Transforming Education with Machine Learning: Practical Applications and Case Studies

Benjamín Garzón

University of Zurich

February 14th, 2025

Learning Goals

- Get acquainted with the main applications of machine learning in education
- Explore two specific applications (student performance prediction and automated grading)
- Acquire hands-on experience with the datasets and models relevant in these two applications

Schedule

2:00 PM - Machine learning in education: an overview

2:30 PM - Prediction of student performance (tutorial)

3:30 PM - Coffee break

4:00 PM - Automated grading (tutorial)

5:00 PM - Summary and discussion

Q&A: at the end of each session or during the hand-on tutorials

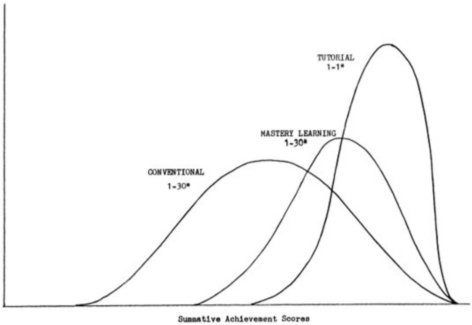
Machine learning in education: an overview

Education

Burgos et al. 2024, https://sdgs.un.org/goals/goal4

Two Sigma Problem

- *The average student working with a **competent human tutor** performs **two standard deviations** above the average control student taught under conventional group methods of instruction*
- With the right tutoring most students have the potential to reach a high level of learning



Bloom 1984; A. Corbett 2001

Automating and Personalizing Educational Activities

- One-to-one (human) tutoring is too costly to be deployed at scale
- Educational technologies can help to automate many aspects of educational activities in order to make them more scalable

Bloom 1984, DALL-E

Educational Contexts

- Early Childhood (Preschool, Kindergarten)
- Primary school
- Secondary school
- Undergraduate (Bachelor's Degree)
- Graduate (Master's/Doctoral Degree)
- Vocational
- Special
- Non-Formal (workshops, seminars, courses)

DALL-E 3

Educational Data

- Educational data from millions of students has become available through the widespread use of digital technologies in education, e.g.:
  - Learning management systems (LMS)
  - Massive open online courses (MOOCS)
- Many educational datasets are freely available:
  - **DataShop** (pslcdatashop.web.cmu.edu)
  - **Kaggle** (kaggle.com)
  - **Papers with Code** (paperswithcode.com/datasets)
  - **ML Repository** (archive.ics.uci.edu)
  - **Assistments** (sites.google.com/site/assistmentsdata)

Educational Data

- Login/logout Times
- Time spent on different exercises/activities/topics
- Responses
- Scores
- Grades
- Forum activity (posts, replies)
- Progress tracking
- Clickstream and mouse movement
- Questions and answers
- Essays
- Dialogue data
- ...

Educational Data

- This wealth of data can be used to improve instruction and learning outcomes
- Machine learning can play a critical role in this task of automating and personalizing educational activities and processes
- Three main research disciplines:
  - **Educational Data Mining (EDM)**: aims to develop **methods for exploring educational data**, understanding students and improving teaching
  - **Learning Analytics (LA)**: the measurement and analysis of data about learners and their contexts to **optimize learning and learning environments**
  - **Psychometrics**: branch of psychology concerned with the **quantification and measurement of psychological attributes**

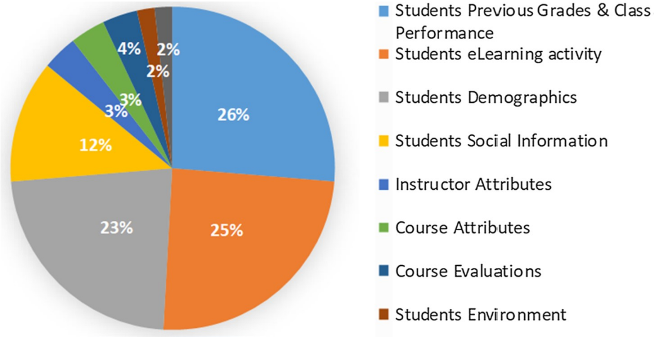
Siemens and R. S. J. D. Baker 2012

Classical ML in Education

- Supervised models for regression and classification (decision trees and ensembles, support vector machines, neural networks)
  - Student grades
  - Student dropout
  - Academic failure
  - ...
- Unsupervised models (K-means clustering, hierarchical clustering, ...)
  - Grouping similar students according to behavioral information or learning patterns
  - ...

