ROYAL INSTITUTE
OF TECHNOLOGY

II2202: Second group meet

Magnus Boman (mab@kth.se)

September 28, 2021

Agenda

- Course deliverables
- Research Plan feedback
- Ethics and Sustainability feedback
- Subgroup formation
- Time plan
- Q&A

Course deliverables (from Canvas)

- Project plan
- Presentation of your proposed research: Ethics & Sustainability
- First draft: Research plan
- Presentation and peer review draft project plans
- Qualitative exercise
- Quantitative exercise
- First draft of report
- Presentation and peer review of draft report
- Opposition before final seminar
- Final Seminar
- Final written report

Research plan feedback

- Time plan mystery
- Criteria for choice (important!)
- Grammar check
- Let both members be visible in all key steps
- Check your status (complete/incomplete)

Method

- Hypothesis-driven vs. Data-driven
- Structure it like a scientific paper?
- Multi-phase project, e.g. Waterfall model?
- Exploratory?
- Qualitative data and/or quantitative data?
- Machine learning:
  Data already available and enough?
  Why deep?
  Bias, overfitting, black boxes (e.g. LSTMs),...
- Empirical work:
  Interviews take a LOT of time, to do and to analyse (and possibly document)

Hypothesis
Bad (but not worst) example

6 Hypothesis

Deep learning algorithms are capable of outperforming typical machine learning algorithms like logistic regression and random forest in certain cases. This project aims to use deep learning algorithms to the dataset to see if they perform better than models mentioned in [2] and [3].

Division of labour
Two earlier examples

Allocation of responsibilities

As we see this as a collective effort and want to maximize both our learning outcomes from this project, many of the responsibilities will overlap. We think we can learn from each other and are prepared to meet regularly to allocate responsibilities during the project.

Alejandro González is responsible for developing and testing the viability of the system, including the evaluation of different approaches for guaranteeing the achievement of the goals described in the project plan. Also he will write certain sections in the report regarding the implementation and the solution adopted.

Daan Knoors is responsible for conducting a literary research on chatbots, voting systems and blockchain solutions. Moreover, he will develop certain parts of the system, write specific sections in the report and presents intermediate and final results together.

Niklas Fürderer <furderer@kth.se> is a master student for data science with working experience in linguistic machine learning applications with both, large multinational technology and startup companies.

Tobias Lindener <lindener@kth.se> is a master student for data science, previously he worked as an IT-Consultant for Internet of Things at IBM.

Allocation of responsibilities

Niklas Fürderer

- Project management
- Risk analysis
- Experimentation & Testing
- Research
- Development
- Evaluation

Tobias Lindener

- System architecture
- Care of industry expert contacts
- Experimentation & Testing
- Research
- Development
- Publication

References
Earlier example

References

[1] J. Jacob, B. Bartholmai, S. Rajagopalan, C. V. van Moorsel, H. V. van Es, F. V. van Beek, M. H. L. Struik, M. Kokosi, R. Egashira, A. L. Brun, A. Nair, S. Walsh, G. Cross, J. Barnett, A. de Lauretis, E. Judge, S. Desai, R. Karwoski, S. Ourselin, E. Renzoni, T. Maher, A. Altmann, and A. Wells, "Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis," *American Journal of Respiratory and Critical Care Medicine*, vol. 198, p. 767–776, 2018.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] T. S. Lee and D. Mumford, "Hierarchical bayesian inference in the visual cortex," *J. Opt. Soc. Am. A*, vol. 20, no. 7, pp. 1434–1448, Jul 2003. doi: 10.1364/JOSAA.20.001434. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-20-7-1434

[4] F. V. Farahani, A. Ahmadi, and M. H. Fazel Zarandi, "Lung nodule diagnosis from ct images based on ensemble learning," in *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2015, pp. 1–7.

Ethics and Sustainability feedback

- Bring in the United Nations' Sustainable Development Goals (UN SDGs)
- Use the SDG subheadings under each goal for inspiration and detail
- Think about energy consumption
- How do you document, store, share your information?
- Do not consider this module as the last word on ethics and sustainability, keep developing your thoughts and include in your final report

Subgroup formation

1
Analysis of Remote Working Productivity in professional workers after COVID-19
Sebastiano Castellan, Paloma Domínguez Sánchez

2
TBC
Håkan Samanci

3
Application and comparison of ranking systems used in gaming
DANIEL WORKINN, DIANA CRISTINA CULINCU

4
Predictive Analysis of Length of Hospital Stay for Diabetic Patients: Analysis of Clinical Data
EMIL DANIEL CIOVICA, PIRIYA SURESHKUMAR

5
Application of Optical Character Recognition in ancient pictographs
Wanqi Jin, Zilong LI

6
Emotional impact of emotions on Wikipedia editors
Remi Chierchia, Nicola Toscan

7
3D Reconstruction of Human Body
Jiaqi Geng, Teng Ma

8
Exploring the characteristic of imported overseas videos in Bilibili
Jiangfan Feng, Shuyi Chen

10
Stock Price Forecasting using Machine Learning Methods
Levent Güner, Erdem Demir

11
Detecting depression from voice and facial actions
Disen Ling, Zhenghong Xiao

12
Studying the most relevant risk factors for heart disease
FRANCESCO DI FLUMERI, PABLO LASO

13
The Unreasonable Virality of Misinformation
PRADHAN PARTHA SARATHI, ANUROOP DIVAKARA

Subgroup 1: Data analytics

5
Application of Optical Character Recognition in ancient pictographs
Wanqi Jin, Zilong LI

8
Exploring the characteristic of imported overseas videos in Bilibili
Jiangfan Feng, Shuyi Chen

10
Stock Price Forecasting using Machine Learning Methods
Levent Güner, Erdem Demir

13
The Unreasonable Virality of Misinformation
PRADHAN PARTHA SARATHI, ANUROOP DIVAKARA

Subgroup 2: Data work processes

1
Analysis of Remote Working Productivity in professional workers after COVID-19
Sebastiano Castellan, Paloma Domínguez Sánchez

