Skoltech

Course Syllabus

Course Title	Models of Sequential Data
Course Title (in Russian)	Модели последовательных данных
Lead Instructor	Zaytsev, Alexey

1. Annotation

Course Description

In this course, we discuss the forefront of modern research in learning from sequence data. The course takes a walk from the basics of sequence processing to modern deep learning approaches. We aim to cover both fundamental and modern advances in this area, which are not commonly discussed in undergraduate or graduate Machine Learning and Deep Learning classes.

Over multiple weeks, we will investigate how researchers and practitioners use these methods and algorithms to analyze time series data, text data, or medical sequences. We explore how experts use these concepts for time series forecasting, machine failure detection, assessing the similarity of oil wells, and more. To provide a full picture of the sequential data models, we provide methods, foundations and their usage in specific applications.

We assume that the course participants are familiar with machine learning and deep learning basics. However, we will strive to ensure the course is accessible to a broad audience of Skoltech MSc and PhD students.

Course Description (in Russian)

В этом курсе мы обсудим основные направления современных исследований в машинном обучении на основе последовательных данных. Курс охватывает как базовые алгоритмы обработки последовательностей, так и современные подходы на основе глубинного обучения, которые обычно не обсуждаются на курсах машинного обучения и глубинного обучения.

Кроме того, мы сделаем акцент на том, как с помощью таких подходов может решаться широкий набор прикладных задач. Мы исследуем, как эксперты используют эти концепции для прогнозирования временных рядов, обнаружения отказов механизмов, оценки сходства нефтяных скважин и многое другое. Чтобы предоставить полную картину последовательных моделей данных, мы приводим методы, основы и их использование в конкретных приложениях.

Мы предполагаем, что участники курса знают основы машинного и глубинного обучения, но попытаемся сделать курс как можно более доступным для широкой аудитории магистров и аспирантов Сколтеха.

2. Basic Information

Course Academic Level	MSc, PhD
Number of ECTS credits	6

Course Prerequisites / Recommendations

- Familiarity with Python and Machine learning
- Experience with Deep Learning frameworks such as Pytorch
- Mathematical competence

Type of Assessment	Graded

Mapping from grades to percentage:

A:	86
B:	76
C:	66
D:	56
E:	46
F:	0

Students of Which Programs do You Recommend to Consider this Course as an Elective?

BSc Programs	Masters Programs	PhD Programs
	Data Science All Master Programs	Computational and Data Science and Engineering All PhD Programs

Maximum Number of Students

	Maximum Number of Students
Overall:	40
Per Group (for seminars and labs):	

Course Stream	Science, Technology and Engineering (STE)

3. Course Content

Topic	Summary of Topic	Contact Hours: Lectures	Contact Hours: Seminars	Contact Hours: Labs	Non-contact Hours: Student's Independent Study
Introduction	In the recent past, there has been limited exploration of deep learning applications for the processing of time series data. We discuss how students can start processing time series and the research's greenfield areas.	2	2		10
ARIMA and stationarity. Decomposition of time series	Classic approaches can help us to understand the internal data dependencies. One of the most known traditional approaches for time series is the ARIMA model, which we consider in detail with its modifications.	3	3		3
Recurrent Neural Networks	The course provides an introduction to Recurrent Neural Networks (RNNs) and their two main architectures: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). We discuss their specific characteristics, implementations, and various applications.	1.5	1.5		10
One-dimensional CNNs	We consider the adaptation of convolutional neural networks (CNNs) for time series processing to address common issues with RNNs.	1.5	1.5		10
Attention. Transformers	The attention mechanism has proven to be a breakthrough in sequence modeling. We discuss how it works, delve into the Transformer architecture, and consider its advantages and disadvantages.	1.5	1.5		10
Modeling of long sequences	Transformers require high computational and memory costs, if we aim for long attention spans. We consider methods that enable long sequences processing and how they evolved over the last few years.	1.5	1.5		5
Deep state-space models. Existing benchmarks	Long sequences can be prohibitively expensive for Transformers. We discuss here an alternative based on the representation of a sequence as an ODE, state-space models.	1.5	1.5		5
Self-supervised learning for sequential data	We consider the concept and implementation details of the self-supervised approach, discussing its two main types: generative and contrastive. We also examine their advantages and disadvantages, as well as modern solutions and limitations.	1.5	1.5		10