Regression: Student Grades

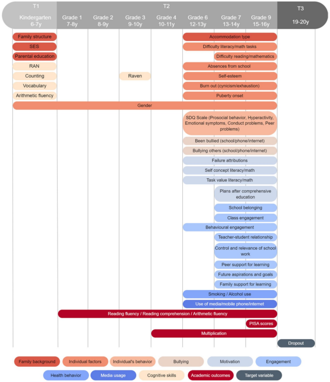
- Important for personalized support, resource allocation, identification of struggling students
- Common predictors:



Abu Saa, Al-Emran, and Shaalan 2019; Batool et al. 2023

Classification: Student Dropout and Academic Failure

- Academic failure is associated with adverse outcomes (lower income, life expectancy; higher unemployment, criminality, poverty risk, institutional costs)
- Longitudinal datasets enable **Early Warning Systems**



Lee and Chung 2019; Psyridou et al. 2024

Affect Prediction

- *Affective states* are a consequential aspect of the learning experience (is the student bored, interested, motivated, confused, frustrated...?)
- Some affective states can be inferred from students' interactions with learning platforms with the aim of improving engagement

Engaged Concentration
The minimum number of previous incorrect actions and help requests for any skill in the clip.
Among the skills involved in the clip, the minimum value for previous incorrect actions and help requests for that skill.
The duration (in seconds) of the fastest action in the clip.
The percentage of clip actions involving a hint followed by an error.

Boredom
The average time the student took to respond on the current step prior to the clip, averaged across all the actions with a clip.
The average time the student took to respond, unitized across time taken by all students on the same problem steps, within sequences of three actions in a row.
The maximum product of the probability of moment-by-moment learning P(J) [9], and the probability of guess P(G) calculated using the contextual guess model [4] for any action in the clip. This can be interpreted as actions where the student learned after guessing.
The maximum number of previous incorrect actions and help. requests for any skill in the clip.

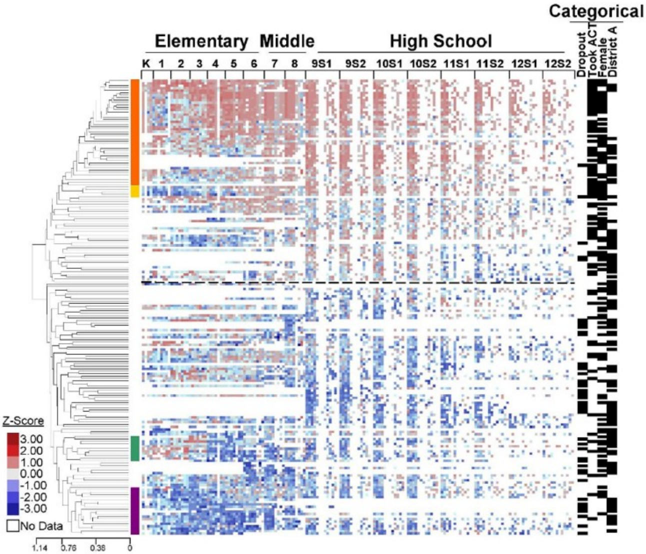
R. Baker 2012; Botelho, R. S. Baker, and Heffernan 2017

Affect Prediction

- Biometric and physiological signals can improve the prediction of affective states
  - Facial emotion recognition
  - Eye-tracking
  - EEG
- Serious ethical concerns

The Wall Street Journal

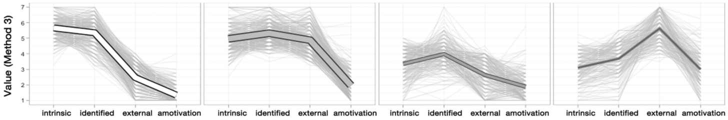
Cluster Analysis: Grading History



Bowers 2010

Cluster Analysis: Motivational Profiles

- Survey on 4 dimensions of motivation in undergraduate engineering students



Dillon and Stolk 2012

Adaptive Educational Systems

- Personalize learning by adjusting to the individual needs of the student
  - Adjust difficulty
  - Score answers
  - Provide feedback
  - Recommend content
  - Pace activities
- Rely on **learner modeling** to predict student performance → See today's **first tutorial session** (14:30)
- **Intelligent tutoring systems**: simulate a human tutor that provides explanations and guides the student to find the solutions step by step

Conversational Tutors

I see your solution. You seem to have made some errors. How about we start fresh and do it step by step?

