Maschinelles Lernen für sicherheitsrelevante Entscheidungen in der Eisenbahnautomatisierung?

Machine learning for safety related decisions in railway automation?

Jens Braband

Trotz des großen Hypes um Künstliche Intelligenz (KI) gibt es noch keine Variante, die komplett automatisiert sicherheitsrelevante Entscheidungen treffen darf. Zwar gibt es teilweise Fortschritte in der Automobiltechnik, diese betreffen aber entweder Assistenzsysteme oder sehr eingeschränkte Betriebsszenarien z. B. auf der Autobahn bei geringer Geschwindigkeit und gutem Wetter. Für die Eisenbahnautomatisierung ist es aber wichtig, Algorithmen zu haben, die sicherheitsrelevante Entscheidungen unter allen Randbedingungen automatisiert treffen dürfen, z. B. weil dies mit konventioneller Technik schon erlaubt ist.

Die populären Techniken der KI, die Verfahren zum Maschinellen Lernen (ML) sind einerseits sehr komplex und schwer erklärbar, andererseits sind Daten oft auch sehr hochdimensional und die Umgebungsbedingungen schwierig, sodass häufig mehrfache Probleme beim Einsatz von ML vorliegen und schwer erkennbar ist, welche die wesentlichen Probleme darstellen.
Daher wurde ein Versuch gemacht, nach dem Einstein zugeschriebenen Motto „Man muss die Dinge so einfach wie möglich machen. Aber nicht einfacher" eine Aufgabenstellung zu finden, die einerseits einfach ist, andererseits aber auch nicht trivial, sodass die wesentlichen Probleme der ML-Anwendung für sicherheitsrelevante Entscheidungen deutlich werden, und gleichzeitig störende Randeffekte minimiert werden. Dabei wurde schnell klar, dass es sich um eine künstliche Aufgabenstellung mit bekanntem Ergebnis handeln muss, insbesondere auch, um die Qualität der Ergebnisse eindeutig bewerten zu können.

1 Ein einfaches Entscheidungsproblem

Die Wahl fiel dabei schnell auf ein einfaches Entscheidungsproblem, wie es in den meisten Lehrbüchern zur KI, z. B. [1], schon auf den ersten Seiten vorkommt und das jedem Experten wohlbekannt ist, da meistens die Grundprinzipien daran erläutert werden: Es liegt ein Datensatz mit einer zusätzlichen binären Klassifikation, d. h. einem Label, vor, z. B. „Rot" für ein haltzeigendes Signal und „Grün" für ein fahrtzeigendes Signal. Dies ist das einfachste Problem einer Signalerkennung, z. B. an einer Ampel oder einem einfachen Eisenbahnsignal. In der Regel sind die Daten, z. B.Bilddaten, hochdimensional, aber durch Extraktion von Features kann man bei einfachen Entscheidungsproblemen oft die Dimension reduzieren. Im vorgeschlagenen Beispiel wird in den Lehrbüchern nur von zwei Dimensionen ausgegangen, dies hat den Vorteil, dass man die Daten einfach visualisieren kann.

Despite the great deal of hype about artificial intelligence (AI), there is still no variant that can make completely automated safety-relevant decisions. Although there have been some advances in automotive technology, these concern either assistance systems or very limited operating scenarios, e.g. on motorways at low speeds and good weather. For railway automation, however, it is important to have algorithms that are allowed to make safety-relevant decisions automatically under all the boundary conditions, i.e. because this is already permitted with conventional technology.

On the one hand, the popular AI techniques, i.e. machine learning (ML) methods, are highly complex and difficult to explain, while on the other hand the data is often very high dimensional and the environmental conditions complex, so that multiple problems often arise during the use of ML and it is difficult to recognise what the main problems are.
An attempt was therefore made to find a definition of the task that is simple, but also not trivial, in line with the motto attributed to Einstein, "Things have to be made as simple as possible. But not simpler", so that the essential problems of the ML application for safety-relevant decisions would become clear and any disturbing side effects would be minimised at the same time. It quickly became clear that this had to be a synthetic task with a known result, especially in order to be able to clearly evaluate the quality of the results.

