M05c - Embeddings

Stand: 03.2025

Damit Künstliche Intelligenz (KI) sinnvoll mit Sprache, Bildern oder anderen Inhalten arbeiten kann, muss sie deren Bedeutung erfassen. Allerdings verarbeitet ein Computer keine Wörter oder Bilder direkt, sondern nur Zahlen. **Embeddings** sind eine Methode, um solche Inhalte als Zahlen zu kodieren, sodass die KI Zusammenhänge und Bedeutungen erkennen kann.

1 Was sind Embeddings?

Ein **Embedding** ist eine mathematische Darstellung eines Wortes, Satzes oder Bildes in Form eines Vektors, also einer Zahlenliste. Diese Zahlen erfassen Ähnlichkeiten und Zusammenhänge zwischen verschiedenen Konzepten.

Beispiel für Sprache:

- Das Wort **„King“** könnte als Zahlenvektor **[0.96, 0.92, 0.08, 0.67]** dargestellt werden.
- Das Wort **„Queen“** könnte **[0.98, 0.07, 0.93, 0.71]** haben.
- Das Wort **„Girl“** könnte **[0.56, 0.09, 0.91, 0.11]** haben.

→ Die Zahlen von „King“ und „Queen“ sind **ähnlicher** als die von „Man“ und „Girl“. Dies zeigt, dass die KI die inhaltliche Nähe dieser Begriffe versteht.

Beispiel für Bilder:

- Ein Bild von einem Hund wird in Zahlen umgewandelt.
- Ein ähnliches Bild erhält einen ähnlichen Zahlenvektor.
- Dadurch kann die KI visuelle Ähnlichkeiten erkennen.

Embeddings werden nicht nur für Sprache und Bilder genutzt, sondern auch in Empfehlungssystemen für Musik, Filme oder sogar in der medizinischen Forschung zur Mustererkennung.

Hypothetisches Beispiel für die Embeddings:

Hypothetische Merkmale zum Verständnis von Worteinbettungen

Quelle: Ein Leitfaden für Anfänger zu Word2Vec. Grundlagen von Word2Vec + Implementierung… | von Manan Suri | Medium

2 Wie entstehen Embeddings?

Embeddings werden mit **künstlichen neuronalen Netzen** oder **statistischen Methoden** erzeugt. Dabei durchläuft der Prozess mehrere Schritte:

1 Daten sammeln

- Sprachmodelle nutzen große Mengen an Texten aus Büchern, Webseiten oder Artikeln.
- Bilderkennungsmodelle analysieren Millionen von Fotos mit passenden Beschreibungen.
- Musik- oder Videoplattformen sammeln Daten zu Nutzerverhalten und Inhaltsmerkmalen.

2 Daten in Zahlen umwandeln

- Wörter werden als **Vektoren** dargestellt, die Bedeutungsähnlichkeiten widerspiegeln.
- Bilder werden in **Pixelwerte** und Merkmale wie Kanten oder Farben umgerechnet.
- Musik wird anhand von Frequenzmustern und Rhythmen analysiert.

3 Neuronale Netze trainieren

- Modelle wie **Word2Vec, GloVe oder FastText** für Sprache sowie **ResNet oder VGG** für Bilder lernen, welche Begriffe oder Objekte ähnlich sind.
- Empfehlungssysteme analysieren, welche Songs oder Filme Nutzer häufig zusammen konsumieren.

4 Ähnlichkeiten erkennen

- Begriffe mit ähnlicher Bedeutung liegen im Zahlenraum nahe beieinander.
- Beispiel: Das Embedding für „König“ liegt näher an „Königin“ als an „Banane“.
- Bilder von Hunden liegen näher an Wölfen als an Autos.

5️⃣ Feinabstimmung (Fine-Tuning)

- Embeddings können für spezifische Anwendungen optimiert werden.
- Beispiel: Eine KI für medizinische Analysen trainiert spezielle Embeddings für Fachbegriffe.
- Streaming-Dienste passen ihre Embeddings an individuelle Nutzerpräferenzen an.

---

3 Positional Encoding

Die Positionskodierung fügt jedem Token-Vektor (aus der Einbettungsmatrix) Informationen über seine Position in der Sequenz hinzu. Dies geschieht durch die Kombination von Positionsinformationen und den ursprünglichen Token-Einbettungen. Ohne zusätzliche Information gäbe es keinen Unterschied zwischen:

- *Die Katze jagt den Hund* und
- *Den Hund jagt die Katze*

Die Positionskodierung ist wie ein kleiner Hinweiszettel, der sagt, welches Wort an welcher Stelle steht.

