KI für Alle 2: Verstehen, Bewerten, Reflektieren

Themenblock Generative KI: 08_05Modellauswahl_Transformer

Transformer

Erarbeitet von
Dr. Christian Geishauser

Lernziele

- Du kannst mit Beispielen wiedergeben, wie der Transformer kontextabhängige Repräsentationen erstellt
- Du kannst den Aufmerksamkeitsmechanismus erläutern
- Du kannst erläutern, wie Transformer schrittweise das nächste Wort vorhersagen
- Du kannst Anwendungen des Transformers nennen

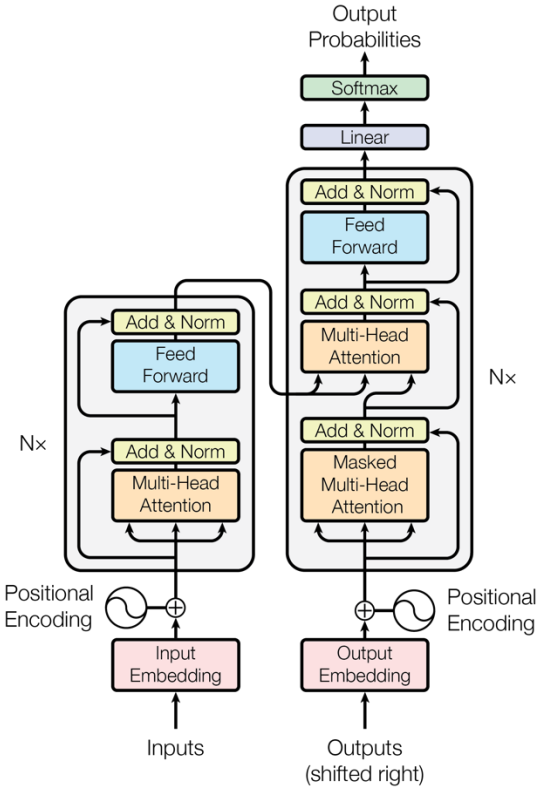
Inhalt

Vielleicht hast du schon von Large Language Models wie ChatGPT oder Bard und modernen Bildgeneratoren wie Midjourney oder Dall-E gehört, oder sogar benutzt. Vielleicht hast du auch schon Übersetzungsdienste wie Google Translate oder DeepL verwendet.
In den Nachrichten wird in diesem Zusammenhang oft davon gesprochen, dass es sich bei den Modellen um große neuronale Netze handelt. Doch um welches neuronale Netz handelt es sich hier überhaupt und wie funktioniert es? Und wie generiert zum Beispiel ChatGPT neuen Text?
In diesem Video möchten wir das grundlegende neuronale Netz dahinter besprechen, den sogenannten Transformer.

Dabei handelt es sich nicht um humanoide Roboter, die sich in Autos verwandeln können,

Einblendung Transformation aus der Transformers-Zeichentrickserie (**Quelle [1]**)

sondern um eine besondere Art von Deep-Learning-Architektur.



Quelle [2]

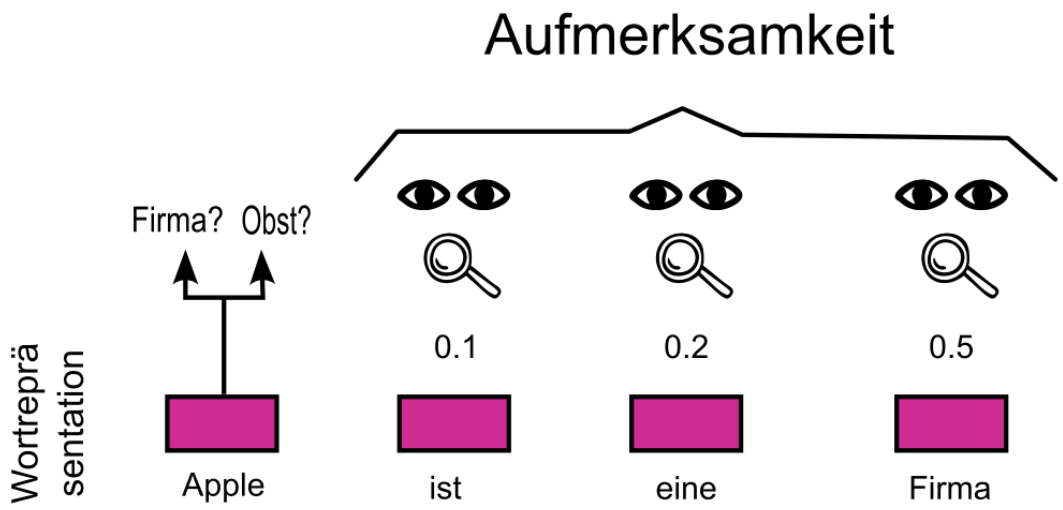
Das sieht auf der Grafik jetzt ein wenig kompliziert aus. Fangen wir also von vorne mit einem einfachen Beispiel an.

