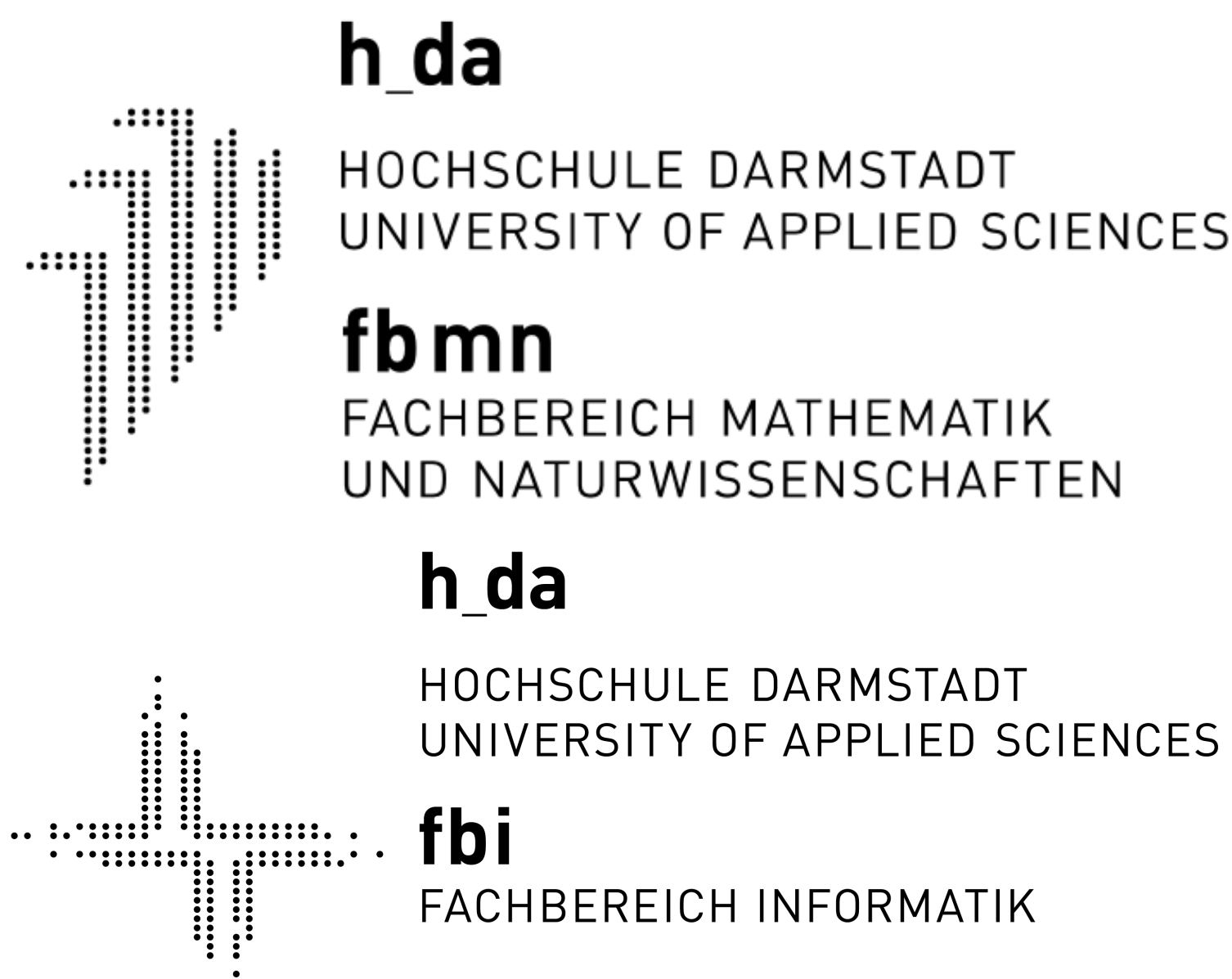


Stimmungsquantifizierung für den Preis von Bitcoin mit Deep Learning

Bernhard Preisler

Hochschule Darmstadt, Fachbereiche Mathematik und Naturwissenschaften & Informatik

