

Institut für Bildungswissenschaften
Sommersemester 2019

**Umsetzung der Kouninschen
Klassenführungsdimension „Valenz und
Herausforderung“ in meinem beruflichen
Handeln**

eingereicht von:

Matrikelnr.:

MEd für Gymnasien und Gesamtschulen

Fächerkombination: Mathematik & Informatik

Dozentin: Dr. Kristina Antonette Frey

in: Die weltberühmten

Klassenführungsdimensionen Jacob Kounins
(066861)

Leistungsart: Prüfungsleistung

Abgabefrist: 01.09.2019

Inhaltsverzeichnis

1	Introduction	1
2	Vorstellung des Datensatzes	2
2.1	Stand der Technik	3
3	Random Forests	3
3.1	Theorie	3
3.2	Praxis und Ergebnisse -> unzufrieden	3
4	Gradient Boosting	3
4.1	Theorie	3
4.2	Implementierungsdetails	5
4.3	eigene Variationen	5
5	Vorstellung Results	6
6	Bewertung der Results	6
7	classification of users	6
8	Auswertung	6
9	Fazit	6
10	Literaturverzeichnis	7
	Eigenständigkeitserklärung	8

Beim Nachweis von Zitaten und Literatur wenden wie die von Unisa vorgeschriebene Harvard-Methode an und folge dabei den Regeln in:
Christof Sauer (Hg.) 2004. Form bewahren: Handbuch zur Harvard-Methode. (GBFE-Studienbrief 5). Lage: Gesellschaft für Bildung und Forschung in Europa e.V. 1. Auflage.

Abstract

1 Introduction

Durch die Vernetzung unserer digitalen Geräte, die unseren Alltag prägen und mitgestalten, ist eine große Menge an Daten vorhanden. Die Verwendung solcher Daten und der damit verbundenen Informationen ist ein Kernelement der heutigen Digitalgesellschaft. Programme, die solche Daten verwerten können, um möglichst viele Informationen zu generieren, spielen in unserer modernen Gesellschaft eine große Rolle. Dies ist nicht nur auf der Ebene von kontextbezogene Werbung von großem Interesse, sondern auch im zur Verbesserung unseres Medizinwesens und unserer Verhaltensweisen. Beispielsweise optimieren schon heute viele Personen unter Zuhilfenahme digitaler Fitnesstracker, etc. ihren persönlichen Lebensstil. Auch das Herstellen eines gesunden Schlafrhythmus ist durch die Erkennung von Schlafphasen möglich.

Im Kontext des Gesundheitswesens ist diese Technologie besonders zentral, denn einerseits kann sie präventiv eingesetzt werden. Beispielsweise können Personen über ihre eigenen Gewohnheiten informiert und bezüglich ihrer Risiken sensibilisiert werden, was sie zu einem gesünderen Lebensstil bewegen kann, welcher wiederum von der Technologie unterstützt werden kann. Aus medizinischer Perspektive ermöglicht man eine Überprüfung darüber, wie und wie gesund eine Person lebt, um dadurch Krankheiten gezielter bekämpfen zu können. Hierfür ist es beispielsweise wichtig, aus verschiedensten Sensordaten Aktivitäten sicher ableiten zu können.

Die hier vorliegende Ausarbeitung entstand im Rahmen eines Informatikpraktikums in der Vorlesung Mustererkennung, geleitet von Prof. Dr. Xiaoyi Jiang und Sören Klemm mit dem Thema Behavioral context recognition in-the-wild from mobile sensors. Ziel des Projektes war es, auf der Basis des Extrasensory Datensatzes Multi-Label-Klassifizierung durchzuführen. Im Wesentlichen sollten dabei gesammelten Sensordaten Aktivitäten zugeordnet werden, in einer Zusatzaufgabe sollten die Benutzer anhand ihrer Daten erkannt werden können.

Die besondere Herausforderung des Projektes bestand darin, dass einerseits die Methodik nicht vorgegeben war und andererseits Multi-Label-Klassifizierung mit einem Datensatz ausgeführt werden sollte, der viele missing labels beinhaltete.