Funktionsweise von Transformern

Um den Transformer zu motivieren, stell dir vor, wir haben folgendes Problem: Wir haben das Wort Apple gegeben und haben eine Wort-Repräsentation dazu erstellt. Wir möchten anhand dieser Repräsentation gerne bestimmen, ob mit Apple die Firma oder das englische Wort für Apfel gemeint ist.

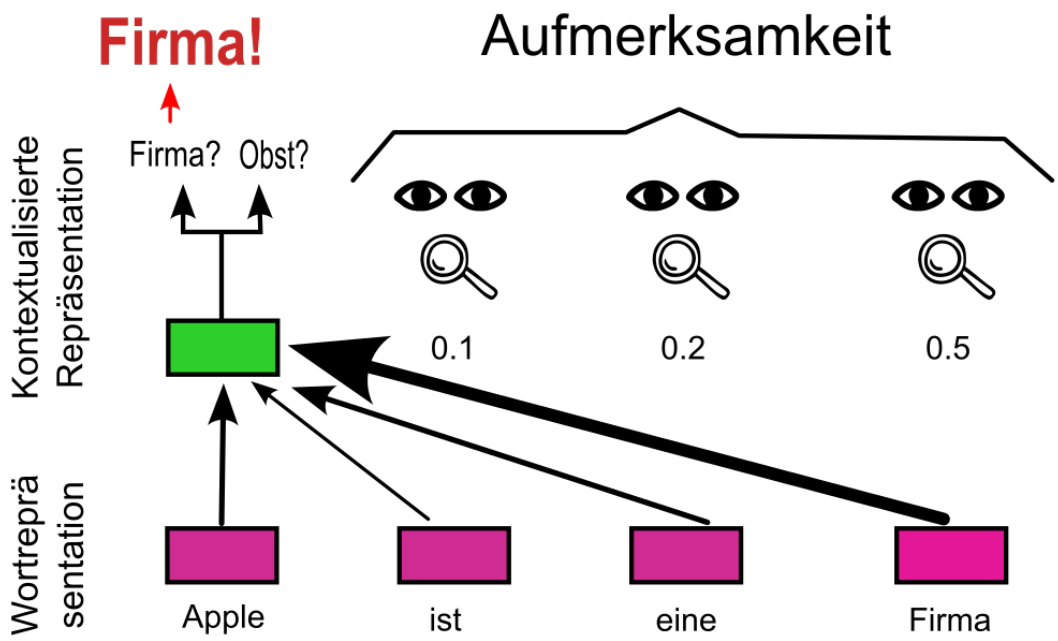
Ohne Kontext, also z. B. einen Satz, in dem das Wort Apple vorkommt, ist dies leider nicht zu entscheiden. Falls wir den Satz „Apple ist eine Firma" gegeben haben, so ist es für uns Menschen leicht zu entscheiden, dass es sich um die Firma handelt. Wie können wir den Kontext nun in die Wort-Repräsentation von „Apple" einbeziehen? Wie können wir also eine „Kontext-abhängige" Repräsentation erstellen? Eine Lösung dazu bietet die Transformer-Architektur.

Ein Schlüsselelement im Transformer ist sein Aufmerksamkeitsmechanismus. Für die Erstellung der Kontext-abhängigen Repräsentation von „Apple" kann der Transformer seine Aufmerksamkeit auf alle anderen Worte richten. Für jedes Wort berechnet der Transformer sodann die Wichtigkeit oder Ähnlichkeit zu „Apple", also wie viel Aufmerksamkeit er darauf richten soll. Die Aufmerksamkeit wird durch eine Zahl zwischen 0 und 1 gegeben. Im Beispiel berechnet der Transformer eine hohe Aufmerksamkeit von 0.5 auf das Wort Firma.

