

Hochautomatisierte Datenauswertung für KI

Vergleich von manuellen und hochautomatischen Verfahren zur Datenerzeugung auf Basis des öffentlichen Tsinghua-Tencent 100k Datensatzes

Abstrakt: Die Kosten für Trainingsdaten verhindern aktuell die Einführung von künstlicher Intelligenz in der Breite. Hochautomatische Verfahren zur Datengewinnung versprechen eine erhebliche Kostenreduktion. Nach einer kurzen Einführung zum aktuellen Stand von KI und einer Analyse der unterschiedlichen Lernprozesse, vergleichen wir daher die Leistung von Mensch und Maschine bei der Kennzeichnung visueller Daten auf Kosten-, Qualitäts- und Quantitätsebene. Die Auswertung basiert auf dem öffentlichen Datensatz Tsinghua-Tencent 100k und analysiert die typischen Fehler von Maschine und Mensch.

| 1 Der Datenaspekt | 2 |
|---|----|
| 1.1 Der aktuelle Stand von Kl | 2 |
| 1.2 Daten sind der Schlüssel | 2 |
| 1.3 Bessere Ergebnisse mit weniger Daten? | 2 |
| 1.4 Wie lernen Menschen? | |
| 1.5 Wie lernen Maschinen? | 2 |
| 1.6 Menschliche Datenauswertung | 3 |
| 1.7 Hochautomatische Datenauswertung | 3 |
| 1.8 Der Wert von Daten | 3 |
| 2 Tsinghua-Tencent 100k Datensatz | 4 |
| 2.1 Informationsgehalt | 4 |
| 2.2 Datenverteilung | 4 |
| 2.3 Hochautomatische Datenauswertung der Positivdaten | 5 |
| 2.3.1 Quantität | 5 |
| 2.3.2 Qualität | 7 |
| 2.4 Auswertung des Negativ-Datensatz | 7 |
| 3 Analyse | |
| 3.1 Wo treten menschliche Fehler auf? | |
| 3.1.1 Gesamtkomplexität und subjektive Bewertung | |
| 3.1.2 Erwartungsgesteuerte Wahrnehmung | |
| 3.1.3 Kleine Objekte in komplexen Szenen | |
| 3.1.4 Geringer Kontrast | |
| 3.2 Wo treten die Fehler bei Maschinen auf? | |
| 3.2.1 Sehr kleine Objekte | |
| 3.2.2 Unbekannte Gründe | |
| 3.3 Was steckt noch in den Daten? | |
| 4 Zusammenfassung | |
| 5 Evoteara GmbH | 13 |



1 Der Datenaspekt

1.1 Der aktuelle Stand von KI

Die heutige KI-Technologie ist mittlerweile für die Verwendung in produktiven Systemen geeignet. Neuronale Netzwerke können zuverlässig trainiert und ausgeführt werden, C++ und Netzwerkoptimierung ermöglichen den zuverlässigen Betrieb und die Prozessintegration. Eine breite Palette verfügbarer Hardware, von eingebetteten Systemen bis hin zu High-End-Rechenzentrumslösungen, ermöglicht den Einsatz von KI-Lösungen in einer Vielzahl von Anwendungsfällen.

1.2 Daten sind der Schlüssel

Deep-Learning Lösungen skalieren nahezu unbegrenzt mit zusätzlichen Daten. Sind jedoch zu wenig Daten vorhanden, weisen sie in der Regel eine geringere Genauigkeit als klassische Ansätze auf. Um ein allgemeines Erkennungsproblem mit Deep Learning zu lösen, empfehlen wir normalerweise 1000-10.000 Trainingsbeispiele *pro Klasse*.

Während alle anderen Anforderungen erfüllt sind, bleibt somit die erforderliche Datenmenge das Haupthindernis für die Einführung von künstlicher Intelligenz.

Derzeit sehen wir 3 unterschiedliche Ansätze, um das Datenproblem zu lösen:

- 1.) Sparse Modeling
- 2.) Transfer learning
- 3.) Deep learning mit hochautomatischer Datenauswertung

1.3 Bessere Ergebnisse mit weniger Daten?

Sowohl Sparse-Modelling als auch Transferlernen versuchen, die erforderliche Datenmenge zum trainieren einer KI deutlich zu reduzieren. Die Grundidee dabei ist das menschliche Lernen nachzuahmen. Zur Einschätzung des Potentials dieser Methoden ist es jedoch erforderlich die wichtigsten Unterschiede im Lernprozess von Mensch und Maschine zu verstehen.