Okay, I'm not sure where to start though.

What should be the first step?

I think Cody's starting amount of tickets, which is 49, should be the first step.

Tell me this instead: How many tickets does Cody have left after spending 25 on a beanie from his original 49 tickets?

Oh, he would have 24 tickets left.

24 is correct, well done. What is the next step?

Next, he won 6 more tickets, so I think we add that to the 24 he already has.

Okay, can you do that calculation?

Yes, 24 + 6 is 30, so Cody would have 30 tickets in total.

30 is the correct answer to the problem. Well done!!! Don't hesitate to contack me again if you have any troubles.

Chowdhury, Zouhar, and Sachan 2024, khanmigo.ai, livehint.com

Automated Grading

- Unlike closed-ended questions, grading answers to open-ended questions or essays is not trivial
- We need to ensure that the grading meets certain conditions and reflects only relevant aspects of the answer
- → See today's **second tutorial session** (16:00)

gemini

Automated Content Generation

- Creating questions and exercises is costly and labor-intensive
- Requires the work of domain experts
- The content must follow given constraints (format, length,...)
- Examples:
  - Generate multiple choice questions from a text
  - Generate questions on a given topic with particular characteristics (e.g., difficulty)
  - Generate questions with a specific answer

Equations: x=y; 2x+4y=48; **Topic**: Livestock
Entities: x: Chicken; y: Rabbit;
1. Rabbits and chicken are in one cage. The number of rabbits is 0 less than that of chickens. They have 48 legs in total. How many rabbits and chickens in cage?
2. There are the same number of chickens and rabbits in the yard, and the total number of legs is 48. How many rabbits and chickens are in the yard?
3. Chicken and rabbits are in the same cage. Xiaoming counted the number of heads of the two animals and found that the number of chicken heads was 0 more than the number of rabbit heads. There are 48 legs in total. May I ask how many chickens are there?
4. Xiaojun is very good at math, but today there is a difficult problem for him: A chicken has 1 head and 2 legs, and a rabbit has 1 head and 4 legs. There are chickens and rabbits in the same cage, and the number of chickens is equal to the number of rabbits. There are 48 feet in total, so how many chickens and how many rabbits are there?

T. Liu et al. 2021

Language Learning

- Text-to-speech models (realistic and expressive voice generation)
- Speech-to-text models (speaking practice)
- Large language models (dialogues, content generation in different languages, ...)

duolingo

univerbal

Talkpal AI

Summary: Applications

- Digital technologies are providing a wealth of educational data, creating opportunities to use machine learning to improve educational outcomes
- Classical supervised and unsupervised machine learning methods can be used to analyze educational datasets, but there are also domain-specific problems that arise from educational activities
- Examples of educational applications that can be powered by machine learning algorithms:
  - Adaptive educational systems
  - Automated grading
  - Automated content generation
  - Language learning
  - ...

Prediction of Student Performance

Predicting Student Performance

- *"What you cannot measure, you cannot improve"*
- To be able to provide **recommendations** and **tailored instructional options** to improve **learning outcomes**, first we want to know where the student stands in different areas
- Equivalently, we want to be able to **predict a student's performance** when they encounter a new exercise in a particular area

Item

- In an assessment, smallest unit contributing to the total score:
  - multiple choice question
  - open-ended question
  - exercise
  - ...