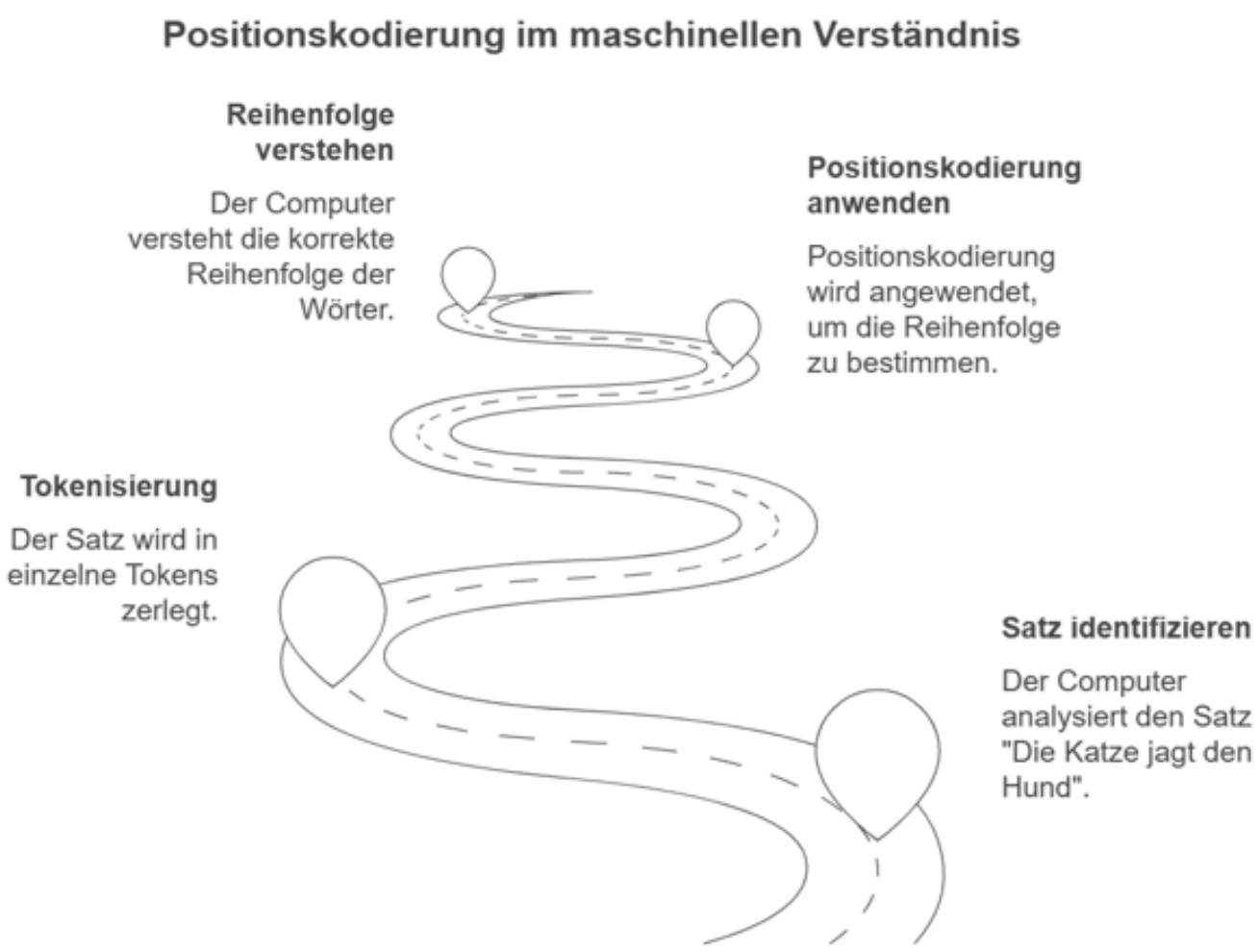

Bild mit napkin.ai erstellt

---

4 Embedding-Modelle

Es gibt verschiedene Einbettungsmodelle wie Word2Vec, GloVe und FastText für Wortrepräsentationen, BERT für kontextuelle Einbettungen sowie Node2Vec und LSTM-basierte Modelle für Netzwerke und Sequenzen, die jeweils auf spezifische Anwendungsfälle und Datenstrukturen optimiert sind.