1.4 Wie lernen Menschen?

Das menschliche Gehirn hat ungefähr 86 Milliarden Neuronen in einer dreidimensionalen Anordnung. Während Menschen Jahre brauchen, um ein grundlegendes Verständnis ihrer Umwelt zu entwickeln, ist der menschliche Lernprozess während ihres gesamten Lebens kontinuierlich. Basierend auf ihrem Vorwissen können Menschen als Erwachsene daher neue Inhalte aus einer geringen Anzahl neuer Beispiele erlernen.

1.5 Wie lernen Maschinen?

Die aktuell verwendete "schwache KI" hat typischerweise 5-150 Millionen Neuronen in einer zweidimensionalen unidirektionalen Anordnung. Mit weniger als 0,17% der Kapazität des menschlichen Gehirns verfügt die KI im Gegensatz zum Mensch über keinerlei Vorkenntnisse. Daher kann eine KI ihr gesamtes Wissen ausschließlich aus den Trainingsdaten beziehen und der Lernprozess ist insgesamt auf wenige Tage während der Trainingsphase begrenzt.

Seite 2 von 13 22.11.2019 V1.02 Author: Tobias Manthey



Lean-Data-Methoden sind daher mathematisch durch den Informationsgehalt oder die Entropie der Trainingsdaten beschränkt. Je niedriger die Entropie oder die Datenmenge ist, desto höher ist das Risiko, dass die KI Pseudomerkmale lernt. Pseudomerkmale beschreiben die verschiedenen Klassen innerhalb der Trainingsdaten, jedoch nicht in der Praxis.

1.6 Menschliche Datenauswertung

Nach unserer Schätzung kann eine einzelne Person bis zu 500 Bilder pro Tag auswerten. Problematisch hierbei ist die monotone Arbeitsbelastung sowie die Subjektivität der Individuen. Die menschliche Wahrnehmung ist von Faktoren wie Annahmen, Müdigkeit und Stimmung beeinflusst. Zusätzliche Faktoren sind sowohl die limitierte Aufmerksamkeitsspanne als auch die begrenzte Fähigkeit zum Verständnis von komplexen Zusammenhängen. Aus diesen Gründen ist zu erwarten, dass die Qualität der menschlichen Datenauswertung mit steigender Komplexität, d.h. mit einer höheren Anzahl von Personen und Klassen abnimmt.

Der Kostenfaktor bei der menschlichen Datenauswertung spielt dabei eine erhebliche Rolle. Für die Bereitstellung eines Datensatzes, der den Mindestanforderungen für die produktive Nutzung basierend auf den 157 Klassen des Tsinghua-Tencent 100k Datensatzes entspricht, ist mehr als ein Mannjahr Aufwand erforderlich.

1.7 Hochautomatische Datenauswertung

Das Ziel der hochautomatischen Datenauswertung ist die Erzeugung von Datensätzen in hoher Qualität in einer Größenordnung von Millionen Trainingsbeispielen und hunderten individuellen Klassen. Mit einer Stückzahl von typischerweise je 100.000 unbeschrifteten Bildern ist die hochautomatisierte Auswertung ein iterativer Prozess, der von Menschen überwacht wird. Ausgangspunkt für den Prozess ist in der Regel ein öffentlicher Datensatz oder manuell beschriftete Daten.

Bevor jedoch Daten automatisch ausgewertet werden können, muss die automatische Verarbeitung eingerichtet werden. Abhängig von der Projektphase und den Anforderungen werden bei der hochautomatisierten Datenauswertung verschiedene Techniken eingesetzt, die auf verschiedenen KI Ansätzen und sogar klassischen Algorithmen basieren. Der Aufwand zum Ausführen einer Datenverarbeitung ist anfangs höher und nimmt mit der Zeit ab. Im Durchschnitt kann jedoch eine einzelne Person die Auswertung von 10.000 bis 20.000 Bildern pro Tag überwachen.

