

Wörter im Vektorraum:

Warum Analogien manchmal scheitern

Begleitmaterial zum KI-Labor „Wörter im Vektorraum“ · Stufe 2 der KI-Fortbildungsreihe

1. „König zu Mann ist wie Frau zu Königin“

Word Embeddings sind Zahlenvektoren, die Wortbedeutungen repräsentieren. Das berühmteste Beispiel für ihre Fähigkeit ist die Analogie-Rechnung: Nimm den Vektor für „König“, subtrahiere „Mann“ und addiere „Frau“ — heraus kommt „Königin“. Das klingt magisch, und in der Tat funktioniert es bei bestimmten Beziehungen zuverlässig. Aber bei vielen anderen scheitert es — und die Gründe dafür sind lehrreich.

In der App „KI-Labor: Wörter im Vektorraum“ arbeiten wir mit **5.000 deutschen Wörtern** aus ConceptNet Numberbatch (Version 19.08). Jedes Wort ist als Vektor mit **50 Dimensionen** gespeichert. Diese Vektoren wurden aus einem sehr viel größeren Modell (300 Dimensionen) durch Dimensionsreduktion gewonnen — ähnlich wie man ein 3D-Objekt auf ein Foto projiziert.

2 Die Ergebnisse im Überblick

Von 13 vordefinierten Analogien in der App liefern mit 50 Dimensionen nur **3 das erwartete Ergebnis auf Platz 1**. Die folgende Tabelle zeigt alle Ergebnisse:

Analogie	Modell sagt	Status	Rang
<i>mann</i> → <i>könig</i> wie <i>frau</i> → königin	königin	✓ Platz 1	1
<i>könig</i> → <i>königin</i> wie <i>prinz</i> → prinzessin	prinzessin	✓ Platz 1	1
<i>wasser</i> → <i>eis</i> wie <i>regen</i> → schnee	schnee	✓ Platz 1	1
<i>sommer</i> → <i>warm</i> wie <i>winter</i> → kalt	heiss	Platz 2	2
<i>hund</i> → <i>tier</i> wie <i>rose</i> → blume	rosa	Platz 2	2
<i>schule</i> → <i>lehrer</i> wie <i>krankenh.</i> → arzt	ärzte	Platz 2	2
<i>fussball</i> → <i>sport</i> wie <i>gitarre</i> → musik	klavier	Platz 3	3
<i>tag</i> → <i>sonne</i> wie <i>nacht</i> → mond	nebel	nicht in Top 5	>5
<i>frankr.</i> → <i>paris</i> wie <i>deutschl.</i> → berlin	neustadt	nicht in Top 5	>5
<i>italien</i> → <i>rom</i> wie <i>england</i> → london	jerusalem	nicht in Top 5	>5
<i>auto</i> → <i>strasse</i> wie <i>schiff</i> → meer	strand	nicht in Top 5	>5
<i>leise</i> → <i>laut</i> wie <i>dunkel</i> → hell	spreche	nicht in Top 5	>5
<i>hund</i> → <i>katze</i> wie <i>gross</i> → klein	grossem	nicht in Top 5	>5

Tabelle: Ergebnisse der Vektor-Analogien mit 50-dimensionalen ConceptNet-Embeddings. „Rang“ zeigt, auf welchem Platz das erwartete Wort liegt (Rang >5 = nicht in den Top 5).

Hinweis für die Fortbildung

Mit 100 Dimensionen verbessern sich einige Ergebnisse (z. B. landet „kalt“ bei „sommer→warm / winter→?“ auf Platz 1). Die Hauptstadt-Analogien bleiben aber auch bei 100 Dimensionen problematisch. Das liegt nicht an der Dimensionszahl, sondern an der Art, wie die Beziehung in den Trainingsdaten kodiert ist (siehe Abschnitt 3).

3 Warum scheitern Analogien?

3.1 Das Kernproblem: Wörter haben keine einzige Bedeutung

Die Analogie-Rechnung $B - A + C = ?$ funktioniert nur, wenn die Richtung von A nach B **dieselbe Beziehung** kodiert wie die erwartete Richtung von C zum Zielwort. Doch Wörter tragen in den Embeddings immer **alle ihre Bedeutungen gleichzeitig**.

