Textgenerering med hjälp av artificiell intelligens

Dan Le

Kandidatavhandling i datavetenskap

Handledare: Marina Waldén

Fakulteten för naturvetenskaper och teknik

Åbo Akademi

Våren 2023

Referat

Innehåll

1. Introduktion

2. Artificiell intelligens

Artificiell intelligens (AI) hänvisar till simulering av mänsklig intelligens i maskiner som är programmerade att utföra uppgifter som vanligtvis skulle kräva mänsklig intelligens för att utföra, såsom inlärning, problemlösning, beslutsfattande och perception. Målet med AI är att utveckla maskiner som beter sig som om de vore intelligenta. [WE2017]

Området AI är stort och täcker ett brett utbud av tekniker och tillvägagångssätt, inklusive maskininlärning, djupinlärning, naturlig språkbehandling, datorseende, robotik och expertsystem. AI-system är designade för att analysera och bearbeta stora mängder data, känna igen mönster, göra förutsägelser och anpassa sig till nya situationer baserat på tidigare erfarenheter.

Maskininlärning är en viktig del inom AI och det innebär att träna algoritmer på stora datamängder för att de ska kunna identifiera mönster och göra förutsägelser baserat på data. Djupinlärning är en delmängd av maskininlärning som använder neurala nätverk för att bearbeta och analysera data på ett sätt som liknar den mänskliga hjärnan.

Naturlig språkbehandling (NLP) är ett annat område inom AI som hanterar interaktionen mellan datorer och mänskligt språk...

2.1 Innehållsgenerering med AI

Innehållsgenerering med AI hänvisar till användningen av maskininlärnings-algoritmer och NLP-tekniker för att automatiskt producera skrivet, ljud eller visuellt innehåll. Denna teknik har potential att revolutionera sättet att skapa innehåll, eftersom den kan generera material av hög kvalitet med en omfattning och hastighet som är omöjlig för en människa att efterlikna.

AI-genererat innehåll används i branscher som marknadsföring, journalistik, underhållning med mera eftersom tekniken kan analysera data om kunders preferenser, beteenden och demografi för att generera innehåll som resonerar med publiken.

Det finns också vissa bekymmer angående AI-genererat innehåll. Ett exempel är gällande upphovsrätten till det genererade innehållet. Det är svårt att avgöra vem som äger innehållet då AI-verktygen använder sig av många olika källor för att generera innehållet. Ibland är det också väldigt tydligt att innehållet i frågan är AI-genererat, så mänsklig redigering kan behövas, och på grund av att AI-verktygen endast använder existerande data för innehållsgenerering så skapas inte nya idéer heller.

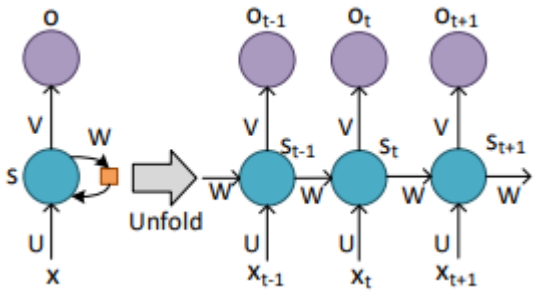
2.1.1 Textgenerering

Textgenerering är processen att använda beräkningsalgoritmer för att producera nya, sammanhängande textstycken som linkar naturligt språk. Textgenerering är ett viktigt forskningsområde inom naturlig språkbehandling och har många tillämpningsmöjligheter som gör det möjligt för datorer att lära sig att uttrycka sig som människor med olika typer av information för att kunna ersätta människor i en rad olika uppgifter. [källa]

En av de vanligaste teknikerna för textgenerering är språkmodellering, som innebär att en beräkningsmodell tränas på en stor textkorpus. Modellen lär sig att förutsäga sannolikheten för ett givet ord eller en ordföljd utifrån det sammanhang där det förekommer. Detta tillvägagångssätt har lett till utvecklingen av flera kraftfulla algoritmer, till exempel återkommande neurala nätverk (eng. Recurrent neural network, RNN) och transformatorer, som avsevärt har förbättrat kvaliteten av textgenerering.