1.8 Der Wert von Daten

Der Wert der Trainingsdaten für künstliche Intelligenz muss auf der Ebene von Quantität, Qualität, Informationsgehalt und Verteilungsebene bewertet werden. Während eine zu geringe Menge an Daten zu einer schlechten Erkennungsleistung führen kann, beeinträchtigen auch falsche oder fehlende Daten den Lernprozess und verhindern so, dass das volle Potenzial der Daten ausgeschöpft werden kann.

Die Erkennungsleistung einzelner Klassen folgt normalerweise der Datenverteilung. Während daher die Gleichverteilung die ideale Verteilung für KI-Trainingsdaten ist, entspricht die Datenverteilung in der Praxis typischerweise einer Gaußschen Verteilung. Eine ungleiche Datenverteilung kann bis zu einem gewissen Grad ausgeglichen werden. Stark unausgewogene Datensätze wirken sich jedoch negativ auf die Erkennungsgenauigkeit von unterrepräsentierten Klassen aus.

Seite 3 von 13 22.11.2019 V1.02 Author: Tobias Manthey



2 Tsinghua-Tencent 100k Datensatz

Nach Angaben der Autoren des Tsinghua-Tencent 100k bietet der Datensatz 100.000 hochwertige Bilder mit einer Auflösung von 4 Megapixeln aus verschiedenen Szenen in China, die 30.000 markierte Verkehrszeichen enthalten. Diese Bilder decken große Unterschiede in der Beleuchtung und den Wetterbedingungen ab. Jedes Verkehrszeichen im Benchmark ist mit einer Klassenbeschriftung, einer äußeren Markierung sowie einer Pixelmaske versehen.

Der Datensatz umfasst 127 Basisklassen und ist in 3 Kategorien:



Abb. 1 Übersicht über den Datensatz

Da mit einem * gekennzeichnete Klassen zusätzlich nach den individuellen Wert des Verkehrszeichen unterteilt werden, beträgt die Gesamtzahl der Klassen im Datensatz 157. Zusätzlich gibt es 3 unspezifische Klassen für "unbekannte" Informations-, Warn- oder Verbotszeichen. Im Datensatz haben wir 22.235 gekennzeichnete Verkehrszeichen in den spezifischen Klassen gefunden. Darüber hinaus haben wir 2067 Verkehrszeichen in den insgesamt 3 Kategorien von unbekannten Verkehrszeichen gefunden.

Der Datensatz ist in 9176 positive sowie 82094 negative Bilder aufgeteilt.

2.1 Informationsgehalt

Die Auflösung, Anzahl, Diversität und Quantität der Bilder im Datensatz ist hoch. Der potenzielle Informationsgehalt des Datensatzes wird daher als sehr gut eingestuft um damit qualitativ hochwertige Deep Learning Netzwerke zu trainieren.

2.2 Datenverteilung

Die Verteilung der Trainingsdaten ist erwartungsgemäß sehr ungleichmäßig. Die durchschnittliche Anzahl von Instanzen pro Klasse beträgt 143 mit einer Standardabweichung von 367. Da 72 Klassen über weniger als 10 verfügbare Trainingsbeipiele verfügen, eignet sich fast die Hälfte der Klassen nicht zum Trainieren eines Deep-Learning-Algorithmus. Weitere 43 Klassen haben weniger als hundert Beispiele, was immer noch eine kritisch niedrige Zahl ist. Nur 5 Klassen erfüllen unsere empfohlene Mindestmenge von 1000 Trainingsbeispielen.

Seite 4 von 13 22.11.2019 V1.02 Author: Tobias Manthey



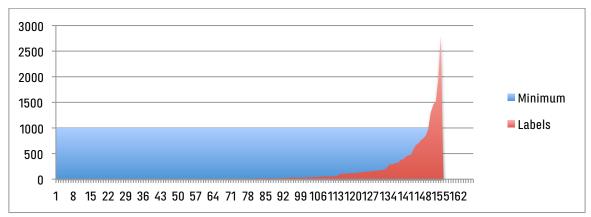


Abb. 2 Datenverteilung über die Klassen

Wenn wir die empfohlenen Mengen auf die insgesamt 157 Klassen des Tsinghua-Tencent-Datensatzes anwenden, würde dies eine Gesamtzahl von 157.000 bis 1.570.000 Instanzen entsprechen. Obwohl die Autoren des Tsinghua-Tencent 100k den Datensatz als "großen Maßstab" betrachten, macht er insgesamt nur 1,5% bis 15% der empfohlenen Datenmenge für einen Datensatz mit Produktionsqualität aus. Die Anzahl der Beispiele in den meisten Klassen liegt unter einer kritischen Schwelle zur objektiven Bewertung verschiedener Deep-Learning-Algorithmen.

