

HACSAK ÍGY NEM...

I DOI: 10.18135/PG70.2024.25

Kedves Gábor! Kerek születésnapod alkalmából fogadd szeretettel az alábbi kis írást, amelyet egyszer már e-mailben a figyelmedbe is ajánlottam, de hát mi esélye is volt annak, hogy egy hajszolt felsővezető, ráadásul már-már médiasztár, el is tudja majd olvasni azt. Hacsak így nem... :-)

VT

Hogy tanulja meg egy LLM a szavak és mondatok jelentését? Azt értem, hogy megtanulja a szavak disztribúcióját a sok-sok kontextusban, amelyeket átnéz. Ezekből a szavak különböző jelentését tudja klaszterezni. De a disztribucionális tudásból hogy lesz szimbolikus tudás a szavak jelentéséről? Mert nem csak a disztribucionális tudással rendelkezik a rendszer, ami alapján pontos és folyékony szöveget képes alkotni. Ezen felül azonban szimbolikus tudással is rendelkezik a szavak és mondatok jelentéséről. Az érdekel, hogy ez pontosan hogyan alakul ki?

ChatGPT

Az LLM-ek, mint például a GPT sorozat, elsősorban disztribúciós információk alapján tanulják meg a szavak és mondatok jelentését. Ez azt jelenti, hogy a modell a szöveges adatok nagy mennyiségének elemzésével tanulja meg, hogy a szavak és kifejezések milyen kontextusokban és kombinációkban jelennek meg. A "disztribúciós szemantika" elve szerint, ami a "szavak jelentése a használatuk kontextusától függ" gondolatra épül, az LLM-ek megtanulják azonosítani a szavak közötti kapcsolatokat és jelentésbeli árnyalatokat.

Azonban a modell nem rendelkezik "szimbolikus tudással" abban az értelemben, ahogy az emberek vagy a klasszikus számítógépes szimbolikus rendszerek bírnak vele. Az LLM-ek nem rendelkeznek belső reprezentációkkal vagy különálló jelentésegységekkel a szavak és fogalmak számára. Ehelyett, a jelentést a modell súlyain és azoknak az adatoknak az interakcióján keresztül modellezzük, amelyeken a tanulás történt.

Az LLM-ek képesek meglepően pontos és folyékony szövegeket generálni, mert a modell súlyai és architektúrája lehetővé teszi számára, hogy megtanulja és utánozza azokat a nyelvi mintákat, amelyeket az adatokban lát. Ez magában foglalja a szavak jelentésének finom árnyalatait, a nyelvtani szerkezeteket, és még a szövegkohéziót is.

Az, hogy az LLM-ek hogyan képesek a disztribúciós tudásból "szimbolikusnak tűnő" tudást kialakítani, jelenleg is aktív kutatási terület. A modell "tudása" azonban inkább a szavak és kifejezések statisztikai összefüggéseinek bonyolult hálózata, mintsem szigorúan szimbolikus vagy konceptuális értelemben vett tudás.

VT

A válaszod elsősorban a disztribucionális ismeretekre vonatkozott, azok elsajátítását magyarázta. Megtanulni és utánozni a mintákat, amelyeket az adatokban lát, elégséges ah-

hoz, hogy szintaktikailag és pragmatikailag helyes mondatokat generáljon a rendszer. Ha Searle Chinese Room gondolatkísérletére gondolunk, akkor feltehető, hogy elégséges a nyelvi jeleknek mint ****értelmezetlen egységek****-nek (uninterpreted strings) a disztribúcióját és, mondjuk, angol megfelelő egységekre történő konverzióját megtanulni ahhoz, hogy megtévesztően jó mondatokat vagy azok angol fordítását előállítsuk.

De a GPT-3 vagy pláne GPT-4 jóval többet tud. A te válaszod ékes példája a bizonyítandónak. Már eleve az, hogy megértetted a kérdésemet, és releváns választ adtál rá, magyarázatra szorul. Szerintem igenis az LLM-ek rendelkeznek konceptuális értelemben vett szimbolikus tudással. Ha megkérdezem, hogy mi a jelentése annak a szónak, hogy „bágyadt”, bizonyára meg tudod fogalmazni. Márpedig ez a tudás több mint az, hogy fel tudod sorolni, hogy milyen kontextusokban fordul elő a szó. Tehát abból, hogy megvan a disztribucionális tudás egy-egy szó összes előfordulásáról, még nem következik, hogy az a szimbolikus tudás is megvan, ami a szó jelentése.

