Seminar
Automated Parameter Tuning and Algorithm Configuration

Frank Hutter & Marius Lindauer

Emmy Noether Research Group on Learning, Optimization, and Automated Algorithm Design
Today’s class

• Overview of seminar
• Introduction to seminar topic
• Brief round of introductions
• Description of available papers
• Tips for giving a good presentation
Overview of the course

Seminar
- Open to BSc, MSc, and even PhD students
- Worth 4 ECTS credits

Meeting times
- Weekly, Thursday 2-4pm
- 8 slots: May 15, May 22, May 29 (Christi Himmelfahrt), June 5, June 12 (Pfingstpause), June 19 (Fronleichnam), June 26, July 3, July 10, July 17 (SAT conference), July 24, July 31

Mechanics
- We discuss research papers (in English; clarifications in German OK)
- You read each paper that is presented
- You present one paper and lead the discussion for that paper
  - Optional: team of 2
    (1 paper presented jointly, longer report with practical part)
- Grades: combination of all aspects of the course
Your part in the course

For one paper:

– Understand it in detail
– Present the paper and lead the discussion; receive anonymous feedback from your peers right after class
– End of term: write a report about the paper or a related topic; receive anonymous reviews from your peers

For each paper being presented:

– Read it, write a brief summary and formulate 3 questions
– Attend the presentation
– Give anonymous constructive feedback to the presenter(s) right after class
– Participate in a lively discussion about the paper

Warning:

– This course will be more work than a standard block seminar
– But you’ll also get more out of it
In detail: preparation for “your” paper

Understand it in detail
- Usually requires reading up on some background material
- It can often help to download the paper’s code and try it out

Plan your presentation (it should take 20-25 minutes)
- What you will present (including background from other papers!)
- What you will skip and why
- Outline: hierarchical bullet points, with time budget for each point
- Optional 1:
  • Meet with one of your peers & discuss outline / draft presentation
- Optional 2:
  • Send slides to Marius 1 week before presentation
  • Meet to discuss the presentation & then adjust it

Practice, practice, practice!
In detail: more about “your paper”

Present the paper and lead the discussion

Open scientific discussion

– Strengths & weaknesses of the paper
  • Typically not everything is perfect
– Relation to other papers we covered
– Interesting future work

Write a report about the paper or a related topic

– In LaTeX, 2-4 pages
– If you work in a team of 2, this will be more involved, e.g.
  • run the optimization procedure on some other interesting data
  • compare an optimization procedure against a different one
  • extensive literature review
In detail: preparation for other papers

Send to Marius max. 48 hours before presentation:

- Brief paper summary (one paragraph)
  - Main contributions
  - **In your own words**, non-specialized language
  - Purpose: learn to **concisely & accurately summarize** work that you don’t understand in every detail

- Three questions
  - E.g., about
    - something you found unclear
    - how the work relates to something else we covered before
    - any potential problems you noticed
  - Purpose: set up our discussion about the paper

Marius accepts/rejects summaries & questions

- Max. 20% missed or rejected summaries, or you won’t pass
What you’ll learn in this course

**Research skills**
- Reading and understanding a specialized research paper
- Exploring the literature for related work & background material
- Assessing strengths & weaknesses of research papers
- Academic writing
- If you work in a team: hands-on experience (mini-project)

**Soft-skills**
- Giving a good oral presentation
- Active participation in a research discussion
- Giving constructive feedback
- Receiving feedback & using it to improve shortcomings
- Communication skills in English
- If you work in a team: team work
The next steps

TODO after this class:

– If you consider presenting next week or the week after
  • Come see us after class
– Email the following to Marius (lindauer@cs.uni-freiburg.de) by Monday, 2pm
  • Bullet points: your course of study, semester & background/interest in the seminar (e.g., what you are hoping to get out of it)
  • A ranked list of 5 papers you’re interested in presenting and reasons why
  • A ranked list of preferred presentation slots
  • If you’d like to work in a team: your partner’s name
  • Send this email if and only if you commit to taking the seminar
– Read the paper(s) for next week (see website)

We will assign the papers by next Thursday

Questions about the mechanics?
Today’s class

- Overview of seminar
- Introduction to seminar topic
- Brief round of introductions
- Description of available papers
- Tips for giving a good presentation
Motivation: AutoML

- Machine Learning has celebrated substantial successes
- But it requires human experts to
  - Preprocess the data
  - Perform feature selection
  - Select a model family
  - Optimize hyperparameters
  - ...
  - Determine the effect of hyperparameters / choice of model

- **AutoML**: taking the human expert out of the loop

- Same problem occurs in many other areas of CS
  
  “Civilization advances by extending the number of important operations which we can perform without thinking of them”
  
