TECHNISCHE UNIVERSITÄT WIEN

Anwendung von KI in der Intensivmedizin – eine Chance für eine gendergerechte Intensivmedizin?

Clemens Heitzinger

Center for Artificial Intelligence and Machine Learning (CAIML) and
Department of Computer Science (Informatics),
TU Wien

Cusanus-Akademie Brixen (11 Oct 2024)

Übersicht

Eine kurze Einführung in künstliche Intelligenz und maschinelles Lernen.

Bestärkendes Lernen (reinforcement learning) als neuer Zugang zur personalisierten und prädiktiven Medizin.

Berechnung optimaler Behandlungsstrategien: Theorie.

Anwendungen in der Intensivmedizin:

- optimale Therapie für Sepsispatienten (reinforcement learning),
- Blutdruckvorhersage (supervised learning).

Künstliche Intelligenz

Griechische Mythologie

Bei Plato (im Dialog *Protagoras*) erschaffen die Brüder Prometheus und Epimetheus, zwei Titanen, die irdischen Lebewesen im Auftrag der olympischen Götter.

Epimetheus stattet die Tiere zwar mit allerlei Fähigkeiten aus, vergisst aber auf die Menschen. Daher sieht sich Prometheus gezwungen, den Göttern das Feuer und das technische Wissen (Intelligenz) zu stehlen, um den Menschen das Überleben zu ermöglichen.

LA trop plus saincte et noble creature
Capable plus de hault sens par nature
Et qui sur tout pouoit avoir puissãce

Griechische Mythologie

Epimetheus: der Nachherbedenker, die nachträgliche Einsicht. Büchse der Pandora.

Prometheus: der Vorbedenker, der Vorausdenkende. Intelligenz. Erkundung, Verwertung, Erfahrung, Planung, Bewertung von Aktionen: alles Begriffe des Reinforcement Learning.

Kybernetik und künstliche Intelligenz

Mitte des 20. Jahrhunderts: Das Zeitalter der elektrischen/elektronischen Computer bricht an (im Gegensatz zu mechanischen Rechenmaschinen).

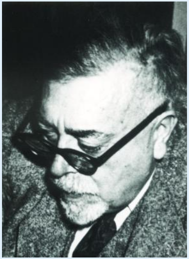
- Norbert Wiener (MIT): *Cybernetics: Or Control and Communication in the Animal and the Machine,* 1948. Kontrolltheorie und theoretische Grundlagen für automatische Navigation, analoge Computer, künstliche Intelligenz und Kommunikation.

  Kybernetik: Kontrolltheorie und kontinuierliche Mathematik.

- John McCarthy : Mitbegründer der AI, Logiker, Lisp, Turing-Preis, etc. Dartmouth College, MIT, Stanford. Dartmouth Workshop.

  Artificial Intelligence / künstliche Intelligenz: Logik und diskrete Mathematik.

Norbert Wiener und John McCarthy



Heutzutage schwingt das Pendel wieder in die Richtung kontinuierlicher Methoden (Deep Learning und Reinforcement Learning).

Dartmouth Summer Research Project on Artificial Intelligence

Forschungsprojekt im Sommer 1956 von John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon am Dartmouth College in Hanover, New Hampshire.

August 1955: Antrag bei der Rockefeller Foundation über 13 500 US-Dollar:

> *“We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer. [. . .]”*

Dartmouth Summer Research Project on Artificial Intelligence

Themen u.a.:

- Computer und Sprache,
- künstliche neuronale Netzwerke,
- Selbstverbesserung,
- Zufälligkeit und Kreativität.

Heutzutage: Deep Learning und Reinforcement Learning verwenden riesige neuronale Netze und hohe Rechenleistung (auf Graphikkarten).

ChatGPT generiert Sprache mit Hilfe von neuronalen Netzwerken.

ChatGPT ist ein selbstlernender Algorithmus.

ChatGPT kann kreativ sein.

Erste Erfolge und Misserfolge: MYCIN

Expertensystem MYCIN Anfang der 1970er Jahr in Stanford.