Topic	Summary of Topic	Contact Hours: Lectures	Contact Hours: Seminars	Contact Hours: Labs	Non-contact Hours: Student's Independent Study
Metric learning for sequential data	Many machine learning approaches require a measure of distance between data points. Metric learning aims at automatically constructing task-specific distance metrics from (weakly) supervised data in a machine-learning manner. It is especially interesting for time series data.	1.5	1.5		5
Change-point detection	The change point detection (CPD) task involves identifying if and when a series' behavior changes significantly. This is an important problem because it can help us understand the causes of changes in data, such as identify emergency situations. We will discuss the theoretical formulation of the CPD problem, as well as basic and advanced methods for solving this problem.	1.5	1.5		10
Audio and speech processing	Audio data represents one of the largest classes of data types. We devote into separate methods to such data since the direct application of standard sequential data processing algorithms is inferior if compared to the modality-specific approaches.	3	3		5
Event sequences processing. Neural temporal point processes	Temporal Point Process is a collection of mathematical points randomly located in a mathematical space on a timeline. We consider its definition and modeling from a mathematical point of view. We also discuss the classical approaches and deep learning realizations.	1.5	1.5		10
Spatio-temporal data processing. ConvLSTM	Spatio-temporal data is one of the largest classes of data types. We will discuss special tools for their analysis and visualization and models for processing them.	1.5	1.5		5
Uncertainty estimation for sequential data. Bootstrap	Point estimation is not enough in some tasks (e.g., in medicine). Hence, we need to estimate the uncertainty for prediction. Here we discuss, how a model can include uncertainty estimate in the case of dependent sequential data	1.5	1.5		5

4. Learning Outcomes

Skoltech Learning Outcomes are indicated as per [Skoltech Learning Outcomes Framework](#).

1. FUNDAMENTAL KNOWLEDGE

- 1.1. KNOWLEDGE OF MATHEMATICS AND NATURAL SCIENCES
- 1.4. INTERDISCIPLINARY THINKING, KNOWLEDGE STRUCTURE AND INTEGRATION

2.1. COGNITION AND MODES OF REASONING

- 2.1.2. System thinking
- 2.1.3. Creative thinking
- 2.1.4. Decision making (with ambiguity, urgency etc.)
- 2.1.5. Critical thinking and meta-cognition

2.2. ATTITUDES AND LEARNING PROCESS

- 2.2.1. Initiative and the willingness to take appropriate risks
- 2.2.2. Willingness to make decisions in the face of uncertainty
- 2.2.3. Responsibility, intensity, perseverance, urgency and will to deliver
- 2.2.4. Resourcefulness, flexibility and an ability to adapt
- 2.2.5. Self-awareness and a commitment to self-improvement, lifelong learning and educating

2.3. ETHICS, EQUITY AND OTHER RESPONSIBILITIES

- 2.3.6. Commitment to social and professional behavior

3.1. COMMUNICATIONS IN INTERNATIONAL ENVIRONMENTS

- 3.1.3. Oral presentation and discussion
- 3.1.4. Inquiry, listening and dialogue
- 3.1.5. Communications in English in scientific, business and social settings

3.2. TEAMWORK AND LEADERSHIP

- 3.2.2. Team operations and project management
- 3.2.3. Team coordination, decision-making and leadership
- 3.2.4. Team growth and evolution
- 3.2.5. Technical and multidisciplinary teaming