1 A simple decision problem

A simple decision problem was quickly chosen. It is one that appears on the first pages of most AI textbooks, e.g. [1], and which is well-known to every expert, since the basic principles are usually explained by the following problem: there is a data set with an additional binary classification, i.e. a label such as "red" for a stop signal and "green" for a go signal. This is the simplest type of signal detection problem, such as at a traffic light or a simple railway signal. The data, i.e. images, is typically high dimensional, but the dimension can often be reduced for simple decision-making problems by extracting the features. Only two dimensions in the proposed example are assumed in the textbooks. This has the advantage that the data can be easily visualised.
Fig. 1 shows an example of just such a data set with colours representing the classification. This means that, if a new unclassified point is added, the question arises: "Should the new point

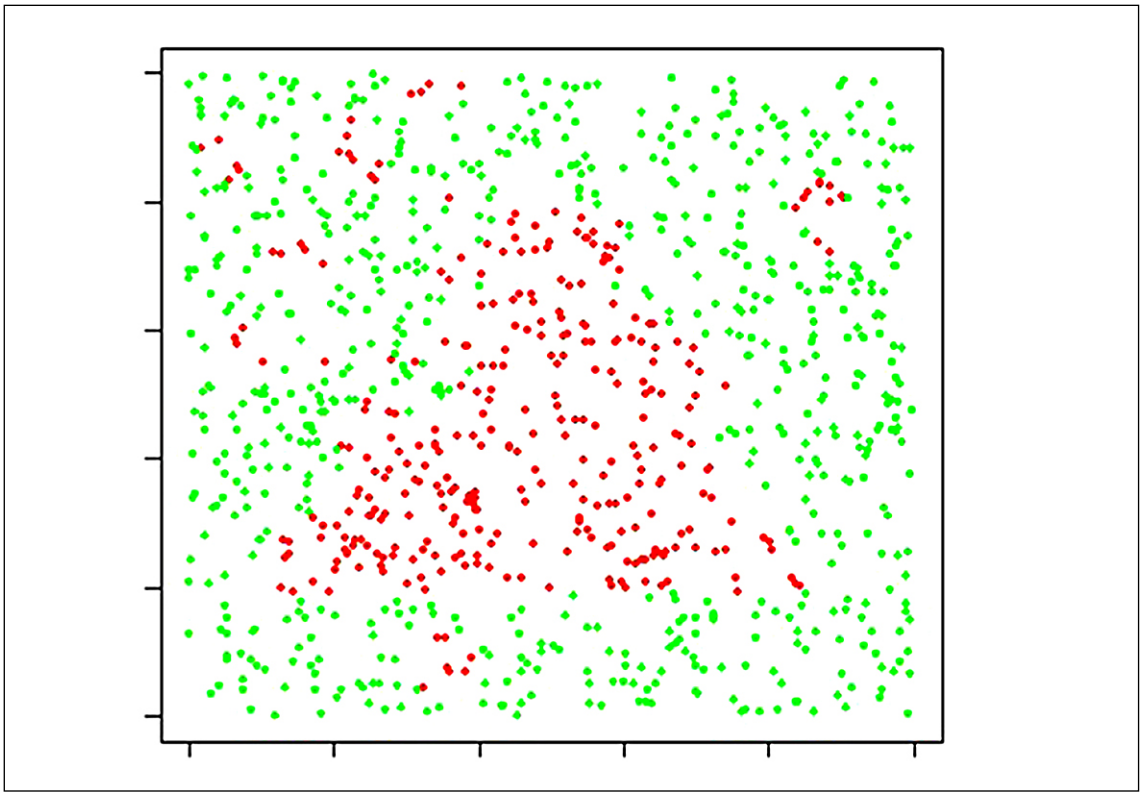

Bild 1: Beispiel-Datensatz
Fig. 1: Example Data Set

Bild 1 zeigt ein Beispiel für einen solchen Datensatz, wobei die Farbe die Klassifikation darstellt. Wenn jetzt ein neuer umklassifizierter Punkt dazukommt, stellt sich die Frage: „Sollte der neue Punkt rot oder grün klassifiziert werden?". Generell können dann zwei Fehler gemacht werden:

- Wird ein eigentlich roter Punkt als grün klassifiziert, so haben wir ein Sicherheitsproblem, denn ein Fahrzeug würde jetzt am eigentlich haltzeigenden Signal vorbeifahren dürfen.
- Wird ein eigentlich grüner Punkt als rot klassifiziert, so gibt es ein Verfügbarkeitsproblem, denn das Fahrzeug stoppt am eigentlich fahrtfreigebenden Signal.