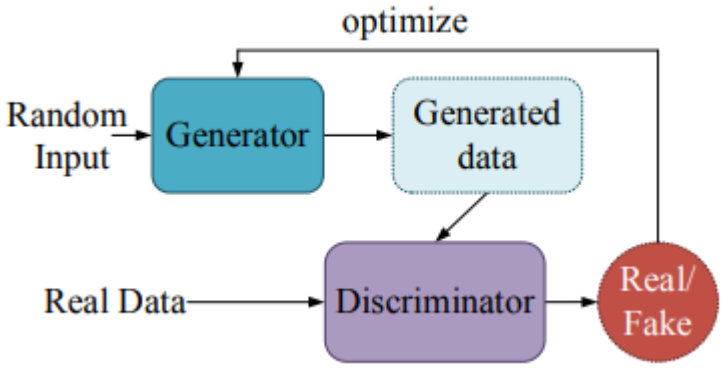
RNN är en speciell neural nätverksstruktur och spelar en betydelsefull roll i textgenerering eftersom de är särskilt utformade för att hantera sekventiella data, såsom text. Jämfört med traditionella neurala nätverk, som bearbetar in- och utmatningar av fast storlek, kan RNN ta emot inmatningar av varierande längd, samt producera utmatningar av varierande längd.

Till skillnad från ”deep neural network” (DNN) och ”convolutional neural network” (CNN), ger nätverket en ”minnesfunktion” för det föregående innehållet. [ZGWLHY2019] Detta minne uppdateras vid varje tidssteg baserat på den aktuella inmatningen och det tidigare tillståndet, vilket gör att nätverket kan modellera beroenden mellan ord i sekvensen. Utmatningen vid varje tidssteg används sedan för att förutsäga nästa ord i sekvensen.



RNN kan komma ihåg tidigare information och använda den för att beräkna det nya nuvarande utdata. Nätverket kan fortsätta generera tills ett stoppvillkor uppfylls. Stoppvillkoret kan till exempel vara en maximal längd eller slutet av en mening.

En annan viktig aspekt inom textgenerering är användningen av ”generative adversarial networks” (GAN), som innebär att man tränar två separata nätverk. Det första nätverket, som kallas för generatorn, skapar falska textprover som ligger närmast den verkliga fördelningen. Det andra nätverket, diskriminatorn, används för att särskilja genererade textprover och verkliga textprover. [ZGWLHY2019] De två nätverken tränas tillsammans med målet att förbättra kvaliteten på den genererade texten med tiden.



De senaste framstegen inom textgenerering har också lett till utvecklingen av storskaliga språkmodeller, till exempel GPT-3, som kan generera högkvalitativ text om ett stort antal ämnen. Dessa modeller är förtränade på stora mängder textdata och kan sedan finjusteras för specifika uppgifter, till exempel språköversättning.

2.2 Förstärkningsinlärning

Förstärkningsinlärning är en gren inom maskininlärning som handlar om beslutsprocesser i dynamiska och osäkra miljöer. Förstärkningsinlärning skiljer sig från andra beräkningsmetoder genom sin betoning på att en agent lär sig från direkt interaktion med sin omgivning utan att förlita sig på exemplarisk övervakning eller kompletta modeller av miljön. [ZGWLHY2019]

Målet med förstärkningsinlärning är att lära AI-systemet att fatta de bästa möjliga besluten för att uppnå ett visst mål. Målen måste vara relaterade till miljöns tillstånd [SB2015] och beroende på hur agenten presterar så får den feedback i form av belöningar eller eventuella straff. Utöver agenten och miljön kan man identifiera fyra huvudelement av ett förstärkningsinlärningssystem: en *policy*, en *belönings-signal*, en *värdefunktion* och en valfri *modell* av miljön. [SB2015]

En policy definierar agentens sätt att bete sig vid en given tidpunkt. [SB2015] Policyn är en kartläggning från stater till åtgärder, och den talar om för agenten vilken åtgärd som ska vidtas i varje tillstånd. I vissa fall kan policyn vara en enkel funktion eller uppslagstabell och i andra fall kan det innebära omfattande beräkningar såsom en sökprocess. Agenten lär sig policyn genom försök och misstag genom att utforska miljön och får sedan feedback beroende på prestation. I allmänhet kan policyer vara stokastiska. [SB2015]

En belöningssignal definierar målet i ett förstärkningsinlärningsproblem. Vid varje tidssteg skickar miljön ett nummer åt agenten. Numret anger belöningens storlek och agentens mål är att maximera den totala belöningen den får på lång sikt. [SB2015] Belöningssignalen definierar alltså vilka händelser som är bra och dåliga för agenten och det enda sättet som agenten kan påverka belöningssignalen är genom sina handlingar, som kan ha en direkt påverkan på belöningen, eller en indirekt effekt genom att förändra miljöns tillstånd.

