Student Research Projects

Study Year 2017/2018

Josif Grabocka, Mohsan Jameel, Nicolas Schilling,
Lydia Voß, Rafael Drumond, Ahmed Rashed, Hadi Jomaa,
Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute for Computer Science
University of Hildesheim, Germany

Outline

1. Aims
2. Research Areas 2017/2018
3. Timeline
4. Proposal

Student Research Projects

Study Year 2017/2018

for whom?

- International Master in Data Analytics (mandatory)
- all IT Master and Bachelor programs (elective)
  - Applied Computer Science
  - Information Management and Information Technology (IMIT)
  - Information Systems

when? — kick-off Thu. 21.12.2017, 12:00 pm

where? — B 037, Spl.

Outline

Aims

1. Students conduct a small well-defined research project
   - in a small group of 4-5 students
   - under supervision of a PhD student, postdoc or professor

2. Students read the literature and comprehend the state-of-the-art in a specific subject of data analytics.

3. Students conduct a computational experiment on their own.

4. Students have the opportunity to extend the state-of-the-art with an own innovation.

More Aims

5. Students learn and practice how to write a short research proposal.
6. Students learn and practice how to conduct a small research project together with partners.
7. Students work on a real problem with real data.
8. Students have fun.

Project Requirements

1. **Problem Setting**:
   - a crisp, specific problem setting
   - that can be tackled with methods from data analytics.

2. **Data Foundation**:
   - data that allows to evaluate and compare different solutions of the problem.

3. **Tangible Outcome**:
   - a workshop paper, an open source software project etc.

Work Load

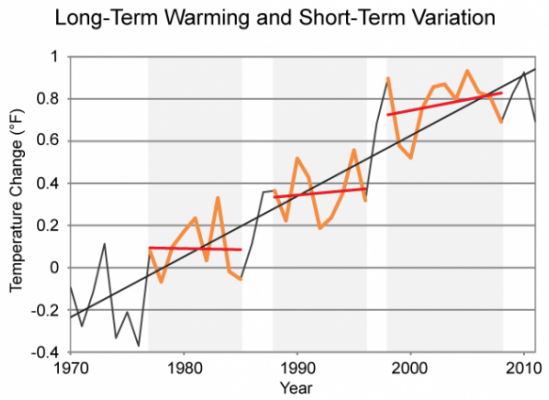
- 15 ECTS, stretched over 2 terms
- 15×30h / student $= 450$h / student
- 1.25 days each week over a year
- for a team of 5 students: 15 person months
- you likely want to organize project work
  - in sprints during term breaks and
  - continuous, but slower progress during terms.

Research Areas and Project Topics

- Every year, we open **research areas**
  - covering interesting actual research topics
  - we know well enough to supervise you
- You can apply for a **topic** within one of these research areas.
  - we do not limit the topics
  - we may point out different example topics within an area, though
  - it is **your job to shape a useful topic** within one of these areas

Outline

Area 1: Multi-Horizon Modeling



Source: `http://www.globalchange.gov/browse/multimedia/long-term-warming-and-short-term-variation`

- Temporal patterns can have several characteristics:
  - Some patterns are short-term related patterns
  - Others a long-term related
  - Some contain both long and short-term pattern

Supervisor: Rafael Rego Drumond

Area 1: Possible Topics & Further Reading

In this project you will explore analyzing time series by looking into short and long term patterns. Possible topics within this research area are:

- Speech Recognition
- Speaker Identification
- Weather Forecasting
- Stock Market Prediction

Points to start investigating this area would be to learn about

- Deep Learning [Goodfellow et al., 2016].
- Time-Series Methods. [Chen et al., 2015, Hamilton, 1994]
- Current Short and Long-Term Pattern Temporal Modeling papers [Goel et al., 2017] and [Lai et al., 2017].

Area 1: Long and Short-Term / References

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G. (2015).
The ucr time series classification archive.
www.cs.ucr.edu/~eamonn/time_series_data/.

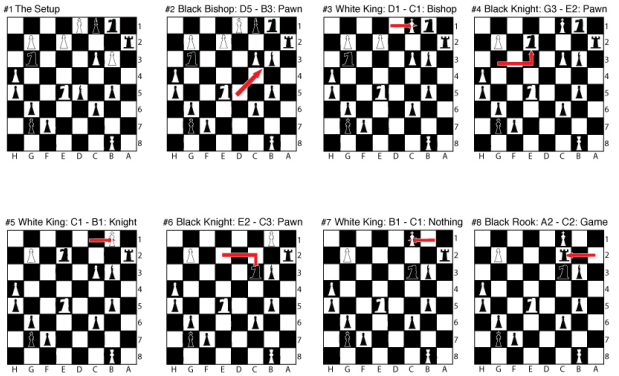
Goel, H., Melnyk, I., and Banerjee, A. (2017).
R2n2: Residual recurrent neural networks for multivariate time series forecasting.
arXiv preprint arXiv:1709.03159.

Goodfellow, I., Bengio, Y., and Courville, A. (2016).
Deep Learning.
MIT Press.
http://www.deeplearningbook.org.

Hamilton, J. D. (1994).
Time series analysis, volume 2.
Princeton university press Princeton.

Lai, G., Chang, W.-C., Yang, Y., and Liu, H. (2017).
Modeling long-and short-term temporal patterns with deep neural networks.
arXiv preprint arXiv:1703.07015.