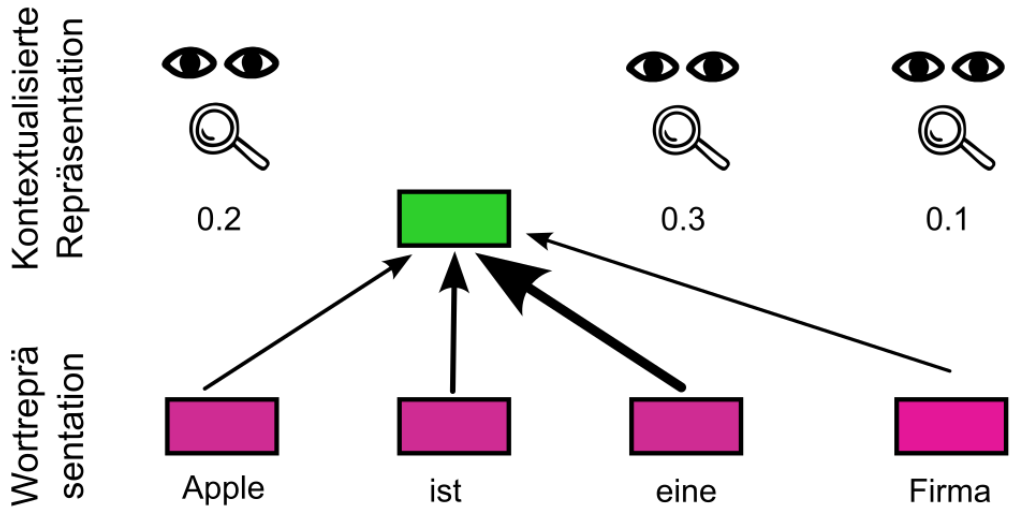


Die Kontext-abhängige Repräsentation für „Apple" in Grün enthält nun Information über alle Wörter in dem Satz. Der Anteil eines Wortes hängt dabei von der Aufmerksamkeit ab. Hier dargestellt durch dickere oder dünnere Pfeile. Da wir nun Informationen über das Wort

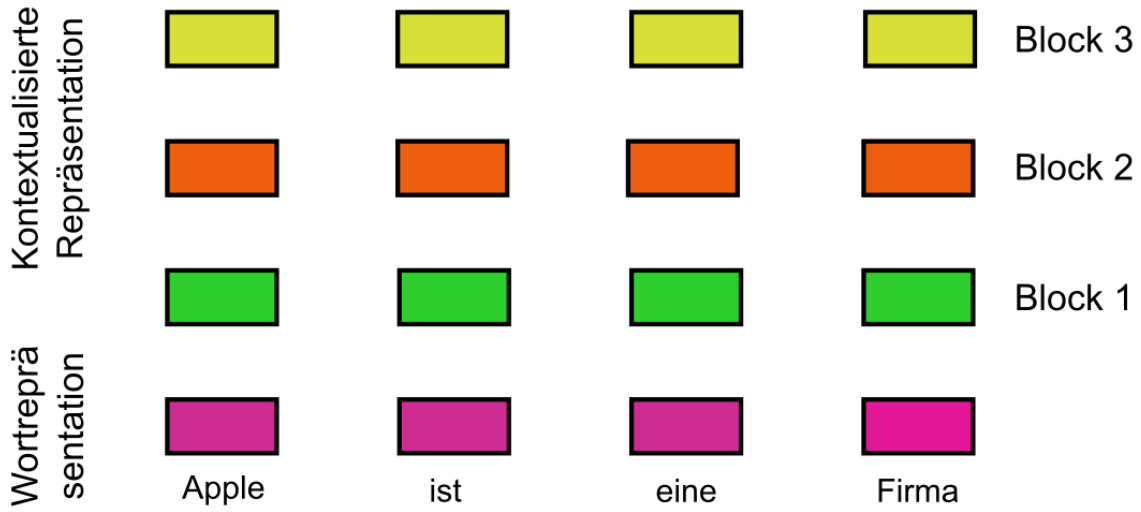
„Firma" miteinbezogen haben, können wir durch die Wort-Repräsentation von Apple entscheiden, dass mit „Apple" die Firma gemeint ist!



Der Transformer berechnet eine Kontext-abhängige Repräsentation, nicht nur für das Wort „Apple", sondern für alle Wörter im Satz … also auch für das Wort „ist". Jede neue Repräsentation enthält somit Information ALLER Wörter, wobei der Beitrag eines Wortes davon abhängt, wie viel Aufmerksamkeit auf das Wort gerichtet wurde.