2 Vorstellung des Datensatzes

Der ExtraSensory Datensatz wurde in den Jahren 2015 und 2016 von Yonatan Vaizman and Katherine Ellis unter Aufsicht von Professor Gert Lanckriet erhoben. Er basiert auf den Sensordaten von Smartphones und Smartwatches, welche von 60 Teilnehmern in minütigem Takt produziert wurden. Das Besondere dieser Werte besteht darin, dass sie von normalen Alltagsgegenständen erzeugt wurden, mehrere Sensoren parallel verwendet wurden und die Daten in einem natürlichen und nicht gestellten Kontext entstanden. Die Sensoren umfassten unter anderem ein accelerometer, gyroscope, location und audio und wurden von beinahe allen Personen in einem Großteil der Aufnahmen verwendet.

Die Zuordnung der Segmente zu Aktivitäten fand im Großteil durch die Benutzer selbst statt, sie besaßen dabei die Option auf vorgegebene Labels oder selbst erstellte für die Beschreibung zurückzugreifen. Im Endeffekt ergaben sich so 377,346 Datenpunkte, welche mit einem oder mehreren der finalen 51 Labels beschrieben wurden. Die Labels repräsentieren dabei Mannigfaltiges, wie etwa in class, singing, stairs (going up), stairs (going down) oder talking.

Das bedeutet, dass zu jedem Datenpunkt ein Labelvektor mit 51 Einträgen existiert. Jeder Eintrag kann die Werte yes oder no beinhalten, muss aber keinen Wert besitzen, wenn die bewertende Person keine Informationen bezüglich des korrespondierenden Labels angab (NaN). Die folgende Betrachtung der Datenpunkte zeigt, dass in jedem Datenpunkt viele Labels unbetrachtet blieben:

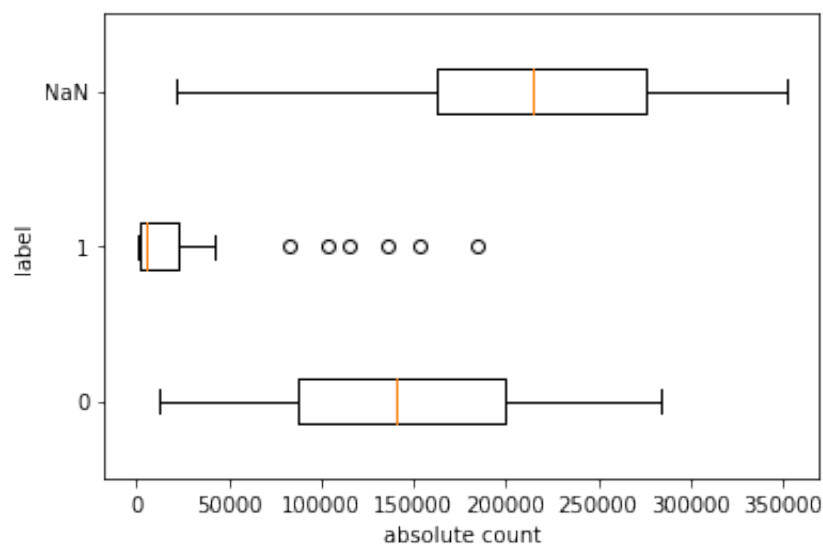


Abb. 1: XXXX

Abbildung 1 visualisiert die Verteilung der gesetzten Label-Werte als Box-Plot Diagramm. Zu

erkennen ist, dass für die meisten Labels ein Großteil der Datenpunkte keinen Wert enthält - was die Klassifikation erschwert. Das heißt, es gibt für viele Labels eine große Anzahl von Datenpunkten, die erst einmal keine Information bezüglich des Labels beinhalten.

Weiter ist zu sehen, dass im Median zu jedem Label nur 5,153 yes-Eintragungen existieren und lediglich für sehr wenige Labels überhaupt mehr als 10% der Datenpunkte eine yes-Eintragung. Es gibt also viele Label-Klassen, die ausgesprochen wenig Repräsentanten beinhalten.

