

INTERVIEW

Analyse von CEO-Stimmen zur Prognose von Gewinnerwartungen mithilfe von KI-Modellen

CL — Prof. Dr. Christian Langmann

DR — Dr. Doron Reichmann

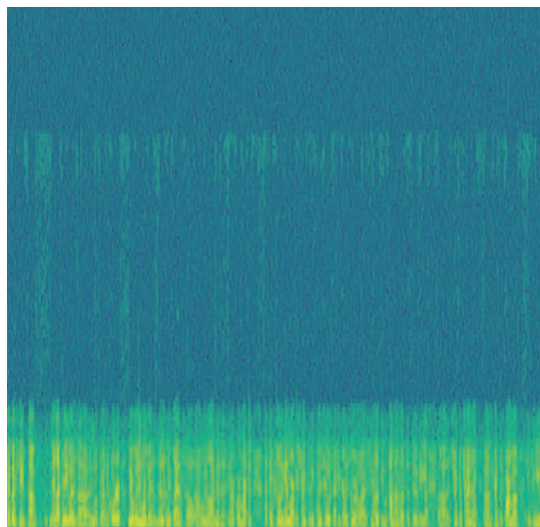
CL Lieber Herr Reichmann, Sie haben mit Ihrer Forschergruppe eine vielbeachtete Studie zu CEO-Stimmen auf Bilanzpressekonferenzen durchgeführt, um daraus Rückschlüsse auf die Gewinnerwartungen zu schließen. Könnten Sie für unsere Leserschaft den Aufbau und die zentralen Ergebnisse der Studie zusammenfassen?

DR Ja, sehr gerne. Die zentrale Frage, die wir uns in der von Ihnen angesprochenen Studie stellen, ist, ob wir über die Stimme von CEOs Gewinnerwartung prognostizieren können. Dabei kann man sich zunächst die Frage stellen: Warum machen wir das überhaupt? Aus der Psychologie wissen wir, dass die Stimme immer auch Emotion kommuniziert und damit mehr Informationen preisgibt als nur das gesprochene Wort. Letzteres haben viele von uns schon in Situationen erlebt. Man sieht zum Beispiel einen Vortrag und bei einer Zwischenfrage kommt die Vortragende ins Schwitzen. Die Vortragende will natürlich nicht preisgeben, dass sie gestresst ist oder die Frage nicht wirklich beantworten kann. Dennoch merken wir anhand nonverbaler Kommunikation, dass die Person verunsichert oder gestresst ist. An dem Beispiel sieht man aber, dass nonverbale Kommunikation in diesem Moment Informationen preisgibt, die wir gar nicht preisgeben wollen.

Hierbei gilt die Stimme als besonders ausdrucksstarkes Medium der nonverbalen Kommunikation. Die Stimme ist sehr komplex, wir können sie zudem nur sehr schwer kontrollieren. In der Medizin gilt der Sprachprozess als eine der kompliziertesten Aktivitäten, die wir überhaupt im menschlichen Körper tagtäglich ausüben.

Das angesprochene Wissen über die Stimme können wir nun nehmen und auf die Situation in Bilanzpressekonferenzen übertragen. Die Idee dabei ist, dass die Stimme von CEOs möglicherweise weitere Informationen (z.B. über die Zukunftsaussichten des Unternehmens) preisgibt, und zwar über das gesprochene Wort des Managements hinaus. Und um diese Frage zu analysieren, haben wir ein Sample von über 8.000 Bilanzpressekonferenzen zusammengestellt und deren Audiospuren – also MP3-Aufnahmen – ausgewertet. Wir hatten quasi für jede Pressekonferenz eine eigene Audiodatei und haben darin jeweils die Sprache des Managements separiert. Insbesondere die Frage-Antwort-Phasen von den Bilanzpressekonferenzen waren dabei interessant, weil diese unvorbereiteter sind und man da wirklich in eine echte Interaktion gerät. Die Präsentationen zu Beginn werden dagegen oft einfach nur abgespult.

Abbildung 1: Spektrogramm



Die Audiodaten wurden dann zunächst in ein Spektrogramm umgewandelt (siehe Abbildung 1). Ein Spektrogramm stellt im Wesentlichen eine Visualisierung von Audiospuren dar. Die X-Achse zeigt den Zeitpunkt der Audiospur an, die Y-Achse zeigt die Frequenz (also die Stimmhöhe) der Audiospur an und die Farbe gibt die Intensität (also Lautstärke) der jeweiligen Frequenz wieder. Die Daten aus den Spektrogrammen haben wir dann in ein KI-Modell auf Basis eines neuronalen Netzes gespeist, das die Spektrogramme mithilfe von Algorithmen aus dem Bereich image recognition analysiert hat. Das sind ähnliche Algorithmen wie diejenigen, mit denen Sie zum Beispiel Bilder erkennen. Beispielsweise könnte ein derartiger Algorithmus auch dafür genutzt werden, um in einem Bild einen Husky von einem Wolf zu unterscheiden. In unserem Beispiel mit den Bilanzpressekonferenzen wissen wir aus der Historie, welches der untersuchten Unternehmen im darauffolgenden Jahr besser oder schlechter performt hat. Diese Erkenntnis zur Performance verbindet das Modell mit den Spektrogrammen. Das Modell sucht nun nach Mustern in den Spektrogrammen, um eine Prognose für die Unternehmensperformance abzuleiten.