Knowledge Component (KC)

- Fundamental building block of knowledge in a particular domain (concept, fact, skill, principle, ...)
- Examples:
  - **Rule from name:** "Circle area formula" $\rightarrow$ "$A = \pi r^2$"
  - **Experimental design:** Test hypothesis <Var1> causes <Var2> $\rightarrow$ run an experiment that varies only <Var1> and nothing else and measure differences in <Var2>
  - **Plural English spelling:** <Noun> ending <consonant> + "y" $\rightarrow$ remove "y" from <Noun> and add "ies"

Koedinger, A. T. Corbett, and Perfetti 2012, learnlab.org

Knowledge Component (KC)

- **Item - KC mapping**: different items can refer to the same KC; a particular item may require several KCs
- **Q-matrix**:

	KC1	KC2	KC3	KC4
Item 1	0	1	0	0
Item 2	0	0	1	0
Item 3	0	1	0	0
Item 4	1	0	0	0
Item 5	1	0	0	1

- Normally provided by educational experts but can also be inferred from data

Latent Knowledge State

- Actual knowledge is not directly observable
- Inferred from the student's performance over time
- **Latent Knowledge State**: Model representation of a student's level of understanding
  - Binary (a KC is mastered or not mastered)
  - Continuous (vector)
- **Learner Modeling**: estimate the student's current knowledge state and track it as it changes as they interact with the learning materials

Learner Modeling Applications

- **Personalized recommendations**: informing an instructional policy (stopping criterion, item order)
- **Open learner model**: visualize estimated knowledge state for learners (reflection, self-regulated learning, engagement, social comparison)
- **Actionable insights**: for teachers, system developers, ... (problematic items, structure of KCs, identification of learner groups)

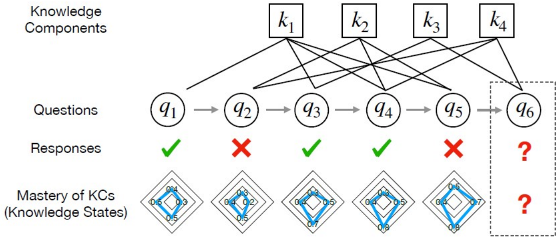
Pelánek 2017

Knowledge Tracing (KT)

- *Considering a student's past history of interactions with a learning system, what is their current knowledge state in given KCs?*

  $x_t = (e_t, r_t)$: interaction at time t
  e_t: item responded at time t (item ID)
  r_t: response at time t (1=correct, 0=incorrect)

- Given the sequence of interactions $(x_1, x_2, ..., x_{t-1})$ before the current time, t, obtain the probability of answering correctly a new item e_t: $P(r_t = 1 | x_1, x_2, ... x_{t-1}, e_t)$

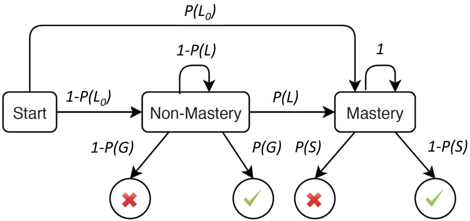


Z. Liu et al. 2023

KT Approaches

- **Bayesian Knowledge Tracing (BKT)**: Notebook 1
- **Deep Learning Knowledge Tracing (DLKT)**: Notebook 2

Bayesian Knowledge Tracing (BKT)



***L_0*:** *Initial knowledge state (1 = mastery)*

L: *Current knowledge state*

***$P(L_0)$*:** *Initial probability of mastery*

P(L): *Probability of transitioning to mastery*

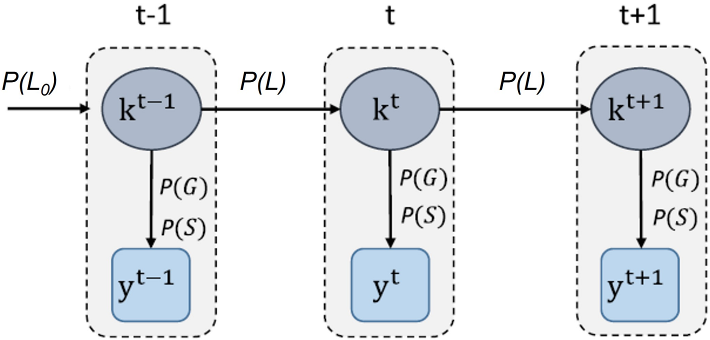
P(G): *Probability of guessing the correct answer when not in the mastery state*

P(S): *Probability of slipping (making a mistake) when in the mastery state*

A. T. Corbett and Anderson 1995; Gervet et al. 2020

Bayesian Knowledge Tracing (BKT)