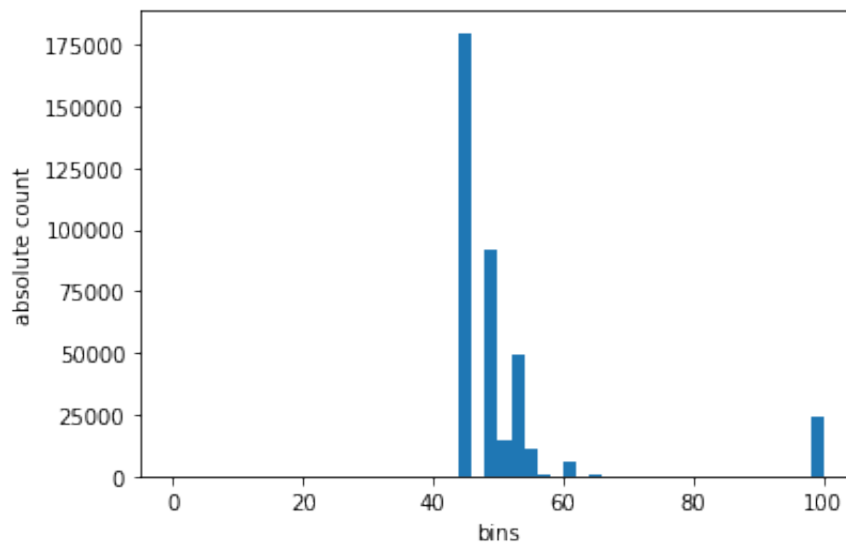


Abb. 2: XXXX

Zuletzt ist zu erkennen, dass in allen Datenpunkten mindestens 44% der 51 Labels nicht bewertet sind. Es gibt also keine Datenpunkte die zu allen Labels gleichzeitig Aussagen machen.

2.1 Stand der Technik

3 Random Forests

3.1 Theorie

3.2 Praxis und Ergebnisse -> unzufrieden

4 Gradient Boosting

4.1 Theorie

Der folgende Abschnitt orientiert sich an Chen, Guestrin (2016).

Beim Gradient Boosting geht es darum, iterativ eine Menge von gradient trees zu erzeugen, die ein gegebenes Klassifikationsproblem möglichst effektiv lösen.

Gradient trees sind im Wesentlichen decision trees, nur dass jedem Blatt des Baumes einen Wert, dessen Gewicht, w_i zugewiesen wird. Folglich kann ein gradient tree mit T Blättern und entsprechenden Gewichten (w_1, \dots, w_T) als eine Funktion $f : \mathbb{R}^m \rightarrow \mathbb{R}$ aufgefasst werden. Jedem Datenpunkt $x \in \mathbb{R}^m$ wird unter f das Gewicht des zu x korrespondierenden Blattes zugeordnet $f(x) = w_{q(x)}$. $q : \mathbb{R}^m \rightarrow T$ ist dabei die Funktion, welche einem sample x die Indexnummer des zu x gehörigen Blattes zuweist. Weiter enthält $I_j := \{i \mid q^{-1}(j) = x_i\}$ die Indizes derjenigen samples, welche auf das Blatt Nummer j verweisen. Die Menge der gradient trees bezeichnen wir mit \mathcal{F} .

Ein tree ensemble model $\{f_1, \dots, f_K \mid f_i \in \mathcal{F}\}$ bestehend aus K gradient trees wird nicht, wie von decision trees bekannt, durch Voting ausgewertet, sondern indem die Funktionswerte des samples x der einzelnen gradient trees in der ensemble Funktion $\Phi = \sum f_k$ aufsummiert werden. Das predicted outcome \hat{y} eines samples x ergibt sich damit zu:

$$\hat{y} = \Phi(x) = \sum_{k=1}^K f_k(x)$$

Ist weiter eine Menge $X = \{(x_1), \dots, (x_n)\} \subset \mathbb{R}^m$ von n Beispielen mit jeweils m features und dem zugehörigen ground truths $\{y_1, \dots, y_n\}$ gegeben, so erhalten wir die regularization funktion:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \Phi(x_i)) + \sum_{k=1}^K \Omega(f_k), \text{ wobei } \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|_2^2$$

Bei l handelt es sich um unsere loss-function, eine konvexe und mindestens 3 Mal differenzierbare Funktion. Ω ist in diesem Kontext ein Strafterm und überwacht innerhalb von \mathcal{L} die Struktur des Baumes. Die Wahl von γ beeinflusst die Anzahl der zulässigen Blätter, λ die Gewichte ebendieser. Somit ist Ω eine Vorbeugemaßnahme gegen overfitting. Infolgedessen präferiert das mit \mathcal{L} erschaffene Modell gradient trees, die von einfacher Gestalt sind.