Area 2: Time-Series Forecasting



- Time-series are omnipresent
  - Digital communication
  - Audio data
  - Sensor data of humans, machines, processes
- They can present themselves as a sequences such as:
  - Word sequence
  - Weather
  - Decisions in a game of chess

Source of the picture: `https://mrciaranlowry.wordpress.com/2013/03/28/44-planning-the-perfect-chess-game/`

Supervisor: Rafael Rego Drumond

Area 2: Possible Topics & Further Reading

In this project you will explore forecasting time-series sequences by looking into previous sequence data. Your job is to model one kind of T.S. with an architecture that is capable of predicting the next instances of the current sequence instances that has not been seen yet. Possible topics within this research area are:

- Word completion
- Weather forecasting
- Human Decision Prediction

Points to start investigating this area would be to learn about

- Deep Learning [Goodfellow et al., 2016].
- Time-Series Methods. [Chen et al., 2015, Hamilton, 1994]
- Current Short and Long-Term Pattern Temporal Modeling papers [Borovykh et al., 2017].

Area 2: Time-Series Forecasting / References

Borovykh, A., Bohte, S., and Oosterlee, C. W. (2017).
Conditional time series forecasting with convolutional neural networks.
arXiv preprint arXiv:1703.04691.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G. (2015).
The ucr time series classification archive.
www.cs.ucr.edu/~eamonn/time_series_data/.

Goodfellow, I., Bengio, Y., and Courville, A. (2016).
Deep Learning.
MIT Press.
http://www.deeplearningbook.org.

Hamilton, J. D. (1994).
Time series analysis, volume 2.
Princeton university press Princeton.

Area 3: Time Aware Recommender Systems

- Recommender systems are used in various domains, for example
  - Online Shopping
  - Social Networks
  - Search Engines
  - Banking
  - Insurance
  - Health

Recommender systems data is **usually** represented as multi relational graph with different entities interacting with each other.

Area 3: Possible Topics & Further Reading

Possible topics within this research area are:

- Recurrent Neural networks for recommender systems
- Deep multi relational classification models
- Personalized Recommendation Using Time Series Analysis
- and many more

Points to start investigating this area would be to learn about

- Time-aware recommender systems: a comprehensive survey [Campos et al., 2014]
- Recurrent Recommender Systems [Wu et al., 2017]
- Personalized Recommendation Using Time Series Analysis [Zhang et al., 2015]

Area 3: Time Aware Recommender Systems / References

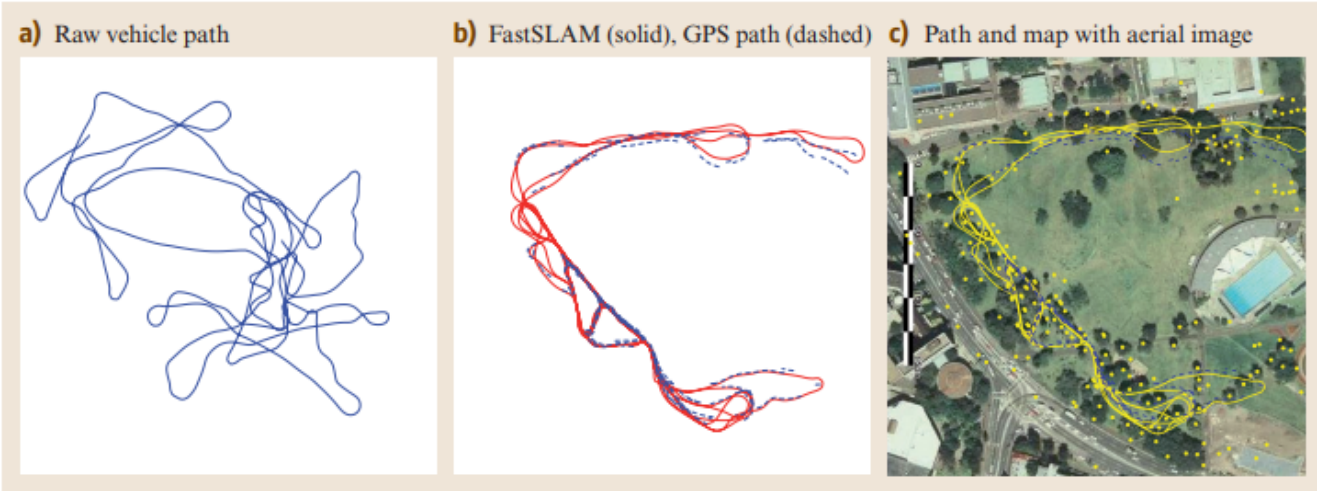
Campos, P. G., Díez, F., and Cantador, I. (2014).
Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols.
User Modeling and User-Adapted Interaction, 24(1-2):67–119.

Wu, C.-Y., Ahmed, A., Beutel, A., Smola, A. J., and Jing, H. (2017).
Recurrent recommender networks.
pages 495–503.

Zhang, Y., Zhang, M., Zhang, Y., Lai, G., Liu, Y., Zhang, H., and Ma, S. (2015).
Daily-aware personalized recommendation based on feature-level time series analysis.
pages 1373–1383.