- Hidden Markov Model
- Binary hidden states and responses



Abdelrahman, Wang, and Nunes 2023

Model Equations

- Probability of knowledge state conditional on correct/incorrect answer:

$$P(L_t|r_t = 1) = \frac{P(L_t)(1 - P(S))}{P(L_t)(1 - P(S)) + (1 - P(L_t))P(G)}$$

$$P(L_t|r_t = 0) = \frac{P(L_t)P(S)}{P(L_t)P(S) + (1 - P(L_t))(1 - P(G))}$$

- Updated prior for next step:

$$P(L_{t+1}) = P(L_t|r_t) + (1 - P(L_t|r_t))P(L)$$

BKT Assumptions

- Vanilla version:
  - Binary responses (correct/incorrect)
  - Binary knowledge states (mastered or not)
  - Each item corresponds to a single KC
  - Separate parameters for each KC
  - Forgetting not possible
- Several extensions (*Khajah, Lindsey, and Mozer 2016; Pardos and Heffernan 2011; Yudelson, Koedinger, and Gordon 2013*)

Datasets

Dataset	#Questions	#Students	#Interactions	#KCs	Public available
ASSISTments2009	26, 688	4, 217	346, 860	123	Yes
ASSISTments2012	179, 999	46, 674	6, 123, 270	265	Yes
ASSISTments2015	100	19, 917	708, 631	-	Yes
ASSISTChall	3, 162	1, 709	942, 816	102	Yes
STATICS2011	1, 224	335	361, 092	80	No
Junyi Academy	722	247, 606	25, 925, 992	41	Yes
Simulated-5 (Synthetic)	50	4, 000	200, 000	5	Yes
Algebra 2005-2006	1, 084	575	813, 661	112	Yes
Algebra 2006-2007	90, 831	1, 840	2, 289, 726	523	Yes
Bridge to Algebra	19, 258	1, 146	3, 686, 871	493	Yes
EdNet-KT1	13, 169	784, 309	95, 293, 926	188	Yes
EdNet-KT2	13, 169	297, 444	56, 360, 602	188	Yes
EdNet-KT3	13, 169	297, 915	89, 270, 654	293	Yes
EdNet-KT4	13, 169	297, 915	131, 441, 538	293	Yes

https://sites.google.com/site/assistmentsdata/home/2015-assistments-skill-builder-data
https://sites.google.com/view/assistmentsdatamining
https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=507
https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=1198
https://www.kaggle.com/junyiacademy/learning-activity-public-dataset-by-junyi-academy/tasks
https://github.com/chrispiech/DeepKnowledgeTracing/tree/master/data/synthetic
https://pslcdatashop.web.cmu.edu/KDDCup
https://github.com/riiid/ednet
https://company.riiid.co/en/product

Abdelrahman, Wang, and Nunes 2023

Tutorial 1

Bayesian Knowledge Tracing

https://tinyurl.com/AMLD25
BKT_tutorial.ipynb

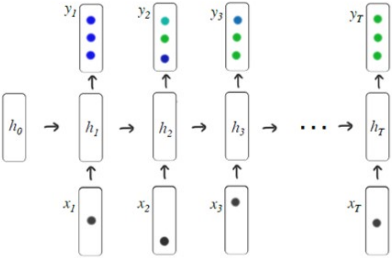
Deep KT (DKT)

- Based on simple RNNs or LSTMs

  x_t: encodes interaction at time t, i.e., current exercise + corresponding response (correct/incorrect)

  h_t: student latent knowledge state given past observations (multidimensional)

  y_t: predicted probability of responding correctly to each item, $P(r_t = 1 | x_1, x_2, ... x_{t-1}, e_t)$



Piech et al. 2015

Loss

- Binary cross-entropy:

$$L = \frac{1}{N}\sum_{i=1}^{N}[\hat{y}_i \log(y_i) + (1-\hat{y}_i)\log(1-y_i)]$$

  N: number of observations
  y_i: model prediction (i.e., probability of correct) $y_i \in [0,1]$
  $\hat{y}_i$: true response (0=incorrect, 1=correct)