Der Transformer kann außerdem aus mehreren Transformer-Blöcken bestehen. Jeder Block berechnet eine neue Repräsentation basierend auf der vorherigen. Je mehr Blöcke, desto „tiefer“ ist der Transformer. Tiefere Transformer können komplexere Zusammenhänge herausfinden. Das Sprachmodell „BERT“ zum Beispiel besteht aus 12 solcher Blöcke.



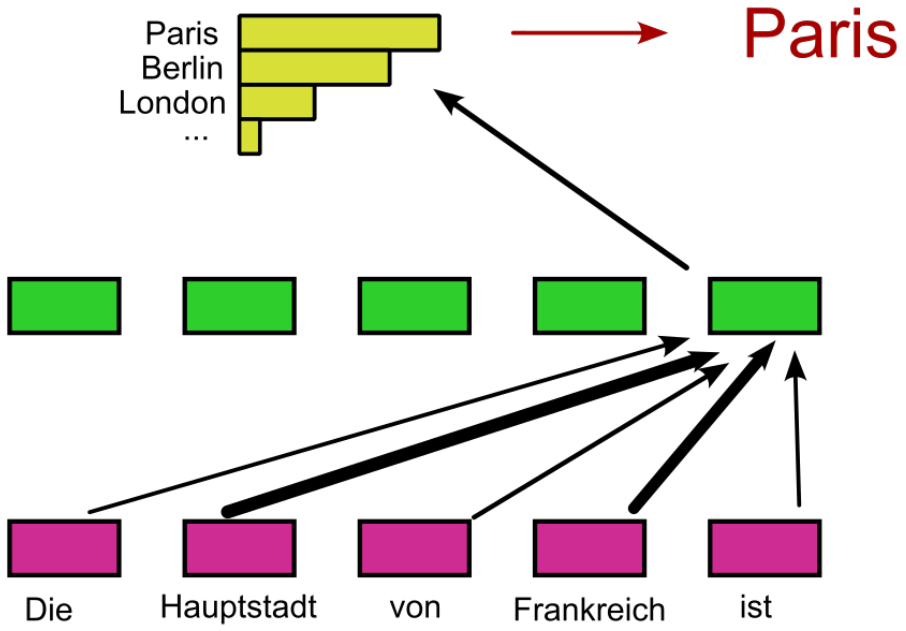
Zudem kann der Transformer alle Wort-Repräsentationen innerhalb eines Blockes gleichzeitig berechnen, was enorme Rechenzeit spart. Außerdem hat er für die Berechnung einer neuen Wort-Repräsentation Zugriff auf alle anderen Wörter, egal wie weit diese Wörter im Text entfernt liegen. Dadurch kann der Transformer sehr gut Zusammenhänge zwischen auch entfernten Wörtern lernen.

Texterzeugung

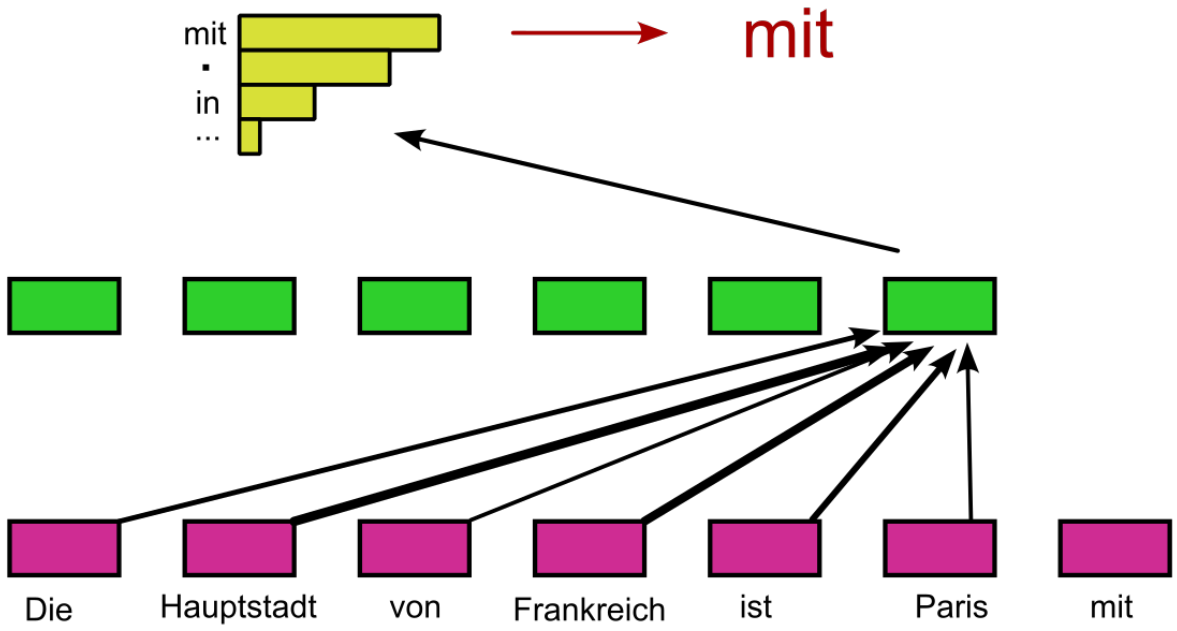
Gerade hast du gelernt, wie der Transformer eine Kontext-abhängige Wort-Repräsentation erzeugen kann. Wir konnten damit entscheiden, ob mit dem Wort „Apple" die Firma oder das Obst gemeint war. Der Transformer kann jedoch noch mehr, nämlich neuen Text erzeugen! Der Transformer berechnet dafür schrittweise das nächste Wort im Text. Das nächste Wort wird vorhergesagt basierend auf der Repräsentation des letzten Wortes im Text.