Area 4: Simultaneous Localization and Mapping

SLAM "is a process by which a mobile robot can build a map of an environment and at the same time use this map to deduce its location" [Durrant-Whyte and Bailey, 2006]



Source:[Thrun and Leonard, 2008]

Supervisor: Hadi Samer Jomaa

Area 4: Possible topics within this research area

With the growing success of Neural Networks in the field of computer vision, and building on existing SLAM paradigms, this grand project has the potential to branch out into several interesting projects, that when merged together result in a robust localization model.

- Object Recognition[Liang and Hu, 2015]
- Object Tracking [Wang et al., 2016]
- Depth Estimation [Zhu et al., 2016]
- Place Recognition [Lowry et al., 2016]

Area 4: References

Durrant-Whyte, H. and Bailey, T. (2006).
Simultaneous localization and mapping: part i.
IEEE robotics & automation magazine, 13(2):99–110.

Liang, M. and Hu, X. (2015).
Recurrent convolutional neural network for object recognition.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3367–3375.

Lowry, S., Sünderhauf, N., Newman, P., Leonard, J. J., Cox, D., Corke, P., and Milford, M. J. (2016).
Visual place recognition: A survey.
IEEE Transactions on Robotics, 32(1):1–19.

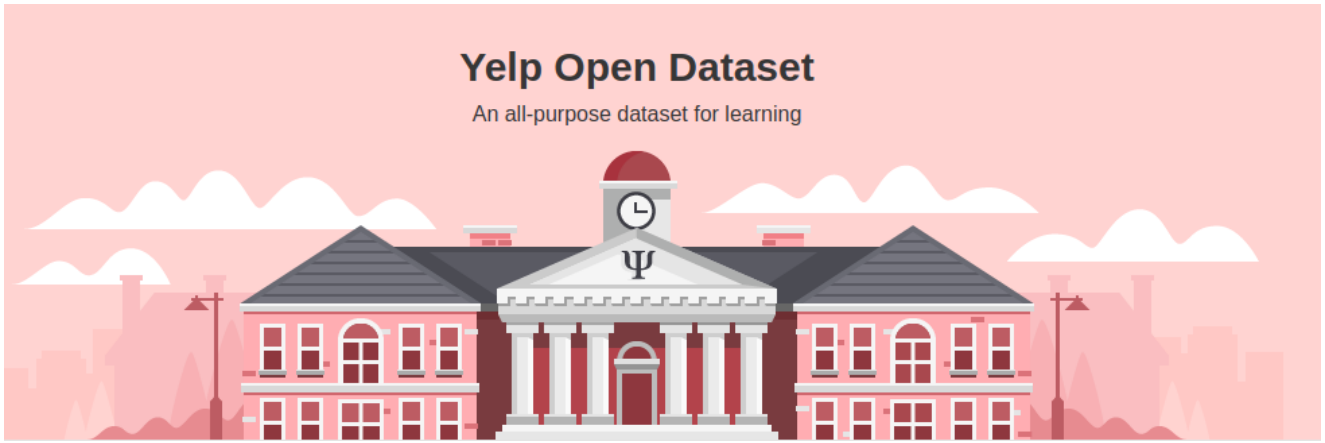
Thrun, S. and Leonard, J. J. (2008).
Simultaneous localization and mapping.
In *Springer handbook of robotics*, pages 871–889. Springer.

Wang, X., Türetken, E., Fleuret, F., and Fua, P. (2016).
Tracking interacting objects using intertwined flows.
IEEE transactions on pattern analysis and machine intelligence, 38(11):2312–2326.

Zhu, L., Wang, X., Wang, D., and Wang, H. (2016).
Single image depth estimation based on convolutional neural network and sparse connected conditional random field.
Optical Engineering, 55(10):103101–103101.

Area 5: Yelp Challenge

"The challenge is a chance for students to conduct research or analysis on our data and share their discoveries with us. Whether you're trying to figure out how food trends start or identify the impact of different connections from the local graph, you'll have a chance to win cash prizes for your work!" `https://www.yelp.com/dataset/challenge`



Source:`https://www.yelp.com/dataset/challenge`

Supervisor: Hadi Samer Jomaa

Area 5: Possible topics within this research area

The Yelp Dataset has been cited in hundreds of research papers due to its rich content. It comprises of thousands of photos, reviews, businesses including millions of features. This leaves room for the following topics to be investigated:

- Recommender Systems [Yu et al., 2013]
- Sentiment Analysis [Zhang, 2015]
- Data Mining [Hu et al., 2014]
- Food Identification
  `https://thenextweb.com/insider/2017/06/30/yelp-created-the-hotdog-app-from-silicon-valley-only-bett`

Area 5: References

Hu, L., Sun, A., and Liu, Y. (2014).
Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction.
In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 345–354. ACM.

Yu, X., Ren, X., Sun, Y., Sturt, B., Khandelwal, U., Gu, Q., Norick, B., and Han, J. (2013).
Recommendation in heterogeneous information networks with implicit user feedback.
In *Proceedings of the 7th ACM conference on Recommender systems*, pages 347–350. ACM.

Zhang, Y. (2015).
Incorporating phrase-level sentiment analysis on textual reviews for personalized recommendation.
In *Proceedings of the eighth ACM international conference on web search and data mining*, pages 435–440. ACM.