Der Vorteil ist, dass man bei einem solchen synthetischen Datensatz alle Parameter in der Hand hat, d. h. man kennt die „Wahrheit" (ground truth) und kann auch einige Probleme, die sich sonst in praktischen Anwendungen ergeben, einfach ausblenden:

- Die Daten stehen in beliebiger Anzahl zur Verfügung, sowohl zum Training als auch zur Validierung.
- Alle Daten sind korrekt klassifiziert.
- Die Daten sind identisch und unabhängig verteilt, d. h. sie sind repräsentativ, und die Verteilung ändert sich nicht mit der Zeit, z. B. weil sich die Umgebung verändert.
- Es wurde angenommen, dass die Datensätze separabel sind, d. h. es gibt eine ideale Trennung der Bereiche.

Damit kann man sich auf die folgenden wesentlichen Aufgaben konzentrieren:

- Unter welchen Annahmen kann man eine verlässliche Abschätzung für die Fehlklassifikationswahrscheinlichkeit erhalten?
- Mit welcher Konfidenz kann man die Fehler abschätzen?
- Wie kann man die Sicherheitsargumentation führen?

Wenn man diese Aufgaben selbst für diese einfache Aufgabenstellung nicht so lösen könnte, dass ein Einsatz in sicherheitsrelevanten Anwendungen möglich ist, dann müsste man den Einsatz von ML generell für solche Anwendungen in Frage stellen. Man könnte dann allerdings genau eingrenzen, warum man gescheitert ist.

be classified as red or green?". In general, two mistakes can be made:

- If a red dot is classified as green, we have a safety problem, because a vehicle will be allowed to drive past the signal that actually shows a stop.
- If a green dot is classified as red, there is an availability problem, because the vehicle will stop at a signal that is actually releasing it for travel.

The advantage lies in the fact that one has all the parameters to hand with such a synthetic data set, i.e. the "truth" (ground truth) is known and some problems that otherwise arise in practical applications can also be easily dismissed:

- the data is available in any sample size, both for training and validation
- all the data is correctly labelled
- the data is identically and independently distributed, i.e. it is representative and the distribution does not change over time, e.g. because the environment changes
- it was assumed that the data sets were separable, i.e. there was an ideal separation of the areas.

This made it possible to concentrate on the following essential tasks:

- Under what assumptions can a reliable estimate of the probability of misclassification be obtained?
- With what degree of confidence can the errors be estimated?
- How can the safety argument be made?

If even such a simple task was unable to be solved in this way so that it could be used in safety-relevant applications, it would prove necessary to question the use of ML for such applications in general. However, it would then be possible to narrow down exactly why it had failed. If, however, the task was able to be solved, it would still be necessary to resolve the problems dismissed above for concrete applications, but it would be a start.

tert ist. Wenn man diese Aufgabe lösen könnte, müsste man allerdings für konkrete Anwendungen noch die oben ausgeblendeten Probleme lösen, aber ein Anfang wäre gemacht.

2 Der Wettbewerb

Natürlich müsste man jetzt dieses Problem mit möglichst vielen Methoden, unter verschiedenen Annahmen etc. angehen. Um hier eine möglichst große Diversität zu erreichen und den Aufwand begrenzt zu halten, hat sich die Siemens Mobility GmbH entschieden, diese Aufgabe als Studentenwettbewerb offen auszuschreiben. Der sogenannte „AI – Dependability Assessment – Student's Challenge" (AI-DA-SC) genannte Wettbewerb lief vom 15. Februar 2021 bis zum 16. Juli 2021. Teilnahmeberechtigt waren Teams von bis zu drei Studierenden, die noch keinen Doktortitel erworben hatten.

