Avoid playing learner and system off against each other

Ji-Ung Lee, Christian M. Meyer, Iryna Gurevych

Ubiquitous Knowledge Processing Lab
Technische Universität Darmstadt
https://www.informatik.tu-darmstadt.de/ukp/

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Self-directed language learning is gaining popularity
- Challenges with the increasing number of learners

Hand-crafted exercises
- ❏ Requires experts
- ❏ Limited capacities
- ❏ High-quality exercises

vs

Automatically generated exercises
- ❏ No user interaction required
- ❏ Large capacities
- ❏ Lesser-quality exercises

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Self-directed language learning is gaining popularity
- Challenges with the increasing number of learners

Hand-crafted exercises vs **Automatically generated exercises**

Hand-crafted exercises		Automatically generated exercises
❑ Requires experts	vs	❑ No user interaction required
❑ Limited capacities		❑ Large capacities
❑ High-quality exercises		❑ Lesser-quality exercises

Automatically generate high-quality exercises that

- ❑ **Match a learner's skills and interests**
- ❑ **Allow better system estimates**

Automated Exercise Generation

Automatically generate high-quality exercises that

- Match a learner’s skills and interests
- Allow better system estimates

How to create such an exercise generation system?

- Train an ML model which estimates the suitedness of an exercise according to a given learner profile
- This requires annotated training data consisting of:
  - An input exercise
  - A suitedness annotation for a given learner profile

Data Acquisition Scenarios

Teacher-based data acquisition

- Teachers annotate exercises by assessing them
- Expensive / time-consuming
- Difficult to get a lot of teachers
- Teachers can assess exercises of all proficiency levels

Learner-based data acquisition

- Learners “annotate” exercises by solving them
- Cheap / quick
- Easy to get a lot of learners
- Learners only can give implicit feedback about their own proficiency level (e.g. error-rate)

Data Acquisition Scenarios

Teacher-based data acquisition

- Teachers annotate exercises by assessing them
- Expensive / time-consuming
- Difficult to get a lot of teachers
- Teachers can assess exercises of all proficiency levels

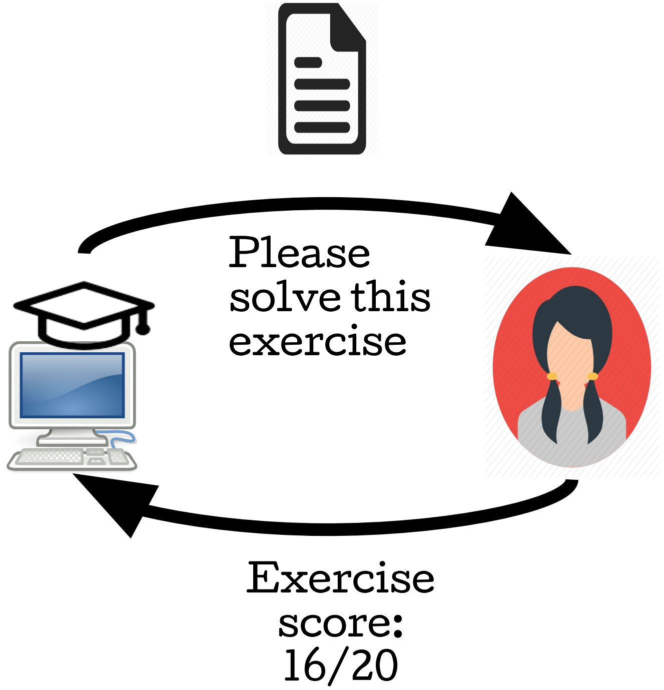
Learner-based data acquisition

- Learners “annotate” exercises by solving them
- Cheap / quick
- Easy to get a lot of learners
- Learners only can give implicit feedback about their own proficiency level (e.g. error-rate)