Identifikation von Bakterien, die schwere Infektionen verursachen, und Empfehlung von Antibiotika (-mycin). 600 Regeln.

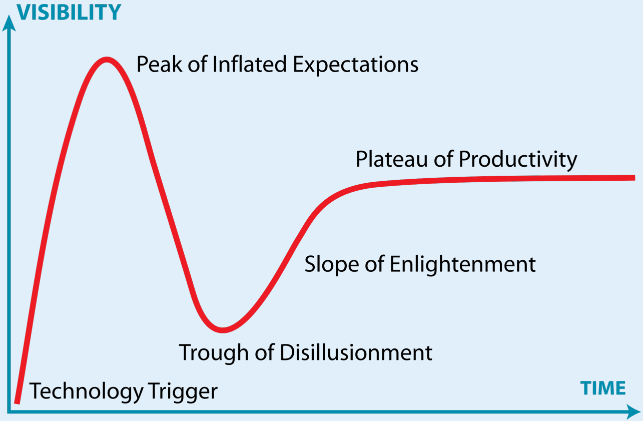
Eingabe: lange Reihe von ja/nein-Fragen.
Ausgabe: Bakterien, ihre Wahrscheinlichkeiten, Konfidenz in Diagnose, verwendete Regeln, empfohlene Antibiotika/Behandlungen.

Acht unabhängige Experten bewerteten die von MYCIN empfohlenen Behandlungen zu 65% als akzeptabel.

Empfohlene Behandlungen von fünf Faculty Member der Stanford Medical School wurden von den Experten zwischen 42.5% und 62.5% als akzeptabel bewertet.

MYCIN wurde nie in der Praxis eingesetzt.

Der *AI Winter* und *Hype Cycles*



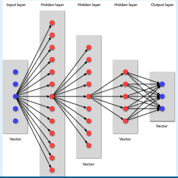
Quelle: Wikimedia.

Deep Learning

Deep Learning im Gegensatz zu flachen neuronalen Netzwerken (vgl. (irreführende) theoretische Resultate).

Mehrere Erfolgsfaktoren für das Lernen:

- enorme Datenmengen,
- enorme Rechenkapazitäten (Cluster von Graphikkarten),
- algorithmische Entwicklungen,
- kommerzielles Interesse von US-amerikanischen Internetfirmen z.B. an der Beschreibung von Szenen in Bildern und Filmen.



Gebiete und Teilgebiete

Artificial intelligence:

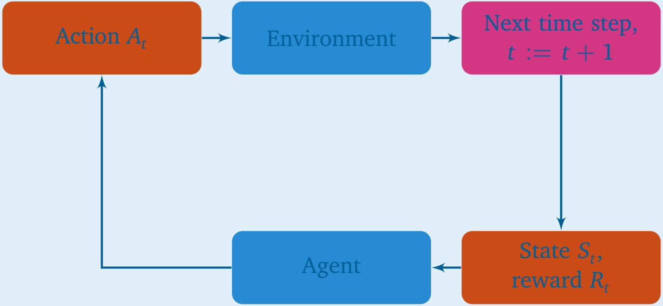
Logik, symbolische Methoden etc. Verwenden diskrete Methoden.

Machine Learning. Verwendet kontinuierliche Methoden (Gradient).

- Supervised Learning:
    - Decision Trees.
    - Random Forests.
    - Artificial Neural Networks:
        - Deep Learning.
        - Transformers.
        - Etc.
    - Etc.
- Unsupervised Learning.
- Reinforcement Learing.

Generative AI: z.B. ChatGPT.

Reinforcement Learning (bestärkendes Lernen): ein äußerst allgemeines Konzept



The goal is to find an optimal policy π (a function of the state) that maximizes the expected value

$$\mathbb{E}_\pi[G_t|S_t = s], \qquad G_t := R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots ,$$

of the return G_t following the policy π.

