Summary of discussion session at AI / Big Data seminar 190917

Some 70 people attended the half-day seminar held at four locations. The seminar concluded with discussion groups that came up with suggestions for how SLU could better support this promising and rapidly developing area of research. The questions and the notes from the different discussion groups are attached. Below is an extract of thoughts and suggestions, most of which came up in more than one group.

- SLU needs **more computer scientists,** working alongside other scientists to increase the quality of the AI work being done. Interdisciplinary collaborations is key for future activities.

- **More courses** in data collection, data management, artificial intelligence, bioinformatics, programming etc should be included already at undergraduate level. We need to collaborate better between departments and faculties on this. Courses could be open to different levels, both undergraduate, PhD and even for PIs. The courses at SLU should be "bridging courses" to create an understanding of the area and learn how to communicate with the computer specialists rather than becoming one yourself. SLUs "Forskarskolor" could take responsibility for this at the graduate level. The centre of statistics (Statistics@SLU) can be a key actor at all educational levels.

- SLU should **collaborate with other universities** where they have deeper knowledge in mathematics, computing etc – for example LU, UU, HiS. Both for research and education.

- A **closer connection between researchers – IT division – the library's Data Curation Unit (DCU)** and **Statistics@SLU** should be established. Early contact, even in the grant application phase in order to structure data collection, processing and storage. More information about existing resources. Perhaps establish a combined, centralized support function?

- Create a **meeting ground** for SLU staff interested in AI/Big Data/Machine Learning/Statistical learning. Open seminars and discussions, to learn more of what is going on at SLU – for coordination and inspiration.

- **Support/collaboration with the IT division:** Help with short and long term data storage (present resources are not sufficient for future needs), data processing, security issues, designing and building workstations for AI, a better, higher capacity network, and entry points to and information about national services like SNIC

- **Support/collaboration with the DCU:** Data management plans (preferably different models depending on funding body), archiving solutions in SLU databases, help with metadata.

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Big Data and Artificial Intelligence: harnessing the power for SLU

September 17th, 9.00 – 12.00
video-connected meetings in Ultuna (Sal O1), Umeå (Holly Wood), Alnarp (Articum Spiltan) and Skara (Nonnenrummet)

The opportunities and challenges of the rapidly increasing availability of data is one of the few topics on which almost everyone agrees on in research and education. But how should SLU address this revolution-in-progress? Welcome to a half-day to chart a course together, with inspiration provided from both within and outside our university.

Pro vice-chancellors Kevin Bishop and Ylva Hillbur, organizers

---

Program

9.00 – 9.10: Introduction and goals for the day (Kevin Bishop)

Session1: External inspiration
9.10 – 9.30 Karl Åström, Matematikcentrum, Lunds universitet:
"Artificial Intelligence, Machine Learning and Computer Vision"
9.30 – 9.50 Erik Billing, Högskolan i Skövde:
"*From Big Data to Small Robots - Current Trends of AI and our Place as Human Users"*

Session 2: Research examples from SLU
9.50 – 10.00 Kristin Piiki (NJ): *"Big data and small data in decision support for crop production"*
10.00 – 10.10 William Lidberg (S): *"Digital soil mapping with machine learning"*
10.10 – 10.20 Harald Klein (LTV): *"Mapping of new methods for inventory of Urban Green Spaces"*
10.20 – 10.35 Erik Bongcam Rudloff (VH): "A*rtificial Intelligence: Benefits and Challenges in the field of Bioinformatics"*

10 min break

Session 3: How can SLU support use of AI and Big Data in research and FOMA?
10.45 – 10.55 Petra Lagerkvist, IT Director: *"What can the IT Division do, now and in the future"*
11.55 – 11.05 Hanna Lindroos, DCU: *"What can the Data Curation Unit do, now and in the future*"

Session 4: What strategies can we pursue and how do we organize ourselves?
11.10 – 12.00 Group discussions at the different campuses
A list of key questions to discuss will be provided. A rapporteur for each group/campus will be asked to submit written summary of the discussion to the organizers (no common oral reporting).

Group discussions at AI/Big Data seminar 190917

Suggested issues for discussion. You may chose only to discuss some of them, and other issues that come up in the group can be added.

A rapporteur for each group/campus is asked to submit a written summary of the discussion to the organizers.

- What is your vision for the use of Artificial Intelligence, Big Data and Machine Learning at SLU in five years?

- What can SLU do to make the most of the possibilities created by Artificial Intelligence, Big Data and Machine Learning?

- How can we prepare SLU students for using Artificial Intelligence, Big Data and Machine Learning in the future?