Till skillnad från belöningssignalen som indikerar vad som är bra i en omedelbar mening, så indikerar värdefunktionen vad som är bra på lång sikt. [SB2015]

Modellen efterliknar beteendet hos miljön och gör det möjligt att dra slutsatser om hur miljön beter sig. Modeller används för att planera hur framtida situationer ska hanteras innan de har upplevts. [SB2015] Det finns två tillvägagångssätt för att lösa förstärkningsinlärningsproblem: modellbaserad och modellfri. Modellbaserad förstärkningsinlärning innebär att man lär sig modell av miljön medan modellfri lär sig via försök och misstag.

3. Textgenererare

AI-textgeneratorer har på blivit väldigt populära under de senaste åren tack vare deras förmåga att generera människoliknande text till följd av olika uppgifter gällande behandling av naturligt språk. Dessa textgeneratorer baseras på modeller för djupinlärning som kan lära sig det naturliga språkets struktur från stora mängder textdata, och använda denna kunskap för at generera ny text som är mycket lik mänsklig skrift. Två populära exempel på textgenererare är ChatGPT och AI Dungeon, som är utformade för olika ändamål men delar många gemensamma funktioner.

3.1 ChatGPT

ChatGPT (Chat Generative Pre-trained Transformer) är en språkmodell utvecklad av OpenAI som lanserades i november 2022. Den är konstruerad för att bearbeta och förstå naturlig språkinmatning för att sedan generera mänskliga svar och föra konversationer med användare. Den stora mängden data som programmet har tränats med kommer från en mängd olika källor, inklusive böcker, webbsidor och andra typer av texter för att lära sig mönster och samband mellan ord och meningar. Detta gör det möjligt för programmet att generera en mängd olika svar samt konversera med användaren gällande ett stort antal olika ämnen.

ChatGPT använder sig av en djupinlärningsalgoritm som kallas för transformatormodell (eng. transformer model). Transformatormodellen är utformad för att bearbeta sekvenser av tokens, såsom ord eller tecken, parallellt. Modellen består i sin grund av en serie kodnings- och avkodningslager. [MP2022] Kodningslagren tar emot en sekvens av inmatningstoken och kodar dem till en sekvens av dolda tillstånd med hjälp av en självuppmärksamhetsmekanism (eng. self-attention). [VSPUJGKP2017] Avkodningslagret tar sedan in de kodade tillstånden och genererar en sekvens av utmatningstoken med hjälp av en kombination av självuppmärksamhet och uppmärksamhet från kodare och avkodare.

En av de största fördelarna med transformatormodellen är dess förmåga att hantera avlägsna beroenden utan att behöva använda RNN, som kan vara långsamma och beräkningsmässigt kostsamma. Detta uppnås genom självuppmärksamhets-mekanismen, som gör det möjligt för modellen att selektivt fokusera på olika delar av inmatningssekvensen när den genererar utmatningssekvensen.

För att generera svar på användarinmatning använder ChatGPT transformator-modellen. Den är förtränad på en stor korpus av textdata och finjusteras sedan på en specifik uppgift, som att svara på frågor eller föra en konversation. När en användare matar in ett meddelande använder modellen inmatningen för att generera en sekvens av dolda tillstånd, som sedan används för att generera ett svar med hjälp av avkodningslagren. Svaren returneras sedan till användaren och processen upprepas för varje efterföljande meddelande. Modellen uppdateras kontinuerligt med nya data, vilket gör att den kan anpassas och förbättras över tid.