Area 6: Video Analysis for Regression

- Extracting key features from video frames
  [Tran et al., 2015, Feichtenhofer et al., 2016, Varol et al., 2016]
- to predict box office returns of movies
- by using **movie trailers** and meta data
  `https://www.opusdata.com/data.php`

The tasks covered within this project can be narrowed down to:

- Video Processing
- Regression

Area 6: References

Feichtenhofer, C., Pinz, A., and Zisserman, A. (2016).
Convolutional two-stream network fusion for video action recognition.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1933–1941.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015).
Learning spatiotemporal features with 3d convolutional networks.
In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497.

Varol, G., Laptev, I., and Schmid, C. (2016).
Long-term temporal convolutions for action recognition.
arXiv preprint arXiv:1604.04494.

Area 7: Distributed Machine Learning

Source: `https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/`

- Data is growing in magnitude as well as complexity.
  - Criteo released a 24 days advertisement dataset (1 TeraByte)
  - Training a deep learning model using 1.3 billion images on a single GPU requires days.
  - Internet of Things (IoT) will produce one-tenth of Earth's data by 2020 (44 ZetaBytes) [Gartner, 2017].
- Main objective of big data analytics is to build models that scales with large, complex and heterogeneous data.

Area 7: Distributed Machine Learning

- **Possible directions:**
  - Scaling a decentralized machine learning algorithm in a cloud environment [Zhang et al., 2016].
  - Scaling Deep learning models to a large number of GPUs [Goyal et al., 2017].
    - Area to look at: mini-batch and learning rate coupling.
    - Adjusting the delayed gradient updates.
  - Large scale recommender system in the cloud [Zhong et al., 2016, Rendle et al., 2016]
    - Implement recommender system models using big data frameworks (Apache Spark, Apache Hadoop, etc)
    - Optimize performance and reduce communication cost.
- **Contact Person:** Mohsan Jameel

Area 7: Distributed Machine Learning

Gartner (2017).
Gartner says 8.4 billion connected "things" will be in use in 2017, up 31 percent from 2016.
In *https://www.gartner.com/newsroom/id/3598917*.

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., and He, K. (2017).
Accurate, large minibatch sgd: Training imagenet in 1 hour.
arXiv preprint arXiv:1706.02677.

Rendle, S., Fetterly, D., Shekita, E. J., and Su, B.-y. (2016).
Robust large-scale machine learning in the cloud.
In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1125–1134. ACM.

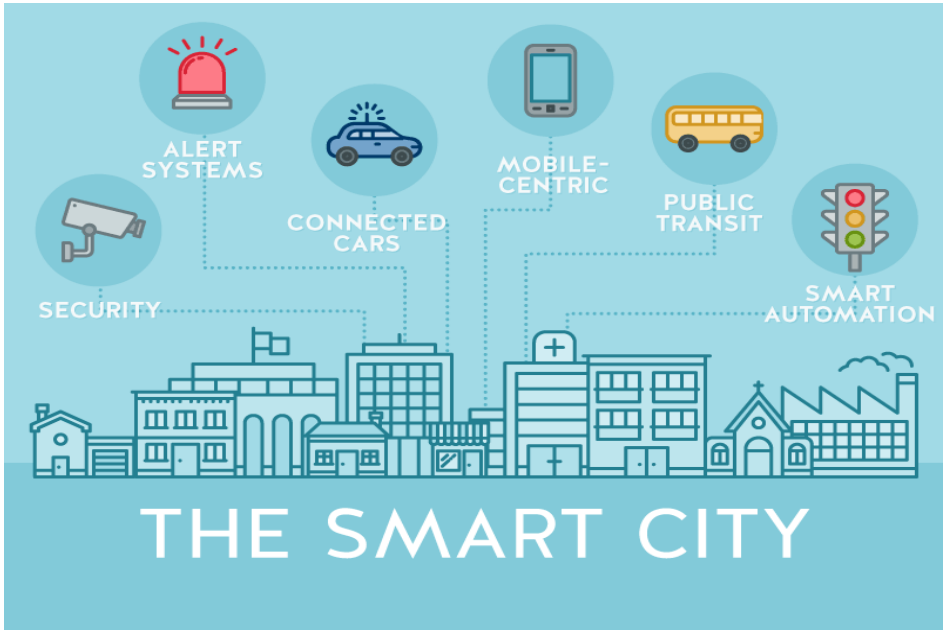
Zhang, H., Hsieh, C.-J., and Akella, V. (2016).
Hogwild++: A new mechanism for decentralized asynchronous stochastic gradient descent.
In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 629–638. IEEE.

Zhong, E., Shi, Y., Liu, N., and Rajan, S. (2016).
Scaling factorization machines with parameter server.
In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1583–1592. ACM.

Area 8: Internet of Things (IoT)

- the internet of things is a global infrastructure
- where physical and virtual entities are linked
- so they can interact to support humans in their daily life

Area 8a: IoT - Smart Cities



Source: `http://www.satprnews.com/wp-content/uploads/2017/09/Internet-of-Things-IoT-in-Smart-Cities-768x576.png`

- **Smart City.** A city that uses electronic data to augment the planning and resource management processes.
  - Town, traffic planning.
  - Reduce carbon and energy footprints of a city.
  - Event management through electronic data streams.