- As a researcher, what kind of support or collaboration would you like to see with the IT Division?

- As a researcher, what kind of support or collaboration would you like to see with the Data Curation Unit?

Summary of discussions in the "statistics and data analysis" discussion group

Participants:

Juliette Hayer, researcher in bioinformatics, SLU Bioinformatics infrastructure (SLUBI)

Lars Rönnegård, professor in quantitative genetics, Beijerlab

Niclas Högberg, Ph.D. student working with sensor data on cattle

Guillaume Chapron, Wildlife Ecology Unit

Ali Moazzami, researcher in organic chemistry, SLU metabolomics platform

Staffan Betnér, research assistant in statistics, Statistics@SLU

Jesper Rydén, senior lecturer in statistics, Statistics@SLU

Claudia von Brömssen, senior lecturer in statistics, Statistics@SLU

Summary:

We agreed that, in the long run, only providing support within the area (Big data, AI, statistical/machine learning) is definitely not enough. It is necessary to identify how other universities in life science have addressed the issue. For example, the Norwegian University of Life Sciences has a Masterprogramme in data science.

In the short run, there is a need to provide relevant courses for SLU Ph.D. students. Niclas mentioned his problem to find courses within the machine learning and big data area, which did not require several mathematics, statistics and data science courses as prerequisites. Most of the current Ph.D. students at SLU will not become professionals within the area, but many need to learn to communicate with people who are, i.e. programmers, system designers, statisticians, mathematicians,… SLU can provide such bridging courses.

Guillaume also pointed out that many Ph.D. student lack basic understanding of statistics and programming. Maybe elementary courses should be made mandatory.

We also discussed the different approaches to data analysis, i.e. the traditional hypothesis approach versus pattern recognition and agreed on that hypotheses, sound research questions and biological knowledge is still essential for research in the big data /machine learning area. We should not fall for buzz words, but identify which will be the tools for tomorrow. Then we can identify the expertise and organisation needed to move forward.

Additionally we discussed that expertise in, e.g. data science and programming might be needed at various department to obtain the right level of involvement, to learn a common language and to provide consistency. A risk with placement at various department is, clearly, that it is difficult to establish a 'critical mass' and the career paths might not be obvious..

Several participants of the groups are involved in platforms/organisations/centres and the collaboration within fields works well over faculties and sites. However, we raised the concern that also these platforms stand as islands. More collaboration is needed, but is difficult as long as much of the work is focused on support. A bigger umbrella organisation could be considered.

We concluded that we strongly hope that this is not the only initiative from the SLU management to discuss and proceed with this topic. We are looking forward to a follow-up or feedback from SLU and hope that we will meet again soon. As initiator and leader of the session, Claudia also hopes that next time the scheduled program will not end with producing and storing data, but that the handling and analysis of data will be a significant part.

What is your vision for the use of Artificial Intelligence, Big Data and Machine Learning at SLU in five years?

- Att alla känner till konceptet - general awareness – way of working.
- Teaching basic statistics.
- Skaffa en math department.
- Grow SLU statistics
- Inkludera redan på grundnivå för studenterna.

- What can SLU do to make the most of the possibilities created by Artificial Intelligence, Big Data and Machine Learning?

- Integrera i utbildningen för studenter
- Se till att forskare lär sig om konceptet
- Se till att data tas om hand (DCU-perspektiv) och hanteras rätt under hela processen– inte bara data utan även kod.
- Technical infrastructure for machine learning in research.
- Tillsätt en koordinator för att hålla koll på AI och omvärldsbevakning
- Hands-on workshop för att lära forskare om hur de kan inkorporera AI i sina ansökningar ex.
- webinarier

- How can we prepare SLU students for using Artificial Intelligence, Big Data and Machine Learning in the future?

- incorporate in the education

- As a researcher, what kind of support or collaboration would you like to see with the IT Division?

- Better functioning Network
- Storage
- Computing areas – like ex. SNIC – somewhere to analyse/work with data
- Using virtual environments for computing and courses.

- As a researcher, what kind of support or collaboration would you like to see with the Data Curation Unit?

- Help with metadata for DNAsequences before uploading to repository

What is your vision for the use of Artificial intelligence, Big Data and Machine learning at SLU in five years?

The number of people starting to use AI and Big Data is rapidly increasing and SLU are working with important natural resources and have collected large amounts of data already. National inventory's will be more important and we need to make sure that the quality of those inventories keep a high standard.

What can SLU do to make the most of the possibilities created by Artificial Intelligence, Big Data and Machine Learning?