Die Ausschreibung und Durchführung des Wettbewerbs erfolgte über eine Webseite ([2], siehe Link zur Website weiter unten), auf der drei Datensätze unterschiedlichen Umfangs und unterschiedlichen Charakters bereitgestellt wurden. Die Teams sollten als Wettbewerbsbeitrag eine Ausarbeitung mit dem gewählten Modell, der Abschätzung der Fehlerrate sowie einem Sicherheitsargument einreichen. Dabei wurde ein eher moderates Sicherheitsziel von 1:1000 für eine sicherheitsgerichtete Fehlklassifikation vorgegeben, da gleichzeitig auch die Anzahl der Datensätze relativ klein gehalten wurde (1000-50 000). Dies entspricht einerseits eher der typischen Größe von Datensätzen in der Eisenbahnautomatisierung, andererseits sollten die Berechnungen auch ohne Spezialhardware ausgeführt werden können.

Diese Ausarbeitungen wurden von einer Expertenjury bewertet, die aus

- Prof. Dr.-Ing. Corinna Salander (Leiterin des DZSF, Honorarprofessorin Universität Stuttgart),
- Prof. Dr. Martin Fränzle (Carl-von-Ossietzky-Universität Oldenburg),
- Prof. Dr. Volker Tresp (Siemens AG, Honorarprofessor LMU München)

sowie dem Autor bestand. Jede Einreichung wurde von mindestens drei (teilweise zusätzlichen) Experten bewertet. Anschließend erhielten die Teams zusätzliche, unklassifizierte Validierungsdatensätze, anhand derer die Fehlerabschätzungen praktisch überprüft wurden.

Insgesamt wurden 32 Arbeiten von 18 Universitäten aus 15 Ländern eingereicht. Diese Ausarbeitungen deckten nicht nur praktisch alle bekannten ML-Modelle ab, von statistischen Ansätzen bis zu Neuronalen Netzwerken, sondern auch Kombinationen von Modellen oder sogar eigene innovative Ansätze.

Aufgrund dieser Bandbreite beschloss die Jury, drei erste Preise zu verteilen, und zwar für

- den besten Original-Beitrag,
- die beste Balance zwischen Sicherheit und Verfügbarkeit,
- das beste praktische Ergebnis

sowie fünf zweite Preise.

3 Die Ergebnisse

Um es gleich vorwegzunehmen: Kein Team, auch die Gewinner nicht, konnte eine theoretische Abschätzung abgegeben, mit der das Sicherheitsziel erreicht werden konnte. Hinzu kam, dass in über der Hälfte der Fälle die praktische Validierung schlechter ausgefallen ist als die eigene theoretische Abschätzung. Praktisch konnten nur wenige Teams das Sicherheitsziel erreichen oder ihm sehr nahekommen, einige davon aber nur auf Kosten einer inakzeptablen Verfügbarkeit. Bild 2 illustriert dies für einen Datensatz:

2 The competition

Of course, this problem would now have to be tackled with as many methods as possible, under different assumptions, etc. In order to achieve a wide degree of diversity and to limit the effort, Siemens Mobility GmbH decided to openly advertise this problem as a student competition. The so-called "AI – Dependability Assessment – Student's Challenge" (AI-DA-SC) competition ran from 15th February 2021 to 16th July 2021. Teams of up to three students who had not yet obtained a doctorate degree were eligible to participate.

The publication of the problem and the implementation of the competition took place via a website ([2], see link to website below) on which three data sets of different sizes and properties were made available. The teams were required to submit a paper with their chosen model, the estimated error rate and a safety argument as their competition entry. A rather moderate safety target of 1:1000 was specified for a safety-related misclassification, since the number of data records was also kept relatively small (1000-10,000). On the one hand, this more or less corresponds to the typical size of data sets in railway automation, while, on the other hand, the calculations were also supposed to have been able to be carried out without the need for any special hardware.

The submissions were evaluated by a jury of experts consisting of

- Prof Dr-Ing Corinna Salander (Head of DZSF, Honorary Professor University Stuttgart),
- Prof Dr Martin Fränzle (Carl-von-Ossietzky-University Oldenburg),
- Prof Dr Volker Tresp (Siemens AG, Honorary Professor LMU Munich)

and the author. Each submission was evaluated by at least three (sometimes more) experts. The teams then received additional, unclassified validation datasets that were used to practically verify their error estimates.