Area 8a: IoT - Smart Cities

- **Possible directions:**
  - Event prediction and management i.e. social events, emergencies or anomalies [Panagiotou et al., 2016]
    - handling multi-modal data streams
    - Predictive analytics
  - **Datasets:** Data Analytics for Smart Cities [Macsin, 2016]
  - More projects [Murray, 2017]
- **Contact Person:** Mohsan Jameel

Area 8a: IoT — Smart Cities

Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., and Albrecht, J. (2012).
Smart*: An open data set and tools for enabling research in sustainable homes.
SustKDD, August, 111:112.

Das, B. (2014).
Machine Learning Challenges for Automated Prompting in Smart Homes.
PhD thesis, WASHINGTON STATE UNIVERSITY.

Jakkula, V. and Cook, D. J. (2011).
Detecting anomalous sensor events in smart home data for enhancing the living experience.
In *Proceedings of the 7th AAAI Conference on Artificial Intelligence and Smarter Living: The Conquest of Complexity*, pages 33–37. AAAI Press.

Lachut, D., Banerjee, N., and Rollins, S. (2014).
Predictability of energy use in homes.
In *Green Computing Conference (IGCC), 2014 International*, pages 1–10. IEEE.

Macsin, A. (2016).
Great iot, sensor and other data sets repositories.
In *https://www.iotcentral.io/blog/great-iot-sensor-and-other-data-sets-repositories*.

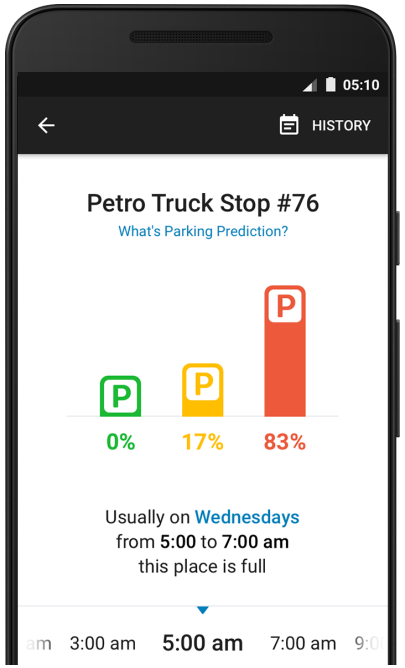
Murray, P. (2017).
40 brilliant open data projects preparing smart cities for 2018.
In *https://carto.com/blog/forty-brilliant-open-data-projects-preparing-smart-cities-2018/*.

Panagiotou, N., Zygouras, N., Katakis, I., Gunopulos, D., Zacheilas, N., Boutsis, I., Kalogeraki, V., Lynch, S., and O'Brien, B. (2016).
Intelligent urban data monitoring for smart cities.
In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 177–192. Springer.

Rasch, K. (2013).

Area 8b: IoT - Parking Spot Prediction

- IoT consits of a variety of devices or objects which have the ability to compute, sense and interact with their surroundings and generate a vast amount of data
  - Sensors
  - Actuator
  - Mobile phones
  - Parking meters
- Parking is limited in almost every major city in the world
  - finding an available parking spot could be troublesome
  - 30% of traffic in urban scenarios is generated by vehicles looking for a parking space [**?**]



Source of the picture: `https://truckerpath.com/uploads/2017/10/7.-parking-prediction-actual-screen-new.png`

Supervisor: Lydia Voß

Area 8b: Possible Topics & Further Reading

Your job is to predict the probability of having a free parking space in the area of interest. Possible topics within this research area are:

- survival analysis: finding an available space with relation to type of day, time-period, etc.
- predict the overall occupancy of parking spaces using past information.

Points to start investigating this area would be to learn about

- Deep Learning [Goodfellow et al., 2016].
- Time-Series Methods[Chen et al., 2015, Hamilton, 1994]
- Current parking prediction papers [**?**, **?**].

Area 8b: Possible datasets

You can start with investigating these datasets:

- http://iot.ee.surrey.ac.uk:8080/datasets.html
- https://truckerpath.carto.com/tables/truck_parking_availability_for_weekdays_july_2016/public
- https://data.gov.uk/dataset/parking-data, https://data.gov.ie/dataset/parking
- https://data.melbourne.vic.gov.au/Transport-Movement/Parking-bay-arrivals-and-departures-2014/mq3i-cbxd

Area 8b: IoT - Parking Spot Prediction / References

Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., and Albrecht, J. (2012).
Smart*: An open data set and tools for enabling research in sustainable homes.
SustKDD, August, 111:112.

Das, B. (2014).
Machine Learning Challenges for Automated Prompting in Smart Homes.
PhD thesis, WASHINGTON STATE UNIVERSITY.

Jakkula, V. and Cook, D. J. (2011).
Detecting anomalous sensor events in smart home data for enhancing the living experience.
In *Proceedings of the 7th AAAI Conference on Artificial Intelligence and Smarter Living: The Conquest of Complexity*, pages 33–37. AAAI Press.