preisler.bernhard@gmail.com



Zusammenfassung

Unser Ziel ist mittels Deep Learning-Modellen Sentimentindizes aus Tweets von Twitter zu bilden, um den Einfluss dieser Stimmung auf die Preise von Bitcoin zu messen. Diese Hypothese weisen wir über einem Zeitraum vom 10.01. bis 31.08.2018 nach, indem wir zu Beginn die Tweets herunterladen und aufbereiten. Danach lassen wir 1.857 Tweets mittels Amazon Mechanical Turk nach ihrer Stimmung annotieren und wenden diese auf unterschiedliche maschinelle Verfahren an. Im Fokus steht der Vergleich von herkömmlichen Modellen zu den Deep Learning-Ansätzen, die ähnliche oder schlechtere Ergebnisse hinsichtlich der Klassifikationsgüte liefern. Mittels den Modellen generieren wir unsere Sentimentindizes, wovon wir einen vielversprechenden Index entnehmen und diesen detaillierter untersuchen. Dabei finden wir erste hochsignifikante Hinweise auf den Einfluss der Stimmung von Twitter auf den Kurs von Bitcoin.

Einführung

In einer Zeit mit einer quasi unbegrenzten Möglichkeit an frei zugänglichen Daten spielt die Textanalyse eine immer größer werdende Rolle. Bereits ohne Computer legten erfolgreiche Unternehmen Wert auf die Meinung anderer und versuchten diese aus Zeitungen oder Meinungsumfragen zu gewinnen und mit ihrer persönlichen Meinung abzugleichen. Unsere initiale Hypothese wurde durch Cindy K. [7] inspiriert. Als übergeordnetes Thema nimmt sie sich das Nachweisen von Animal Spirits, dt. animalische Instinkte oder das Herdentreiben, heraus, das Keynes J. im Jahr 1936 [2] prägt. Cindy untersucht in ihrer Veröffentlichung das Sentiment von 34 unterschiedlichen Lokalzeitschriften in Bezug zu deren Städten verteilt über den USA von 2000 bis 2013. Dabei entdeckt sie, dass der Sentiment der Lokalzeitschriften eine um zwei Jahre versetzte hohe Korrelation in Bezug zum Immobilienindex aufweist.

Zu Beginn beschreiben wir unsere Datensätze. Um die Stimmung auf Twitter zu gewinnen, laden wir als erstes Tweets mit Hashtags über Kryptowährungen herunter. Die Tweets bereiten wir so auf, dass die maschinellen Lernverfahren diese verarbeiten können. Im nächsten Schritt generieren wir uns einen Trainingsdatensatz mit Amazon Mechanical Turk (AMT). Dieser Datensatz dient als Grundlage für unsere Modellbildung. Zum Vergleich ziehen wir erprobte Verfahren aus dem Bereich des maschinellen Lernens heran. Ein großer Teil der Thesis beschäftigt sich mit der Modellbildung mittels Deep Learning (DL), die wir vergleichen und verbessern. Gegen Ende der Thesis erstellen wir mit unseren Modellen eine Zeitreihe, die wir diskutieren. Weiterhin versuchen wir mit dieser Zeitreihe ein Verhältnis zwischen der Stimmung auf Twitter und dem Bitcoin-Niveau herzustellen. Abschließend besprechen wir weitere Anknüpfungspunkte an diese Arbeit.

Datensätze

Als *Datengrundlage* dienen uns die Tweets aus dem Zeitraum vom 10.01. bis 31.08.2018, die wir mittels der Twitter-API herunterladen. Dazu verwenden wir PHP in Verbindung mit einem Cronjob, der nach Hashtags sucht. Die hochfrequentierten Hashtags, bspw. #btc oder #blockchain, werden jede Minute und die niedrigfrequentierten Tweets alle acht Minuten abfragt. Der Datensatz besteht aus 59,14 Millionen Tweets mit insgesamt 1,36 Milliarden Wörtern.

