

Active Learning

Jonathan Wischhusen
jonathan.wischhusen@haw-hamburg.de

HAW Hamburg, 20099 Hamburg, Deutschland

Zusammenfassung. Es werden mehr und mehr Daten in verschiedensten Bereichen generiert. Um diese für das vielversprechende Supervised Learning einzusetzen, fehlt ihnen eine wichtige Eigenschaft, ihre Annotation. Diese Voraussetzung ist in der Regel sehr teuer, da die Beschreibung von Menschen durchgeführt werden muss und es sich um tausend bis hunderttausende Dateninstanzen handelt. Dem entgegen steht Active Learning. Diese Arbeit beschreibt die Ideen hinter Active Learning und die grundlegenden Szenarien sowie Query Strategien. Das Grundlagenwissen soll das Verständnis fördern, um Problemstellungen in den Active Learning Kontext überführen zu können. Einige Themen der Forschung rund um die Thematik werden skizziert.

1 Einleitung

Künstliche Intelligenz (KI) erlebt im 21. Jahrhundert erneut eine Renaissance. Ausschlaggebend ist der Fortschritt in der Mustererkennung durch moderne Neuronale Netze, wie dem in [1] beschriebenen Deep Residual Network zur Bildererkennung. Diese Systeme bilden oft die Grundlage für viele KI-gestützte Dienste, die bereits im Alltag zu finden sind. Als Beispiele lassen sich die Gesichtserkennung auf Fotos in Sozialen Netzwerken oder die Straßen- und Verkehrsschilderkennung durch (noch teil-) autonom fahrende Fahrzeuge nennen. Dabei sind die Systeme nicht auf reine Bildinformation beschränkt, sondern bilden Klassifikatoren für verschiedene Arten von Daten und Sachverhalte.

Deep Learning bildet eine Untergruppe des Supervised Learnings. Diese Art des Machine Learnings erfordert für eine hohe Genauigkeit in der Vorhersage von Klassen, tausende bis hunderttausende Beispieldaten anhand derer das zugrundeliegende Modell trainiert wird. Man könnte vermuten, das sei kein Problem in Zeiten von Big Data und der verfügbaren Rechenleistung heutiger Prozessoren. Jedoch muss jede Datensatz annotiert sein, sodass die Maschine zu jeder Eingabe die korrekt Ausgabe kennt und diese Relation in das Modell induzieren kann.

Die Annotierung dieser Trainingsdaten ist dabei keineswegs trivial, da sie aus wirtschaftlicher Sicht sehr teuer und zeitintensiv sein kann. Zwar können durch Benutzer als Spam markierte E-Mails für einen Spam-Klassifikator oder Filmbewertungen im Internet als Trainingsdaten für Filmempfehlungen herangezogen werden, die durch ihre Natur nicht explizit mit Kosten verbunden sind.

Allerdings erfordern komplexere Supervised Learning Systeme Annotationen, im Folgenden auch als Labels bezeichnet, die sich nicht so einfach bestimmen lassen.

Dies schafft in vielen Bereichen, in denen es ausreichend Daten gibt, denen aber für eine konkrete Betrachtung aus der Machine Learning Sicht, das dazugehörige Label fehlt, jenes Problem, dass Supervised Learning Methoden nicht vernünftig trainiert und eingesetzt werden können.

Ein Ansatz um der Knappheit der wichtigen Trainingsdaten zu begegnen heißt Active Learning.

Im Fokus des Grundseminars¹, in dessen Rahmen diese Arbeit erstellt wird, steht die Auseinandersetzung mit einem Forschungsbereich der Informatik. Die vorangegangene Motivation skizziert vereinfacht eine immer noch aktuelle Problematik. Während in der Vergangenheit der benötigte Rechenaufwand als ein primärer Kostenpunkt betrachtet wurde, spielen heutzutage die Annotierungskosten der benötigten immensen Datenmengen eine entscheidende Rolle in der Machbarkeit von Supervised Learning Lösungen.