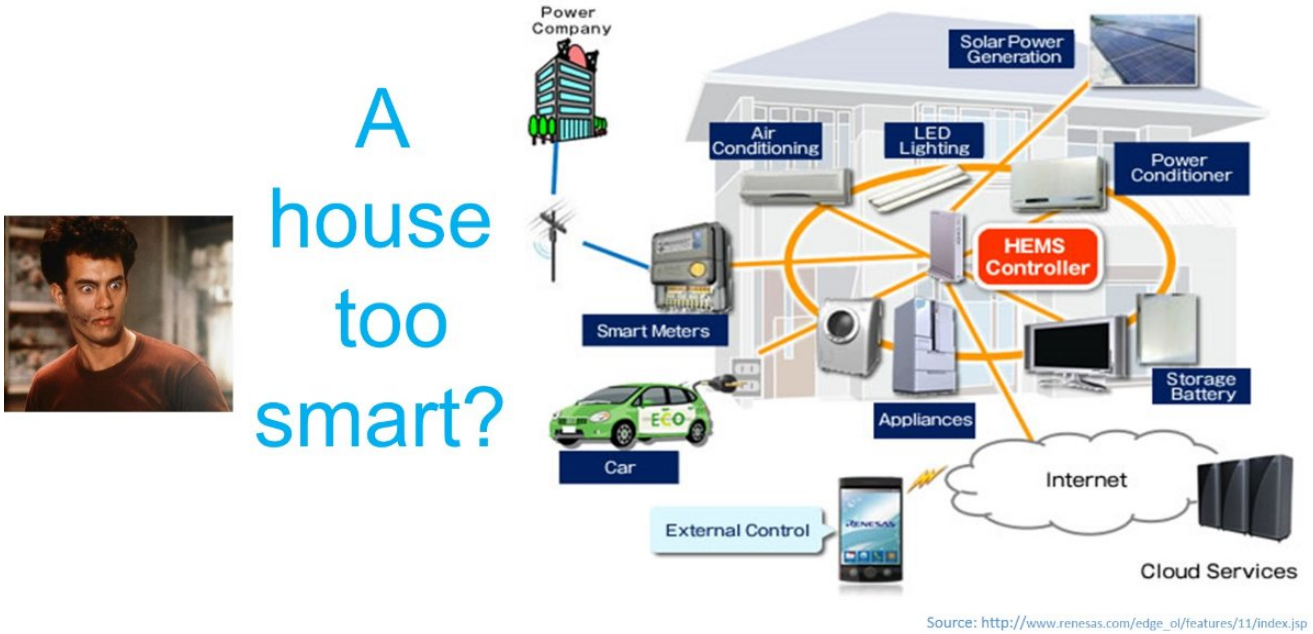
Lachut, D., Banerjee, N., and Rollins, S. (2014).
Predictability of energy use in homes.
In *Green Computing Conference (IGCC), 2014 International*, pages 1–10. IEEE.

Macsin, A. (2016).
Great iot, sensor and other data sets repositories.
In *https://www.iotcentral.io/blog/great-iot-sensor-and-other-data-sets-repositories*.

Murray, P. (2017).
40 brilliant open data projects preparing smart cities for 2018.
In *https://carto.com/blog/forty-brilliant-open-data-projects-preparing-smart-cities-2018/*.

Panagiotou, N., Zygouras, N., Katakis, I., Gunopulos, D., Zacheilas, N., Boutsis, I., Kalogeraki, V., Lynch, S., and O'Brien, B. (2016).
Intelligent urban data monitoring for smart cities.
In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 177–192. Springer.

Area 8c: IoT — Smart Home



Source: `https://media.licdn.com/mpr/mpr/p/2/005/08a/244/20b89a9.jpg`

- **Smart Home.** A sensor enrich entity with capacity to log and control household activities:
  - Optimize and Forecast energy consumption of a household.
  - Adapt temperature and lighting to the comfort level of inhabitants.
  - Reduce carbon and energy footprint of a household.

Area 8c: IoT — Smart Homes

- **Possible directions:**
  - Forecasting [Truong et al., 2013, Lachut et al., 2014]
  - Activity recognizing and anomaly detection [Das, 2014, Jakkula and Cook, 2011]
  - Recommender System [Rasch, 2013]
  - **Datasets:** Smart* Data Set for Sustainability [Barker et al., 2012]
- **Contact Person:** Mohsan Jameel

Area 8c: IoT — Smart Home

Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., and Albrecht, J. (2012).
Smart*: An open data set and tools for enabling research in sustainable homes.
SustKDD, August, 111:112.

Das, B. (2014).
Machine Learning Challenges for Automated Prompting in Smart Homes.
PhD thesis, WASHINGTON STATE UNIVERSITY.

Jakkula, V. and Cook, D. J. (2011).
Detecting anomalous sensor events in smart home data for enhancing the living experience.
In *Proceedings of the 7th AAAI Conference on Artificial Intelligence and Smarter Living: The Conquest of Complexity*, pages 33–37. AAAI Press.

Lachut, D., Banerjee, N., and Rollins, S. (2014).
Predictability of energy use in homes.
In *Green Computing Conference (IGCC), 2014 International*, pages 1–10. IEEE.

Macsin, A. (2016).
Great iot, sensor and other data sets repositories.
In *https://www.iotcentral.io/blog/great-iot-sensor-and-other-data-sets-repositories*.