Bevor wir die Tweets quantifizieren, werden sie *bereinigt*, indem wir die doppelten Tweets, Retweets, Stoppwörter, Hashtags, Nutzer, HTML und URLs entfernen und alle Wörter in die Kleinschreibung konvertieren. Als Messwert dient uns die Feature Importance.

Supervised Learning benötigt ein Ziel, auch Target oder Output genannt. Weswegen wir ausgewählte Tweets mit Amazon Mechanical Turk annotieren lassen. Die Tweets werden mittels dem vaderSentiment [1] vorselektiert, sodass wir Tweets mit einer starken Stimmung erhalten. Da jeder Algorithmus bzw. jedes Modell Fehlklassifikation erzeugt, verwenden wir Amazon Mechanical Turk, um diesen zu entfernen. Insgesamt lassen wir 1.857 Tweets von sieben unterschiedlichen Bewertern annotieren. Die *annotierten Tweets* bestehen aus 1.042 positive, 727 negative und 88 neutrale Tweets. Diese werden für die Evaluation der Modelle verwendet.

Nachdem Vergleich der Modelle verwenden wir diese, um die Stimmung zu quantifizieren. Dies geschieht, indem die Modelle die Stimmung des bereinigten Datensatzes tageweise berechnen. Die daraus entstehenden Sentimentindizes reichern wir mit den dementsprechenden Preisen von Bitcoin an. Dadurch erhalten wir eine *Zeitreihe*, die wir nutzen, um einen Einfluss der Stimmung von Twitter auf den Kurs von Bitcoin nachzuweisen.

Deep Learning

In diesem Abschnitt stellen wir die Verfahren vor, die wir zum Bilden der Sentimentindizes verwenden. Grob können wir die Verfahren in die zwei Gruppen Vergleichsverfahren und DL-Verfahren einteilen. Ersteres dient zum Vergleich der Güte zu den DL-Verfahren und bestehen aus den zwei lexikalischen Verfahren vaderSentiment und sentimentr und zwei maschinellen Lernverfahren RandomForest und Support-Vector-Machines (SVM).

Das erste DL-Modell entnehmen wir dem Paper *Hierarchical Deep Learning for Text Classification (HDLTex)* [4]. Das Künstliche Neuronale Netz (KNN) besteht aus acht Hidden Layern mit jeweils 1.024 Neuronen. Bis auf den Dropout, der auf 0,5 gesetzt wurde, und die Batchsize von 128 werden die Standardinstellungen von Keras verwendet. Da das Modell mit ca. 8 Millionen einstellbaren Parametern für eine Klassifizierung extrem groß ist, liegt unser Fokus auf der Reduzierung der Parameter. Für unser angepasstes Modell reduzieren wir die Anzahl der Neuronen pro Layer auf 128, 256, 512, 1024, 1024, 512, 256 und 128. Außerdem versuchen wir das Overfitting mittels höheren Dropout-Raten von 0,75 und weiteren L2-Restriktionen im Hidden Layer 4 mit 0,01, im Hidden Layer 5 mit 0,015 und im Hidden Layer 6 mit 0,01 zu verringern. Dabei sinkt unsere Parameteranzahl auf 2,1 Millionen.

Unser zweites Modell entnehmen wir dem Paper *Convolutional Neural Networks for Sentence Classification (CNNSC)* [3]. Als Input nutzt das Netz die Word2Vec-Vektoren, die aus dem Google News-Archiv generiert wurden. Das Netz besteht aus einem Convolutional Neural Networks-Layer (CNN-

# Methoden	F1	Acc
1. CNNSC	0,78	0,80
2. ang. CNNSC	0,83	0,85
3. HDLTex	0,72	0,75
4. ang. HDLTex	0,75	0,77
5. RandomForest	0,76	0,73
6. SVM	0,77	0,79
7. vaderSentiment	0,83	0,85
8. sentimentr	0,80	0,79

Tabelle 1: Prognoseergebnisse aller Verfahren. Die Zeilen 1-4 repräsentieren die Ergebnisse der DL-Modelle und die Zeilen 5-8 die der maschinellen und der lexikalischen Vergleichsverfahren.