Das lässt sich am Beispiel „**Italien** → **Rom** wie **England** → ?“ sehr gut zeigen. Logisch erwarten wir „London“. Das Modell sagt aber „Jerusalem“. Warum?

Was steckt in der Richtung „Rom minus Italien“?

Wenn wir berechnen, welche Wörter am besten zur Richtung $\text{rom} - \text{italien}$ passen, finden wir: **Kirche, Dom, Münster, evangelische, katholische, Christus, Luther**. Das heißt: Im deutschsprachigen Trainingskorpus kommt „Rom“ vor allem im Kontext von **Papsttum, Kirchengeschichte und Antike** vor — weit häufiger als im Kontext „Hauptstadt von Italien“.

Addiert man diese „Kirchen-Richtung“ auf „England“, landet man folgerichtig bei **Jerusalem** (heilige Stadt), **Papst** und **katholische** — nicht bei London.

Was steckt in „London minus England“?

Die Richtung $\text{london} - \text{england}$ zeigt dagegen auf: **Bahnhof, Innenstadt, Chicago, Werkstatt, Paris**. Hier kodiert das Modell die Bedeutung „Großstadt / urbaner Raum“. London ist für das Modell primär eine Metropole, nicht „die Hauptstadt von England“.

Schlüsselzahl

Die Cosinus-Ähnlichkeit zwischen den Richtungen $\text{rom} - \text{italien}$ und $\text{london} - \text{england}$ beträgt nur **0,20** (bei perfekter Parallelität wäre es 1,0). Die beiden „Hauptstadt-Beziehungen“ zeigen also in fast völlig verschiedene Richtungen — weil jedes Stadt-Land-Paar seine eigene Semantik aus dem Trainingskorpus mitbringt.

3.2 Die drei häufigsten Fehlerquellen

Fehlerquelle 1: Polysemie (Mehrdeutigkeit)

„Rom“ ist gleichzeitig eine moderne Hauptstadt, der Sitz des Papstes und das Zentrum eines antiken Weltreichs. Das Embedding ist ein *Durchschnitt* all dieser Bedeutungen. Je nach Kontext „gewinnt“ eine andere Facette. Bei der Analogie-Rechnung kann sich die falsche Bedeutung durchsetzen.

Weiteres Beispiel: „Hund → Tier / Rose → ?“ ergibt „rosa“ statt „Blume“, weil „Rose“ im Deutschen auch eine Farbe assoziiert und die orthografische Nähe von „rose“ zu „rosa“ im Embedding-Raum stark wirkt.

Fehlerquelle 2: Flexionsformen als Konkurrenten

Bei „Schule → Lehrer / Krankenhaus → ?“ liefert das Modell „ärzte“ (Plural) vor „arzt“ (Singular). Im Vokabular existieren beide Formen als separate Einträge, und der Pluralvektor liegt zufällig minimal näher am Ziel. Für das Modell sind „arzt“ und „ärzte“ zwei verschiedene Wörter — es „weiß“ nicht, dass es Flexionsformen desselben Wortes sind.

Fehlerquelle 3: Inkonsistente Beziehungsrichtungen

Die Analogie-Rechnung nimmt an, dass die Beziehung „Land → Hauptstadt“ immer in dieselbe Richtung zeigt. Tatsächlich kodiert jedes Stadt-Land-Paar eine andere Mischung aus Bedeutungen:

Richtung	Top-5 assoziierte Wörter	Bedeutung
rom - italien	Kirche, Dom, HL., Kirchen, Münster	<i>Religion & Antike</i>
london - england	Bahnhof, Innenstadt, Chicago, Werkstatt, Paris	<i>Großstadt / urban</i>
paris - frankreich	Innenstadt, Werkstatt, Bahnhof, Altstadt, Paris	<i>Großstadt / urban</i>
berlin - deutschland	BD, CD, Amsterdam, Spreche, Paris	<i>unklar / Rauschen</i>

Tabelle: Was verschiedene Stadt-Land-Richtungen tatsächlich kodieren. Nur „London“ und „Paris“ zeigen in eine ähnliche Richtung („urban“). „Rom“ geht in eine völlig andere (religiöse) Richtung.