More likely for crowd-sourcing or online learning

Online Learning Scenario

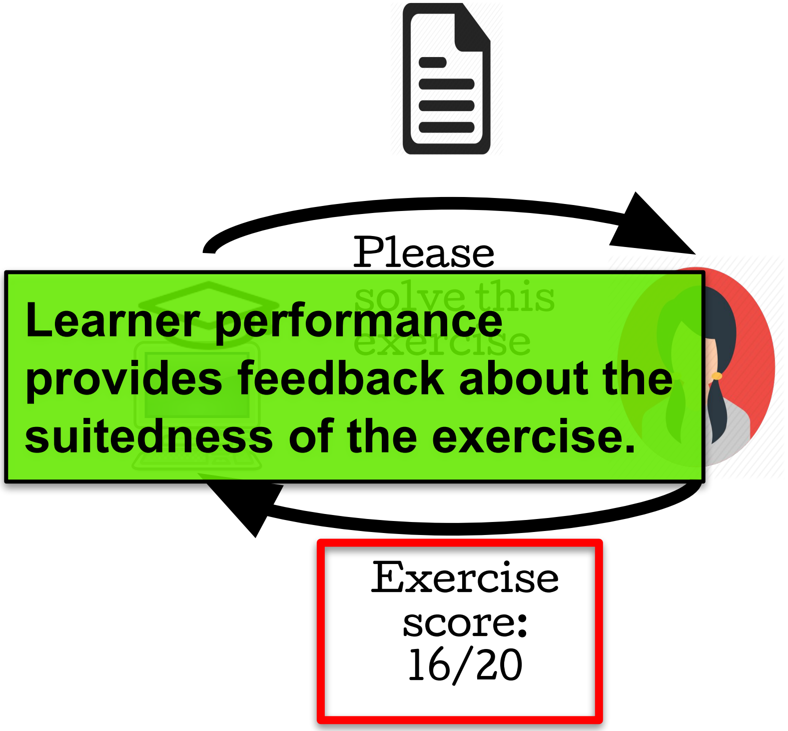
1) Automated tutor generates exercises
2) Learner solves exercises
3) Learner performance evaluation is used to improve the machine learning model
(is the exercise suited?)
4) Re-iterate from 1)



Online Learning Scenario

TECHNISCHE UNIVERSITÄT DARMSTADT

1) Automated tutor generates exercises
2) Learner solves exercises
3) Learner performance evaluation is used to improve the machine learning model (is the exercise suited?)
4) Re-iterate from 1)



Acquisition Bottleneck

TECHNISCHE UNIVERSITÄT DARMSTADT

Requirements for creating (sufficiently) huge datasets

- Big source of unlabeled data (any arbitrary exercise)

- Crowd-sourcing
  - Learners as annotators: **Learnersourcing**
  - Every learner is an expert of their own proficiency

- **Active machine learning**
  - Reduce the number necessary data by sampling intelligently
  - Has been shown to be effective for crowd-sourcing

Acquisition Bottleneck

Requirements for creating (sufficiently) huge datasets

- Big source of un
- Crowd-sourcing
  - Learners as ar
  - Every learner i
- **Active machine**
  - Reduce the nu ... tly
  - Has been shov