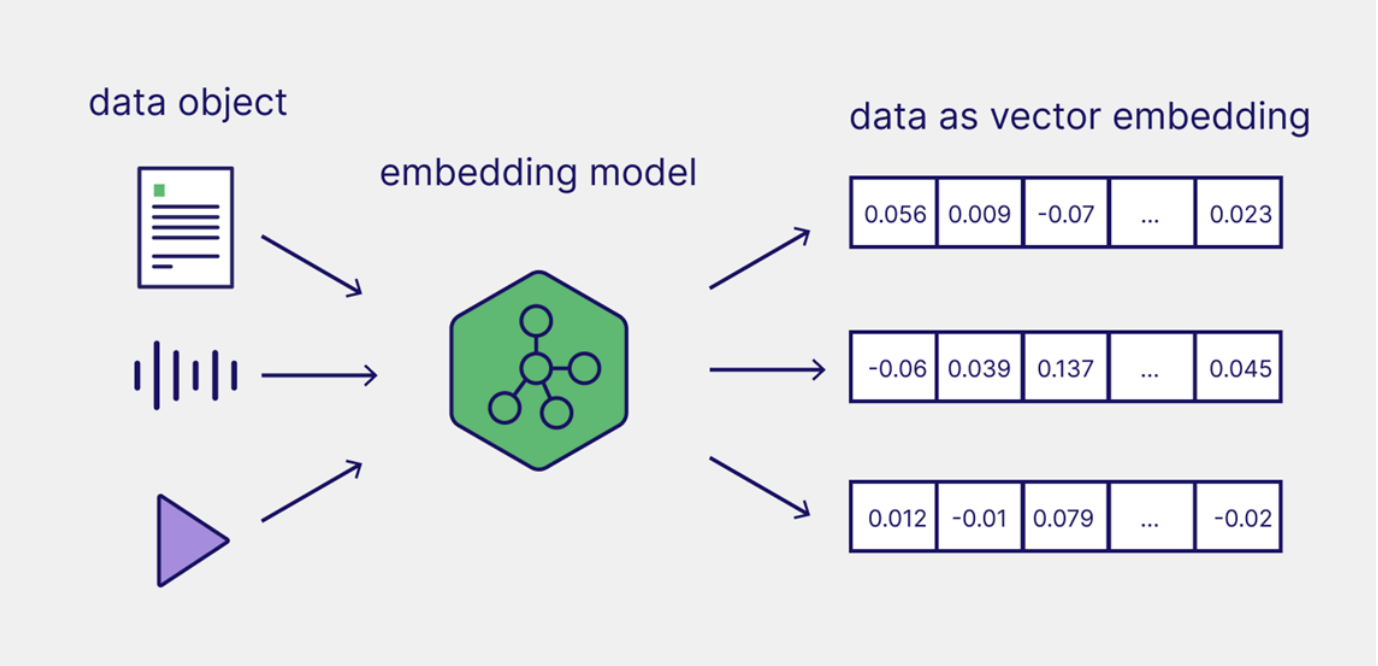

Bild: Step-by-Step Guide to Choosing the Best Embedding Model for Your Application | Weaviate - Vector Database

Übersicht Einbettungsmodelle:

Einbettungsvektor	Typische Größen	Einsatzbereich
Word2Vec	100-300 Dimensionen	Wort- und Satzähnlichkeiten, NLP
GloVe	50, 100, 200, 300 Dimensionen	Semantische Wortbeziehungen, NLP
FastText	100-300 Dimensionen	OOV-Wortbehandlung, NLP
BERT (Basisversion)	768 Dimensionen	Kontextuelle Textverarbeitung, NLP
BERT (Large-Version)	1024 Dimensionen	Fortgeschrittene NLP-Anwendungen
text-embedding-ada-002 (OpenAI)	1536 Dimensionen	Hochqualitative semantische Suche, RAG
sentence-transformers	384-768 Dimensionen (modellabhängig)	Semantische Ähnlichkeit, Clustering, RAG
Benutzer- und Produkteinbettungen	50-200 Dimensionen	Empfehlungssysteme, Personalisierung
Einbettungen aus CNNs (VGG16)	4096 Dimensionen (für FC-Schichten)	Bildverarbeitung, Objekterkennung
Einbettungen aus CNNs (ResNet)	Variiert (tiefer mit unterschiedlichen Größen)	Bildanalyse, Feature-Extraktion
Node2Vec	64-256 Dimensionen	Graphenanalysen, soziale Netzwerke

Einbettungsvektor	Typische Größen	Einsatzbereich
LSTM-basierte Sequenzeinbettungen	50-500 Dimensionen	Zeitreihen, Sprachmodellierung, NLP