- Same for the remaining models

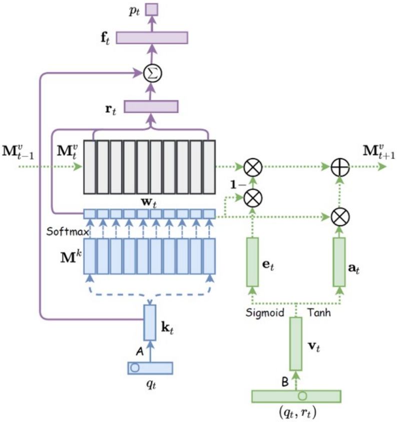
DKT Code

```
# __init__
...
self.interaction_emb = Embedding(self.num_q * 2, self.
    emb_size)
self.lstm_layer = LSTM(self.emb_size, self.hidden_size,
batch_first=True)
self.out_layer = Linear(self.hidden_size, self.num_q)
self.dropout_layer = Dropout()
...

# forward
...
x = self.interaction_emb(q + self.num_q * r)
h, _ = self.lstm_layer(x)
y = self.out_layer(h)
y = self.dropout_layer(y)
y = torch.sigmoid(y)
...
```

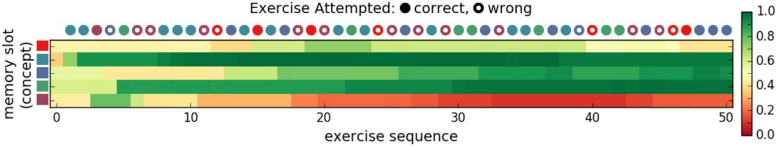
Dynamic Key-Value Memory Networks (DKVMN)

- Encodes separate knowledge states for different KCs simultaneously
  - **Static** key memory for (automatically learned) KCs
  - **Dynamic** value memory (erase/add) for knowledge states for the different KCs



Zhang et al. 2017

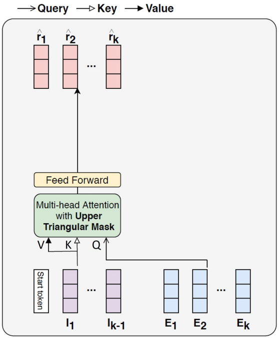
Predicted Response Probability

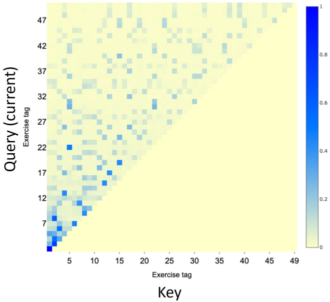


Zhang et al. 2017

Transformer-Based Approaches

- Learn the **attention weights** of items in a sequence of interactions in a way that reflects their relative importance to predict the probability of answering correctly a new item e_t:
  $P(r_t = 1 | x_1, x_2, ... x_{t-1}, e_t)$
- Multi-head attention layer(s)
- Parallelism to speed up training

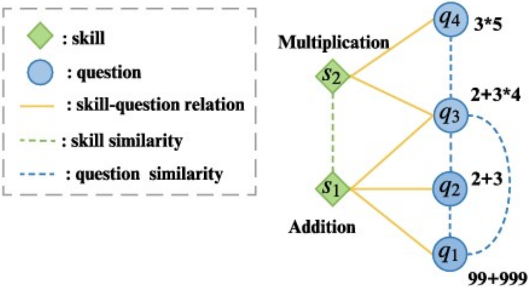




Choi et al. 2021; Pandey and Karypis 2019; Shin et al. 2021

Graph-Based Approaches

- Use graph neural networks (GNNs) to learn relationships from the data to improve prediction
  - Similarity item-KC
  - Similarity KC-KC
- Compensate data sparsity
- Bipartite graph:



Y. Liu et al. 2020; Yang et al. 2020

Model Performance

- DLKT models are most suitable for large datasets; for small datasets they tend to overfit, and simpler learner models fare better
- BKT models generally tend to perform worse than DLKT models, but some variants can still be competitive for certain datasets
- There are many DLKT variants, but they tend to perform similarly

Gervet et al. 2020; Khajah, Lindsey, and Mozer 2016; Z. Liu et al. 2023

Summary: Knowledge Tracing

- Learner models aim to track the knowledge state of a student as it evolves over time, ideally for different **knowledge components** (KCs) and predict responses to new **items**
- Learners models allow providing adequate **recommendations**, **feedback** and **insights** on the learner
- Knowledge tracing (KT) is a popular framework for learner modeling, with two main approaches: Bayesian Knowledge Tracing and Deep Learning Knowledge Tracing
- Different flavors of deep learning models have been developed based on different deep learning architectures (key-value memory, transformers, graph neural networks)
- Deep learning-based models are more difficult to interpret, but perform better on **large datasets** typical of digital learning applications