2.3 Hochautomatische Datenauswertung der Positivdaten

Die Stückgröße für eine einzelne Iteration bei der hochautomatischen Datenauswertung ist normalerweise genau so groß wie der gesamte Tsinghua-Tencent 100k-Datensatz. Im Rahmen dieser Auswertung möchten wir das Potenzial der hochautomatischen Datenauswertung auf große Datensätze (> 1 Million) aufzeigen. Daher wählen wir ein vorab trainiertes Netzwerk mit einer Qualität, die normalerweise nach einigen Iterationen erreicht wird.

Der Gesamtaufwand für die hochautomatische Auswertung der 9176 Positivbilder in 157 Klassen betrug 2 Tage.

2.3.1 Quantität

Im Vergleich zu 22.235 spezifischen Verkehrszeichen beim Menschen, fand die hochautomatische Auswertung 24.863 Verkehrszeichen. Dies entspricht einem Plus von 2.628 Markierungen oder 12%. Der Trend war für jede Gruppe von Verkehrszeichen gültig.

Seite 5 von 13 22.11.2019 V1.02 Author: Tobias Manthey



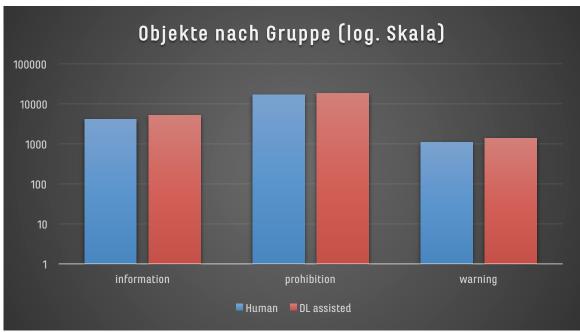


Abb. 3 Markierte Instanzen nach Gruppe

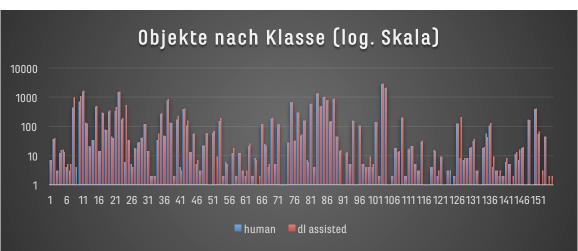


Abb. 4 Markierte Instanzen nach Klasse



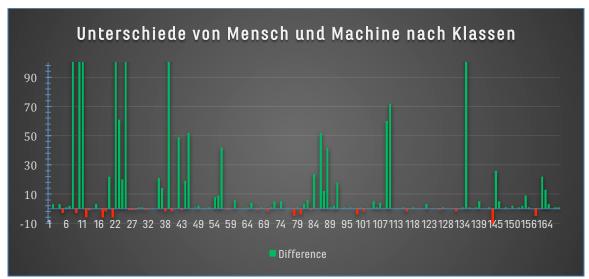
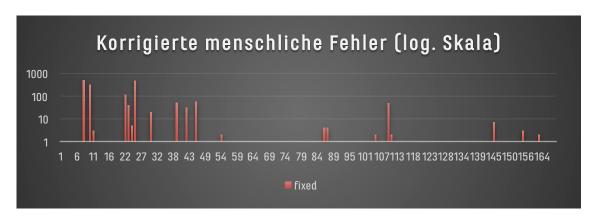


Abb. 5 Unterschiede begrenzt auf 100

2.3.2 Qualität

Zur automatischen Überprüfung der Qualität der menschlichen Datenauswertung wurde ein Deep Learning-Klassifikator verwendet. Dabei fand der Deep Learning-Klassifikator insgesamt 1699 falsche Markierungen in den Positivdaten. Dies entspricht einer menschlichen Genauigkeit von 93%. Der vorherrschende Fehler bestand darin, bekannte Klassen als "unbekannte" Verkehrszeichen in der entsprechenden Kategorie zu markieren. Die KI Klassifizierung kam hingegen auf 98,1% Genauigkeit.