Die Prognose, die das Modell ausgespuckt hat, war überraschend gut und hat besser funktioniert, als wenn wir Bilanzkennzahlen oder Texte aus den Bilanzpressekonferenzen eingespeist haben. Die gewonnenen Prognosen könnten auch dazu genutzt werden, um Trading-Strategien zu

optimieren und Forecasts von professionellen Analysten zu verbessern. Das sind unsere zentralen Ergebnisse.

CL Mithilfe von neuronalen Netzen haben Sie also die Strukturen und Muster in den Spektrogrammen untersucht. Soweit mir bekannt ist, lassen sich mit neuronalen Netzen aber keine linearen Zusammenhänge und damit keine Faktoren herauskristallisieren. Das heißt, wir könnten jetzt nicht sagen, welche Rolle das Geschlecht im Sinne einer Männer- oder Frauenstimme für die Prognose gespielt hat, richtig?

DR Ja, genau. Wir können heute nicht genau sagen, welches die bestimmenden Faktoren sind. Da muss man aber auch gleichzeitig sagen, generell wissen wir in der Forschung, also abseits der Wirtschaftswissenschaften, noch gar nicht genau, wie Emotionen in Stimmen entstehen. Nehmen wir als Beispiel eine Person, die wütend wird. Es gibt Personen, die werden dann leise, kommunizieren ruhig. Andere werden laut, erhöhen die Lautstärke der Stimme und ihre Stimmhöhe. Dennoch verstehen wir als Menschen, dass diese Person gerade wütend ist. Wir wissen aber gar nicht genau, warum wir das wissen bzw. wie wir das als Menschen messen können. Die Stimme wirkt hierbei anders als Emotion im Gesicht. Im Gesicht kann man anhand von sogenannten facial landmarks gut vorhersagen, wie Mimik wirkt. Wenn sich also diese oder jene Landmark nach oben bewegt, z.B. jemand die Augenbrauen nach oben zusammenzieht, spiegeln sie eine bestimmte Emotion wider. Das können wir bei der Stimme derzeit noch nicht direkt in Beziehung setzen. Dennoch verstehen wir als Mensch die emotionale Kommunikation über die Stimme. Da wir also nicht genau wissen, welche spezifischen Merkmale für die Identifikation von Emotionen besonders wichtig sind, füttern wir unsere Modelle mit den Spektrogrammen, die im Wesentlichen das gesamte Frequenzspektrum abbilden, was wir als Menschen hören können. In den nächsten Schritten könnten wir zum Beispiel hingehen und untersuchen, ob die Performance des Modells besser oder schlechter wäre, wenn wir nur Männer- oder nur Frauenstimmen in die Studie nehmen. Das haben wir tatsächlich auch probiert: Die Prognose hat insbesondere bei Männern besser funktioniert. Allerdings muss man dazu sagen, dass das Sample von Frauen in unserer Studie auch sehr klein war. Der Bedarf



Die Stimme ist sehr komplex, wir können sie zudem nur sehr schwer kontrollieren.

für Folgestudien in dieser Art ist groß, da wir leider bisher sehr wenig über die Stimmanalyse in solchen Settings wissen.

CL Verstanden. Wie würden Sie die menschliche Intuition gegenüber einem KI-Modell auf Basis eines neuronalen Netzes wie dem Ihren einschätzen?

DR Es ist schwierig zu sagen, wie unser Modell gegen Leute abschneidet, die eine gute Intuition haben. Ich würde vermuten, dass der Mensch auch gut abschneiden kann. Die Frage ist für mich aber eher, wie viel Gewicht ich in meine Intuition, in mein Gespür, also in mein Bauchgefühl lege. Umfragen zeigen, dass zum Beispiel Finanzanalysten genau hinhören, wenn sie mit CEOs telefonieren. Sie versuchen, einen Eindruck davon zu bekommen, wie realistisch das ist, was gerade erzählt wird. Das können sicher einige Leute besser als andere. Die Frage ist aber immer, wie viel Gewicht legt die einzelne Person am Ende hierauf. Ich halte es für schwierig, wenn zum Beispiel Finanzanalysten bei einer falschen Prognose dem Chef erklären müssen, ich hatte so ein Bauchgefühl.

CL Lassen Sie uns einmal die Seite wechseln und in das Unternehmen selbst schauen. Als CEO müsste ich zum einen versuchen, möglichst wenig frei zu sprechen und stattdessen viele vorgefertigte Fragen-Antwort-Situationen üben. Zum anderen müsste ich doch versuchen, meine Stimme inhaltlich, stimmlich und emotional zu trainieren und mich zu konditionieren, damit ich ruhig reagiere. Könnte man das Modell so, also mit gutem Training, überlisten oder zumindest beeinflussen?