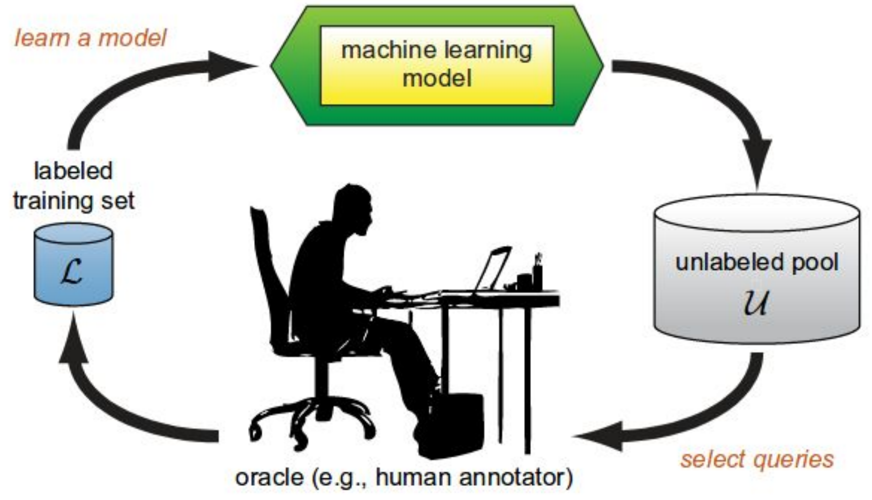


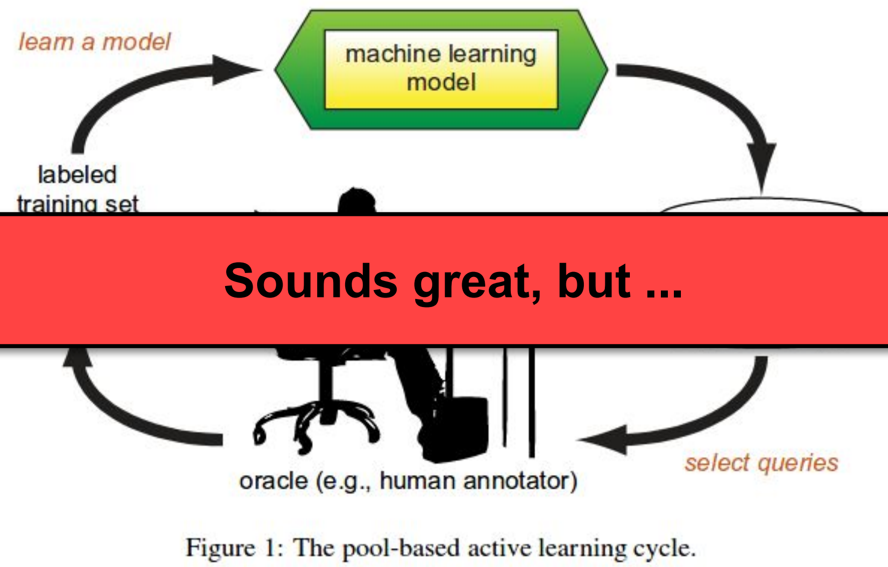
Figure 1: The pool-based active learning cycle.

Active learning in a nutshell (picture from Settles, 2009)

Acquisition Bottleneck

Requirements for creating (sufficiently) huge datasets

- Big source of un
- Crowd-sourcing
  - Learners as
  - Every learne
- **Active machine**
  - Reduce the nu   tly
  - Has been sho



Active learning in a nutshell (picture from Settles, 2009)

Sounds great, but ...

… avoid playing learner and system off against each other

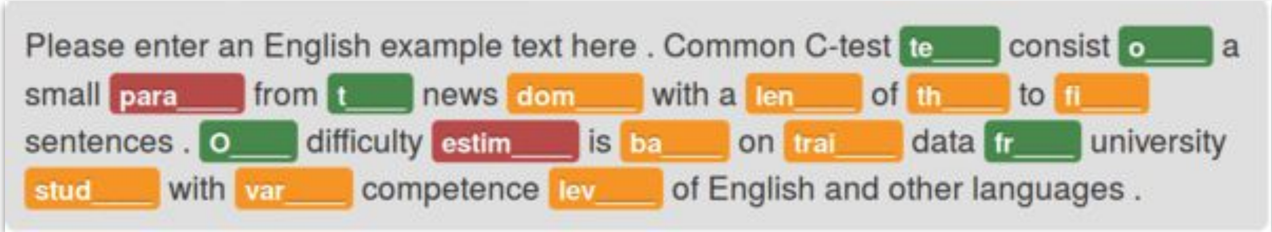
Advantages of Active machine learning

- Active machine learning aims to improve the learning efficiency of the ML **model**
  - Models can be trained more efficiently with less data
  - This reduces the amount of required data

Consequences for crowd-sourcing

- Does not necessarily reduce annotation time
- More difficult annotations lead to more errors of the annotators
  - We may end up **hindering a learner's learning process**