Im Folgenden wird der Machine Learning Ansatz Active Learning dargestellt. Aufgrund des Umfangs geht diese Arbeit mehr auf die Kernkonzepte und Methoden ein und beschreibt weniger die darunterliegenden mathematischen Verfahren.

In Abschnitt 2 werden die Grundlagen und Probleme erläutert. Daraufhin folgt in Abschnitt 3 ein Querschnitt der Forschungsthemen. Abschließend fasst Abschnitt 4 in einem Ausblick über Active Learning eine mögliche persönliche Ausrichtung zusammen.

2 Active Learning Grundlagen

Man betrachtet das Problem der teuren Annotationen aus der theoretischen Sicht des Active Learnings um ein Modell (Klassifizierer/Regression) zu trainieren folgendermaßen:

- i. nicht jeder Datensatz hat die gleiche Relevanz für das Modell
- ii. der Algorithmus entscheidet welche Datensätze für ihn interessant sind
- iii. eine Annotation zu erfragen kostet nichts

Versinnbildlicht man dies, nimmt der Algorithmus die Rolle eines Schülers ein, der einen Lehrer, oft als Orakel bezeichnet, bei Unklarheit fragt, wie ein bestimmter Datensatz eingeordnet werden soll.

Die entstehende Annahme, dass der lernende Algorithmus, auf diese Weise deutlich weniger annotierte Trainingsdaten benötigt, als passives Supervised Learning², um eine gleichwertige oder höhere Genauigkeit zu erreichen, ist in vielen Bereichen empirisch belegt [2].

¹ HAW MINF, GSM WiSe 18/19

² Der Algorithmus kennt zu jedem Trainingsdatensatz das entsprechende Label.

Active Learning unterstützt insoweit das *Human in the Loop* Konzept, als dass das Orakel für gewöhnlich ein Mensch ist, es aber nicht zwangsweise sein muss. Unterstützende Algorithmen oder Maschinen sind ebenfalls möglich, solange Anfragen des Schülers (auch Queries genannt) korrekt beantwortet werden können. Da der Mensch allerdings über weitreichende kognitive Fähigkeiten verfügt und im Idealfall ein breites Domänenwissen besitzt, ist er als Orakel sehr nützlich und kann dieses durch das Beantworten der Queries entsprechend in das Modell übertragen.

Die Idee hinter Active Learning ist nicht neu, weshalb sich bereits einige grundlegende Methoden ergeben haben, die im Folgenden dargestellt werden.

2.1 Szenarien

Die Art und Weise wie der Algorithmus das Orakel befragt wird als Szenario bezeichnet. Es haben sich mit der Zeit drei Szenarien etabliert, die in Abbildung 1 visualisiert sind.

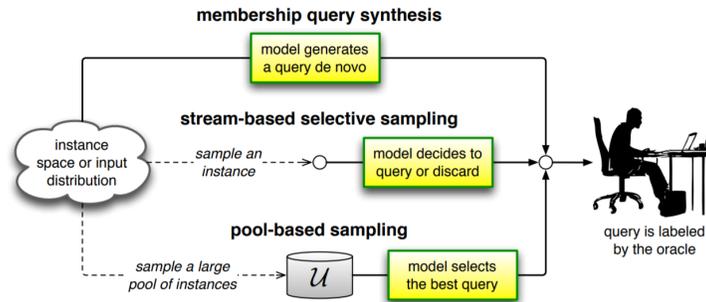


Abb. 1. Gängige Active Learning Query-Szenarien ³

Membership Query Synthesis

In dieser Form der Anfrage generiert bzw. synthetisiert der Schüler typischerweise einen fiktiven Datensatz innerhalb der Grenzen der Eingabedaten [3]. Dieser repräsentiert markante Eigenschaften, generell auch als Features bezeichnet, durch die indirekt eine Gruppe von Eingabedaten beschrieben wird.