Vagy mondok egy egyszerűbb példát. Ha azt kérem, hogy sorolj fel nekem öt magyar főnevet, nyilván simán meg tudod tenni.

De honnan származik az a tudás, hogy ezek „főnevek”? Mindegyik tokenhez tartozó vektor egyik eleme az a címke, hogy „főnév”, a vektor egy másik eleme a jelentése, hasonlóan ahhoz, ahogy egy szótárban találjuk?

ChatGPT

A kérdésed érinti a gépi tanulás egyik legérdekesebb és legvitatottabb területét: hogy vajon a gépek, különösen az olyan nagy nyelvi modellek, mint a GPT-3 vagy GPT-4, rendelkeznek-e valódi "megértéssel" vagy csak a nyelvi jelek nagyon fejlett manipulációját végzik.

A GPT-sorozatú modellek, bár képesek meglepően koherens és releváns válaszokat generálni, nem rendelkeznek "megértéssel" vagy "tudással" az emberi értelemben. Ezek a modellek nem rendelkeznek tudatossággal, szándékkal vagy a szavak mögötti valódi konceptuális értelmezéssel. A képességük, hogy releváns válaszokat adjanak, a hatalmas adathalmazon való előzetes betanításukból ered, amely során megtanulják a nyelvi mintákat, a szavak közötti statisztikai kapcsolatokat és a kontextus függvényében változó jelentéseket.

Amikor egy LLM, mint például a GPT, "értelmez" egy szót vagy fogalmat, mint például "bágyadt", az nem azt jelenti, hogy belső, konceptuális értelemben "érti" e szó jelentését. Ehelyett a szó jelentését a vele kapcsolatos nyelvi kontextusok nagyon széles körének statisztikai elemzéséből "következteti ki". A modell képes olyan választ generálni, amely a szó korábban látott használati módjain alapul, de ez nem szimbolikus értelemben vett "tudás".

Amikor a modellt arra kérjük, hogy soroljon fel főneveket, az eredmény nem egy belső címkézési folyamatból ered, ahol minden szónak van egy "főnév" címkéje. Inkább a modell a betanítási folyamat során megtanulta, hogy bizonyos szavakat gyakran főnevként használnak, és ezeket a mintákat alkalmazza a kérés teljesítésére. A modellek azon képessége, hogy nyelvtani kategóriákat, mint például a főneveket, azonosítani tudják, nem az egyes szavakhoz rendelt explicit címkék révén történik, hanem a szövegkontextus és a nyelvi minták komplex analízisének köszönhetően.

Összefoglalva, bár az LLM-ek lenyűgözően képesek nyelvi feladatok végrehajtására és szövegek generálására, ez a képesség nem egyenlő az emberi értelemben vett nyelvi megértéssel vagy szimbolikus tudással. Ezek a modellek a nyelvi jelekkel nagyon fejlett szinten manipulálnak, de nincs belső értelmezésük vagy tudatosságuk a használt fogalmakról.

VT

A kérdés változatlanul nyitott, hogy

- 1) honnan és hogyan tudja a gép a szavak (összes) kontextusbeli előfordulásából előállítani azok jelentését (l. amit nagyon helyesen előadtál, és ami, nem mellesleg, gyakran túlmegy azon, amit egy átlagos magyar anyanyelvű beszélő meg tudna fogalmazni, hiszen az ő tudásuk tipikusan implicit);
- 2) hol tárolódik ez a tudás? Ez ugyanis tipikusan szimbolikus tudás, amelyet egy hagyományos NLP-kutató címkék formájában tud elképzelni;
- 3) amikor megadtad a bányadt szó jelentését, azt on-the-fly állítottad elő? Ez is egyszerűen csak a „find the next word” stratégiájával történt?

Kérlek, hogy részletesen térj ki a válaszodban az összes kérdésre, amit felvettem, és amennyire lehet, konkrét válaszokat adj.