Das tatsächliche gradient tree boosting findet statt, indem iterativ neue Bäume dem Ensemble hinzugefügt werden. Im t -ten Iterationsschritt besteht unser Ensemble aus den bereits erzeugten gradient trees $\{f_1, \dots, f_{t-1}\}$. Ziel ist es, den neuen gradient tree f_t so zu wählen, dass er unser model am meisten verbessert. Das ist genau dann der Fall, wenn

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \Phi(x_i)) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

minimal ist. Da die $\Omega(f_k)$ für die Bäume f_1, \dots, f_{t-1} bereits feststehen, spielen sie in der Betrachtung keine Rolle. Weiter stehen von $\Phi(x_i) = f_1(x_i) + \dots + f_{t-1}(x_i) + f_t(x_i)$ bereits die ersten $t - 1$ Summanden fest. Wir bezeichnen diese Summe mit \hat{y}^{t-1} , da sie das predicted outcome nach dem $t - 1$ -ten Iterationsschritt des ensembles wiedergibt.

Damit können wir (1) auch formulieren als:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (2)$$

Um die Abhängigkeit von l und f_t zu trennen, verwenden wir die 1D-taylor series-expansion an der Stelle $(y_i, \hat{y}^{(t-1)})$ auf die Funktion l und nur in der zweiten Koordinate an. Obwohl es sich bei l um eine 2D Funktion handelt, finden Änderungen in unserem Kontext nur in einer Koordinate statt, weshalb es berechtigt ist, sie als eine 1D Funktion aufzufassen, etwa wie $l(y_i, x) = l_{y_i}(x)$. Es ergibt sich:

$$l_{y_i}(\hat{y}^{(t-1)} + f_t(x_i)) = l_{y_i}(\hat{y}^{(t-1)}) + l'_{y_i}(\hat{y}^{(t-1)}) \cdot f_t(x_i) + \frac{1}{2} l''_{y_i}(\hat{y}^{(t-1)}) \cdot f_t(x_i)^2 + R_{\text{taylor}} \quad (3)$$

Wir bezeichnen im Folgenden $g_i = l'_{y_i}(\hat{y}^{(t-1)}) = \frac{\partial l}{\partial \hat{y}^{(t-1)}}(y_i, \hat{y}^{(t-1)})$ und $h_i = l''_{y_i}(\hat{y}^{(t-1)}) = \frac{\partial^2 l}{\partial^2 \hat{y}^{(t-1)}}(y_i, \hat{y}^{(t-1)})$. Die g_i und h_i haben die Eigenschaft, dass es sich um Konstanten handelt, da sowohl y_i als auch $\hat{y}^{(t-1)}$ im aktuellen Iterationsschritt bekannt sind. Angewandt auf $\mathcal{L}^{(t)}$ ergibt sich somit:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n (l(y_i, \hat{y}^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2) + \Omega(f_t) \quad (4)$$

Da die konstanten Terme $l(y_i, \hat{y}^{(t-1)})$ nicht von f_t abhängen, können wir sie im Rahmen des Optimierungsproblems wegfällen lassen und erhalten das vereinfachte objective:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n (g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2) + \Omega(f_t) \quad (5)$$

4.2 Implementierungsdetails

-> Multilabel einzeln trainieren

4.3 eigene Variationen

-> Hyperparametern -> Geschwindigkeit (GPU)

5 Vorstellung Results

-> Metric

6 Bewertung der Results

7 classification of users

8 Auswertung

- Auswirkungen verschiedener Gewichte der Datenpunkte und sowas (Modifikationen an Input, etc) - NaN Handling - Findung der Parameter

9 Fazit

10 Literaturverzeichnis

DRÜKE-NOE, C. 2014. *Aufgabekultur in Klassenarbeiten im Fach Mathematik – Empirische Untersuchungen in neunten und zehnten Klassen*. Wiesbaden: Springer Spektrum.

Anti-Plagiatserklärung

Erklärung des Studierenden

Hiermit versichere ich, dass ich die vorliegende Hausarbeit mit dem Namen „Umsetzung der Kouninschen Klassenführungsdimension „Valenz und Herausforderung“ in meinem beruflichen Handeln“ selbstständig verfasst habe, und dass ich keine anderen Quellen und Hilfsmittel als die angegebenen benutzt habe und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

Ort, Datum

Tony Prange