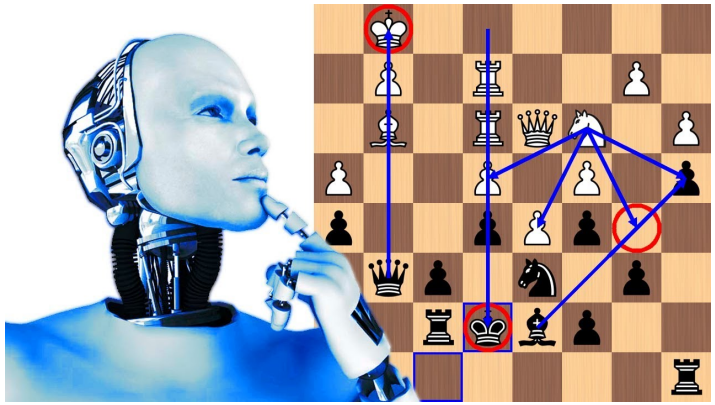
Murray, P. (2017).
40 brilliant open data projects preparing smart cities for 2018.
In *https://carto.com/blog/forty-brilliant-open-data-projects-preparing-smart-cities-2018/*.

Panagiotou, N., Zygouras, N., Katakis, I., Gunopulos, D., Zacheilas, N., Boutsis, I., Kalogeraki, V., Lynch, S., and O'Brien, B. (2016).
Intelligent urban data monitoring for smart cities.
In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 177–192. Springer.

Rasch, K. (2013)

Area 9: Reinforcement Learning in classics games

- The goal of reinforcement learning is to learn good policies for sequential decision problems, e. g.
  - Chess
  - Checkers
  - Bridge
  - Go



- In games as varied as Chess, Checkers, Backgammon or scrabble, computers have exceed human levels of performance. Go has long been viewed as the most challenging game for artificial intelligence. Recently Google developed a program (AlphaGo) defeating the human European Go champion [Silver et al., 2016].
  - Novel combination of supervised learning and reinforcemnet learning
  - Using deep neural networks

Source of the picture: `https://www.youtube.com/watch?v=pcdpgn9OINs`

Supervisor: Lydia Voß

Area 9: Possible Topics & Further Reading

The idea of this project is to learn a model by reinforcement learning from self-play; as in the Google AlphaGo example; for a different game. Possible topics within this research area are:

- Self-play learning
- Move prediction
- Policy improvement

Points to start investigating this area would be to learn about

- Reinforcement Learning Methods[Sutton and Barto, 1998]
- Deep Learning [Goodfellow et al., 2016].
- Current papers about methods beating human world champion in the game Go [Silver et al., 2017, Silver et al., 2016].

Area 9: Reinforcement Learning in games / References

Goodfellow, I., Bengio, Y., and Courville, A. (2016).
Deep Learning.
MIT Press.
http://www.deeplearningbook.org.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016).
Mastering the game of go with deep neural networks and tree search.
Nature, 529(7587):484–489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017).
Mastering the game of go without human knowledge.
Nature, 550(7676):354.

Sutton, R. S. and Barto, A. G. (1998).
Reinforcement learning: An introduction, volume 1.
MIT press Cambridge.

Area O: Open Innovation

- Any topic that involves data analytics & machine learning
  - autonomous driving with a fleet of small robots
  - speech interfaces
    - like Apple's Siri or Amazon's Echo
  - extreme classification (with 100.000 of classes)
  - tag recommender systems
  - joint image segmentation and labeling
  - opinion mining
  - ...
- invest more time into related work at proposal stage
  - open innovation proposals should have between 5–10 pages.
- discuss your idea with one of us early

Outline

Timeline

21.12.2017 — Introduction to student research projects

- chose your area
- build your team and
- write your research proposal

13.03.2018 — Deadline for proposals

20.03.2018 — Notification & start of projects

- work on your project

16.06.2018 — 1st interim presentation

18.10.2018 — 2nd interim presentation

- work on your project
- prepare a final presentation

13.12.2018 — Closing conference

(& Introduction to student research projects 2018/2019)