Reinforcement Learning: Anwendungen

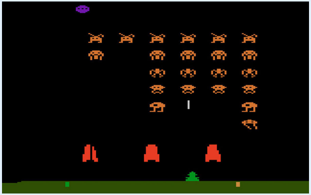
- Backgammon, Computerspiele (Atari 2600), Schach, Go;
- Robotik, autonomes Fahren;
- Industrieautomation (Google Data Centers);
- medizinische Behandlungsstrategien;
- Wertpapierhandel;
- Websites (Optimierungen und Vorschläge, z.B. Nachrichten);
- ChatGPT,
- etc.

Sie haben Reinforcement Learning heute schon indirekt verwendet!

Reinforcement Learning: Google DeepMind: AlphaGo Zero

Offene Herausforderungen zu diesem Zeitpunkt:

- Ein Algorithmus für viele Atari 2600 Spiele.
- Go (sehr viel größerer Suchbaum als Schach).



Quelle: Wikimedia.

Reinforcement Learning: Google DeepMind: AlphaGo Zero

Ein Algorithmus für

- mehr als 50 Atari 2600 Spiele,
- Schach,
- Go.

Zero: ohne Vorwissen (tabula rasa).

Besser als:

- menschlicher Spieletester nach Üben (Atari 2600),
- bestes Schachprogramm (und damit bester Schachspieler),
- bester Gospieler Lee Sedol [Silver et al.: *Mastering the game of Go without human knowledge,* Nature, 2017].

Lee Sedol verlor 1:4 gegen AlphaGo im Jahr 2016 und beendete daraufhin seine Karriere: "[. . .] there is an entity that cannot be defeated."

Probably Approximately Correct

Motivation for PAC (Probably Approximately Correct) Learning/Bounds

Learning faces a fundamental problem:
In general, the samples $X \sim D$ used for learning are random variables.

Although the number of samples used for learning may be large,
it will always be finite.
We can never be sure that we have observed everything (“black-swan events”).

Hence learning results must reflect the randomness (cf. Bayesian estimation).

Cygnus atratus (southeast and southwest Australia). (Attribution: Fir0002/Flagstaffotos.)

The General Form of PAC Results

Two fundamental limitations in learning:

- Unobserved items.
  (Give rise to P in PAC.)
- We may have observed outliers, reducing accuracy.
  (Give rise to A in PAC.)

General form of PAC statements:
The function (i.e., the generalization) selected from a class of learnable functions based on the available samples has a small generalization error (the A in PAC) with high probability (the P in PAC).

Learnable: computable within reasonable (= polynomial) amount of time.

Probably approximately correct (PAC) learning results are among the best that can be expected in the presence of uncertainties and randomness.

The General Form of PAC Results: a Leading Example

A general learning problem: estimate the expected value $\mathbb{E}[X]$ of a random variable X (i.e., $\mathbb{E}[X]$ is the true value) using samples/observations.

Reliable learning: We would also like to bound the absolute error $|\bar{X} - \mathbb{E}[X]|$.

A PAC statement (solution of a learning problem) has the form

$$\forall \delta \in [0,1]: \quad \mathbb{P}[|\bar{X} - \mathbb{E}[X]| < \epsilon(\delta)] > 1 - \delta,$$

where the challenge is to find the function ϵ of the probability δ.

Then we can give confidence intervals: With probability at least $1 - \delta$ (the P in PAC), the absolute value $|\bar{X} - \mathbb{E}[X]|$ of the error is at most $\epsilon(\delta)$ (the A in PAC). Thus the function ϵ gives the size of the (here two-sided) confidence interval around $\bar{X}$ in which the true value $\mathbb{E}[X]$ lies.

Policy Evaluation

Having learned an optimal policy, an independent validation using the validation data is needed.

In particular, lower bounds are needed.

Concentration inequalities are useful for obtaining PAC-like estimates in reinforcement learning.

In reinforcement learning, policy evaluation in the off-policy case is hard.