  (Alfred North Whitehead)
Blackbox function optimization

• Optimize a function $f$ over a domain $X$:

  – **Only mode of interaction**: query $f(x)$ at arbitrary $x \in X$

  $$x \rightarrow \text{black rectangle} \rightarrow f(x)$$

• Special characteristics
  – No gradient information
  – Typically, $f$ is not convex
  – Evaluations can be noisy: we observe $f(x) + \varepsilon$, with random $\varepsilon$
Generality of the problem definition

• Function can be implicitly defined
  – All you need is a way to evaluate your function with different input parameters $x \in X$
  – E.g., run an algorithm with parameters $x$ and measure its performance
  – E.g., run a physical process with control parameters $x$ and measure a quantity to be optimized

• General performance measures
  – “Anything that can be measured”
  – E.g., algorithm runtime, approximation error, agreement between output and target output, solution quality, energy consumption, memory consumption, latency, ...
Method of Choice: Bayesian Optimization

• Prominent approach to optimize expensive blackbox functions [Mockus et al., '78]

• Efficient in the number of function evaluations

• Works when objective is nonconvex, noisy, has unknown derivatives, etc

• Recent convergence results [Srinivas et al, '10; Bull '11; de Freitas, Smola, Zoghi, '12]
Algorithm parameters

Decisions that are **left open during algorithm design**

- E.g., real-valued thresholds
- E.g., which heuristic or which optimizer to use

Parameter types

- Continuous, integer, ordinal
- **Categorical**: finite domain, unordered, e.g. \{A,B,C\}

Parameter space has **structure**

- Parameters of sub-algorithm A are only active if A is used
  - E.g., parameters of a specific kernel in an SVM
  - E.g., hyperparameters for layer k of a deep network

Parameters give rise to a **space of algorithms**

- Many “configurations” (e.g. $10^{47}$)
- Configurations often yield qualitatively different behaviour
  → **Algorithm configuration** (as opposed to “parameter tuning”)
The Algorithm Configuration Problem

Definition

– Given:
  • Runnable algorithm $\mathcal{A}$ with configuration space $\Theta = \Theta_1 \times \cdots \times \Theta_n$
  • Distribution $D$ over problem instances $\Pi$
  • Performance metric $m : \Theta \times \Pi \rightarrow \mathbb{R}$

– Find:

$$\theta^* \in \arg \min_{\theta \in \Theta} E_{\pi \sim D}[m(\theta, \pi)]$$

Motivation

Customize versatile algorithms for different application domains

– Fully automated improvements
– Optimize speed, accuracy, memory, energy consumption, …

Very large space of configurations
Crucial question in practice: which distribution do you want to optimize for?

- Goal of parameter tuning: solve future problems better → need *distribution over future problems*

- Example 1: learning a regression model that fits my 100 training data points vs. learning a model that will generalize to new data points

- Example 2: quickly sorting a single list with 1 billion entries vs. quickly sorting arbitrary lists with 1 billion entries

- Example 3: shortest path finding on a compute cluster vs. shortest path finding on a smart phone
Generalization of performance (2)

The dark ages

– Student tweaks the parameters manually on 1 problem until it works
– Supervisor may not even know about the tuning
– Results get published without acknowledging the tuning
– Of course, the approach does not generalize

A step further

– Optimize parameters on a training set
– Evaluate generalization on a test set

What you should do: also avoid “peeking” at the test set

– Put test set into a vault (i.e., never look at it)
– Split training set again into training and validation set
– Use validation set to assess generalization during development (or, better yet, perform cross-validation)
– Only use test set in the very end to generate results for publication
Research from several fields is converging

Until recently: each community used their own methods

- Evolutionary algorithms to tune evolutionary algorithms
- Gradient-based optimizers to tune gradient-based optimizers
- Machine learning to tune machine learning algorithms
- Local search to tune local search

We advocate: choose the right optimizer for the task at hand

- Are the parameters discrete, continuous, or mixed?
- How many parameters are there?
- How much noise is there?
- Etc...
This seminar

Focus: Foundations
   – Machine learning: regression, stochastic processes
   – Statistics: experimental design, statistical tests
   – Optimization: global, stochastic, mixed continuous/discrete
   – AI: local search, population-based methods
   – All of them “on-the-fly”, in the context of parameter tuning

Possible applications
   – Machine Learning
   – Computer vision
   – Natural language processing
   – Formal verification
   – AI planning
   – Robotics
   – Graphics
   – Algorithm Engineering
   – High-Performance Computing
   – ...