Tutorial 2

Deep Learning Knowledge Tracing

https://tinyurl.com/AMLD25
DLKT_tutorial.ipynb

Thanks for your attention!

benjamin.garzon@ife.uzh.ch

References

Abdelrahman, Ghodai, Qing Wang, and Bernardo Nunes (Nov. 2023). “Knowledge Tracing: A Survey”. en. In: *ACM Comput. Surv.* 55.11, pp. 1–37. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3569576.

Abu Saa, Amjed, Mostafa Al-Emran, and Khaled Shaalan (Dec. 2019). “Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques”. en. In: *Tech Know Learn* 24.4, pp. 567–598. ISSN: 2211-1662, 2211-1670. DOI: 10.1007/s10758-019-09408-7.

Baker, Ryan (2012). “Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra”. In: *Proceedings of the 5th International Conference on Educational Data Mining*.

Batool, Saba et al. (Jan. 2023). “Educational data mining to predict students' academic performance: A survey study”. en. In: *Educ Inf Technol* 28.1, pp. 905–971. ISSN: 1360-2357, 1573-7608. DOI: 10.1007/s10639-022-11152-y.

Bloom, Benjamin S. (June 1984). “The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring”. In: *Educational Researcher* 13.6, p. 4. ISSN: 0013189X. DOI: 10.2307/1175554.

Botelho, Anthony F., Ryan S. Baker, and Neil T. Heffernan (2017). “Improving Sensor-Free Affect Detection Using Deep Learning”. en. In: *Artificial Intelligence in Education*. Ed. by Elisabeth André et al. Vol. 10331. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 40–51. ISBN: 978-3-319-61424-3 978-3-319-61425-0. DOI: 10.1007/978-3-319-61425-0_4.

Bowers, Alex J (2010). “Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis”. en. In.

Burgos, Daniel et al., eds. (2024). *Radical Solutions for Artificial Intelligence and Digital Transformation in Education: Utilising Disruptive Technology for a Better Society*. en. Lecture Notes in Educational Technology. Singapore: Springer Nature Singapore. ISBN: 978-981-97-8637-4 978-981-97-8638-1. DOI: 10.1007/978-981-97-8638-1.

Choi, Youngduck et al. (Apr. 2021). “Towards an Appropriate Query, Key, and Value Computation for Knowledge Tracing”. en. In: *LAK21: 11th International Learning Analytics and Knowledge Conference*. arXiv:2002.07033 [cs], pp. 490–496. DOI: 10.1145/3448139.3448188.

Chowdhury, Sankalan Pal, Vilém Zouhar, and Mrinmaya Sachan (Apr. 2024). *AutoTutor meets Large Language Models: A Language Model Tutor with Rich Pedagogy and Guardrails*. en. arXiv:2402.09216 [cs]. DOI: 10.48550/arXiv.2402.09216.

Corbett, Albert (2001). “Cognitive Computer Tutors: Solving the Two-Sigma Problem”. en. In: *User Modeling 2001*. Ed. by G. Goos et al. Vol. 2109. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 137–147. ISBN: 978-3-540-42325-6 978-3-540-44566-1. DOI: 10.1007/3-540-44566-8_14.

Corbett, Albert T. and John R. Anderson (1995). “Knowledge tracing: Modeling the acquisition of procedural knowledge”. en. In: *User Model User-Adap Inter* 4.4, pp. 253–278. ISSN: 0924-1868, 1573-1391. DOI: 10.1007/BF01099821.

References

Dillon, Alex and Jonathan Stolk (Oct. 2012). "The students are unstable! Cluster analysis of motivation and early implications for educational research and practice". en. In: *2012 Frontiers in Education Conference Proceedings*. Seattle, WA, USA: IEEE, pp. 1–6. ISBN: 978-1-4673-1352-0 978-1-4673-1353-7 978-1-4673-1351-3. DOI: 10.1109/FIE.2012.6462485.