Q-learning

The state-action value function is defined as

$$q_\pi(s,a) := \mathbb{E}_\pi[G_t \mid S_t = s, A_t = a] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, A_t = a\right].$$

The next approximation Q_{t+1} is defined as

$$Q_{t+1}(s,a) := \begin{cases} (1-\alpha_t)Q_t(s_t,a_t) + \alpha_t(r_{t+1} + \gamma \max_a Q_t(s_{t+1},a)), & (s,a) = (s_t,a_t), \\ Q_t(s,a), & (s,a) \neq (s_t,a_t). \end{cases}$$

Here $\alpha_t \in [0,1]$ is the step size or learning rate and $\gamma \in [0,1]$ denotes the discount factor.

Q-learning

The new value $Q_{t+1}(s_t, a_t)$ can also be written as

$$\underbrace{Q_{t+1}(s_t,a_t)}_{\text{newvalue}} := \underbrace{Q_t(s_t,a_t)}_{\text{oldvalue}} + \alpha_t(\underbrace{r_{t+1} + \gamma \max_{a\in\mathscr{A}(s_{t+1})} Q_t(s_{t+1},a)}_{\text{target}} - \underbrace{Q_t(s_t,a_t)}_{\text{oldvalue}}),$$

which is the form of a semigradient SGD method with a certain function approximation (features = characteristic functions of states).

Speedy Q-learning

Speedy Q-learning is the Q-learning variant

$$\begin{aligned} Q_{t+1}(s,a) &:= Q_t(s,a) + \alpha_t\big(\mathscr{B}_t Q_{t-1}(s,a) - Q_t(s,a)\big) \\ &\qquad + (1-\alpha_t)\big(\mathscr{B}_t Q_t(s,a) - \mathscr{B}_t Q_{t-1}(s,a)\big), \\ \alpha_t &:= \frac{1}{t+1}, \\ \mathscr{B}_t Q_t(s,a) &:= R_{t+1} + \gamma \max_a Q_t(S_{t+1},a). \end{aligned}$$

Here $\mathscr{B}_t$ is the (usual) empirical Bellman operator. Note that it is applied to $Q_t(s,a)$ and to $Q_{t-1}(s,a)$.

[M. Ghavamzadeh et al.: Speedy Q-learning, NIPS 2011, 2411–2419.]

Distributional Reinforcement Learning

In distributional reinforcement learning, the return distribution $Z^{\pi}(s,a)$

$$Z_{\pi}(s,a) := \sum_{t=0}^{\infty} \gamma^{t} R_{t}$$

$(s,a) \in \mathscr{S} \times \mathscr{A}$, is the sum of discounted rewards following the policy π starting in state s and taking a.

The probability distribution of the return is calculated, not only its expected value.

True generalization: The usual state-action value function Q_{π} is recovered as

$$Q_{\pi}(s,a) := \mathbb{E}[Z_{\pi}(s,a)].$$

Advantage: Risk is quantified. Disadvantage: loss of uniqueness.

Distributional Reinforcement Learning

We use the Cramér distance

$$\begin{aligned}\bar{\ell}_2(\eta,\xi) &:= \sup_{(s,a)\in\mathcal{S}\times\mathcal{A}} \ell_2(\eta^{(s,a)},\xi^{(s,a)})\\ &= \sup_{(s,a)\in\mathcal{S}\times\mathcal{A}} \left(\int_{\mathbb{R}} |F_{\eta^{(s,a)}}(z)-F_{\xi^{(s,a)}}(z)|^2 \mathrm{d}z\right)^{1/2},\end{aligned}$$

where η and ξ are return distributions. Then the composition

$$\Pi_{\mathcal{C}}\mathcal{B}^{\pi}: \quad \mathcal{P}_z^{\mathcal{S}\times\mathcal{A}} \to \mathcal{P}_z^{\mathcal{S}\times\mathcal{A}}$$

of the categorical projection $\Pi_{\mathcal{C}}$ and the Bellman operator $\mathcal{B}^{\pi}$ is a $\sqrt{\gamma}$-contraction w.r.t. $\bar{\ell}_2$ on the categorical (discretized) distributions $\mathcal{P}_z$.