A total of 32 submissions from 18 universities and 15 countries were submitted. These submissions covered not only virtually all the known ML models, from statistical approaches to neural networks, but also combinations of models or even new innovative approaches.

Based on this diversity, the jury decided to award three first prizes in the following categories

- the best original approach,
- the best balance between safety and availability,
- the best practical result

and five second prizes.

3 The results

In order to break the suspense, no team, not even the winners, was able to give a theoretical argument that achieved the safety goal. In addition, the practical validation was worse than our own theoretical forecast in more than half the cases. In practice, only a few teams were able to achieve the safety goal or to come quite close to it, but even some of them only did so at the expense of unacceptable availability. Fig. 2 illustrates this for a data set: the safety target corresponded to 0.1 % misclassifications and the hidden goal was to minimise unavailability. Each point represents the results of a team.

- Almost all the teams avoided making any additional assumptions, even though this was sometimes obvious. For example, the uniform distribution of the data points in fig. 1 seems obvious and could also have been statistically checked.

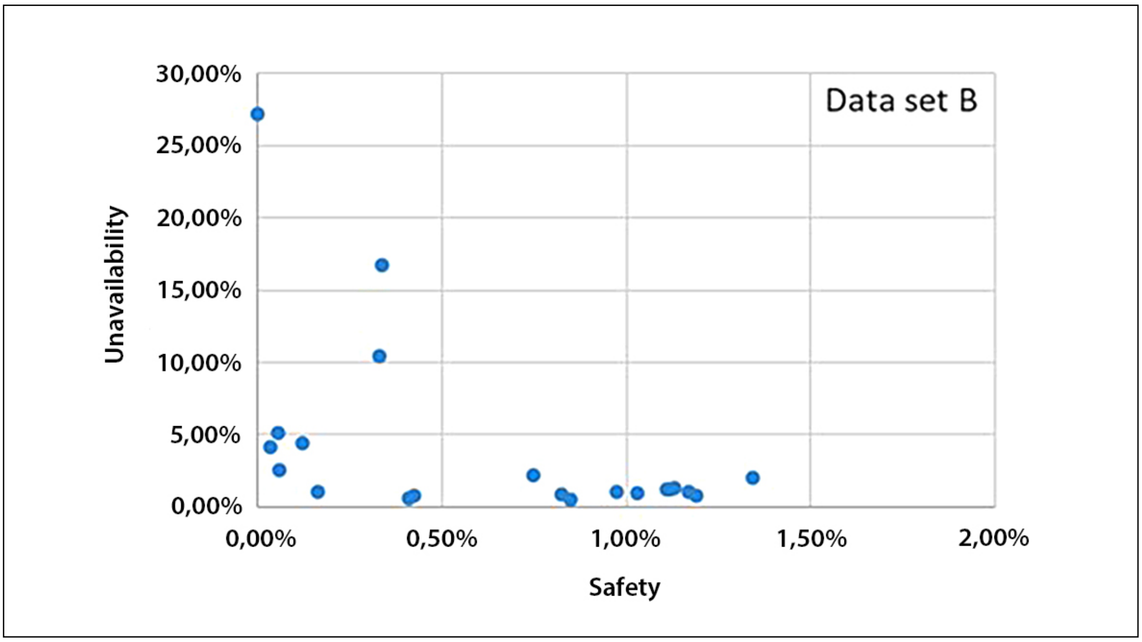

Bild 2: Das Verhältnis von Sicherheit und Verfügbarkeit
Fig. 2: Relationship between Safety and Availability

Das Sicherheitsziel entspräche 0,1 % Fehlklassifikationen, und dabei sollte die Nichtverfügbarkeit minimiert werden. Jeder Punkt repräsentiert die Ergebnisse eines Teams.
Über die Gründe dafür kann man eher nur spekulieren, aber auffällig war:

- Fast alle Teams haben es vermieden, zusätzliche Annahmen zu machen, obwohl dies manchmal auf der Hand lag. Z. B. in Bild 1 scheint eine Gleichverteilung aller Datenpunkte offensichtlich und hätte auch statistisch plausibilisiert werden können.
- Einige Teams haben nicht alle Annahmen, die bez. der Datensätze vorgegeben waren, ausgenutzt, z. B. die Separation der Daten.
- In der Aufgabenstellung wurde bewusst offengelassen, auf welchem statistischen Vertrauensniveau die Untersuchungen erfolgen sollten. Viele Teams haben sich für deutlich höhere Niveaus als 95 % entschieden (wie z. B. bei wichtigen medizinischen Stu-

- Some teams did not take advantage of all the assumptions that were given, such as the separation of the data.
- The level of statistical confidence at which the investigations should be carried out was deliberately left open in the task. Many teams opted for levels significantly higher than 95 % (such as in major medical studies). However, a higher level of confidence worsens the estimates, i.e. it would also be necessary to agree on a uniform level for reasons of the comparability of the results.