Outline

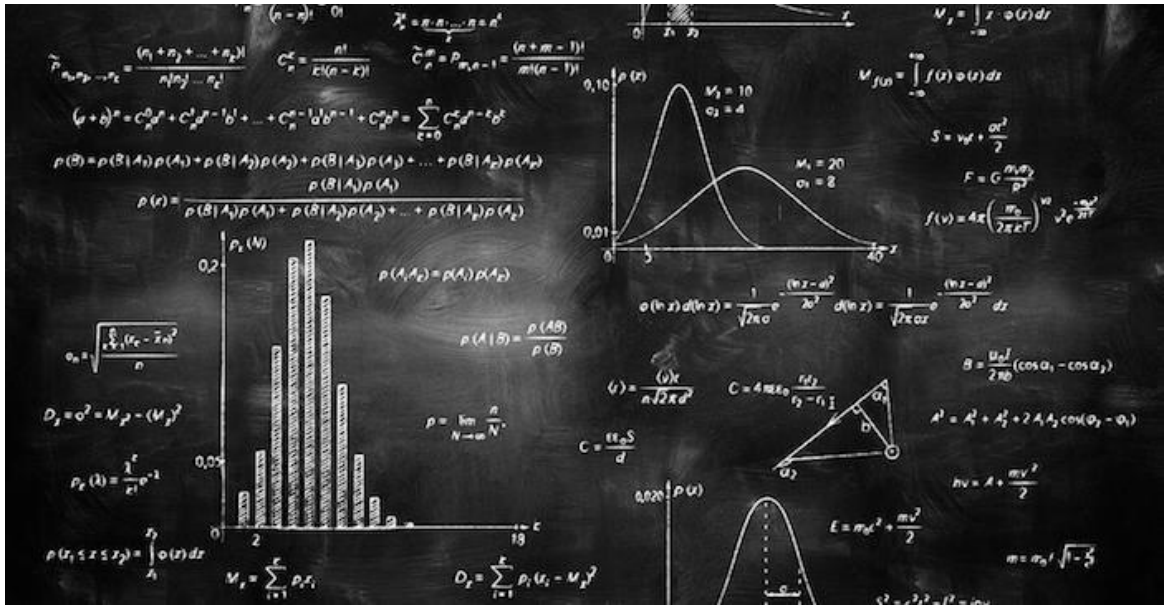
Proposal

section	length
1. Problem Setting	0.5 – 1 page
2. State-of-the-Art	0.5 – 1 page
3. Data Foundation	0.25 – 0.5 page
4. Research Idea	0.5 – 1 page
5. Tangible Outcomes	1 sentence – 0.5 page
6. Work Plan	0.25 – 0.5 page
7. Resources	1 sentence – 0.25 page
8. Team	0.25 – 0.5 page
A. References	no limit
	3 – 5 pages

- Sections are recommendations, you can section in a different way.
  - but make sure you provide clear answers to the questions w.r.t. these 8 aspects
- Page limits are indicative, you can write more or less.

Proposal / 1. Problem Setting

- What is the problem you want to solve?
- Describe the problem in words and
- formally
  - given x, find an instance of type y with properties z



Proposal / 2. State-of-the-Art

- If others have tackled the problem already:
  - Which solutions exist?
  - What are their properties? What their limitations?
- If the problem is completely novel:
  - What are simple/straight-forward solutions and what are their limitations?
  - What are the most closely related problems and how are they different?
- Provide complete references.

Proposal / 3. Data Foundation

- What data is (publicly) available for your problem?
  - provide references
  - provide brief summary statistics
- Do you plan to collect data as part of your project?

Proposal / 4. Research Idea

- What do you plan to do? e.g.,
  - reproduce an experiment from the literature
  - combine two methods from the literature
  - research a new idea / method
- Which experiments do you plan to run?

Proposal / 5. Tangible Outcomes (1/2)

- What tangible results will your project have?
- All projects should result in some written **documentation** (pick one)
  - a workshop paper submission
    - usually 8-16 very compact pages
    - identify a workshop or conference already
  - software documentation
    - not just API documentation, but a story about requirements, design, implementation etc.
    - approx. 30 pages
  - a business plan
    - for a start-up company
  - a project report
    - describe what you did, argue your choices etc.
    - approx. 40 pages

Proposal / 5. Tangible Outcomes (2/2)

- Most projects also should result in some **software prototype**
  - open source software project
  - an internal prototype just for you and us
- but your project could have other types of tangible outcomes, too:
  - a demo
  - a tutorial
    - as webpage or as video
  - a website or a webservice
  - a MOOC

Proposal / 6. Work Plan

- Structure work in tasks or work packages.
- Provide a time-wise planning.
- Describe task dependencies.
- A rough planning should be fine
  - maybe 4-5 tasks
- if you plan to write some software:
  - will you build on top of an existing software?
    - identify what is still missing
  - which libraries are you using?
  - have you decided about the programming language already?

Proposal / 7. Resources

- Which resources do you need?
  - computing time
  - hardware, conference fees
  - conference fees
- Estimate total costs in euros.
- We likely cannot provide very large sums.

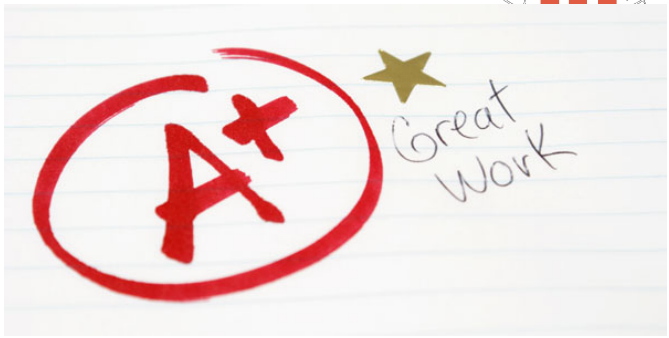
Proposal / 8. Team

- Who is in the team with which role?
- What are your prior expertises?
  - Machine Learning 1 is a formal requirement for **all** team members.
- We expect each team to bring members from 3 different countries.
- Why are you a good team to conduct the project?
- Provide a contact email.

Submitting Your Proposal

- you can discuss an idea and a draft of your proposal with potential supervisors up front
- the submission deadline is strict.
- we will assess your proposal and either
  - accept it as it is,
  - propose some modifications that should help you to stay on track or
  - reject it, esp. proposals
    - that make absolutely no sense,
    - are very vague,
    - are written in a careless way and
    - without any prior consultation
    - we may offer specific replacement topics on a take-or-leave-it basis

A Word About Grading



- final grading will depend on
  - did you address a challenging problem or a more down-to-earth one?
  - how clever the solution is you finally found
  - the quality of your proposal
  - the quality of your tangible results
    - how well is a workshop paper written?
    - is an open source software used by others?
    - does a software prototype work well or segfault?
  - how well you worked
    - did you flexibly deal with issues on the way?
    - a project is not about sticking to the initial plan.


ANY
QUESTIONS
?