DR Aktuell wissen wir nicht, ob Leute, die ihre Stimme trainieren und zum Beispiel aktiv viel Ausdruck in ihre Stimme bringen, anders in unserem Modell abschneiden. Umfragen zeigen aber, dass wir die Stimme tatsächlich nur sehr schwer trainieren können. Das heißt jedoch nicht unbedingt, dass wir sie gar nicht trainieren können oder dass das in der Praxis nicht stattfindet.

Zu bedenken ist im Setting von Bilanzpressekonferenzen aber: Wenn solche „Abwehrmaßnahmen“ ergriffen und bekannt werden, könnte das vom Finanzmarkt als negatives Signal gewertet werden. Der Finanzmarkt und seine Akteure sind sehr sensibel, was solche Maßnahmen angeht.

Ohnehin stellt sich die Frage, inwieweit ein Training der Stimme, um möglichst keine Emotionen zu transportieren, sinnvoll ist. Wenn ich ohne Emotionen sprechen will, also möglichst neutral, dann spreche ich auch langweilig. Damit fehlt mir aber als CEO wiederum ein wichtiges Mittel, um Überzeugungsarbeit zu leisten und damit die Stimme auch strategisch zu nutzen. Nicht nur an Finanzmärkten, sondern auch in vielen anderen Bereichen wie in der Politik wird viel mit Überzeugung in Reden bewegt. Hier überzeugen Leute, die rhetorisch gut sprechen, charismatisch sind und intoniert reden. Daher würde sich eine CEO in vielerlei Hinsicht selbst beschneiden, wenn sie versucht, möglichst monoton und emotionslos zu bleiben.

CL Ich würde gerne den Blick auf die Kommunikation ins Unternehmen werfen. Welche Empfehlungen aus den Ihnen bekannten Studien würden Sie Führungskräften in Unternehmen geben, damit deren Stimme überzeugend wirkt?

DR Wir arbeiten gerade an einer Studie dazu. Erste Erkenntnisse haben wir zum Beispiel aus der Krisenkommunikation. Gerade bei dieser Art von Kommunikation sollten die Akteure auch auf stimmliche Merkmale achten. Wie kommuniziere und spreche ich, wenn ich schlechte Neuigkeiten überbringen muss? Hier sollte ich eher empathisch wirken, damit meine Kommunikation überzeugt. Wenn ich also 30 Leute entlassen muss, dann sollte ich das nicht in einem fröhlichen Ton sagen. Das kennen wir auch aus dem Bereich der Medizin, wo Ärzte regelmäßig schlechte Neuigkeiten überbringen müssen. Gleichzeitig gibt es Indizien, die darauf hinweisen, dass man in schwierigen Zeiten nicht


Dr. Doron Reichmann

leitet die Forschungsgruppe Finance/Accounting/Auditing/Controlling/Taxation der Fakultät für Wirtschaftswissenschaften an der Ruhr-Universität Bochum.

mit einer sehr hohen Stimme sprechen sollte. Vielmehr sollte man versuchen, eine gewisse Dominanz zu signalisieren und zu signalisieren, wir kriegen das alles hin. Vereinfacht sollte man Ruhe ausstrahlen. Und das kann man eben auch mit der Stimme.

CL Wie schätzen Sie die Gefahren von KI-Modellen wie dem Ihren ein? Könnte es nicht zu einem Bias kommen, so dass aus Stimmen künftig die falschen Schlüsse gezogen werden, nur weil sie in der Vergangenheit auftrafen?

DR Ich würde unser Modell sicherlich nicht für Personaleinstellungen oder vor Gericht nutzen. Das Schöne am Finanzmarkt ist dagegen, dass meine Entscheidung im Wesentlichen immer eine 50:50-Chance ist. Einmal liege ich richtig, einmal falsch. Mithilfe des Modells will ich den Anteil meiner richtigen Entscheidungen nur über die 50 Prozent bringen und schon hat das Modell einen Mehrwert für mich. Kurzum, mit dem Modell wird man auch viele Fehlkäufe machen, aber die hätten wir wahrscheinlich sowieso gemacht.

CL Ich gehe davon aus, dass die Algorithmen in Ihrer Studie eine entsprechende Rechenleistung benötigen bzw. benötigt haben. Vor dem Hintergrund zunehmender Performance mobiler Endgeräte oder Laptops: Wäre eine Live-Stimmenanalyse im Meeting, z.B. dem internen Monats-Meeting zum Finanzreport, aus Ihrer Sicht möglich?

DR Eine solche Analyse wäre heute wahrscheinlich schon machbar. Der Aufwand bei den KI-Modellen liegt darin, sie zu trainieren. Das Trainieren der Daten dauert eine Weile, aber wenn das Modell einmal alle Daten verarbeitet hat, kann es relativ einfach auf neue (d.h. ungesehene) Daten angewandt werden. Wenn ich dem trainierten Modell eine Stimme zur Analyse gebe, bekomme ich sofort eine Antwort. Aus meiner Sicht wäre es daher durchaus möglich, schon jetzt während Bilanzpressekonferenzen die Stimme live zu analysieren.

CL Vielen Dank für das interessante Gespräch, Herr Dr. Reichmann.

Das Interview führte Prof. Dr. Christian Langmann. ■