4 The findings

Even though the task could not be solved, the competition brought many interesting insights. The multitude of ideas, some of them very creative, was also impressive. Since the prob-

dien). Ein höheres Vertrauensniveau verschlechtert allerdings die Abschätzungen, d. h. es wäre notwendig, sich auch aus Gründen der Vergleichbarkeit der Ergebnisse auf ein einheitliches Niveau zu einigen.

4 Erkenntnisse

Der Wettbewerb hat, obwohl die Aufgabe nicht gelöst werden konnte, viele interessante Erkenntnisse gebracht. Auch die Vielzahl teilweise sehr kreativer Ideen war beeindruckend. Da das Problem nicht gelöst wurde, wurde auch der Prozess, mit dem die Daten generiert wurden, nicht veröffentlicht. D. h. es ist weiter möglich, sich an der Lösung des Problems zu versuchen [2].

Bezüglich der angewandten ML-Algorithmen war kein klarer Trend erkennbar. Es gab sehr gute Ergebnisse sowohl mit klassischen Methoden wie Support Vector Machines als auch mit komplexeren Methoden wie Neuronalen Netzwerken. Dies könnte ggf. aber auch mit der relativen Einfachheit des Problems zusammenhängen, dass es hier keine Vorteile bezüglich der verwendeten Methode gab.

Es war erstaunlich, dass so viele theoretische Abschätzungen falsch waren bzw. sich in der Praxis als falsch herausgestellt haben, insbesondere, da viele Abschätzungen einfach auf statistischen Konfidenzintervallen beruhten. Vielleicht wurde diese Aufgabe von den Teams unterschätzt. Nur wenige Abschätzungen beruhten auf statistischer Lerntheorie, wobei auch dort eher auf klassischen Resultaten aufgesetzt wurde [3]. Allerdings wurde bei den Abschätzungen häufig lax mit den Annahmen umgegangen, wie z. B. Unabhängigkeit.

Alle Abschätzungen für die Fehlerwahrscheinlichkeit p_E für einen Klassifikator F folgten allerdings der folgenden Form (etwas vereinfacht nach [3], nicht alle Parameter waren in allen Varianten vorhanden):

$$p_E = R_{emp}[F] + c\sqrt{\frac{S_{\mathcal{F},n} + \ln 1/\delta}{n}} \quad (1)$$

Die grundsätzlichen Einflussfaktoren sind [3]

- die empirische Fehlerwahrscheinlichkeit $R_{emp}[F]$. Hier muss das Ergebnis der Validierung angesetzt werden, nicht das Ergebnis des Trainings.
- $S_{F,n}$ hängt von der strukturellen Komplexität des ML-Algorithmus F ab und dem Stichprobenumfang n. In der Regel ist dieser Term nur sehr schwer abschätzbar und meistens auch nur sehr konservativ.
- $1\text{-}\delta$ ist das angestrebte statistische Vertrauensniveau. Wenn dieses Niveau einmal festgelegt ist, geht es als Konstante ein.
- c ist eine Konstante, die für jede Abschätzung spezifisch ist und die alle anderen Einflüsse sammelt, z. B. von Verteilungsannahmen etc.

Man erkennt in Formel (1), dass man offensichtlich nicht besser werden kann als das Ergebnis der Validierung $R_{emp}[F]$. Dies ist nicht erstaunlich, sollte aber immer bedacht werden, da hier der Stichprobenumfang n linear eingeht. Im besten Fall wäre dieser Term null, d. h. es gibt keine Fehlklassifikationen in der Validierung. Aber auch in diesem Fall geht im zweiten Term die Wurzel des Stichprobenumfangs n in die klassischen Abschätzungen ein. D. h. um diesen Term um eine Größenordnung zu verringern, muss man den Stichprobenumfang verhundertfachen! Dies erklärt auch, warum die klassischen Abschätzungen im Wettbewerb schwach ausfielen.