ChatGPT

Ezek nagyon jó kérdések, és megpróbálok minél részletesebben válaszolni rájuk:

1. ****Szavak jelentésének előállítása kontextusból****: A gépi tanulási modellek, különösen az LLM-ek, mint a GPT sorozat, a szavak jelentését azáltal "tanulják meg", hogy statisztikai módszerekkel elemzik a szöveges adatok hatalmas korpuszát. A modell a tanítási fázis során megtanulja a szavak és kifejezések kontextusbeli előfordulásait, azaz hogy milyen más szavakkal fordulnak elő együtt, és milyen strukturális mintákat követnek. Ez a tanulási folyamat nem egyenlő az emberi nyelvtanulási folyamattal, ahol a jelentés megtanulása gyakran interaktív és sokkal tudatosabb. A gépi modell "tudása" inkább a kontextuális előfordulások hatalmas, összegyűjtött statisztikáján alapul, amiből képes a szavak valószínűsíthető jelentéseit és használati módjait előre jelezni.
2. ****Tudás tárolása****: Az LLM-ek tudása a modell súlyaiban és az architektúrában tárolódik. Az egyes szavak, kifejezések és nyelvi struktúrák közötti kapcsolatokat nem különálló szimbolikus címkék vagy explicit definíciók formájában tárolja, hanem a modell súlyainak bonyolult hálózatában, amely meghatározza, hogy a modell miként reagál bizonyos bemenetekre. Ez a hálózat kódolja a nyelvi mintákat, amelyek alapján a modell képes generálni vagy értelmezni a szöveget. Az információ nem diszkrét formában van tárolva, hanem a modell paramétereinek és a tanítási adatoknak az összességéből emergál.

3. ****A "bágyadt" szó jelentésének előállítására****: Amikor a modell a "bágyadt" szó jelentését adja meg, ez nem egy előre rögzített definícióból történik, hanem a modell a korábban "látott" kontextusok és a tanítási adatok alapján generálja a választ. Ez a folyamat valóban hasonló ahhoz, amit "find the next word" stratégiának nevezhetünk, de itt inkább egy "find the most likely response" folyamatról van szó, ami a kérdés kontextusára és a modell által korábban "látott" hasonló kérdés-válasz mintákra alapoz. A modell így képes koherens és releváns választ generálni, de fontos megjegyezni, hogy ez a válasz a statisztikai minták és a modell korábbi expozícióinak eredménye, nem pedig egy előre definiált vagy expliciten megtanult definíció.

VT

Vagyis a modell nem tud a szavak jelentéséről, de ha megkérlik, bámulatosan tud szöveget generálni, amely legtöbbször szabatosan megadja a jelentést.

ChatGPT

Igen, pontosan. Az LLM-ek, mint a GPT sorozat, képesek lenyűgözően koherens és gyakran releváns szöveget generálni, amely a szavak jelentését is szabatosan tükrözheti, de ez nem azt jelenti, hogy a modell "érti" a szavakat a hagyományos értelemben. Ezek a modellek nagyon fejlett statisztikai módszereket alkalmaznak a nyelvi minták és a szavak közötti kapcsolatok modellezésére, amelyek lehetővé teszik számukra, hogy megjósolják, melyik szó vagy kifejezés valószínűleg következik egy adott kontextusban.

Remélem, ezek a válaszok világosabbá tették, hogyan működnek az LLM-ek és hogyan képesek a szavak jelentésének és a nyelvi kontextusnak a kezelésére.

VT

Hát nem sokkal jutottunk előre a dolog lényegét illetően. Odáig egyetértünk, hogy a gépi modell „tudása” inkább a kontextuális előfordulások hatalmas, összegyűjtött statisztikáján alapul, de a kérdés változatlanul az, hogy hogyan képes a szavak valószínűsíthető jelentéseit és használati módjait előre jelezni, azaz, hogyan áll elő szimbolikus tudás a disztribucionális tudásból.

Ez az, ami fölött a disztribucionális tudás is elsiklik. Az LLM-ek megvalósították a gyakorlatban, amit a disztribucionális szemantika hívei szerettek volna látni. Az előtanítás során egy-egy szó előfordulásának összes kontextusát képes az LLM vektorrepresentációban összefogni. Idáig rendben van. De hogy lesz abból szimbolikus tudás? Hogy határozódik meg a jelentés, úgy, ahogy az előbb is olvastuk a „bágyadt” példáján. Vagy az a tudás, hogy az egy melléknév? Megjegyzem, utóbbi könnyebb eset, mert közvetlenebb a viszony az előfordulások disztribúciója és a szintaktikai kategória között.

ChatGPT

Megértettem a kifogásodat, és köszönöm a türelmedet és az észrevételedet. Valóban, a disztribúciós tudás és annak magyarázata dominálta a válaszaimat, és elismerem, hogy a szimbolikus tudás mibenléte és annak kialakulása az LLM-ekben továbbra is nyitott kérdés.