Perhaps create a network SLU for people who work with AI and Big Data where they can share experience and solve problems. There is a small group (AI coffe) at the forest faculty that meet every thusday to discuss papers and methods. A lot of researchers working with AI at SLU are self-learned or happy amateurs. A network/group could help create more confidence.

SLU will also need to make an effort to hire real data scientists to increase quality and reproducibility of the AI work being done.

How can we prepare SLU students for using Artificial Intelligence, Big Data and Machine Learning in the future?

Everyone don't need to know how to use machine learning but we suggest that SLU introduces a Phd course in statistical learning.

As a researcher, what kind of support or collaboration would you like to see with the IT Division?

The IT department need to have some time set aside to be involved in research projects at an early stage. They can give important suggestions and feedback during the planning phase. There is a problem with EU projects where EU don't allow for internal billing that we need to solve.

Processing power and storage are going to be big problems moving forward. While SLU have a collaboration with the super computer HPCN2 we also need to have access to powerful workstations for everyday use. Most work with big data and machine learning is trial and error and preprocessing data. In addition to the standard SLU computers the IT department could help with designing and building workstations for AI researchers.

In terms of storage there are two types for storage. Short term and long term. The short term storage is needed as a workspace where researchers can work with large datasets on NASes or internal hard drives. This storage needs to be fast. This storage would not have the same requirements for backups and safety. The long term storage is for achieving the final dataset and need backup and safety.

Sharing Big Data is also difficult right now. There is no way to share large datasets (several TBs) within SLU or between SLU and other parts of the world. An internal / external FTP could be created

for this purpose. These FTPs would be temporary storage that is automatically cleaned once every month or so.

As a researcher, what kind of support or collaboration would you like to see with the Data Curation Unit?

Reproducibility is important and sometimes it’s more important to share the original data and well documented methods instead of all created by a project.

Notes from discussion AI-seminar 190917 (Ultuna)

What is your vision for the use of Artificial intelligence, Big Data and Machine learning at SLU in five years?

In five years, we need a connection between computer scientists and other scientists. You can teach biologists to use the tools, but we must also master development together with computer specialists.

KoN evaluation mentioned need of center for Big Data (animal breeding and genetics).

What can SLU do to make the most of the possibilities created by Artificial Intelligence, Big Data and Machine Learning?

We need a cross-organisation solution to coordinate the different roles, for example the library and the IT division needs more collaboration.

A non-formalized meeting ground to keep awareness of what is going on, discuss common interests and possibilities to collaborate, share systems etc. Perhaps in the form of “drop-in sessions” in the interface between “BIO” and “TECHNOLOGY”.

How can we prepare SLU students for using Artificial Intelligence, Big Data and Machine Learning in the future?

Start at undergraduate level when it comes to data management, bioinformatics, artificial intelligence etc. Here we must collaborate between different departments and faculties. Right now there are forces to conserve program curricula instead of opening up to new (perhaps compulsory) courses.

Share master and PhD courses between the levels – (could even be for PIs) and present an open list of courses, collaborate.

Camilla.soderquist@slu.se can be contacted regarding the courses at the library.

As a researcher, what kind of support or collaboration would you like to see with the IT Division?

Need for more generic data skills at the departments – researchers can not and should not do everything themselves (even though they are not always good at asking for help)

More formal collaboration between departments and IT-division – for example in security issues

As a researcher, what kind of support or collaboration would you like to see with the Data Curation Unit?

Archiving solutions, data management plans.

Discussion Skara AI seminar September 17, 2019

- What is your vision for the use of Artificial Intelligence, Big Data and Machine Learning at SLU in five years?

- Increased use of databases in the departments. Showing good examples of how to use databases can help us get there. Perhaps the DCU can help.

- Better communication between subject experts and data experts. Perhaps courses/workshops can help.

- What can SLU do to make the most of the possibilities created by Artificial Intelligence, Big Data and Machine Learning?

- How can we prepare SLU students for using Artificial Intelligence, Big Data and Machine Learning in the future?

- There is a possibility and a need to better integrate data collection/management/modelling /evaluation, into many of SLUs undergraduate courses. It would be very useful with computer exercises in this context. Dept. of Soil and Environment, Division of Precision Agriculture would be happy to contribute.

- Perhaps it would be possible to start an undergraduate and/or a graduate course in programming.

- University of Skövde (HiS) has courses in data mining and human computer interaction. Master students at HiS may be interested in master projects at SLU.

- A workshops like the one today (or even a workshop series), but perhaps more practically oriented, would be good way for subject experts to learn new tools. The workshops shall preferably involve all subject areas and all campuses of SLU. Both introductory and more advanced workshops are welcome. A workshop series does not only have to be restricted to SLU but could be organized in a national collaboration with other universities.