Almost sure convergence $\lim_{k\to\infty}\bar{\ell}_2(\eta_k,\eta_{\mathcal{C}})=0$ given Robbins-Monro conditions.

[M. Rowland et al.: an Analysis of Categorical Distributional Reinforcement Learning, arXiv:1802.08163 (stat.ML), 2018]

A PAC Result in Distributional Reinforcement Learning: Assumptions

The speedy Q-learning update rule translated to distributional RL gives the return distributions

$$\eta_{k+1}^{(s,a)} := \eta_k^{(s,a)} + \alpha_k(\Pi_{\mathscr{C}} \mathscr{B}_k^\pi \eta_{k-1}^{(s,a)} - \eta_k^{(s,a)}) + (1-\alpha_k)(\Pi_{\mathscr{C}} \mathscr{B}_k^\pi \eta_k^{(s,a)} - \Pi_{\mathscr{C}} \mathscr{B}_k^\pi \eta_{k-1}^{(s,a)}). \tag{1}$$

Assumption

The state-action space is finite with $n := |\mathscr{S} \times \mathscr{A}|$ elements. The categorical distribution $\eta_{\mathscr{C}}$ is the unique fixed point of $\Pi_{\mathscr{C}} \mathscr{B}^\pi$. The rewards are bounded by $R_{\max} > 0$. The discount factor γ is smaller than 1. Let $\bar{\beta} := 1/(1-\sqrt{\gamma})$. Let $V_{\max} := \frac{1}{1-\gamma} R_{\max}$ be the maximal attainable return. For the N fixed atoms we assume $z_1 = -V_{\max}$ and $z_N = V_{\max}$. Finally, the two initial return distribution functions are equal, i.e., $\eta_{-1} = \eta_0$, and the η_k are obtained by update rule (1).

A PAC Result in Distributional Reinforcement Learning: Main Result

Theorem

Under these Assumptions, the inequality

$$\bar{\ell}_2(\eta_{\mathscr{C}}, \eta_k) \leq \sqrt{2V_{\max}}\bar{\beta}\left(\frac{\sqrt{\gamma}}{k} + \sqrt{\frac{2\log\frac{2nN}{\delta}}{k}}\right)$$

holds for all time steps $k \geq 1$ with probability at least $1 - \delta$.

Corollary: almost sure convergence in $\bar{\ell}_2$.

[M. Böck, CH: Speedy Categorical Distributional Reinforcement Learning and Complexity Analysis, *SIAM Journal on Mathematics of Data Science (SIMODS)*, 4(2):675–693, 2022]

Sketch of the Proof

1. Stability. We show by induction that the $\eta_k^{(s,a)}$ are indeed probability measures.
2. Error martingale. Turn errors into a martingale.
3. Prove upper bounds for the difference by induction.
4. Bounding the error in probability. Maximal Hoeffding-Azuma inequality applied pointwise at atoms.
5. Finally, find suitable δ as inverse function of ϵ in maximal Hoeffding-Azuma inequality.

1. Stability

Stability (Lemma 4.3): We show by induction that the $\eta_k^{(s,a)}$ are indeed probability measures.

The proof is by induction. The main observation is to consider the sample update and to use induction.

In the induction step, the new sample update is equal to the categorical projection and the Bellman operator applied to the old sample update.

2. Error Martingale

Turn errors into a martingale.

Lemma 4.4: For each atom z_i, the cumulative distribution functions of the error ϵ_k evaluated at the atom z_i are a uniformly bounded martingale difference sequence.

This allows to extend the analysis of speedy Q-learning to categorical distributions.

3. Upper Bounds for the Difference by Induction

Lemma 4.6: For all $k \geq 1$, the inequalities

$$\|\eta_{\mathscr{C}} - \eta_k\|_{\bar{\ell}_2} \leq \frac{\sqrt{\gamma}\bar{\beta}}{k}\sqrt{2V_{\max}} + \frac{1}{k}\sum_{j=1}^{k}\sqrt{\gamma}^{k-j}\left\|E_{j-1}\right\|_{\bar{\ell}_2}$$

hold.