Layer), Max-Pooling-Layer und einem KNN-Layer. Der CNN-Layer besteht aus 100 Kernels der Größe 3x3, einen Drop-Out von 50 % und einer Regularisierung von 3. Die Anzahl der einstellbaren Parameter summiert sich auf ca. 2 Millionen. Passen wir das Modell auf unseren Datensatz an, erhalten wir die besten Ergebnisse mit den Twitter-Vektoren als Input, entfernen den Regularisierer, erhöhen die Anzahl der Kernel von 100 auf 128 und verringern die Größe der Kernels von 3x3 auf 1x1. Die Anzahl der einstellbaren Parameter summiert sich auf ca. 1,3 Millionen.

Tabelle 1 zeigt die Prognoseergebnisse der Verfahren. Wir finden ein Modell (CNNSC), das ähnlich gut abschneidet, wie unser bestes Vergleichsverfahren vaderSentiment. Die Güte der weiteren Verfahren fällt schlechter aus.

Bitcoin und Sentiment

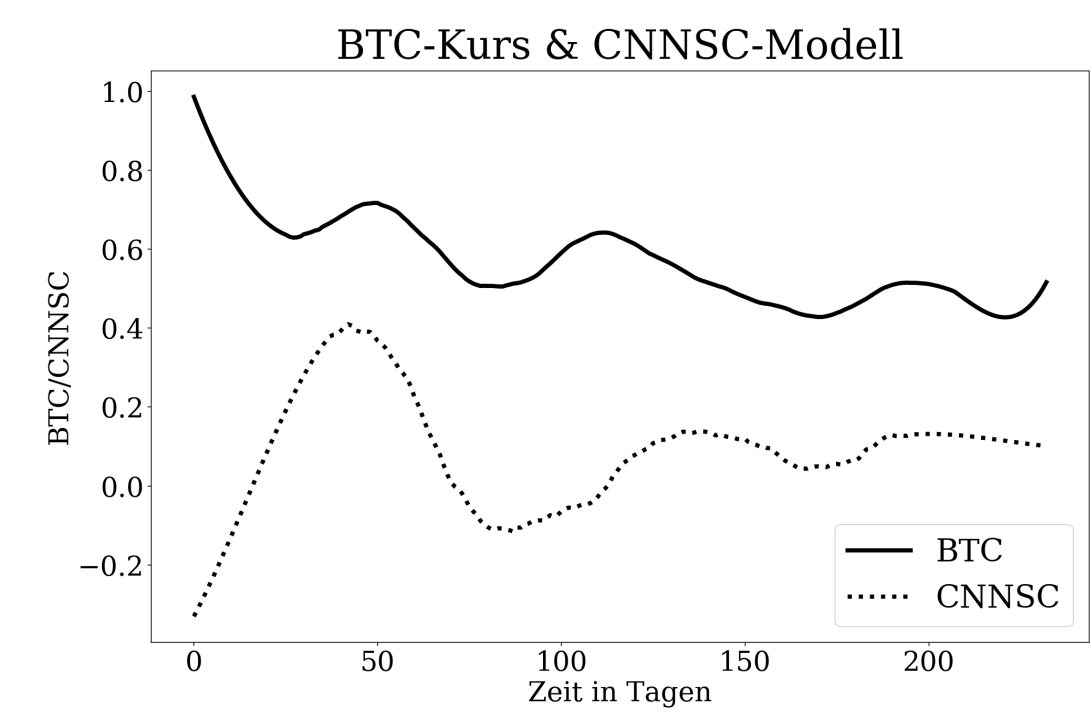


Abbildung 1: BTC-Kurs und CNNSC-Modell. Die x-Achse repräsentiert die Tage in einem Zeitraum vom 10.01. - 31.08.2018 und die y-Achse das Bitcoin-Niveau und den Sentiment vom CNNSC-Modell.