- As a researcher, what kind of support or collaboration would you like to see with the IT Division?

- As a researcher, what kind of support or collaboration would you like to see with the Data Curation Unit?

- How to collect big data and how to critically pre-process and clean the data? Can IT Dept. or DCU assist in data collection and pre-processing.

- We want much more information from the DCU on what they can assist us with. It would also be good with a promotion campaign for researchers to increase:

    i. Publishing of data

ii. Storing of data in a SLU databases

We fear that a lot of data is only stored in individual researcher's own computers and is not properly documented, which means that the data can be lost or useless when a researcher leaves SLU. That is a waste of resources.

- Today´s event was well organized day and with well chosen invited lecturers. It was nice that four campuses could meet together.

Some other random notes from the discussion

- AI and ML algorithms are tools. They do not do anything for us but we can have good use of them.

- Concerning robot equipment. We can have an experimental lab with rented equipment (robots) or we can collaborate on equipment with other universities/research institutes. We do not need to own everything ourselves. University of Skövde is building Assar, a robotic arena in collaboration with Volvo (arms, social robots etc). Perhaps there is an opportunity for collaboration? Is there a need for this? Is anyone at SLU working on robotics?

Group discussions at AI/Big Data seminar 190917

Suggested issues for discussion. You may chose only to discuss some of them, and other issues that come up in the group can be added.
A rapporteur for each group/campus is asked to submit a written summary of the discussion to the organizers.

- What is your vision for the use of Artificial Intelligence, Big Data and Machine Learning at SLU in five years?

Our success in the future depend on how well we can apply these techniques.

AI techniques are replacing everything else in forest research. We have stopped doing our previous techniques.

AI can help us understand the complexities of People environment interactions.

There might be problems mapping people, unlike to mapping plants etc. Could be a backlash.
Security issues. How do you safely store data with a lot of personal information?
People are more aware of privacy issues nowadays. It is difficult to recruit people for research studies.

- What can SLU do to make the most of the possibilities created by Artificial Intelligence, Big Data and Machine Learning?

People at SLU generally need more knowledge in AI.

Each scientist need to have knowledge.

We are in need of computer scientist to include in our research.
mathematicians, statisticians, datascientists. Are in Lund. Make use of that.
They have techniques. We have questions and knowledge in e.g. agriculture.
We have good experiences of collaborations with LU from PlantLink.

We could do a workshop on agriculture. With AIML@LU

We should not try to build a computer science unit at Alnarp. We should buy or merge.

We know very little about each other at SLU. Even less about people outside.

There needs to be a coordinator on a SLU central level, maybe faculty level. Mapping and collaborating.

We have good knowledge on big data. Less on AI.

The PhDs should be trained in data analysis. Don’t let them handle it to a pro.

- How can we prepare SLU students for using Artificial Intelligence, Big Data and Machine Learning in the future?

Anders Carlsson had quite a lot of thoughts on preparing students. Contact him if you want input on this issue.

There is a problem with the basic education. Students are not skilled enough. We have to take people from outside. You start here because you are interested in plants, not because you are interested in computers.

Introducing techniques in an early stage. Lectures from LU.

- As a researcher, what kind of support or collaboration would you like to see with the IT Division?

The storage is problematic.

Processing data might be a bottleneck. Too few people know data processing at SLU.

There are three problems which we need to solve:
Data storage (today they are being solved locally – not sustainable)
Data Management
Data processing.

- As a researcher, what kind of support or collaboration would you like to see with the Data Curation Unit?

- Can DCU create templates for creating ”datahanteringsplaner” for all our different research funds like EU, Formas. They all want different information there does not seem to be one datahanteringsplan.

- Continue informing ab out what kind of resources are already existing at the SLU, e.g. DCU and IT who can help with AI/ Big Data.

- Big Data changes the way research is done, researchers need to understand this and need education.

- It was interesting to see what research is going on at SLU that involves AI/ Big Data. We need to spread information about it.

- The interdisciplinary approach and even with involving IT is a must in the future.

- Should AI/ Big Data be part of our courses

- As a continuation of this seminarian:
    - Create a course in Big Data/ AI for researchers and students as well.
    - We need to create a centralized function that supports AI/ Big Data covering IT research infrastructe, data scientists, help with interpreting data. Smaller institutions or researchers that are new to Big Data need help to start.
    - Create a platform an arena to create more seminaries like this one and spread information about it
    - Start with creating a webpage about this seminarian and continue to update it with new happenings