3.3. COLLABORATION AND CHANGE

- 3.3.1. Establishing diverse connections and networking
- 3.3.2. Appreciating different roles, perspectives and interests
- 3.3.3. Negotiation and conflict resolution
- 3.3.4. Stakeholders' engagement
- 3.3.5. Bringing about intentional change

4.1. MAKING SENSE OF GLOBAL SOCIETAL ENVIRONMENTAL AND BUSINESS CONTEXT

- 4.1.3. Understanding the technical products, systems and infrastructure of the sector

4.2. VISIONING — INVENTING NEW TECHNOLOGIES THROUGH RESEARCH

- 4.2.2. Basic research leading to new scientific discovery
- 4.2.3. Research aimed at developing new technologies
- 4.2.4. Imagining utility of new science and technology
- 4.2.5. Developing concepts and reducing to practice

4.3. VISIONING — CONCEIVING AND DESIGNING SUSTAINABLE SYSTEMS

- 4.3.2. Identifying and formulating objectives and goals
- 4.3.6. Design process management, including planning, project judgment and effective decision-making

5. Assignments and Grading

Physical Attendance Requirement (% of classes) 80

Assignment Type	Assignment Summary	% of Final Course Grade
Test/Quiz	At the end of almost all even classes, students will have either an in-class quiz or a take-home test. The quizzes are multiple-choice and are solved online on Canvas. The objective of these quizzes is to solidify the knowledge acquired during the class. The total number of quizzes equals 6.	20

Assignment Type	Assignment Summary	% of Final Course Grade
Report	Early Report. At the beginning of the term, the instructor presents a set of projects to students. Students form groups of three to four participants and work on selected projects. Projects can consist of reproduction studies, novel work, or analytical tasks. Students must start working on projects by preparing several sections, which will be included in the Report. Early Report must include at least the following sections: Introduction, Related Work, Methods (including Data Description and Methodology with models' description), and Experiments (including at least a description of experiments' setup).	5
Report	Students must write an 8-page double-column report describing their work and results, including plots and bibliography. This report must be of high quality and comparable to a publication at a small venue for machine learning research. The objective is to submit the reports for publication to machine learning conferences.	15
Presentation	One of the final classes will be devoted to oral project presentations. Students must prepare a 10-minute presentation describing the idea of their project and the obtained results. They will then have a 10-minute follow-up question-and-answer session.	10
Report	Peer-review. Students must do a peer-review and grade the work of their peers.	5
Homework Assignments	Our course will have two graded assignments corresponding to classic models (e.g., ARIMA) and Transformers. The code should correspond to the PEP-8 standard, be reproducible (requirements.txt is mandatory), and be described in detail.	25
Midterm Exam	Midterm exam	20

6. Assessment Criteria

Assignment 1 Type

Test/Quiz

Sample of Assignment 1

Are the Transformer architectures RNNs?
- Yes
- No

What are some downsides of RNNs?
- They are not suited for long sequences
- They do not generalize well
- They can only be used for forecasting tasks
- They do not have downsides

Assessment Criteria for Assignment 1

At the end of some classes, participants must complete a quiz in Canvas. The test will be short, multiple-choice, and graded. Each of them can only be taken once. The examination is completed individually. No forms of academic dishonesty will be tolerated.
Most exams will have a duration of around 45 minutes. The questionnaire contains questions related to the content covered on the same day and during the previous class.

Assignment 2 Type

Report

Sample of Assignment 2

Early Report. This deliverable consists of a report proving a deep understanding of the aim of the selected project. The report must match the template provided by the course organizers and contain at least the following sections: Introduction, Related Work, Methods (including Data Description and Methodology with models' description), and Experiments (including at least a description of experiments' setup).

The report must be submitted in Canvas. It must not contain direct formulations from articles, and all the theoretical basics must be based on existing articles with the correct citation design. The report must contain at least the following sections: Introduction, Related Work, Methods (including Data Description and Methodology with models' description), and Experiments (including at least a description of experiments' setup).

