

The Tyranny of Metrics

eine erweiterte Rezension^[1]

Ronald Hasenberger



Ich möchte hier mit einem disclaimer starten: ich bin ein zahlenorientierter Mensch und betrachte Metriken als (üblicherweise) wertvolle Unterstützung bei Entscheidungen bzw. zur Optimierung.

Trotzdem (oder gerade deshalb) hat mich ein Buch mit diesem Titel gereizt, um andere Sichtweise(n) kennen zu lernen - in der Erwartung/Hoffnung damit in der Anwendung von Metriken lernen zu können.

Diese Erwartung wurde durch das Buch auch weitgehend erfüllt, wenngleich ich den Eindruck nicht vermeiden kann, dass das Buch eine gewisse Tendenz hat das Kind mit dem Bade auszuschütten. Insbesondere auf den ersten Seiten wird, oft anhand von Beispielen aus TV-Serien (es sei einmal dahingestellt, ob/wie sehr diese die Realität abbilden; ungeachtet dessen erscheinen sie plausibel), gezeigt wo Metriken zu Fehloptimierungen führen können, nicht ohne hervorzuheben, dass mit „gesundem Menschenverstand“ alles doch

viel besser zu managen sei. Diese sehr radikale Sichtweise wird im weiteren Text zwar teilweise relativiert, ist aber doch die dominante Darstellung im Buch (und aus meiner Sicht problematisch).

In den folgenden Abschnitten möchte ich auf die aus meiner Sicht wesentlichen Aussagen etwas detaillierter eingehen.

Vergleichbarkeit / externe Einflüsse

Ein wesentlicher Kritikpunkt ist der Vergleich von Metriken zwischen Organisationen, die in unterschiedlichen Umgebungen aktiv sind. Als Beispiel hierfür werden Metriken von Krankenhäusern verwendet, wobei darauf hingewiesen wird, dass die Metriken eines Krankenhauses in einer wohlhabenden Gegend nicht mit den Metriken eines Krankenhauses verglichen werden kann, welches in einer armen Umgebung liegt. Dieser Umstand ist für mich nachgerade offensichtlich (siehe auch ^[2] wo unter anderem dieser Aspekt und mögliche Ansätze zur Abhilfe für eine *Net-*

work-Operations Umgebung diskutiert werden) und daher nicht zu bestreiten.

Allerdings schießt das Buch aus meiner Sicht über das Ziel hinaus, wenn gesundheitspolitische Metriken in vergleichbarer Weise kritisiert werden. Durch den geänderten Perimeter der Aufgabenstellung sehe ich in diesem Szenario die Beeinflussung der Rahmenbedingungen als Teil der (gesundheitspolitischen) Aufgabenstellung, womit ich (diese) Kritik als unangemessen wahrnehme. Die Frage der Angemessenheit wird damit zu einer Frage der Wahl der Grenzen des betrachteten Systems. Im System „Krankenhaus“ ist es legitim, die sozialpolitischen Rahmenbedingungen als unbeeinflussbare Umgebungsvariablen abzubilden; im System „nationale Gesundheitspolitik“ sind aus meiner Sicht die sozialpolitischen Rah-

→

→ Objekte in Echtzeit Videos

Der nächste Schritt war es, die Kenntnisse der Einzelbilderkennung in Echtzeitvideos anzuwenden.

Die Herausforderung dabei ist, vieles zu erkennen und gleichzeitig darzustellen. Dabei an vielen Rädchen zu drehen, um zu verhindern, dass Objekte falsch erkannt werden.

Bei der Darstellung werden die Objekte mit der % Zahl umrahmt, die die Wahrscheinlichkeit anzeigt, welches Objekt erkannt wurde.

Viele Anwendungsgebiete

Personen zählen

Wenn der Computer einmal Objekte als solches erkannt hat, kann er diese auch zählen. Einerseits wie viele Objekte gleichzeitig sichtbar sind, als auch das Überqueren eines Objekts über eine virtuelle Linie.