2.4 Auswertung des Negativ-Datensatz

Der Negativ-Datensatz sollte der Definition entsprechend keine Verkehrsschilder enthalten. Die Analyse der 82.094 Bilder aus dem Negativ-Datensatz dauerte 2 Tage. Im Ergebnis wurden 2137 Schilder gefunden. Dies entspricht einer menschlichen falsch-negativ Rate von 2,5% bzw. einer Trefferquote von 97,5%. Die KI hingegen kommt auf eine falsch-negativ Rate von 1% bzw. auf eine Trefferquote von 99%.

Seite 7 von 13 22.11.2019 V1.02 Author: Tobias Manthey



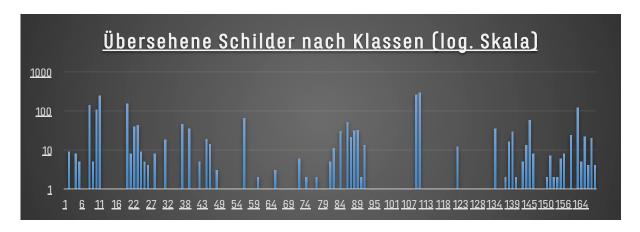




Abb. 6 Beispiele von falschen Markierungen

3 Analyse

3.1 Wo treten menschliche Fehler auf?

3.1.1 Gesamtkomplexität und subjektive Bewertung

Den Überblick über 157 Klassen ist für Menschen keine leichte Aufgabe. Dabei werden insbesondere seltene bekannte Verkehrszeichen, als "unbekannt" gekennzeichnet. Ein weiterer Grund ist die subjektive Beurteilung in Bezug auf geringfügige Variationen bekannter Zeichen, wie z.B. kleine Pfeile unter dem Symbol bekannter Zeichen.



3.1.2 Erwartungsgesteuerte Wahrnehmung

Während Menschen ein Sichtfeld von etwa 175 Grad haben, beträgt der Bereich zum scharfen Sehen nur 0,016 Grad. Nur durch schnelle Augenbewegungen wird dieser Bereich größer wahrgenommen. Um die Auswertung effizienter zu lösen, konzentriert sich der Mensch daher auf Bereiche, in denen

Seite 8 von 13 22.11.2019 V1.02 Author: Tobias Manthey



Verkehrszeichen normalerweise zu erwarten sind. Objekte, die an ungewöhnlichen Orten auftreten, werden daher häufiger übersehen.



Abb. 8 Menschen übersehen die Schilder in der Seitenstraße



Abb. 9 Maschinen suchen unabhängig vom Kontext

3.1.3 Kleine Objekte in komplexen Szenen

Kleine Objekte in komplexen oder verrauschten Szenen haben eine erhöhte Wahrscheinlichkeit von Menschen übersehen zu werden.

Seite 9 von 13 22.11.2019 V1.02 Author: Tobias Manthey





Abb. 10 Beispiel eines kleinen Warnschilds in einer komplexen Szene

3.1.4 Geringer Kontrast

Im Vergleich zu Menschen erkennt eine KI-Erkennung Objekte in Szenen mit geringem Kontrast in der Regel besser.



Abb. 11 Geringer Kontrast

3.2 Wo treten die Fehler bei Maschinen auf?

3.2.1 Sehr kleine Objekte

Menschen markieren teilweise sehr kleine Objekte bis zu 10x10 Pixel, obwohl diese nicht mehr mit Sicherheit erkennbar sind. Da Bilder typischerweise in Sequenzen vorliegen, kennt der Mensch manchmal ein Schild aus einem früheren Bild. Ein weiterer Effekt ist, dass das menschliche Gehirn teilweise fehlende Informationen mit Erfahrungswerten ergänzt. Dieser Effekt ist zu beobachten, wenn wir Objekte in kleineren Bildern manchmal besser erkennen als in größeren. Es ist jedoch

Seite 10 von 13 22.11.2019 V1.02 Author: Tobias Manthey



0

zweifelhaft, ob sich solche subjektiven Informationen positiv auf den Lernprozess einer Klauswirken. Wir empfehlen daher generell keine sehr kleinen Objekte zu markieren.