nearen Regression (1), dass der Bitcoin-Kurs die Stimmung von Twitter enthält. ΔS_0 steht für die CNNSC-Rendite von gestern auf heute, ΔS_1 von vorgestern auf gestern, ΔS_2 von vor drei Tagen auf vorgestern, ΔB für das Bitcoin-Niveau von gestern auf heute und Δx die ökonomischen Variablen gestern auf heute.

$$\Delta B = \Delta S_0 + \Delta S_1 + \Delta S_2 + \Delta x \quad (1)$$

Tabelle 2 zeigt die hochsignifikanten Koeffizienten. Die Spalte (1) besteht ausschließlich aus den CNNSC-Renditen, wohingegen wir in der Spalte (2) den linearen Zeittrend (LZ) eliminieren, der wenig Einfluss auf unsere Koeffizienten hat. Dadurch können wir behaupten, erste Hinweise im Rahmen der Thesis auf den Einfluss des Kurses von Bitcoin durch die Stimmung von Twitter gefunden zu haben.

Ergebnisse

Unser bestes DL-Modell erreicht die Klassifikationsgüte unserer besten Vergleichsverfahren. Außerdem erkennen wir anhand der restlichen DL-Modelle, dass sie natürlicherweise größere Datensätze benötigen, um klassische Verfahren zu übertreffen. Damit können wir die allgemeine Hypothese bestätigen, dass DL-Verfahren größere Datensätze benötigen, um klassische Verfahren zu übertreffen.

Weiterhin finden wir mittels dem CNNSC-Sentiment in Verbindung mit der multiplen linearen Regression erste hochsignifikante Hinweise auf den Einfluss der Stimmung von Twitter auf das Bitcoin-Niveau.

Ausblick

Neben etlichen Einflussfaktoren wie die Modell- oder die Hyperparamaterwahl, spielt die Menge der annotierten Tweets eine wesentliche Rolle bei der Klassifikation mittels DL-Modellen. Weiterhin sind wir uns sicher, dass in der Datenbereinigung weitere Schritte durchgeführt werden können, um die Güte der Klassifikation und somit die Qualität des Sentiments zu erhöhen. Da aus zeitlichen Gründen nur ein Sentimentindex untersucht wurde, besteht die Chance weitere hochsignifikante Verbindungen von Twitter zum Bitcoin-Kurs mittels den restlichen Sentimentindizes herzustellen.

Literatur

[1] C. J. Hutto. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. Association for the Advancement of Artificial Intelligence, 2014.

[2] John Maynard Keynes. *The General Theory of Employment, Interest and Money*. Springer, 2018.

[3] Yoon Kim. *Convolutional Neural Networks for Sentence Classification*. Association for Computational Linguistics, 2014.

[4] Kamran Kowsari, Donald E. Brown, and Mojtaba Heidarysafa. *Hierarchical Deep Learning for Text Classification*. arXiv, 2017.

[5] E. Martínez-Cámara. *SINAI: Machine Learning and Emotion of the Crowd for Sentiment Analysis in Microblogs*. Association for Computational Linguistics, 2013.

[6] Whitney K Newey and Kenneth D West. *A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelationconsistent Covariance Matrix*. National Bureau of Economic Research Cambridge, Mass., USA, 1986.

[7] Cindy K. Soo. *Quantifying Sentiment with News Media across Local Housing Markets*. The Review of Financial Studies, 2017.

	log-Rendite Bitcoin	
	(1)	(2)
ΔS_0	0,183***	0,183***
ΔS_1	0,217***	0,216***
ΔS_2	-0,171***	-0,174***
LZ		✓
Ang. R ²	0,102	0,117

Notiz: *p<0,1; **p<0,05; ***p<0,01

Tabelle 2: Drei Tage Sentiment-Prognose für die Bitcoin-Rendite. S_0 steht für die heutige Rendite, S_1 die gestrige Rendite und S_2 für die Rendite von vor zwei Tagen des CNNSC-Sentiments. Die Koeffizienten sind mit NeweyWest [6] berechnet. Der *Lineare Zeittrend (LZ)* eliminiert den Trend.