5 Training von Embedding-Modellen

Das Training von Embedding-Modellen wie Word2Vec basiert auf der Idee, dass Wörter, die in ähnlichen Kontexten vorkommen, ähnliche Bedeutungen haben. Hier wird detaillierter beschrieben, wie dieses Prinzip im Training umgesetzt wird:

Algorithmus-Auswahl

Word2Vec bietet zwei grundlegende Modelle zur Generierung von Wort-Embeddings:

1. **CBOW (Continuous Bag of Words)**: Hierbei wird das Zielwort basierend auf einem umgebenden Wortkontext vorhergesagt. Das Modell bekommt mehrere Wörter als Eingabe (den Kontext) und versucht, das Wort in der Mitte (das Zielwort) zu vorherzusagen.
2. **Skip-Gram**: Hier wird der umgekehrte Ansatz verfolgt. Ausgehend von einem Zielwort versucht das Modell, die umgebenden Kontextwörter vorherzusagen.

Training

Das Training von Word2Vec kann wie folgt zusammengefasst werden:

- **Initialisierung**: Zuerst werden Vektoren für jedes Wort zufällig initialisiert.
- **Durchlauf durch den Korpus**: Das Modell geht durch den gesamten Textkorpus, nimmt jedes Wort zusammen mit seinen Nachbarwörtern (innerhalb eines bestimmten Fensters) und führt Trainingsiterationen durch.
- **Verlustfunktion**: Die Hauptaufgabe beim Training ist die Optimierung der Verlustfunktion. Für CBOW und Skip-Gram wird oft eine Funktion verwendet, die die logarithmische Wahrscheinlichkeit maximiert, korrekte Wörter basierend auf ihren Kontexten vorherzusagen.
    - Bei **CBOW** wird der Verlust berechnet, indem die Differenz zwischen dem vorhergesagten Zielwort und dem tatsächlichen Zielwort über die Softmax-Funktion gemessen wird.
    - Beim **Skip-Gram** wird der Verlust für jedes vorhergesagte Kontextwort berechnet.
- **Backpropagation**: Mit Hilfe des Gradientenabstiegs oder ähnlicher Optimierungsalgorithmen werden die Gewichte (Wortvektoren) so angepasst, dass die Verlustfunktion minimiert wird. Dies bedeutet, dass die Wortvektoren nach und nach angepasst werden, um den wahren Kontext besser widerzuspiegeln.

Ergebnis

Das Ergebnis des Trainings ist ein Set von Vektoren, eines für jedes Wort im Vokabular. Wörter, die in ähnlichen Kontexten vorkommen, enden nahe beieinander im Vektorraum, was ihre semantische Ähnlichkeit widerspiegelt. Diese Vektoren können dann in verschiedenen nachgelagerten maschinellen Lernaufgaben verwendet werden, z.B. in der Sentiment-Analyse, bei der Klassifikation von Dokumenten oder anderen NLP-Aufgaben, die eine numerische Repräsentation von Text erfordern.

Evaluierung

Um die Qualität der Embeddings zu überprüfen, werden oft qualitative Tests wie die Suche nach den nächsten Nachbarn (ähnliche Wörter finden) oder quantitative Benchmarks (z.B. auf Datensätzen für analoge Aufgaben) durchgeführt. Diese Evaluierungen helfen dabei festzustellen, ob das Modell die Wortbedeutungen effektiv erfasst hat.

6 Warum sind Embeddings so wichtig?

✅ **Sprachverarbeitung**: Chatbots, Übersetzungen und Textanalysen basieren auf Embeddings.
✅ **Bilderkennung**: KI kann ähnliche Bilder oder Objekte erkennen.
✅ **Suche & Empfehlungssysteme**: Personalisierte Vorschläge auf Plattformen wie Netflix, Spotify oder YouTube nutzen Embeddings.
✅ **Musik- und Videovorschläge**: Streaming-Dienste berechnen Nutzerpräferenzen basierend auf Embeddings.
✅ **Medizinische Diagnosen**: KI analysiert Krankheitsbilder und medizinische Muster durch Embeddings.
✅ **Generative KI**: Sprachmodelle wie ChatGPT nutzen Embeddings, um kontextbezogene Antworten zu generieren.

📝 Fazit

Embeddings sind ein zentrales Konzept in der modernen KI. Sie ermöglichen Maschinen, Bedeutungen zu erfassen, Muster zu erkennen und personalisierte Inhalte zu liefern. Ohne Embeddings wären viele heutige KI-Technologien nicht denkbar – von Chatbots über Bilderkennung bis hin zu Streaming-Diensten. Sie sind das **unsichtbare Gerüst**, das intelligente Systeme erst möglich macht.