Explicit Use Case: C-tests

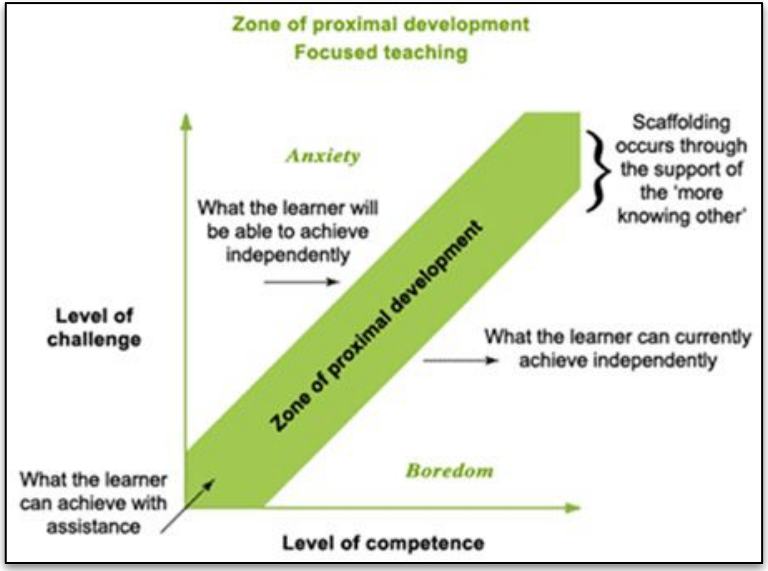


- Proposed by Klein-Braley and Raatz (1982)
- Gap every **second word** in a text by removing the **latter half**
- The first and last sentence have no gaps to provide some context
- Less ambiguous than cloze tests

Explicit Use Case: Suitedness

What is a suitable exercise?

- "Zone of proximal development" (Vygotsky, 1978)
- Guidance zone in which a learner is able to learn optimally
- Measure suitedness implicitly using the learner's error-rate
- **Our goal is to generate exercises fitting into the approximately optimal error-rate**



Previous Work at UKP Lab

Task: Difficulty prediction of C-Tests

- Feature-based approach
- Gap-level difficulty prediction

Data

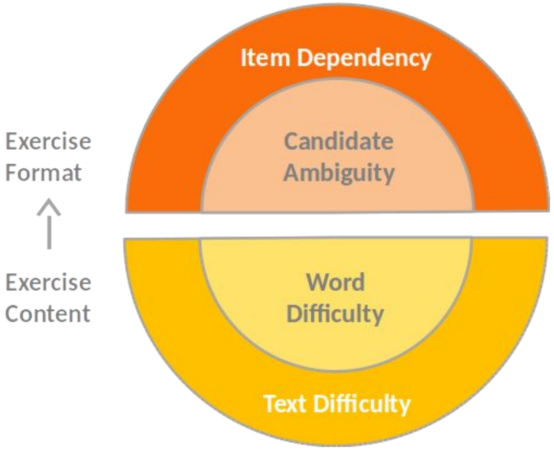
- 77 English C-Tests filled out by over 3,4k participants
- Provided by the Language Center of TU Darmstadt
- Each test has 20 gaps
- Each participant solved 5 tests
- 72 tests for training (1440 gaps)
- 5 tests for testing (100 gaps)

Predicting and manipulating the difficulty of text-completion exercises for language learning (Beinborn, 2016)

TECHNISCHE UNIVERSITÄT DARMSTADT

Features based on four dimensions

- Item dependency
  - Gap difficulty depends on the surrounding gaps
- Candidate ambiguity
  - Inspired from automated solving (Zweig et al, 2012)
- Word difficulty
  - Length, class, singular/plural, …
- Text difficulty
  - Readability features from all linguistic levels (Balakrishna, 2015)



UKP

Task Formalization

Research question

- Can we sample exercises which simultaneously help **learner** and **model**?