ChatGPT fungerar som en chatbot; användaren matar in ett meddelande och programmet behandlar och delar upp texten och meningar i mindre delar, s.k. ”tokens”, vilket hjälper programmet att förstå enskilda ord och hur de fungerar i längre text. Modellen tar sedan in dessa tokens och använder sina språkkunskaper för att generera ett svar baserat på inmatningsmeddelandet och konversationens kontext.

3.2 AI Dungeon

AI Dungeon är ett textbaserat äventyrsspel som använder sig av artificiell intelligens för att skapa innehåll som reflekterar spelarens inmatningar. Spelet skapades av Nick Walton år 2019 och använder OpenAIs GPT-3 modell.

Spelaren börjar med att välja en av ett fåtal färdiga spelvärldar. Beroende på vilken värld som väljs så får spelaren välja mellan ett antal relevanta karaktärer och namnge den. Sedan börjar en berättelse och spelaren kan därefter välja mellan tre olika sätt att engagera sig med deras textinmatning. Spelaren kan skriva vad den vill och AI:n kommer att använda det för att generera nästa del av berättelsen. AI:n kan även lägga till karaktärer och skapa slumpmässiga händelser i berättelsen vilket gör spelet oförutsägbart. Spelet fortsätter så länge spelaren vill eftersom AI:n kommer att generera nya delar baserat på inmatning. [källa]

AI Dungeon använder sig av en kombination mellan behandling av naturligt språk, maskininlärning och djupinlärning för att generera berättelserna. GPT-3 modellen genererar text som är sammanhängande och engagerande.

3.3 Generative pre-trained transformer

Generative pre-trained transformer (GPT) är en autoregressiv språkmodell, utvecklad av OpenAI och baserad på ”transformer” arkitekturen som introducerades av Vaswani et al. 2017 [VSPUJGKP2017]. GPT använder en tvåstegsträningsstrategi av förträning och finjustering. Modellen är förtränad på en stor ”textkorpus” med hjälp av oövervakad inlärning vilket innebär att den lär sig representera språkets mening och struktur utan tydliga instruktioner eller återkopplingar från människor. När GPT-modellen är förtränad kan den finjusteras på en specifik uppgift, till exempel maskinöversättning eller textgenerering. Finjusteringen innebär att modellen tränar på en anpassad och uppgiftsorienterad datamängd så att den kan generera mer exakta och lämpliga svar.

GPT-3 är den tredje iterationen inom GPT-serien, och den är den största språkmodellen hittills. GPT-3-modellen är tränad på datamängder som innehåller 400 miljarder ”tokens” och har över 175 miljarder parametrar.

3.4 Jämförelser och tester

4. AI-genererade textsammanfattningar

AI-genererade textsammanfattningar är en annan viktig forskningsfråga inom textgenerering. Syftet med AI-genererad textsammanfattning är ge en kortfattad beskrivning för användare genom att komprimera och förfina originaltexten och bevara innehållets väsentliga informationskomponenter och betydelse. [ZGWLHY2019]

Textsammanfattning kan ses som en process för informationssyntetisering, där ett eller flera indatadokument integreras i ett kort sammandrag. Textsammanfattningen kan uppnås med två olika metoder – ”hämtningsbaserad” metod och generativ metod. [ZGWLHY2019] [MP2022]

Den hämtningsbaserade metoden innebär att man väljer ut de viktigaste meningarna eller fraserna, som innehåller det mest nödvändiga och lämpliga informationen, från originaltexten och kombinerar dem för att skapa en sammanfattning. [MP2022] Detta tillvägagångssätt är relativt okomplicerat eftersom det inte innebär att man måste generera ny text. Det kan dock resultera i sammanfattningar som saknar sammanhang.

Till skillnad från den hämtningsbaserade metoden så kan den generativa metoden skapa meningar som inte finns i den ursprungliga texten. [MP2022] Den generativa metoden skapar ny text som fångar den viktigaste informationen och innebörden från originaltexten, snarare än att kopiera och klistra in befintlig text. Denna metod är mer utmanande eftersom det kräver att AI-systemet förstår innebörden och sammanhanget i den inmatade texten, samt genererar text som är grammatiskt korrekt och sammanhängande.

5. Likheter mellan AI-genererade resultat och mänskligt skrivande

5.1 Plagiering via textgenerering

5.2 Plagieringsdetektering med AI

6. Sammanfattning