Straßenverkehr

Mit der Kombination aus Erkennen von speziellen Objekten, dem Zählen der Anzahl von Objekten und dem Überschreiten einer virtuellen Linie (ROI = *region of interest*) ergeben sich weitere Anwendungsfälle. Die Kombination aus vorhandenen Verkehrszeichen, dem Überholverbot für LKWs und der Spurkontrolle, das Überfahren einer Sperrlinie mittels Erkennung einer Lane (Fahrspur) und einer virtuellen Linie. Weiteres die Kennzeichenerfassung mit OCR, die Abstandmessung und Geschwindigkeitsmessung mittels fixer Linien auf der Fahrbahn. Einen Geisterfahrer erkennt man, indem man in die andere Richtung zählt. Man kann einen Stau erkennen, oder Gegenstände auf der Fahrbahn die dort nichts verloren haben.

Fazit

Es gibt schon zahlreiche Anwendungsbereiche von KI, einiges existiert bereits und funktioniert relativ gut. KI wird ver-

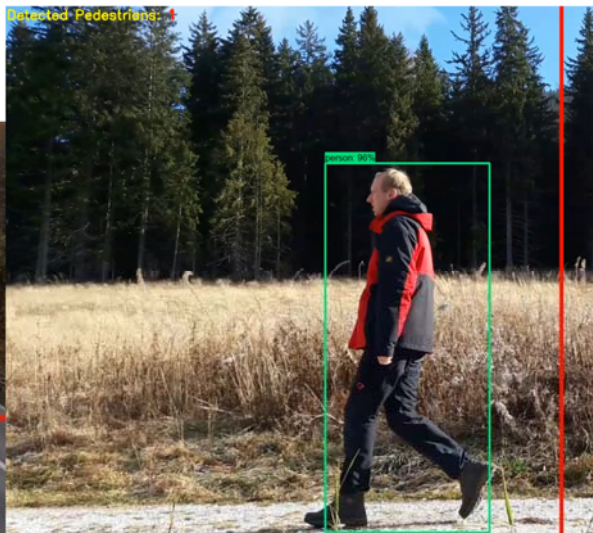
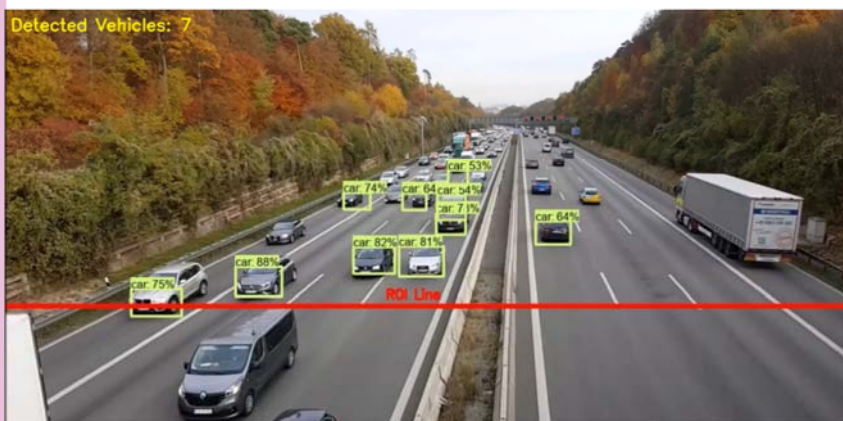
mutlich bei vielen wiederkehrenden Tätigkeiten eingesetzt werden, wo der Computer nicht ermüdet und 24/7 durcharbeitet. Das wird viele zukünftige Jobs beeinflussen und uns alle irgendwie betreffen. Es muss aber nicht negativ sein. Man denke nur an die medizinischen Anwendungsfälle.

Autorenbox

Thomas Reinwart verfügt über umfangreiche Berufserfahrung auf dem IT Sektor. In den letzten 20 Jahren war er in den Bereichen Softwareentwicklung, Softwaredesign, Architekt und als Consultant tätig. Technischer Fokus ist derzeit Microsoft .net und SQL Server, wo er alle aktuellen Microsoft Zertifizierungen hat.



Email: office@reinwart.com



menbedingungen Teil des zu optimierenden Systems.