Learner objective

- Get exercises suited for their current skill level

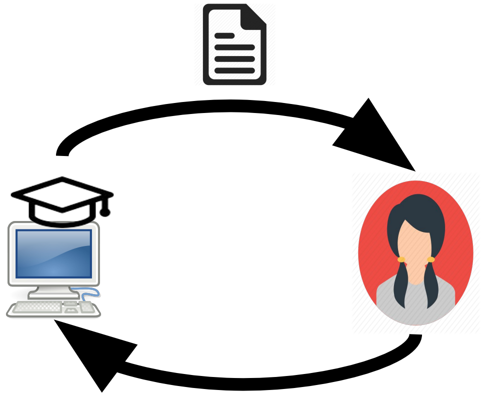
Model objective

- Reduce the number of samples for predicting the gap difficulty

Experimental Set up

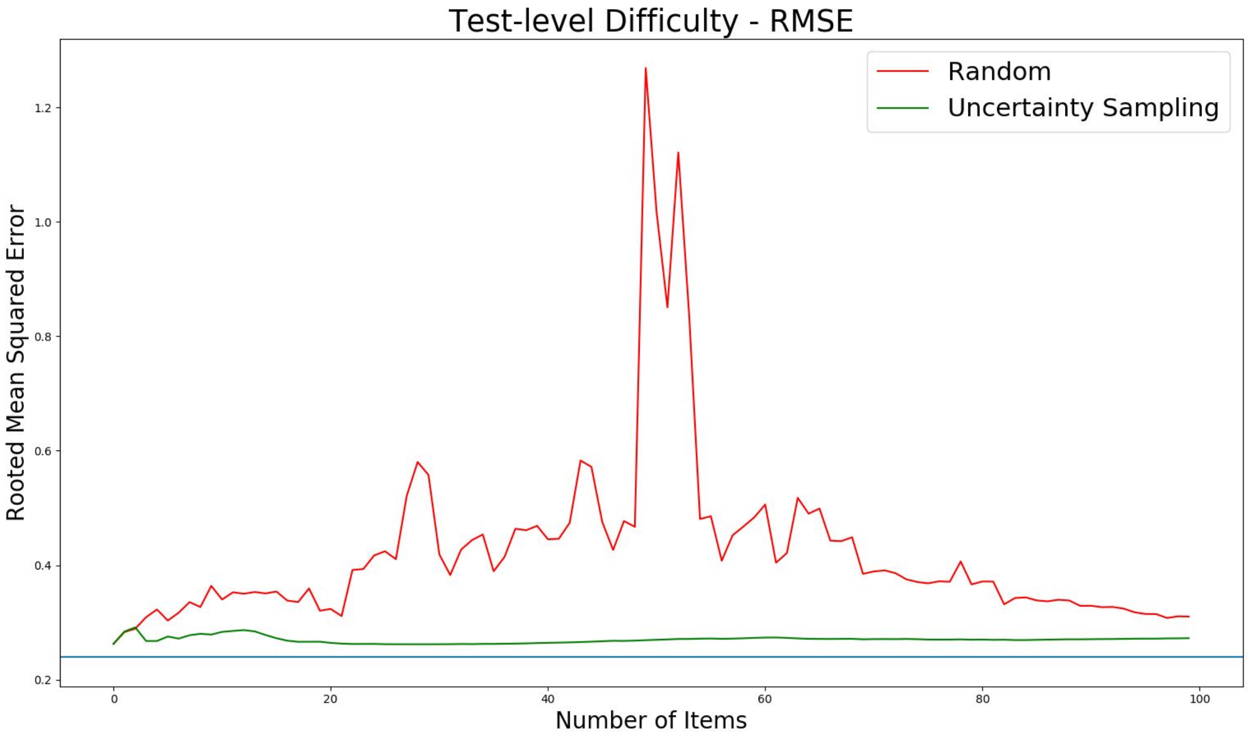
- Linear regression model
- Hand-crafted features (Beinborn, 2016)
- Upper bound performance (trained on all training examples): **0.24** RMSE

- Starts with a single example
- Increase training set by **one example per iteration**
- Sampling strategies:
  - Random sampling (baseline)
  - Uncertainty sampling



