
- Presentation on one to two papers from literature list
- Differences between PS and HS:
  - PS: mostly a **single** paper, **easier**, **30 (to 40) minutes** plus discussion. Potentially shared between two people (then 60 minutes plus discussion)
  - HS: mostly **two** (related) papers, **harder**, **60 minutes** plus discussion

Content (60%), Style (20%), Discussion (20%)

- **Content**
    - Problem definition and originality
    - Has the main point been grasped?
    - Technical and/or linguistic background and methodology
    - Evaluation
    - Paper Comparisons: has the main related work been looked at?
    - Own opinion: Weaknesses and Strengths of Papers, future work, potential

- **Presentation:** Structure, Examples, Background and additional material, slides and graphics, clarity,

- **Leading the discussion:** Is discussion being stimulated? Can questions be answered? How is the interaction with the reviewers?

Content (60%), Style (20%), Discussion (20%)

- **Content**
    - Problem definition and originality
    - Has the main point been grasped?
    - Technical and/or linguistic background and methodology
    - Evaluation
    - Paper Comparisons: has the main related work been looked at?
    - Own opinion: Weaknesses and Strengths of Papers, future work, potential

- **Presentation:** Structure, Examples, Background and additional material, slides and graphics, clarity,

- **Leading the discussion:** Is discussion being stimulated? Can questions be answered? How is the interaction with the reviewers?

Content (60%), Style (20%), Discussion (20%)

- **Content**
    - Problem definition and originality
    - Has the main point been grasped?
    - Technical and/or linguistic background and methodology
    - Evaluation
    - Paper Comparisons: has the main related work been looked at?
    - Own opinion: Weaknesses and Strengths of Papers, future work, potential

- **Presentation:** Structure, Examples, Background and additional material, slides and graphics, clarity,

- **Leading the discussion:** Is discussion being stimulated? Can questions be answered? How is the interaction with the reviewers?

- Send me draft slides **at least 7 days** before own talk (exception first talk)
- Opportunity for feedback: Office hour Thursday before seminar
- Discuss your presentation with reviewers if necessary

- Michael Wiegand has made an excellent overview over presentation dos and donts available here: `https://www.cl.uni-heidelberg.de/courses/ws19/sentiment/material/howToPresent.pdf`
- I especially like from the above presentation
  - Tips on background and related work (page 18/19)
  - Emphasis on examples and diagrams
  - page 27/28 on evaluation and presenting tables
  - Tips on slide design: p. 35-42

- Please **keep us all awake** with
  - change of medium (blackboard, videos)
  - examples and exercises
  - audience interaction . . .
- **Enrich** the paper not just reiterate it:
  - try and present the github demo/code
  - make up your own examples
  - connect papers to each other
  - develop new visualizations of method or results
  - say what you think about the paper . . .

How to pass the module: Term Paper

- Presentation extension
- Puts paper in the research context and evaluates it
- Assessment criteria
  - Technical understanding
  - Extension with and comparison to other papers
  - Structure and clarity
  - Presentation: referencing, English/German, Systematicity . . .
  - Own Opinion and criticism
- MA: Max 14 pages, at least 11pt Font. Only electronic and pdf.
- BA: Max 10 pages, at least 11 pt font. Only electronic and pdf.
- Deadline: 15 April 2020, 17:00.

How to pass the module: Project

- Possibilities: Evaluation, Reimplementation, Further development of a paper . . .
- Can be foundation for master/BA thesis
- Focussed: only 6 weeks planned
- Project description: Plan, including data usage, statistics, description of implementation, evaluation...
- Assessment criteria
    - Research plan
    - Implementation
    - Demo and/or Documentation
    - Evaluation
- Max 8 pages, at least 11pt Font. Only as pdf.
- Additionally: Code with README
- Deadline: 15 April 2020, 17:00.

Delayed term papers/projects

- Extensions without mark change only with doctor's note.
- Otherwise marking with reduced marks:
  - 1 Week: one step ($1.0 \longrightarrow 1.3$, $3.0 \longrightarrow 3.3$)
  - 2 Weeks: two step ($1.0 \longrightarrow 1.7$)
  - 3 Weeks: three steps ($1.0 \longrightarrow 2.0$)
  - 4 Weeks: four steps ($1.0 \longrightarrow 2.3$)
  - even longer: not accepted anymore, 5.0

What's next?

- Questions?
- Literature and topics list
- 5.11: Introduction by me
- First student presentation: 12.11 (assigned today)
- **Registration and topic choice:** Email by Sunday, 3rd of November, 11.59 pm with three ordered topic preferences (Header: PS/HS Bias: Presentation)
- In email: Preference project/term paper or second talk
- In email: Name, Bachelor (%)/Master, percentages

- **For next time (5.11) Exercise I:** Read Obermeyer et al. 2019 plus the news presentation and forum discussion on the paper in `https://www.spiegel.de/netzwelt/apps/usa-algorithmus-benachteiligt-afroamerikanische-patien` `html`
- **Write a 1-2 page answer** to the questions on the next slide and send to me by Monday 4rth of November 12 am. Pairwise submission allowed.

1. What kind of bias does the paper describe (intentional/unintentional; to which protected or minority group; stereotyping or algorithmic decisions)?
2. What is the cause of the bias exhibited?
3. Research the terms *disparate treatment* vs *disparate impact*: are these two terms relevant to the bias exhibited?
4. Has the *Spiegel* article presented the research appropriately?
5. What additional points of the forum participants under the *Spiegel* article are (ir)relevant and which exhibit what kind of misunderstandings? (See for example posts 12,15,28 but please also look at others).