Das bedeutet aber, dass ohne weitere Annahmen ein sehr großer Stichprobenumfang notwendig ist, um einen ML-Algorithmus für

lem was not resolved, the process used to generate the data was not published. This means that it is still possible to try to solve the problem [2].

There was no clear trend with regard to the used ML algorithms. Very good results were achieved both with classical methods such as support vector machines and with more complex methods such as neural networks. However, this could also be related to the relative simplicity of the problem, meaning that there were no advantages with regard to the method used.

It was amazing that so many theoretical estimates were wrong or turned out to be wrong in practice, especially since many estimates were simply based on statistical confidence intervals. Perhaps this task was underestimated by the teams. Only a few estimates were based on statistical learning theory, whereby classical results were also used [3]. However, the estimates were often lax and included assumptions, such as independence.

However, all the estimates for the error probability p_E for a classifier F followed the following form (somewhat simplified - not all the parameters were available in all the variants):

$$p_E = R_{emp}[F] + c\sqrt{\frac{S_{\mathcal{F},n} + \ln 1/\delta}{n}} \quad (1)$$

The basic influencing factors are [3]

- the empirical error probability $R_{emp}[F]$. Here, the result of the validation must be applied, rather than the result of the training.
- $S_{F,n}$ depends on the structural complexity of the ML algorithm F and the sample size n. As a rule, this term is very difficult to estimate and usually only very conservative.
- $1\text{-}\delta$ is the desired level of statistical confidence. Once this level has been set, it enters as a constant.
- c is a constant that is specific to each estimate and collects all other influences, e.g. distribution assumptions, etc.

You can see in formula (1) that it is obviously not possible to get a better result for the validation $R_{emp}[F]$. This is not surprising, but should always be considered, as here the sample size n declines linearly. In the best case, this term would be zero, i.e. if there were no misclassifications in the validation. But even in this case, the root of the sample size n is included in the classic estimates in the second term. This means that in order to reduce this term by an order of magnitude, you have to increase the sample size by a hundredfold! This also explains why the classic assessments were weak.

However, this also means that a very large sample size is necessary to qualify an ML algorithm for safety applications without any further assumptions. This therefore seems hopeless, at least for high safety requirements.

However, if you compare this situation with other complex problems in safety applications, you can see that safety assessment works best on a model-based basis. For example, the Binary Symmetric Channel (BSC) model has proven its worth in safe data transmission, the Arrhenius equation in reliability prediction or Markov models for evaluating safety architectures, etc. To paraphrase George P. Box, however, the following applies generally: "All the models are wrong, but some are useful", because these models never hold exactly. However, they have been supplemented with application rules in standardisation and calculation assumptions have been agreed, so that comparable results can be obtained in comparable applications. AI is completely missing this so far, but it is necessary for the successful safety approval of such procedures.

Sicherheitsanwendungen zu qualifizieren. Zumindest für hohe Sicherheitsanforderungen erscheint dies aussichtslos.
Vergleicht man diese Situation allerdings mit anderen komplexen Problemen in der Sicherheitstechnik, so erkennt man, dass Sicherheitsbewertung am besten modellbasiert funktioniert. Z. B. in der sicheren Datenübertragung hat sich das Modell des Binary Symmetric Channel (BSC) bewährt, oder die Arrhenius-Gleichung in der Zuverlässigkeitsvorhersage oder Markov-Modelle zur Bewertung von Sicherheitsarchitekturen etc. Frei nach George P. Box gilt allerdings für alle: „All the models are wrong, but some are useful", denn exakt gelten diese Modelle niemals. Allerdings sind sie durch Anwendungsregeln in der Normung ergänzt worden, und man hat sich auf die Berechnungsgrundlagen geeinigt, sodass in vergleichbaren Anwendungen vergleichbare Ergebnisse herauskommen. Dies fehlt für KI bisher komplett, ist aber für erfolgreiche Zulassung von solchen Verfahren notwendig.
Um dann noch zu handhabbaren Datenumfängen zu kommen, lohnt es sich, die Bücher vom Turing-Preisträger Judea Pearl [4] zu lesen: Er kommt zur Erkenntnis, dass es notwendig ist *„ein Modell des Prozesses, der die Daten generiert, zu formulieren, oder zumindest einiger Aspekte dieses Prozesses"*. Dies entspricht dem oben formulierten Ansatz, Annahmen über das Datenmodell bzw. die Anwendungsumgebung aufzustellen (und zumindest zu plausibilisieren), um die Komplexität des ML-Modells und den benötigten Datenumfang zu reduzieren.