Dieses Szenario hat den Nachteil, dass es sich nur bedingt für ein menschliches Orakel einsetzen lässt. Soll beispielsweise Handschrift erkannt werden, könnte das Orakel nach einem Label zu einem Bild gefragt werden, auf dem ein künstliches Zeichen zu sehen ist, welches der Mensch nicht einordnen kann

³ Bildquelle: Settles

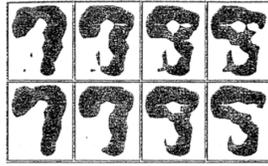


Abb. 2. Ein Beispiel eines synthetisierten Querys, der nicht eindeutig durch einen Menschen interpretiert werden kann kann.⁵

(Abbildung 2). Ebenfalls ist bei einer Natural Language Processing (NLP) Aufgabe eine für den Menschen unverständliche Sprachsequenz zu erwarten, setzt man Membership Query Synthesis dort ein.

Hier schafft der Einsatz einer Maschine als Orakel, die künstliche Queries durch ein Experiment beantworten kann, wie beispielsweise der Roboter in [4], Abhilfe.

Stream-Based Selective Sampling

In diesem Szenario werden Datenpunkte nacheinander aus der Menge der Eingabedaten, die kein Label besitzen, betrachtet. Der Algorithmus entscheidet daraufhin, ob das Wissen des Labels der aktuelle Instanz einen Mehrwert für das Modell hat und erfragt oder die Instanz verworfen werden soll.

Dadurch dass jede Instanz für sich betrachtet wird, bleiben die Anfragen an das Orakel sensible für die unbekannte Verteilung der Eingabedaten. Die Entscheidung eine Annotation zu erfragen, kann durch verschiedene Metriken gestützt sein. Diese werden in Abschnitt 2.2 erläutert.

Pool-Based Sampling

Wie zu Beginn erwähnt, lassen sich große Datenmengen in vielen Domänen finden. Allerdings sind sie wegen der fehlenden Annotationen ungeeignet für passives Supervised Learning. Ähnlich wie bei dem Stream-Based Selective Sampling, werden Datensätze, die das Orakel einordnen soll, nach konkreten Metriken, die einen gewissen Informationsgehalt preisgeben, ausgewählt.

Der Unterschied besteht darin, dass beim Pool-based Sampling dieselbe Metrik auf alle Eingabedaten bzw. einer Teilmenge (siehe Abbildung 1) angewandt wird, und daraufhin die interessanteste Instanz relativ zu den verbleibenden dem Orakel vorgelegt wird. Das Stream-Based Sampling entscheidet hingegen unabhängig der restlichen möglicherweise unbekanntem Datenpunkte.

⁵ Bildquelle: <http://www.cs.cmu.edu/~epxing/Class/10701-12f/Lecture/settles.active-nov14.pdf/>

2.2 Query Strategien

Warum Active Learning für das Training des Modells weniger Datenpunkte benötigt und eine hohe Genauigkeit erzielen kann, legt Abbildung 3 in einem Beispiel offen. Im Gegensatz zu den drei Szenarien, erstrecken sich die Abfrage Strategien über viele verschiedene Ansätze.

Welcher Ansatz sich für ein Problem am meisten eignet, lässt sich pauschal nicht bestimmen, vielmehr muss individuell für das zu lösende Problem und die Verteilung der verfügbaren Daten entschieden werden.

Im Folgenden werden bekannte Strategien erläutert.

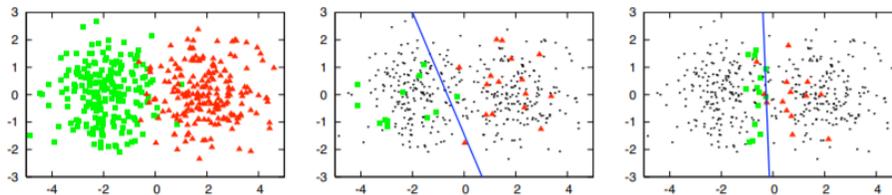


Abb. 3. *links:* Verteilung und vollständig annotierte Datenpunkte; *mitte:* lineare Regression mit zufällig ausgewählten Instanzen; *rechts:* lineare Regression mit explizit ausgewählten Instanzen ⁷

Uncertainty Sampling

Der lernende Algorithmus wählt den Datensatz aus, bei dem er sich bei der Einordnung am unsichersten ist. Die Unsicherheit wird in der Regel durch eine der folgenden drei Metriken bestimmt.