4 Warum funktioniert König → Königin?

Eine naheliegende Frage: Wenn die Richtungen so inkonsistent sind, warum klappt „Mann → König / Frau → Königin“ dann so zuverlässig?

Die Antwort liegt in der **Konsistenz der Beziehung in der Sprache** selbst. Geschlechterbeziehungen (Mann/Frau, König/Königin, Prinz/Prinzessin) tauchen in Texten in extrem konsistenten Kontexten auf. Überall, wo „König“ steht, gibt es ähnliche Sätze mit „Königin“. Diese Parallelität prägt sich geometrisch in die Vektoren ein.

Bei Hauptstädten ist das anders: Die Sätze, in denen „Rom“ und „Italien“ zusammen vorkommen, handeln von völlig anderen Themen als die Sätze mit „London“ und „England“. Geografie ist in natürlicher Sprache nicht so konsistent kodiert wie Geschlecht.

Faustregel für die Fortbildung

Analogien funktionieren gut bei Beziehungen, die in der Sprache systematisch und konsistent vorkommen (Geschlecht, Aggregatzustand, Exemplar/Kategorie). Sie scheitern bei Beziehungen, die in verschiedenen Kontexten unterschiedlich assoziiert sind (Geografie, Gegensätze, abstrakte Zuordnungen).

5 Die Kernbotschaft: Embeddings sind kein Lexikon

Word Embeddings bilden **keine logischen Relationen** ab, sondern **statistische Kookkurrenz-Muster**. Sie fangen ein, welche Wörter in welchen Kontexten zusammen vorkommen — nicht, was diese Wörter „Bedeutung“. Wenn man das versteht, werden die Fehler vorhersagbar:

- **Embeddings sind ein Spiegel der Trainingsdaten.** Wenn „Rom“ im Trainingskorpus überwiegend in religiösen Kontexten vorkommt, zeigt der Vektor in Richtung Religion — nicht in Richtung „Hauptstadt“.
- **Jedes Wort hat genau einen Vektor für alle Bedeutungen.** Neuere Sprachmodelle (BERT, GPT) lösen das, indem sie kontextabhängige Vektoren berechnen — „Rom“ bekommt in einem Satz über Tourismus einen anderen Vektor als in einem Satz über den Papst.
- **Analogie-Rechnung ist ein schönes Demo, aber kein robustes Werkzeug.** Sie funktioniert als Illustration dafür, dass Bedeutung in Vektoren steckt — aber man sollte sie nicht als Beweis für „Verständnis“ des Modells interpretieren.

6 Didaktische Empfehlung für den Unterricht

Die App enthält bewusst **sowohl funktionierende als auch scheiternde Analogien**. Statt die Fehler zu verstecken, empfehlen wir, sie als **Forscherverfragen** einzusetzen:

1. **Einstieg:** Lassen Sie die Teilnehmenden zuerst „Mann → König / Frau → ?“ ausprobieren. Der Wow-Effekt ist garantiert.
2. **Vertiefung:** Dann „Italien → Rom / England → ?“. Die Überraschung, dass „Jerusalem“ statt „London“ herauskommt, ist ein starker Diskussionsanlass.
3. **Reflexion:** Fragen Sie: „Warum klappt König→Königin, aber nicht Italien→Rom auf England→London?“ Die Antwort führt zur Kernbotschaft: Embeddings sind Spiegel der Trainingstexte, kein Weltwissen.
4. **Transfer:** Was bedeutet das für Sprachmodelle wie ChatGPT? Sie arbeiten mit ähnlichen Prinzipien, nur mit kontextabhängigen Vektoren und Milliarden statt Tausenden von Dimensionen.

Tipp

Lassen Sie die Teilnehmenden im „Eigene Analogie“-Bereich der App selbst experimentieren. Besonders spannend: Berufe (Lehrer/Lehrerin, Arzt/Ärztin), Sprachen (Deutschland/Deutsch, Frankreich/Französisch) oder emotionale Gegensätze (Freude/Trauer). Manche funktionieren überraschend gut, andere überraschend schlecht.