Gervet, Theophile et al. (2020). "When is Deep Learning the Best Approach to Knowledge Tracing?" en. In: *Journal of Educational Data Mining* 12.3.

Khajah, Mohammad, Robert V. Lindsey, and Michael C. Mozer (June 2016). *How deep is knowledge tracing?* en. arXiv:1604.02416 [cs]. DOI: 10.48550/arXiv.1604.02416.

Koedinger, Kenneth R., Albert T. Corbett, and Charles Perfetti (July 2012). "The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning". en. In: *Cognitive Science* 36.5, pp. 757–798. ISSN: 0364-0213, 1551-6709. DOI: 10.1111/j.1551-6709.2012.01245.x.

Lee, Sunbok and Jae Young Chung (July 2019). "The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction". en. In: *Applied Sciences* 9.15, p. 3093. ISSN: 2076-3417. DOI: 10.3390/app9153093.

Liu, Tianqiao et al. (Sept. 2021). *Mathematical Word Problem Generation from Commonsense Knowledge Graph and Equations*. en. arXiv:2010.06196 [cs]. DOI: 10.48550/arXiv.2010.06196.

Liu, Yunfei et al. (Dec. 2020). *Improving Knowledge Tracing via Pre-training Question Embeddings*. en. arXiv:2012.05031 [cs]. DOI: 10.48550/arXiv.2012.05031.

Liu, Zitao et al. (Jan. 2023). *pyKT: A Python Library to Benchmark Deep Learning based Knowledge Tracing Models*. en. arXiv:2206.11460 [cs]. DOI: 10.48550/arXiv.2206.11460.

Pandey, Shalini and George Karypis (July 2019). *A Self-Attentive model for Knowledge Tracing*. en. arXiv:1907.06837 [cs]. DOI: 10.48550/arXiv.1907.06837.

Pardos, Zachary A. and Neil T. Heffernan (2011). "KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model". en. In: *User Modeling, Adaption and Personalization*. Ed. by Joseph A. Konstan et al. Vol. 6787. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 243–254. ISBN: 978-3-642-22361-7 978-3-642-22362-4. DOI: 10.1007/978-3-642-22362-4_21.

Pelánek, Radek (Dec. 2017). "Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques". en. In: *User Model User-Adap Inter* 27.3-5, pp. 313–350. ISSN: 0924-1868, 1573-1391. DOI: 10.1007/s11257-017-9193-2.

Piech, Chris et al. (June 2015). *Deep Knowledge Tracing*. en. arXiv:1506.05908 [cs]. DOI: 10.48550/arXiv.1506.05908.

Psyridou, Maria et al. (June 2024). "Machine learning predicts upper secondary education dropout as early as the end of primary school". en. In: *Sci Rep* 14.1, p. 12956. ISSN: 2045-2322. DOI: 10.1038/s41598-024-63629-0.

Shin, Dongmin et al. (Apr. 2021). "SAINT+: Integrating Temporal Features for EdNet Correctness Prediction". en. In: *LAK21: 11th International Learning Analytics and Knowledge Conference*. arXiv:2010.12042 [cs], pp. 490–496. DOI: 10.1145/3448139.3448188.

References

Siemens, George and Ryan S. J. D. Baker (Apr. 2012). "**Learning analytics and educational data mining: towards communication and collaboration**". en. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Vancouver British Columbia Canada: ACM, pp. 252–254. ISBN: 978-1-4503-1111-3. DOI: 10.1145/2330601.2330661.

Yang, Yang et al. (Sept. 2020). ***GIKT: A Graph-based Interaction Model for Knowledge Tracing***. en. arXiv:2009.05991 [cs]. DOI: 10.48550/arXiv.2009.05991.

Yudelson, Michael V., Kenneth R. Koedinger, and Geoffrey J. Gordon (2013). "**Individualized Bayesian Knowledge Tracing Models**". en. In: *Artificial Intelligence in Education*. Ed. by David Hutchison et al. Vol. 7926. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 171–180. ISBN: 978-3-642-39111-8 978-3-642-39112-5. DOI: 10.1007/978-3-642-39112-5_18.

Zhang, Jiani et al. (Feb. 2017). ***Dynamic Key-Value Memory Networks for Knowledge Tracing***. en. arXiv:1611.08108 [cs]. DOI: 10.48550/arXiv.1611.08108.