Assignment 3 Type

Report

Sample of Assignment 3

Provide a project report. This deliverable must be an 8-page double-column report with reliable, logical, high-quality text describing the project and obtained results.

Assessment Criteria for Assignment 3

The report must be submitted in Canvas. The quality of the submission should be equal to that of a publication that can be published in a minor venue.

Assignment 4 Type

Presentation

Sample of Assignment 4

This deliverable consists of a submitted a pdf/pptx file with slides and a 10-minute oral presentation. Each team member must participate in a project presentation.

Assessment Criteria for Assignment 4

Provide a pdf/pptx file with slides in Canvas. In-class participation is mandatory. The grade for the presentation will be assigned based on your presentation skills, quality of the presentation, overall exposition, and presented results.

Assignment 5 Type

Report

Sample of Assignment 5

Peer-review. Peer review of another team's report.

Assessment Criteria for Assignment 5

Peer review of another team's report must be submitted in Canvas and consist of an objective assessment of the team's work with constructive comments and suggestions. This type of activity is similar to the work of reviewers making decisions about papers' publication.

Assignment 6 Type

Homework Assignments

Sample of Assignment 6

Time series forecasting by ARIMA model.

Assessment Criteria for Assignment 6

Our course will have two graded assignments dedicated to classic models (e.g., ARIMA) and Transformers. The code should correspond to the PEP-8 standard, be reproducible (requirements.txt is mandatory), and be described in detail.

Assignment 7 Type

Midterm Exam

Input Sample of Assignment 7

Midterm Exam

Assessment Criteria for Assignment 7

The questions will be related to the course topics.

Assignment 8 Type

Class participation

Sample of Assignment 8

Mandatory

Assessment Criteria for Assignment 8

Mandatory

7. Textbooks and Internet Resources

You can request at most two required textbooks. Additionaly, you can suggest up to nine recommended textbooks.

Required Textbooks	ISBN-13 (or ISBN-10)
-	-

Recommended Textbooks	ISBN-13 (or ISBN-10)
Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning, 2015.	9780262035613
Goldberg, Yoav. "Neural network methods for natural language processing." Synthesis Lectures on Human Language Technologies 10.1 (2017): 1-309.	9781627052986
Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, Danilo P. Mandic and Jonathon A. Chambers. 2001.	9780471495178

Papers	DOI or URL
Dey, R. and Salemt, F.M., 2017, August. Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.	
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).	

Web-resources (links)	Description
Dey, R. and Salemt, F.M., 2017, August. Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.	
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).	

8. Facilities

Software
Python 3.6+
Jupyter
Open source python libraries including machine learning ones: PyTorch, pandas, Matplotlib, seaborn, NumPy

Equipment
Laptop

9. Additional Notes

Comments

We will follow these rules for the final submissions:
(1) If you fail to submit either the PROJECT, REPORT, or PRESENTATION in time, your final grade for all of them will be zero. The three deliverables are the results of multiple weeks of work. As such, it must be a joint submission. We do not accept late submissions. Plan your workload.

(2) The preliminary grade for the PROJECT and the REPORT will be assigned based on the grading scheme approved in your project proposals. This grade can be decreased depending on their quality, such as the structure, completeness, logic, and quality of the code.

(3) Please be aware that even if your grade is 100% according to the grading scheme for the PROJECT and the REPORT, you may finally end with a zero if your experiments are incomplete or the results are non-reproducible, or they lack proper explanations and code.

(4) The grades for PROJECT, REPORT, and PRESENTATION will be the same for all team members in most cases. For this reason, the workload must be equally divided between team members.

(5) In some extreme cases, we can modify individual members' grades proportional to the contributions of each team member.

The proposed course 1) has explicit academic content and requirements for receiving credits, 2) is in alignment with the program's learning outcomes, 3) adheres to policies and Skoltech regulations.

Lead Instructor confirms