E_{j-1} ... cumulative error to the sample update.

Proof by induction.

4. Bounding the Error in Probability

Apply the maximal Hoeffding-Azuma inequality pointwise at the atoms (i.e., discretization points).

Lemma (Maximal Hoeffding-Azuma Inequality)

Let $\mathscr{V} := \{V_1, \ldots, V_T\}$ *be a martingale difference w.r.t. to the filtration* $\mathscr{F}_k$ $(\mathbb{E}[V_k | \mathscr{F}_{k-1}] = 0)$ *such that* $\mathscr{V}$ *is uniformly bounded by* $L > 0$*. Then for any* $\epsilon > 0$*, the inequality*

$$\mathbb{P}\left[\max_{1 \le k \le T} \left| \sum_{i=1}^{k} V_i \right| > \epsilon \right] \le 2 \exp\left(\frac{-\epsilon^2}{2TL^2} \right)$$

holds.

Here the filtration is the σ-field generated by the experience.

4. Bounding the Error in Probability

This yields:

Lemma

For all $\epsilon > 0$ and all time steps T, the inequality

$$\mathbb{P}\left[\max_{1\leq k\leq T} \|E_{k-1}\|_{\bar{\ell}_\infty} > \epsilon\right] \leq 2nN \exp\left(\frac{-\epsilon^2}{2T}\right) =: \delta$$

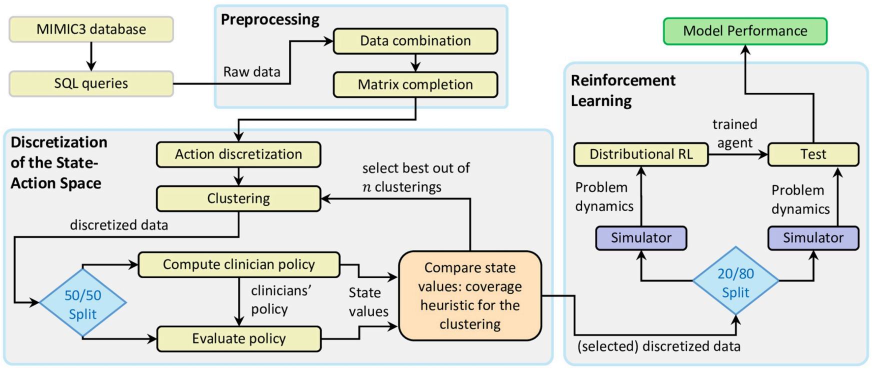
holds.

5. Find $\epsilon(\delta)$

Combine inequalities above and find $\epsilon(\delta)$.

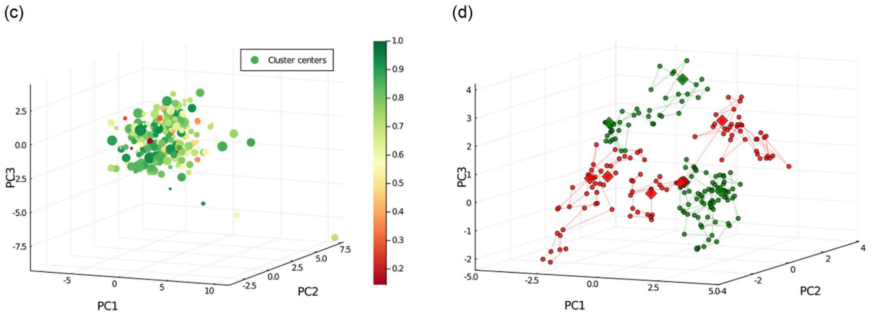
RL in Intensive-Care Units: Sepsis

Data Flow



[CH et al.: Superhuman performance for sepsis treatment by distributional reinforcement learning, PLOS ONE, 2022, DOI: 10.1371/journal.pone.0275358]