Im Beispiel möchten wir das nächste Wort des Satzes „Die Hauptstadt von Frankreich ist" bestimmen. Hierzu möchten wir die kontextabhängige Repräsentation von „ist" verwenden. Wie zuvor berechnen wir kontextabhängige Repräsentationen. Um als nächstes Wort „Paris" vorhersagen zu können, ist es wichtig, die Wörter „Hauptstadt" und „Frankreich" einzubeziehen. Hier wieder dargestellt durch dickere Pfeile. Anhand der neuen Repräsentation von „ist" berechnet der Transformer nun für jedes mögliche Wort eine Wahrscheinlichkeit, dass es das nächste Wort ist. Wir können uns dann für das Wort mit der

höchsten Wahrscheinlichkeit entscheiden, im Beispiel „Paris”. Die zweithöchste Wahrscheinlichkeit im Beispiel wäre „Berlin“.



Das vorhergesagte Wort wird nun der Eingabe hinzugefügt und wir können weitere Wörter vorhersagen! Das nächste Wort könnte zum Beispiel „mit” lauten. Ein Punkt, um den Satz zu beenden, wäre aber auch wahrscheinlich.



So können wir schrittweise den Satz vervollständigen zu „Die Hauptstadt von Frankreich ist Paris mit 2 Millionen Einwohner*innen”. Der Transformer kann mit einem speziellen Zeichen dabei selbst bestimmen, wann die Textgenerierung beendet werden soll. Solange dieses spezielle Zeichen nicht benutzt wird, erzeugt der Transformer weitere Wörter und kann somit auch mehrere Textabschnitte auf einmal erzeugen.

Außer dem Aufmerksamkeitsmechanismus berechnet der Transformer weitere interne Repräsentationen basierend auf kleinen neuronalen Netzen. Dies wurde zum Zweck der Einfachheit hier weggelassen.

Abschluss

Zusammenfassend hast du hier die grundlegende Idee des Transformers kennengelernt: Sein Aufmerksamkeitsmechanismus, der es ihm ermöglicht, Aufmerksamkeit auf bestimmte Wörter im Text zu richten. Wir können damit Kontext-abhängige Wort-Repräsentationen berechnen, und zwar gleichzeitig für alle Wörter im Text. Außerdem hast du kennengelernt, wie der Transformer neuen Text erzeugen kann, in dem er Schritt für Schritt das nächste Wort im Text vorhersagt.

Quellen

Quelle [1] C:G. (2006, 29. Dezember). *Transformers - Transformers transforming!* [Video]. YouTube. https://www.youtube.com/watch?v=e8FgDmo_9sY

Quelle [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*. https://dl.acm.org/doi/10.5555/3295222.3295349

Weiterführendes Material

https://huggingface.co/learn/nlp-course

https://colab.research.google.com/github/tensorflow/text/blob/master/docs/tutorials/transformer.ipynb

https://towardsdatascience.com/transformers-141e32e69591

Disclaimer