Überfokussierung

Ein im Buch behandelte Kritikpunkt an Metriken ist die Fokussierung des Handelns der Verantwortlichen auf jene Metriken, welche in Ihren Leistungsvereinbarungen relevant sind. Dies ist ein bekanntes Problem für das aber auch eine Lösung bekannt ist: die Zielsetzung auf Grundlage einer Balanced Score Card (BSC). Mit der Einführung einer BSC werden unterschiedliche Dimensionen (z.B. Kosten, Kundenzufriedenheit, Mitarbeiter) durch Metriken erfasst und Ziele gesetzt. Die Metriken können sich dabei durchaus „widersprechen“ (Maßnahmen zur Steigerung der Kundenzufriedenheit kosten Geld, können damit im Konflikt zur Kostenreduktion stehen); wesentlich ist daher das Balancieren (daher „*Balanced Score Card*“) der Anforderungen in den unterschiedlichen Dimensionen um eine Überforderung der Organisation zu vermeiden.

Die BSC behebt aber nicht das Problem von „falschen“ Metriken, welche Optimierung in eine grundsätzlich „falsche“ Richtung belohnen, wenn keine moderierenden oder korrigierenden Metriken definiert sind.

Fehlleitung

Durch die Vorgabe von „falschen“ Metriken kann eine Organisation dazu gebracht werden, auf die falschen Aspekte zu achten. Die zitierten Beispiele sind aus einer Polizeiumgebung („*Maximierung der Anzahl der Verhaftungen*“ als Ziel kann dazu führen, dass viele Kleinkriminelle verhaftet werden anstatt sich auf die – wenigen – Großkriminellen zu konzentrieren) als auch aus dem medizinischen Umfeld (Bewertung auf Grundlage von Erfolgsquoten bei Behandlungen kann dazu führen, dass schwierige Fälle nicht mehr angenommen werden, weil das Risiko eines Misserfolgs zu groß ist). Diesen Aussagen ist nichts entgegenzusetzen; letztendlich sind es aber die falschen Metriken, eingesetzt ohne Korrektiv, welche die Ursache für das Problem darstellen.

Egoismus

Der Kritikpunkt hier ist, dass der – üblicherweise ohnehin vorhandene – „*organisatorische Egoismus*“ durch Metriken weiter verschärft wird. Letztendlich handelt es sich dabei um die Anwendung der beiden oben genannten Probleme (Überfokussierung, Fehlleitung) und kann daher mit denselben Korrekturmechanismen behandelt werden.

Management by Metrics

Es wird im Buch dargestellt, dass Metriken, verwendet zur Bemessung von Prämien das Potential für besonders schädliche Auswirkungen haben. Dem kann ich nicht widersprechen und es ist letztendlich eine (mit Hilfe der Prämien) verstärkte Variante der Fehlleitung, die weiter oben diskutiert wurde.

Grundsätzlich erscheint es aus meiner Erfahrung meistens sinnvoll, eine gute Balance zwischen „harten“ (basierend auf Metriken) und „weichen“ (basierend auf der Einschätzung durch die Vorgesetzten – die auch die Eigentümer sein können) Zielen zur Bemessung der Prämien zu haben. Der richtige Mix ist dann aber auch nicht basierend auf einer fixen Regel, sondern hängt von zahlreichen Aspekten ab – um nur einige zu nennen:

- Die Persönlichkeit des Mitarbeiters
- Der Situation des Unternehmens
- Der Dynamik des Marktes

Insbesondere in den dynamischen Umgebungen in denen ich meist tätig war und noch immer bin, kommt dazu, dass jegliche Ziele Gefahr laufen, dass sie knapp nach der Vereinbarung schon wieder obsolet sind. Die korrekte Umgangsweise ist die Anpassung der Ziele an die geänderten Bedingungen. Allerdings habe ich die dafür notwendige Disziplin bisher nur selten erlebt, in welchem Fall es umso wichtiger ist, dass der Vorgesetzte die Möglichkeit hat von mechanistischen Bewertungen der Zielerfüllung^[3] auch abzuweichen.

Aufwand der Messung

Ein weiterer Aspekt ist der Aufwand für die Messung. Grundsätzlich stellen Messungen einen Aufwand dar, womit dem Autor des Buches recht zu geben ist. Andererseits werden gerade in (teil-) automatisierten Umgebungen zahlreiche Messungen als Nebenprodukt der Automatisierung generiert, was den Aufwand der Messung für derartige Metriken gegen 0 gehen lässt. Eine damit einhergehende Gefahr ist allerdings, dass einfach alles gemessen wird, was (einfach) messbar ist – womit ein signifikanter Aufwand bei der Auswertung generiert wird.