Questions about the seminar topic?
Today’s class

- Overview of seminar
- Introduction to seminar topic
- Brief round of introduction
- Description of available papers
- Tips for giving a good presentation
Introductions

• Some information on yourself
  – Your name
  – Your field of study and semester
  – Why you’re interested in this course & what you hope to get out of it

• Less than 1 minute per person

• Purpose: get to know each other’s background, maybe find a team partner
Today’s class

- Overview of seminar
- Introduction to seminar topic
- Brief round of introduction
- Description of available papers
- Tips for giving a good presentation
Tutorials / Introductory material for Bayesian Optimization

- Bayesian optimization tutorial. *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modelling and Hierarchical Reinforcement Learning* (Sections 1-2, i.e., pages 1-23)

- Classic paper on Bayesian optimization for continuous blackbox functions: EGO. *Efficient Global Optimization of Expensive Black-Box Functions*, Journal of Global Optimization, 1998. (Sections 1-4.2, i.e. pages 455-476)

Improvements of GP-based Bayesian Optimization


Papers (2)

Approaches for General Algorithm Configuration


- **Parallel Algorithm Configuration (code)**. *Parallel Algorithm Configuration*. Proceedings of LION 2012: Intl. Conference on Learning and Intelligent Optimization

Hyperparameter Optimization and Structure Search


Combinations of Bayesian Optimization for Hyperparameter Optimization with Transfer Learning, Meta Learning, and Ensemble Learning


- **Meta Learning 1:** [Collaborative hyperparameter tuning](#). Proceedings of ICML 2013: International Conference on Machine Learning.

- **Meta Learning 2:** [Efficient Transfer Learning Method for Automatic Hyperparameter Tuning](#). Proceedings of AISTATS 2014: Artificial Intelligence and Statistics.

- **Ensemble Learning:** [Sequential Model-Based Ensemble Optimization](#). arXiv preprint.

**Scientific Understanding Based on the Models of Bayesian Optimization**

- **Simple Forward Selection:** [Identifying Key Algorithm Parameters and Instance Features using Forward Selection](#). Proceedings of LION 2013: Intl. Conference on Learning and Intelligent Optimization.

Today’s class

• Overview of seminar
• Introduction to seminar topic
• Brief round of introduction
• Description of available papers
• Tips for giving a good presentation
How to give a good presentation

Some parts of this are adapted from Thomas Brox’s slides, with permission.
1. Never present other people’s work as your own
   – Never copy-paste
     (even critical when copying from your own work → self-plagiarism)
   – Clearly mention the material you used for your work
     (e.g. code, data, papers; if unpublished material, ask before you use it)
   – State explicitly what is your contribution

2. Never report false scientific results
   – Do not fake data to get the results you want (of course!)
   – Avoid situations that could easily lead to false results
     • Document what you did
     • Make sure comparisons are fair
     • Double check for mistakes (particularly when results are surprisingly good)

• This holds for this seminar, but also for reports, theses, papers, grant proposals, interviews, personal communication
Examples of how to cite others’ work

• Quotes from other work should have quotation marks:
  – *X and Y [12] define this problem as follows:* “…”

• Provide references for figures

• Mention & clarify contributions from others:
  – *The results reported in this section are based on a joint project with X. While he had the main idea and wrote all the code, I was responsible for the experiments.*

  *For our implementation, we built upon the source code provided by X [13].*
Consequences of bad scientific behavior

• If you cheat in an exam, it will be marked as “failed”

• In severe cases, you can get exmatriculated!

• You can get sued for copyright violations

• You can lose your academic degrees even years after your misbehavior

• You can lose the right to submit grant proposals

• You can lose your job

⇒ Never cheat or plagiarize on purpose, clearly mark your references, adopt best practices for avoiding mistakes
Communication is hard work.
The work can be done either on the side of the sender or on the side of the receiver.
Importance of good presentation skills

• You’ll have to give a lot of presentations in your life (both in academia and industry)

• These presentations can decide whether
  – You get a job
  – Your favourite project gets funded
  – You get the resources you need
  – ...

• Presentation skills and communication skills go together
  – Improving one will help with the other
Getting your points across

• What matters is what your audience gets (not which points you “covered”)
  – Often, the audience is not as interested in the topic as you
  – You’ll have to tell them why they should be care
  – If nobody cares or understands it’s typically your own fault

• At least the key points must get across to everyone
  – Some details may only be for experts, that’s OK
Rule #1: Structure is key

• **High level to low level to high level**
  – Catch your audience’s attention
  – Then tell them what you’ll tell them and why they should care (priming)
  – Then tell it to them
  – Then tell them what you just told them

• **Make transitions clear, don’t forget the “meta-talk”**
  – E.g., “In order to explain X, first I’ll need to explain Y”
  – E.g., “Now that we’ve seen X and Y, we have the ingredients to do Z”
  – Remind the audience where you are in the talk,
    e.g. using a re-occurring outline slide
  – Use meaningful titles