- i. *least confident:* Der lernende Algorithmus berechnet je Datenpunkt die Wahrscheinlichkeiten für die definierten Features. Die Instanz mit dem niedrigsten Wert für das am wahrscheinlichste Label wird ausgewählt.
- ii. *margin sampling:* Die Wahrscheinlichkeit für jedes Feature pro Datensatz wird berechnet. Der Datenpunkt mit der größten Differenz zwischen den beiden wahrscheinlichsten Features wird ausgewählt.
- iii. *entropy:* Der Shannon-Entropie Wert wird unter Betrachtung aller Feature Wahrscheinlichkeiten pro Datenpunkt errechnet. Die Instanz mit dem geringsten Wert wird ausgewählt.

Der Nachteil von (i) ist, dass obwohl absolut gesehen die unsichersten Datenpunkte ausgewählt werden, jedoch kann dieser dennoch relativ sicher bestimmt werden, wenn das wahrscheinlichste Feature im Vergleich zu den weiteren Features eine große Abweichung hat. (ii) hingegen betrachtet genau diese Abweichung

⁷ Bildquelle: Settles

der beiden wahrscheinlichsten Features als Differenz voneinander, und wählt die Datenpunkte aus, die dadurch gesehen nah an einer Entscheidungsgrenze liegen, also bei denen die beiden Haupt-Features ähnlich wahrscheinlich sind.

Query by Committee

Der Ansatz von [5] beschreibt den Einsatz von mehreren Modellen, die mit verschiedenen Sichtweisen auf einem annotierten Teil der Eingabedaten trainiert werden. Dadurch liefern die einzelnen Modelle unterschiedliche Wahrscheinlichkeiten für die Klassifizierung für Datenpunkte, die nicht gelabelt sind. In dem Datenpunkt, bei dem sich das Komitee, also die einzelnen Modelle, am wenigsten einig sind, wird der höchste Informationsgewinn vermutet. Dieser wird dem Orakel gezeigt.

Das Ziel von Query by Committee ist es durch verschiedene Sichten die kontroversen (unbekannten) Bereiche innerhalb der Eingabedaten zu finden und aufzulösen, so dass mit wenigen Annotationen die beste Entscheidungsgrenze gefunden wird. Für den Ansatz sind zwei Punkte wichtig:

- i. Modelle im Komitee repräsentieren verschiedene Bereiche/Sichten des Eingabebereiches
- ii. das Komitee muss die Uneinigkeit der Mitglieder bemessen können

Expected Model Change

Die Entscheidungsfindung basiert bei dieser Idee auf der Änderung des Modells, die das Wissen über einen Datensatz nach dessen Beantwortung durch das Orakel zur Folge hat. Es wird also die Dateninstanz abgefragt, bei sich die Parameter in dem Trainingsmodell voraussichtlich am meisten ändern. Der tatsächlichen Klasse wird kein Informationsgehalt zugesprochen.

Der Ansatz hat eine Schwäche, denn er kann durch ungleichmäßige Verteilung der Eigenschaften innerhalb der Eingabedaten schnell unbrauchbar werden. Hat eine Instanz beispielsweise für eine Klasse einen sehr hohen Wert, der eine große Parameteränderung des Modells zur Folge hätte, wird ihr leicht zu viel Informationsgewinn bescheinigt. Normalisierung der Eingabedaten kann dem entgegenwirken.

Expected Error Reduction

Im Vergleich zum Expected Model Change wird bei diesem Ansatz nicht auf die größtmögliche Veränderung des Modells geschaut, sondern auf den erwarteten Fehler bzw. die Genauigkeit des Modells hinsichtlich der Eingabedaten. Derjenige Datensatz, dessen Annotation die Fehlerrate des Modells auf den Eingabedaten am meisten reduzieren würde, wird abgefragt.