Ähnlich zu meiner Einschätzung weiter oben ist daher auch hier eine sinnvolle Balance gefragt, also: so weit als sinnvoll Ziele mit Metriken balancieren und Optimierungsfehler kompensieren; den Rest durch subjektive Beurteilung ergänzen.

Zusammenfassung

Ich halte es für ein sehr gutes Buch, welches viele Denkanstöße gibt und noch mehr Beispiele enthält, wie man Metriken nicht designen sollte bzw. welche Fallen es in diesem Zusammenhang gibt. Eine der relevantesten Fallen, die im Buch behandelt wird (ohne Widerspruch meinerseits) ist die Verwendung von problematischen Metriken für die Bewertung des Erfolgs von Organisationen und Menschen, da damit erst die Motivation für die Fehlfokussierung bis hin zu bewusster Verfälschung begründet wird^[4].

Wie bereits eingangs erwähnt, kann der Fokus des Buches auf „*was kann alles schiefgehen*“ zur Wahrnehmung führen, dass Metriken im Allgemeinen schlecht sind, was aus meiner Sicht weder die Absicht des Autors noch Realität ist. Wesent-

lich sind (ohne Anspruch auf Vollständigkeit):

- Ein gutes Design der einzelnen Metriken
- Ein gutes Set von (balancierten) Metriken
- Gute Balance von Metriken und qualitativen Aspekten

Insbesondere für zahlenorientierte Menschen (wie ich einer bin), ist das Buch damit ein sehr willkommener Kontrapunkt zu Ihrem inhärenten Zugang – und diese laufen wohl auch nicht Gefahr, das Kind mit dem Bade auszuschütten.

Für Menschen, die Zahlen generell eher zweifelnd gegenüberstehen, sehe ich diese Gefahr allerdings sehr wohl, da das Thema Metriken weitgehend von der kritischen Seite beleuchtet wird ohne auf die positiven Aspekte einzugehen (wobei man argumentieren kann, dass die positive Seite in anderen Büchern ausreichend behandelt wird).

Allerdings – was sind die beiden letzten Paragraphen anderes als der Hinweis darauf, dass es gut ist (wäre), sich mit Meinungen und Argumenten auseinanderzusetzen, die den eigenen Widersprechen. Es ist der aufwändigere Weg, der aber in fast allen Fällen zum qualitativ besseren Ergebnis führt.

Referenzen

The Tyranny of Metrics

Jerry Z. Muller; Princeton University Press (6. Februar 2018); ISBN-13: 978-0691174952

Making Cable Networks Ready for the Future Through Operations Benchmarking

Ronald Hasenberger; SCTE Journal of Network Operations, December 2018; <https://www.scte.org/journals>

Lessons from history on the dangers of blind trust in data

John Thornhill, Financial Times, 31.Dec.2018

Endnoten

^[1] Metriken werden in diesem Buch oft auch für SLAs verwendet, was 2 verbundene aber unterschiedliche Aspekte darstellt. In erster Näherung sind SLAs Metriken mit Zielwerten.

^[2] Sehr im Gegensatz zu Peter Drucker's häufig zitiertem Statement: „*If you can't measure it, you can't manage it.*“

^[3] „Harte“ Ziele, basierend auf Metriken werden oft auch (semi-)automatisch und damit zwangsläufig mechanistisch ausgewertet, was die Berücksichtigung von sich ändernden Rahmenbedingungen schwierig macht. Allerdings wird im Buch der Eindruck erweckt, dass „Betrug“ (womit ich etwas großzügig alles vom „gaming“ von Daten bis hin zur Verfälschung von Zahlen zusammenfasse) erst mit der Fokussierung auf Metriken im Geschäftsleben Einzug gehalten hat.

^[4] Gleichzeitig heißt das aber, dass bloß deshalb, weil etwas „gemessen“ wird nicht weniger Aufwand betrieben werden darf die Ergebnisse auch zu verstehen und Entscheidungen erst nach einer angemessenen Analyse zu treffen.