5 Zusammenfassung

Auch für ein sehr einfaches Entscheidungsproblem unter günstigen Randbedingungen konnte bisher kein ML-Modell gefunden werden, mit dem theoretisch oder zumindest praktisch minimale Sicherheitsanforderungen erfüllt werden können (bei gleichzeitiger Berücksichtigung der Verfügbarkeit). Daher ist es nicht verwunderlich, dass auch für komplexere Probleme bisher keine Erfolge berichtet wurden. Es gibt sogar fundamentale Bedenken, die anhand dieses einfachen Problems erläutert wurden. Sinnvoll wäre es, zunächst solche einfachen Probleme zu lösen. Dies ist zwar nicht hinreichend für die Lösung komplexerer Probleme, aber notwendig.
Ein Weg vorwärts könnte in einem modellorientierten Ansatz bestehen, der schon in anderen komplizierten Gebieten der Sicherheitstechnik erfolgreich gewesen ist. Dazu ist es notwendig, existierende Modelle wie in Formel (1) mit domänenspezifischen Annahmen zu stützen und in der Normung anschließend die Bemaßungsgrundlagen und Anwendungsregeln abzustimmen.
Dies bedeutet auch, dass Fortschritte zunächst bei Aufgaben mit eher geringen Sicherheitsanforderungen zu erwarten sind, für die solche Modelle einfacher formuliert werden können, z. B. weil es physikalisches oder domänenspezifisches Wissen gibt, das in solche Modelle eingebracht werden kann. ■

Link zur Wettbewerbsseite / *Link to the competition website:*

In order to arrive at manageable data volumes, it is worth reading the books by Turing Prize winner Judea Pearl [4]: he arrives at the conclusion that it is necessary to formulate "*a model of the process that generates the data or at least some aspects of this process*". This is in line with the aforementioned approach of making assumptions about the data model or application environment (and at least verifying plausibility) in order to reduce the complexity of the ML model and the required amount of data.

5 Conclusion

Even for a very simple decision problem under favourable boundary conditions, no ML model could be found so far, with which the safety minimum requirements can be theoretically or at least practically satisfied (while at the same time taking availability into account). It is therefore not surprising that no successes have been reported to date, even for more complex problems. There are fundamental concerns that have been explained on the basis of this simple problem. It would make sense to solve such simple problems first. While this is not sufficient to solve more the complex problems, it is necessary.
One way forward could be a model-oriented approach that has already been successful in other complicated areas of safety technology. For this purpose, it is necessary to enrich the existing models as in formula (1) with domain-specific assumptions and then to coordinate the assumptions and application rules in standardisation.
This also means that progress can initially be expected in tasks with rather low safety requirements for which such models can be formulated more easily, i.e. because there is physical or domain-specific knowledge that can be incorporated into them. ■

LITERATUR | LITERATURE
[1] Duda, R. O.; Hart, P. E.; Stork, D. G.: Pattern Classification, Wiley, 2001
[2] Siemens AG: AI-DA Challenge, https://ecosystem.siemens.com/universityrelations/ai-da-challenge-ai-dependability-assessment/overview, letzter Abruf 10.6.2022
[3] Vapnik, V. N.: The Nature of Statistical Learning Theory, Springer, 2010
[4] Pearl, J.; Mackenzie, D.: The Book of Why, Penguin Science, 2018

AUTOR | AUTHOR

Prof. Jens Braband
Principal Key Expert
Siemens Mobility GmbH
Anschrift / *Address*: Ackerstraße 22, D-38126 Braunschweig
E-Mail: jens.braband@siemens.com