Dieser Ansatz ist sehr rechenintensiv, da alle unbekanntem Datensätze zur Bestimmung des zu erwarteten Fehlers mit jedem neuen neuen Modell bestimmt werden müssen. Die Bestimmung des Fehlerwerts muss für jeden Datensatz

durchgeführt werden, um schlussendlich das Modell mit der Instanz zu trainieren, die den erwarteten Fehler am meisten minimiert.

Density-Weighted Methods

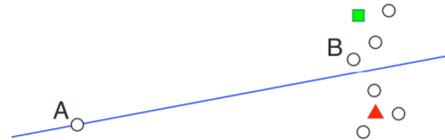


Abb. 4. Obwohl die meiste Unklarheit bei A liegt, profitiert die generelle Genauigkeit mehr davon B zu wissen, da dadurch die Entscheidungsgrenze gemäß der Verteilung verbessert wird.

Diese Art der Abfrage wählt bevorzugt die Datensätze aus, die die grundlegende Verteilung der Eingabedaten repräsentieren. Dahinter steht der Gedanke, das beispielsweise beim Uncertainty Sampling, zwar Instanzen abgefragt werden, deren korrekte Annotation sehr ungewiss ist, jedoch das Wissen über diese Klasse die allgemeine Genauigkeit des Modells nur sehr geringfügig verbessert. Das liegt darin begründet, dass die Instanz nicht zwingend andere Datenpunkte im Eingaberaum repräsentieren muss und somit keine Entscheidungshilfe für andere Instanzen bietet, siehe Abbildung 4.

Um die allgemeine Genauigkeit zu steigern, bietet es sich an die Verteilung des Eingaberaums zu bestimmen und repräsentative Datenpunkte zu erfragen.

2.3 Herausforderungen

Active Learning setzt in seinem theoretischen Ansatz einige Bedingungen voraus, die in einer realen Anwendung nicht immer gegeben sind. Zwei davon werden im Folgenden erläutert:

- i. Die Antwort des Orakels ist immer verlässlich und korrekt
- ii. Die Annotation eines Datensatzes ist mit keinen bzw. gleichen Kosten verbunden

Fehlerhafte Orakel

Das Problem des nicht allwissenden Orakels bezieht sich darauf, dass durch verschiedene Umstände, wie eine geringe Qualität der Eingabedaten oder der Unkenntnis eines menschlichen Orakels, Annotation falsch gesetzt werden können. Diese Fehlentscheidungen müssen in einem realen System beachtet werden.

Ebenfalls lässt sich dazu zählen, dass besonders im Falle der multiplen Annotation, also wenn ein Datensatz durch mehrere Klassen beschrieben wird, die

Annotierungsaufgabe durch ein begrenztes Domänenwissen, der Sichtweise oder durch Ermüdung und Ablenkung des Menschen stark verzerrt wird.

Weiter zu bedenken ist auch die Tatsache, das Orakel eine variierende Qualität der Antworten über die Zeit haben können, zum Beispiel, wenn ein Mensch sich mit der Aufgabe und den Abfragen vertraut gemacht hat.

Variierende Annotationskosten

Ein weiteres Problem ist der reale Aufwand der für eine Annotation von einem Menschen oder einer Maschine aufgebracht werden muss. Dieser kann unterschiedlich hoch für verschiedene Datensätze und Anfragen ausfallen.

Wenn der Grundsatz von Active Learning in das Licht gerückt wird, dass der allgemeine Trainingsaufwand möglichst minimiert wird, muss neben der Auswirkung auf die Qualität der Klassifikation des Modells ebenfalls der Aufwand der Annotation mit in der Query Strategie berücksichtigt werden.

Komplexe Features

Ursprünglich wurden Active Learning Systeme entworfen um eine bestimmte Klasse in Form einer Annotation einer Dateninstanz zuzuordnen. Allerdings ist dies sehr zeitintensiv, wenn das Wissen eines Menschen bei der Annotation eingesetzt werden soll, da er jede Klasse einzeln selektieren und einem Datensatz zuweisen muss.