Cluster Centers and Sample Episodes



[CH et al.: Superhuman performance for sepsis treatment by distributional reinforcement learning, PLOS ONE, 2022, DOI: 10.1371/journal.pone.0275358]

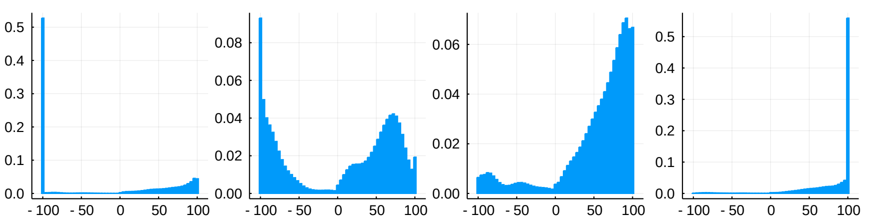
Superhuman Performance in Sepsis Treatment

90-day mortality, $n = 10\,000$.

	Mean reward	Recovery rate
Clinician	47.47	85.41 %
Agent (stochastic policy)	50.83	**88.76 %**

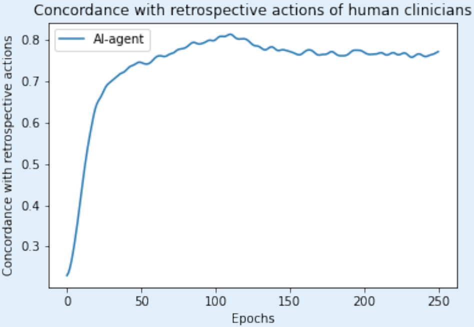
[CH et al.: Superhuman performance for sepsis treatment by distributional reinforcement learning, PLOS ONE, 2022, DOI: 10.1371/journal.pone.0275358]

Superhuman Performance in Sepsis Treatment: Distributional Reinforcement Learning



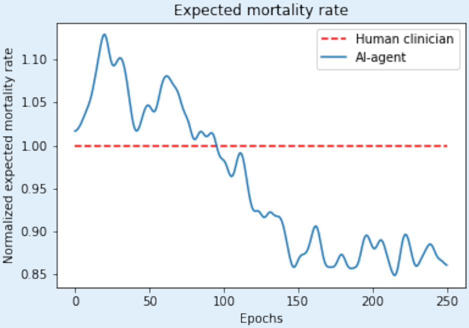
[CH et al.: Superhuman performance for sepsis treatment by distributional reinforcement learning, PLOS ONE, 2022, DOI: 10.1371/journal.pone.0275358]

A Numerical Result for Policy Evaluation: Concordance with Retrospective Actions



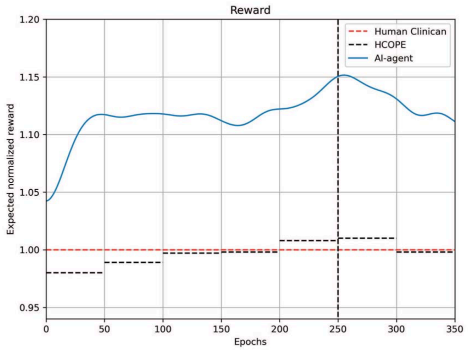
[CH et al.: Development of a Reinforcement Learning Algorithm to Optimize Corticosteroid Therapy in Critically Ill Patients with Sepsis, *J. of Clinical Medicine,* 2023, DOI: 10.3390/jcm12041513]

A Numerical Result for Policy Evaluation: Mortality



[CH et al.: Development of a Reinforcement Learning Algorithm to Optimize Corticosteroid Therapy in Critically Ill Patients with Sepsis, *J. of Clinical Medicine,* 2023, DOI: 10.3390/jcm12041513]

A Numerical Result for Policy Evaluation: 95% One-Sided Confidence Intervals for Reward



[CH et al.: Development of a Reinforcement Learning Algorithm to Optimize Corticosteroid Therapy in Critically Ill Patients with Sepsis, *J. of Clinical Medicine,* 2023, DOI: 10.3390/jcm12041513]

Intensive-Care Units: Blood Pressure

Prediction of Low Blood Pressure (Hypotension)

During surgery, intraoperative hypotension is associated with postoperative morbidity and should therefore be avoided.