Results - Model Objective

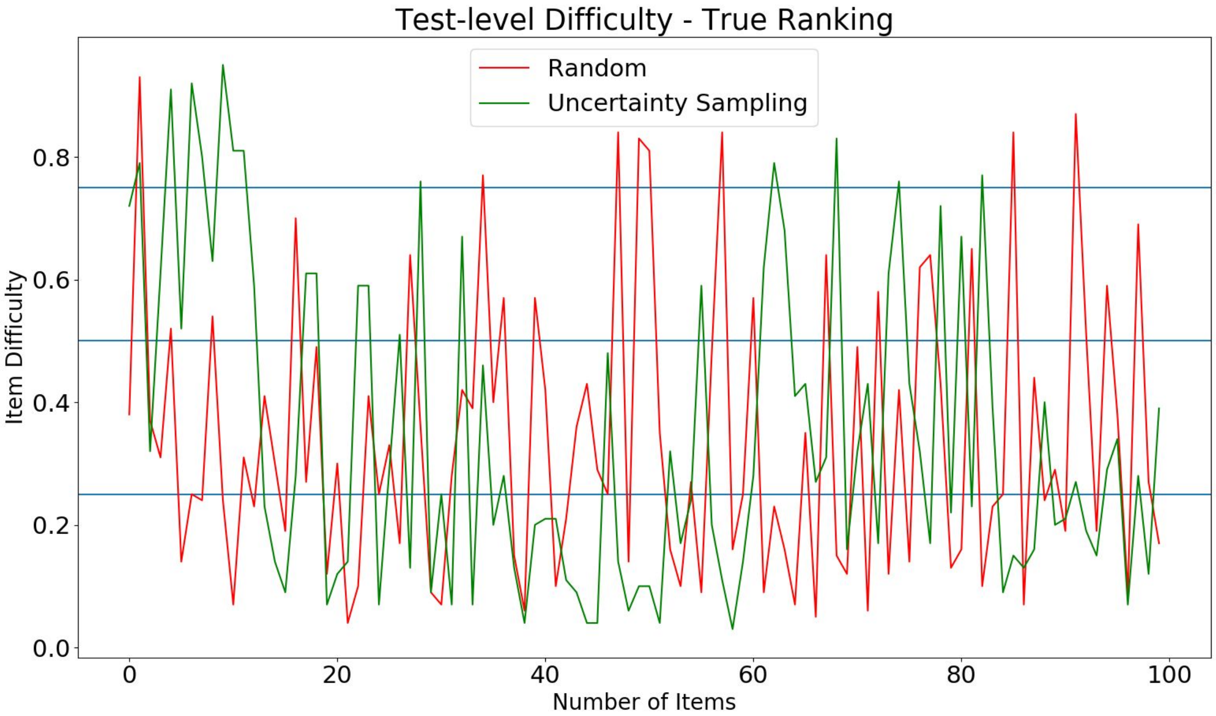
- 100 iterations
- Simple uncertainty sampling already has a positive effect
- Quite close to the upper bound (0.24 RMSE)



Results - Learner Objective

- For both sampling strategies, the ordering of exercise difficulties is nearly random

- This is far from starting with the easiest exercise



Conclusion

TECHNISCHE UNIVERSITÄT DARMSTADT

Advantages

- Learnersourcing may reduce the workload on teachers
- Active machine learning may reduce the amount of required training data

But

- Random sampling and uncertainty sampling both do not care about the learner
- Brings the risk to harm a learner’s learning process

→ **Unethical reduction of learners to mere labelers**

Conclusion

Advantages

- Learnersourcing may reduce the workload on teachers
- Active machine learning may reduce the amount of required training data

But

- Random sampling and uncertainty sampling both do not care about the learner
- Bring

How can we still utilize the benefits of active machine learning?