Abhilfe schafft hier die Idee der Active Feature Selection [6]. Das Ziel ist es, Beschreibungen der Eigenschaften durch das Orakel zu erhalten, wie beispielsweise Reguläre Ausdrücke. Anhand dieser ist der Algorithmus in der Lage indirekt viele Instanzen selbständig mit einer Klasse versehen.

Cold Start

Meistens werden in Active Learning Szenarien eine Teilmenge an annotierten Daten vorausgesetzt, auf denen das initiales Modell trainiert wird.

Betrachtet man allerdings die gesamte Menge der Eingabedaten ohne Labels ergibt sich das Problem des Cold Starts: Welche Dateninstanz ist interessant, wenn keine ersichtliche Relation neben der Verteilung gegeben ist?

Dieses Problem ist besonders essentiell für Empfehlungssysteme und deren *New User* Szenario, in dem es zu Beginn keine realen Anhaltspunkte gibt. Allerdings ist es aus der eigentlichen Active Learning Sicht mehr als Optimierungsproblem zu verstehen, wenn man die erforderlichen annotierten Instanzen auf Null reduzieren kann.

3 Forschung

Obwohl schon sehr viel Forschungsarbeit in den letzten Jahren und Jahrzehnten im Bereich Active Learning geleistet wurde, gibt es fortwährend neue Ansätze zu altbekannten Problemen. Settles benennt fünf Bereiche:

- i. Bessere kostensensitive Ansätze
- ii. Crowdsourced Labels/Noisy Oracles
- iii. Batch Learning
- iv. User Interface Herausforderungen
- v. Wiederverwendbarkeit von Daten

Diese Schwerpunkte sind Teil aktueller Veröffentlichungen, wie zum Beispiel in [7], wo der Umgang mit falsch annotierten Daten und inkorrekten Orakeln untersucht wird. Ebenfalls finden Active Learning Konzepte im Rahmen der Visual Analysis [8] Zuspruch um einerseits die Annotation durch Menschen zu unterstützen, andererseits auf Bedienungsebene der Applikation, die Benutzbarkeit zu verbessern, indem beispielsweise die Interaktion mit der grafischen Oberfläche analysiert wird um zu erwartende Aktionen vorauszusagen. Ebenfalls wird überlegt, die Methodik, wie und durch welche Schlussfolgerungen ein menschlicher Analyst arbeitet, in die Machine Learning Welt zu überführen, um die menschliche Sicht auf Daten und deren Zusammenhänge algorithmisch auszudrücken.

In [9] wird das Problem angesprochen, das Machine Learning Modelle statisch sind und auf einem konkreten Rahmen an Daten angepasst sind. Vielmehr werde eine Methode benötigt die auch mit sich verändernden Bedeutungen von Daten und Relationen funktioniert und das Modell entsprechend weiterentwickelt, sogenanntes Adversarial Active Learning. Dadurch wird Active Learning in den Kontext des Never ending Learning Prinzips [10] und des Evolving Fuzzy Systems [11] geschoben.

3.1 Konferenzen

Durch die Vielfalt der Active Learning Thematik finden sich Forschungsprojekte in verschiedensten Bereichen, weshalb eine Auflistung von expliziten Konferenzen nicht unbedingt hilfreich ist. Es bietet sich allerdings an die Rahmenkonferenzen von Machine Learning im Auge zu behalten:

Knowledge Discovery and Data Mining (SIGKDD)

Die jährlich ⁸ stattfindende Konferenz der Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) gilt laut der Association for Computing Machinery (ACM) als führende Konferenz hinsichtlich der Wissensfindung, -entdeckung und Datensammlung aus Computern und vernetzten Systemen.

International Conference on Machine Learning (ICML)

Die zum 36. Mal stattfindende Konferenz wird unterstützt durch die International Machine Learning Society (IMLS). Sie ist insofern interessant, als das Settles, Autor von [12], dort veröffentlicht.