Abb. 12 Beispiele von sehr kleinen Markierungen





3.2.2 Unbekannte Gründe

In seltenen Fällen übersieht eine KI Objekte ohne ersichtlichen Grund. Dies geschieht häufiger bei Klassen mit einer geringen Anzahl im Datensatz.



Abb. 14 Menschen erkennen beide Schilder



Abb. 15 KI erkennt das rechte Schild nicht



3.3 Was steckt noch in den Daten?

Obwohl der Datensatz bereits 157 Klassen umfasst, fehlen noch zahlreiche Arten chinesischer Verkehrszeichen. Dazu gehören z.B. Richtungs- und Hinweisschilder, die aufgrund ihrer hohen Varianz deutlich schwerer zu erkennen sind als die Verkehrszeichen im Tsinghua-Tencent 100k Datensatz. Ein professionelles Netzwerk für chinesische Verkehrszeichen unterscheidet auf der Basis von mehreren Millionen einzelner Trainingsbeispiele bis zu 400 unterschiedliche Klassen. Im Vergleich zu Tsinghua-Tencent 100k umfassen diese Netzwerk das 2,5-fache der Klassen und sind mit dem 67-fachen der Daten trainiert



4 Zusammenfassung

Mit einer um mehr als 70% verminderten Fehlerrate ermöglicht die hoch-automatisierte Datenauswertung eine Reduktion von Aufwand, Kosten und Dauer um bis zu 90%. Darüber hinaus werden so Datensätze in der Größenordnung von Millionen Trainingsdaten und hunderten Klassen ermöglicht, mit denen Deep Learning Netzwerke eine bisher unerreichte Genauigkeit erzielen können.

Der größte Vorteil dieses Verfahrens ist jedoch die Möglichkeit eines aktiven Änderungsmanagements. Das anfängliche Klassenmodell basiert typischerweise auf menschlicher Erfahrung, bzw. Annahmen. Daher ist die Wahrscheinlichkeit hoch, dass ein Klassenmodell im Laufe des Projekts angepasst werden muss. Mit seinem enormen Vorteil in punkto Quantität ermöglicht die hoch automatisierte Datenauswertung Anpassungen des Klassenmodells während eines Projekts. Neben einer verbesserten Qualität und Quantität trägt dieser Ansatz daher dazu bei, das Projektrisiko erheblich zu reduzieren.

| | Mensch | Maschine | Unterschied |
|----------------------------|--------|----------|-------------|
| Klassifikationsgenauigkeit | 93% | 98,1% | +5,1% |

Seite 12 von 13 22.11.2019 V1.02 Author: Tobias Manthey



| Trefferquote | 97.5% | 99% | +1.5% |
|----------------------|-------------|--------|--------|
| Aufwand Positivdaten | ca. 18 Tage | 2 Tage | +900% |
| Aufwand Negativdaten | ca. 40 Tage | 2 Tage | +4000% |

Basierend auf Erfahrungswerten wurde angenommen, dass ein Mensch in Bezug auf die Positivdaten 500 Bilder und in Bezug auf die Negativdaten 2000 Bilder pro Tag auswerten kann.

5 Evotegra GmbH

Als Komplettlösungsanbieter für kundenspezifische Deep-Learning-Lösungen bietet die Evotegra GmbH das gesamte Spektrum von Beratung und Datenservices bis hin zur Prozessintegration an. Unsere Kernkompetenz ist die kostengünstige Umsetzung von individuellen Lösungen der Bilderkennung. Das Angebot richtet sich insbesondere an kleine und mittlere Unternehmen, die keine eigene KI-Kompetenz aufbauen.

Vielen Dank an Bruno Marotta, Nikolas Markou / Elekti und dem Team von Evotegra.

EvoTegra GmbH Kreuzwingert 11 D - 55296 Gau-Bischofsheim

Tel.: +49 171 / 402 4242 E-Mail: info@evotegra.de Homepage: www.evotegra.de

Seite 13 von 13 22.11.2019 V1.02 Author: Tobias Manthey