→ **Unethical reduction of learners to mere labelers**

Ongoing Work

Methods which satisfy learner and model objective

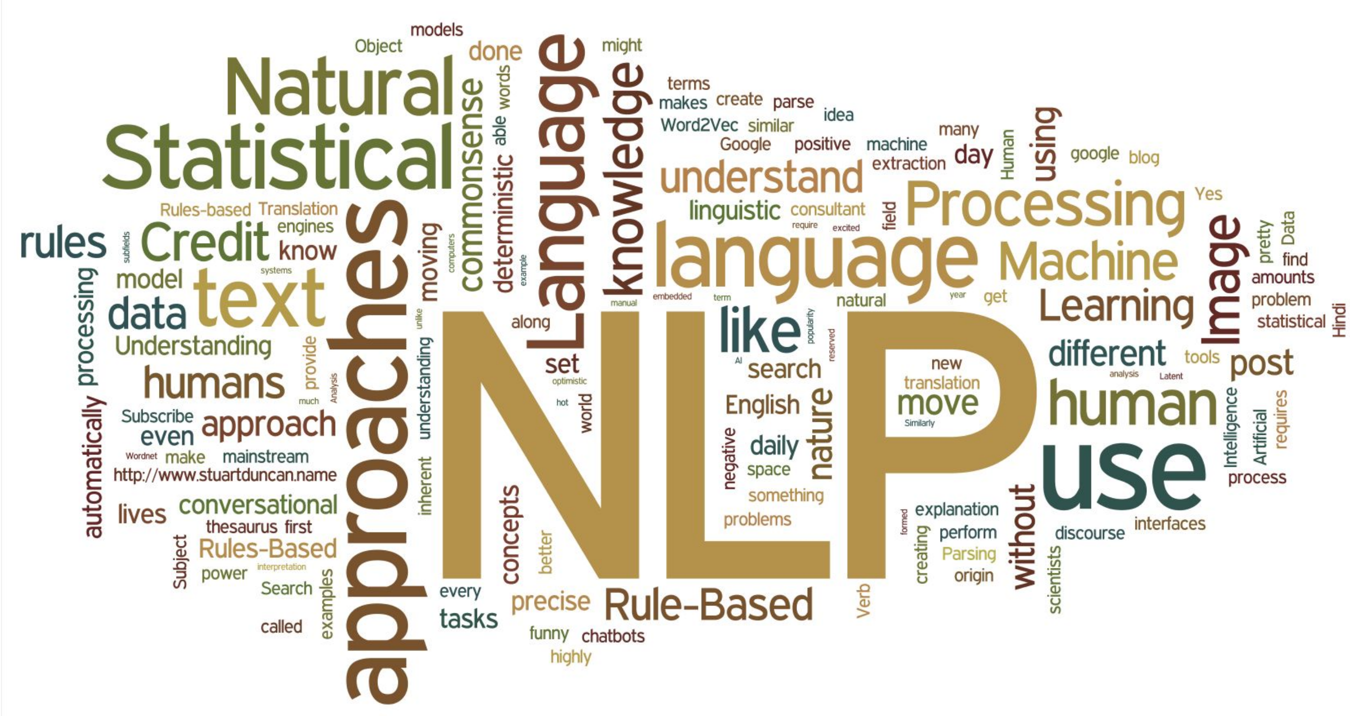
- Improve learner and model in an online learning set up
- This is a difficult challenge

Extension to other use cases

- **Currently:** self-directed language learning
- Means to improve intelligent tutoring systems interactively
- May also be used to train personalized systems (e.g. recommender systems)

Thank you for listening!

TECHNISCHE UNIVERSITÄT DARMSTADT



https://ticary.com/2017/12/12/what-is-nlp.html

UKP

References

[1] Mitkov, R., An Ha, L., and Karamanis, N. (2006). *A computer-aided environment for generating multiple-choice test items.* Natural Language Engineering, 12(2):177–194.

[2] Chinkina, M. and Meurers, D. (2017). *Question Generation for Language Learning: From ensuring texts are read to supporting learning.* In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, Copenhagen, Denmark.

[3] Heffernan, N. T., Ostrow, K. S., Kelly, K., Selent, D., Van Inwegen, E. G., Xiong, X., and Williams, J. J. (2016). *The future of adaptive learning: Does the crowd hold the key?* International Journal of Artificial Intelligence in Education, 26(2):615–644.

References

[4] Yan, Yan, et al. (2011) *Active learning from crowds.* ICML.

[5] Laws, Florian, Christian Scheible, and Hinrich Schütze. (2011) *Active learning with amazon mechanical turk.* Proceedings of the conference on empirical methods in natural language processing. ACL.

[6] Settles, B. (2010) *Active Learning Literature Study*. Computer Science Technical Report. University of Wisconsin. Madison.

[7] Settles, B., Craven, M., and Friedland, L. (2008). *Active learning with real annotation costs.* In Proceedings of the NIPS workshop on cost-sensitive learning, pages 1–10. Vancouver, Canada.

[8] Klein-Braley, C. and Raatz, U. (1982). *Der C-Test: ein neuer Ansatz zur Messung allgemeiner Sprachbeherrschung.* AKS-Rundbrief, 4:23–37.

References

[9] Beinborn, L., Zesch, T., and Gurevych, I. (2014). *Predicting the difficulty of language proficiency tests.* Transactions of the Association of Computational Linguistics, 2(1):517–529.

[10] N. Roy and A. McCallum. (2001) *Toward optimal active learning through sampling estimation of error reduction*. In Proceedings of the International Conference on Machine Learning (ICML), pages 441–448.

Active Machine Learning for Learnersourcing

TECHNISCHE UNIVERSITÄT DARMSTADT

- Standard active machine learning approaches do not care about a learner’s goals

  → **Learners are reduced to mere labelers**

- May sample unfitting exercises for the learner (e.g. too easy, too difficult)

  → **This may harm their learning process**

How can we still utilize the benefits of active machine learning?