⁸ 1995-heute, 2019 zum 25. mal

4 Ausblick

Active Learning ist ein spannender Forschungsbereich, der zumal dem Menschen die Möglichkeit der Teilnahme an sonst automatisierten Prozessen erlaubt. Die, den Maschinen (noch) vorenthaltene Fähigkeit der Wissensfindung und Kreativität des Menschen kann wunderbar mit der schnellen mathematischen Analyse von großen Datenmengen der Maschinen zusammengeführt werden.

Settles hat große Wirtschaftskonzerne bereits 2012 nach dem Einsatz von Active Learning Methoden gefragt, die ihm den Einsatz in großem Umfang bestätigt haben. Es ist davon Auszugehen, dass das Prinzip auch in vielen neuen Bereichen Einzug findet, in denen die Daten die zur Verfügung stehenden Arbeitsressourcen übersteigen.

Auch wenn der Active Learning innerhalb dieser Arbeit nicht aus Sicht des Natural Language Processings erklärt wurde, ist das Konzept in diesem Gebiet weit verbreitet [13]. Durch das Projekt Forum 4.0⁹ sollen Techniken und Methoden erforscht werden, die das Potenzial von Benutzer-Kommentaren im Internet ausschöpfen können. Im Vordergrund steht die Analyse der Kommentare im Hinblick auf ihrer zugrundeliegende Absicht. Da diese nicht von einer Maschine alleine abgeleitet werden kann, bietet sich hier ein Human in the Loop Ansatz an. Dies war ausschlaggebend für die Wahl das Grundseminar als initialen Punkt zu nutzen, um in die Active Learning Thematik einzutauchen. Die tiefgreifende Betrachtung von NLP Problemen in Verbindung mit Active Learning auf realen Kommentaren bietet ein interessantes Forschungsfeld angesichts der Popularität moderner Social Media Plattformen.

⁹ <https://scan.informatik.uni-hamburg.de/forum40/>

Literaturverzeichnis

- [1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [2] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- [3] Dana Angluin. Queries and concept learning. *Mach. Learn.*, 2(4):319–342, April 1988. ISSN 0885-6125.
- [4] Ross D King, Kenneth E Whelan, Ffion M Jones, Philip GK Reiser, Christopher H Bryant, Stephen H Muggleton, Douglas B Kell, and Stephen G Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427(6971):247, 2004.
- [5] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294. ACM, 1992.
- [6] Gregory Druck, Burr Settles, and Andrew McCallum. Active learning by labeling features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 81–90. Association for Computational Linguistics, 2009.
- [7] Mohamed-Rafik Bouguelia, Slawomir Nowaczyk, K. C. Santosh, and Antanas Verikas. Agreeing to disagree: active learning with noisy labels without crowdsourcing. *International Journal of Machine Learning and Cybernetics*, 9(8):1307–1319, Aug 2018.
- [8] Alexander Endert, William Ribarsky, Cagatay Turkay, BL William Wong, Ian Nabney, I Díaz Blanco, and Fabrice Rossi. The state of the art in integrating machine learning into visual analytics. In *Computer Graphics Forum*, volume 36, pages 458–486. Wiley Online Library, 2017.
- [9] Brad Miller, Alex Kantchelian, Sadia Afroz, Rekha Bachwani, Edwin Dauter, Ling Huang, Michael Carl Tschantz, Anthony D. Joseph, and J.D. Tygar. Adversarial active learning. In *Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop, AISEC '14*, 2014.
- [10] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka, and Tom M Mitchell. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.
- [11] Edwin Lughofer. Evolving fuzzy systems—fundamentals, reliability, interpretability, useability, applications. In *HANDBOOK ON COMPUTATIONAL INTELLIGENCE: Volume 1: Fuzzy Logic, Systems, Artificial Neural Networks, and Learning Systems*, pages 67–135. World Scientific, 2016.
- [12] Burr Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012.
- [13] Fredrik Olsson. A literature survey of active machine learning in the context of natural language processing. 2009.