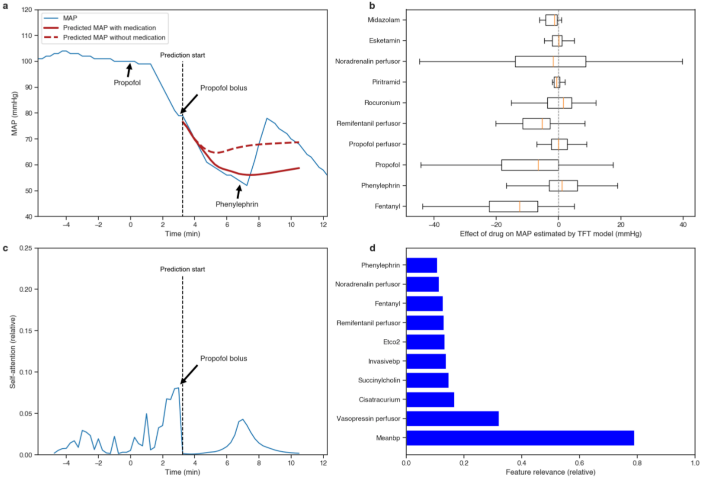
Predicting the occurrence of hypotension in advance may allow timely interventions to prevent hypotension.

We utilized a novel temporal fusion transformer (TFT) algorithm to predict intraoperative blood pressure trajectories seven minutes in advance.

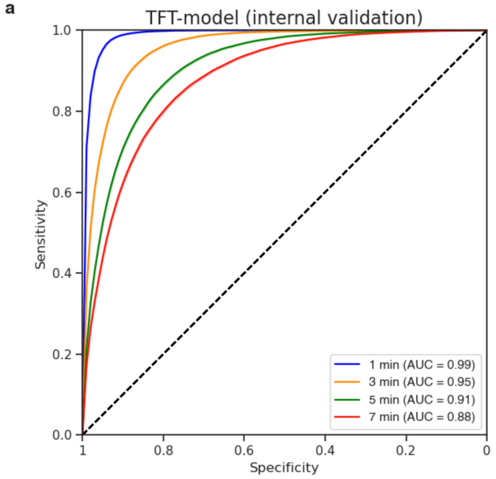
[CH et al.: Temporal fusion transformer models for continuous intraoperative blood pressure forecasting – development and external validation, The Lancet eClinicalMedicine, in print] (5-year impact factor: 9.9)

TFT: [Lim, Bryan, et al.: Temporal fusion transformers for interpretable multi-horizon time series forecasting, International Journal of Forecasting, 17481764, 37:4, 2021]

Prediction of Blood Pressure



Validation (AUC)



Zusammenfassung

Ziel schon seit langer Zeit: personalisierte und prädiktive Medizin.

Wir verfolgen einen in der Medizin neuen Zugang, der optimale Behandlungsstrategien berechnet.

Die Behandlungsstrategien sind vollständig personalisiert und ausgezeichnet interpretierbar. Kein Umweg über die Sprache (large language models).

Entscheidend sind

- die Qualität der Algorithmen,
- die Qualität der Daten.

KI-System werden wohl vermehrt zur Entscheidungsunterstützung eingesetzt werden. Gesammelte Erfahrung.

Thank you!

Email:
`Clemens.Heitzinger@TUWien.ac.at`

Homepage (with list of publications):
`http://Clemens.Heitzinger.name`

CAIML (TU Wien Center for AI and ML):
`http://caiml.org`

Text book *Algorithms with Julia,* Springer, 2022:
`link.springer.com`

Lecture notes on *Reinforcement Learning* (ca. 200 pages)