• **Don’t get lost in details**
  – In case of doubt leave out some details
  – To scientists, some detail is often important; you can use a “T-structure”: combine broad coverage of a topic with depth about one aspect
Rule #2: present in pictures

• Slides full of text are hard to follow
  – The audience will read and not listen to you
  – Reduce text, use more images
  – Use animation sparsely, to guide attention

Method of Choice: Bayesian Optimization

• Prominent approach to optimize expensive blackbox functions [Mockus et al., ’78]
• Approach
  – Observe a few function evaluations
  – Construct a probabilistic model of the objective function, for example a Gaussian process
  – Use that model to compute a so-called acquisition function that quantifies how useful a new data point is, trading off exploitation of areas predicted to be good and exploration of areas where the model is uncertain
  – Use the acquisition function to select the next point to evaluate the function at
  – Evaluate the function there, refit the model, and iterate
• Efficient in the number of function evaluations
• Works when objective is nonconvex, noisy, has unknown derivatives, etc
• Recent convergence results [Srinivas et al., ’10; Bull ’11; de Freitas, Smola, Zoghi, ’12]
Rule #3: Have readable slides

- Can you read this text?
- Also from the back? Remember, the contrast and resolution of your laptop is usually much better than that of the projector.
  - Sometimes the font size is too tiny
  - Sans-serif fonts are easier to read from the back than serif-fonts

Also still quite common is yellow text on white ground

You see this even more often in graphs

Make sure there are no typos in your slides; it’s so unprofessional and unnecessary.

Size up figures to use most of the slide. A slide does not need a big frame.
Rule #4: Practice

• Prepare what you want to say, do not improvise!
  – Have a time budget for each part
  – Write down bullet points of what you want to say in each part
    • Say it out loud a few times & check the timing for the part
    • Then do the part a few times without looking at your notes
  – Write out exactly what you want to say in the first minute and as a closing statement
    • You are most nervous in the beginning
    • You want to end pointedly (also, with a final “Thank you”)
    • Practice first minute and closing statement at least 10 times

• Then put it all together
  – Do the transitions work?
  – Always get stuck at the same point? Change that point!
  – Don’t speak too fast! Speaking too slowly is almost impossible
Rule #5: control you technical equipment

• Prepare and test your equipment before the talk (if possible)

• Checklist:
  – Does your laptop work with the projector?
  – For Mac-Users: do you have the right dongle?
  – Do all videos show properly?
  – Internet connection switched off?
  – Screen saver switched off?
  – Desktop free of too personal items?
  – Enough battery or laptop plugged in?

• Use laser pointer (only) for directing attention
Rule #6: Behave naturally

• Keep eye contact with the audience; don’t turn your back
  – But do not wonder what they might think of your presentation! (now it’s too late)

• Relax
  – Breathing in & out deeply once can help
  – Practice helps building confidence

• Answering questions:
  – First listen to the whole question carefully; don’t interrupt
  – Long/multiple questions: take bullet point notes
  – Think about how you can best answer a question before you answer it
  – Give short and precise answers
Rule #7: Adapt your talk to your audience

• The paper you are presenting is written for a specialized research community

• But your audience has a different background
  – You will need to cover the necessary background
  – We’ll be parameter tuning experts – don’t bore us with what we already know

• For other presentations
  – A talk to the CEO is completely different than one to the tech support group
  – A talk applying method X to problem Y is completely different when you’re talking to community studying X or Y
Rule #8: Learn from the mistakes of others

• You cannot follow someone’s talk?
• You are totally bored?
• You are irritated by a certain behavior of the presenter?

➔ Analyze what the presenter is doing wrong
➔ Make sure to give them (friendly & constructive) feedback and do not make the same mistakes
Giving constructive feedback

• Start with something positive
  – In your own reviews you don’t want to hear only negative things, either
  – People are more receptive to criticism after hearing something positive

• Make concrete suggestions
  – Bad example: “The lecture was bad”
  – Good example:
    “I couldn’t follow the math because I couldn’t read your handwriting on the board – better use a projector or slides”
Today’s class

• Overview of seminar
• Introduction to seminar topic
• Brief round of introduction
• Brief description of available papers
• Tips for giving a good presentation
Reminder: The next steps

TODO after this class:

– If you consider presenting next week or the week after
  • Come see us after class

– Email the following to Marius (lindauer@cs.uni-freiburg.de) by Monday, 2pm
  • Bullet points: your course of study, semester & background/interest in the seminar (e.g., what you are hoping to get out of it)
  • A ranked list of 5 papers you’re interested in presenting and reasons why
  • A ranked list of preferred presentation slots
  • If you’d like to work in a team: your partner’s name
  • Send this email if and only if you commit to taking the seminar

– Read the paper(s) for next week (see website)

We will assign the papers by next Thursday