2
TBC
Håkan Samanci

3
Application and comparison of ranking systems used in gaming
DANIEL WORKINN, DIANA CRISTINA CULINCU

6
Emotional impact of emotions on Wikipedia editors
Remi Chierchia, Nicola Toscan

Subgroup 3: Health data applications

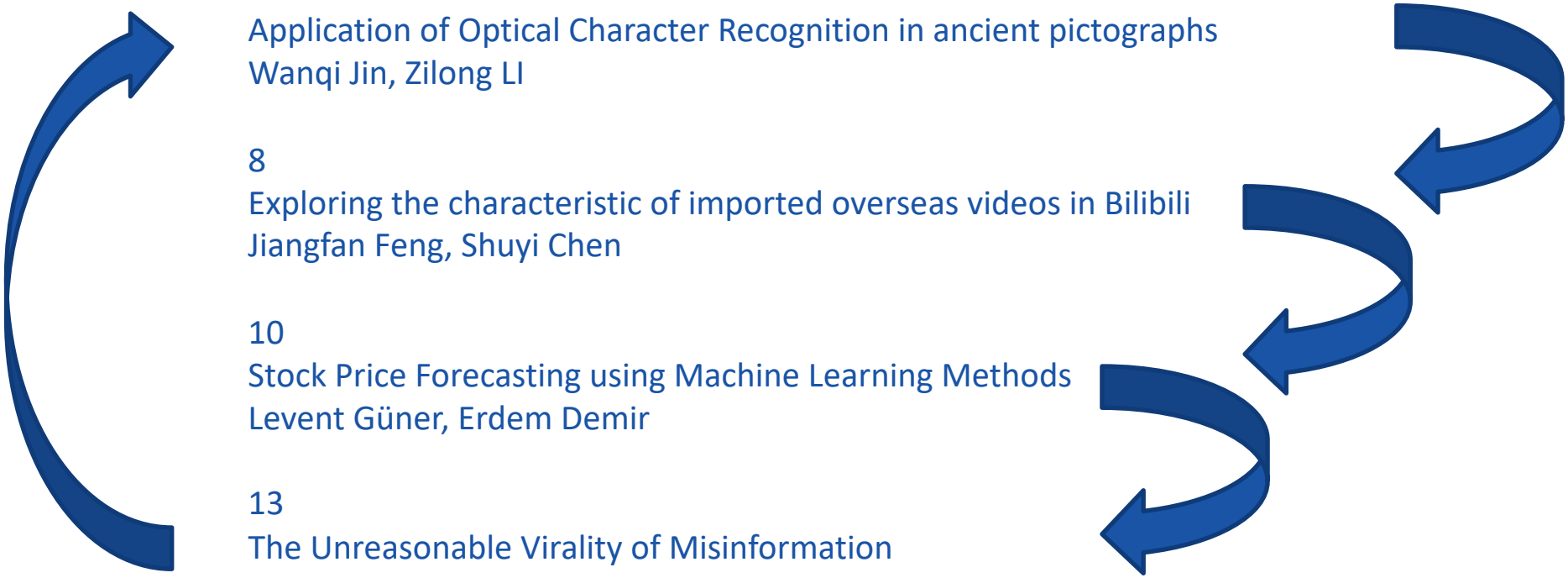
4
Predictive Analysis of Length of Hospital Stay for Diabetic Patients: Analysis of Clinical Data
EMIL DANIEL CIOVICA, PIRIYA SURESHKUMAR

7
3D Reconstruction of Human Body
Jiaqi Geng, Teng Ma

11
Detecting depression from voice and facial actions
Disen Ling, Zhenghong Xiao

12
Studying the most relevant risk factors for heart disease
FRANCESCO DI FLUMERI, PABLO LASO

Subgroup peer review structure



5
Application of Optical Character Recognition in ancient pictographs
Wanqi Jin, Zilong LI

8
Exploring the characteristic of imported overseas videos in Bilibili
Jiangfan Feng, Shuyi Chen

10
Stock Price Forecasting using Machine Learning Methods
Levent Güner, Erdem Demir

13
The Unreasonable Virality of Misinformation
PRADHAN PARTHA SARATHI, ANUROOP DIVAKARA

5 opposes 8 opposes 10 opposes 13 opposes 5

Time plan

- Sep 7 First group meeting (Zoom)
- Sep 28 Second group meeting (Zoom)
- Oct+Nov Individually booked feedback sessions (Zoom, IRL)
- Oct 26 Third group meeting (Zoom)
- Nov 25 Dress rehearsal with 2-min reports (15-17)
- Dec TBC Final seminar and oral examination (3*2h)
- Jan 17 Final report written examination

Questions!?

Magnus Boman
(Professor in Intelligent Software Services)
mab@kth.se
people.kth.se/~mab/
Room: SCS, Electrum, West Wing