|  |  |
| --- | --- |
| **Fakultet elektrotehnike i računarstva** | |
| **Predmet:** Seminar 2  **Ak. god.** 2023/2024  **Mentor:** prof. dr. sc. Tomislav Petković | **Student:** Kristo Palić  **Matični broj:** 0246074767  Zagreb, 26.4.2024 |

|  |
| --- |
| **1. verzija**  **Vizualni transformatori** |

Vizualni transformatori su jedan od najnovijih i najuzbudljivijih pristupa u području računalnog vida, zasnovani na konceptu transformatora koji su prvi put uvedeni u obradi prirodnog jezika.

**Uvod**

Transformatori su se prvi put pojavili kao revolucionarni modeli u obradi prirodnog jezika s radom "Attention is All You Need" iz 2017. godine. Osnova njihove popularnosti leži u mehanizmu pažnje koji omogućava modelima da daju različitu težinu dijelovima ulaznih podataka, čime se povećava efikasnost i dubina analize. Primjenom ovog koncepta na slike, vizualni transformatori koriste isti princip, tretirajući sliku kao niz elemenata (obično piksela ili manjih regija slike) koje analiziraju koristeći mehanizme pažnje.

**Arhitektura vizualnih transformatora**

Arhitektura vizualnih transformatora obično uključuje više ključnih komponenata:

1. **Tokenizacija slike**: Slika se razbija na manje dijelove ili 'tokene', slično razbijanju teksta na riječi. Ovi tokeni mogu biti direktno pikseli ili pak rezultati primjene konvolucijskih mreža.
2. **Embedding tokena**: Svaki token se zatim pretvara u gusti vektor koristeći učene ili fiksne embeddinge.
3. **Mehanizam pažnje**: Koristeći višeslojne strukture pažnje, model obrađuje tokenizirane i embeddane dijelove slike kako bi razumio kontekstualne veze unutar nje.
4. **Klasifikacijski sloj**: Na kraju, transformator koristi izlaz iz slojeva pažnje za klasifikaciju slike ili za neki drugi oblik izlaza, ovisno o zadatku.

**Primjene**

Vizualni transformatori nalaze primjene u širokom spektru područja unutar računalnog vida, uključujući:

* **Prepoznavanje objekata**: Identificiranje i klasificiranje objekata unutar slika.
* **Segmentacija slike**: Razdvajanje slike na semantički smislene regije.
* **Poboljšanje slika**: Automatsko poboljšavanje kvalitete slika koristeći duboke učene modele.

**Prednosti**

Vizualni transformatori donose nekoliko prednosti:

* **Fleksibilnost i skalabilnost**: Zbog svoje arhitekture, vrlo su prilagodljivi i skalabilni na velike skupove podataka i kompleksne zadatke.
* **Bolje razumijevanje konteksta**: Zahvaljujući mehanizmima pažnje, mogu efikasnije razumjeti kontekst unutar slike, što vodi poboljšanim rezultatima.

**Izazovi**

Unatoč mnogim prednostima, vizualni transformatori suočavaju se s nekoliko izazova:

* **Visoki zahtjevi za računalnim resursima**: Transformatori generalno zahtijevaju značajne količine računalne snage, što može biti ograničavajuće.
* **Složenost treniranja**: Prilagodba i optimizacija transformatora može biti zahtjevna, posebno u smislu balansiranja između veličine modela i performansi.

Mehanizam pažnje je ključna komponenta u suvremenim modelima dubokog učenja, posebice u područjima obrade prirodnog jezika i računalnog vida. Temelji se na ideji da modeli mogu poboljšati svoje performanse usmjeravajući fokus na relevantne dijelove ulaznih podataka tijekom obrade. Ovaj pristup omogućuje modelima da obrađuju informacije dinamično, što je posebno korisno u složenim zadaćama poput prijevoda jezika, generiranja teksta, prepoznavanja objekata i analize slika.

**Kako mehanizam pažnje funkcionira**

Mehanizam pažnje funkcionira tako da dinamički određuje važnost različitih dijelova ulaznih podataka i fokusira se na najrelevantnije dijelove za trenutni zadatak. Osnovna ideja je modelirati interakcije između ulaza u svrhu boljeg izvlačenja i sinteze informacija. U praksi, mehanizam pažnje se može opisati kroz nekoliko ključnih koraka:

1. **Query, Key, Value**: Osnovni elementi mehanizma pažnje su query (upit), key (ključ) i value (vrijednost). U kontekstu obrade jezika, na primjer, query bi mogao biti trenutna riječ koju model pokušava obraditi, dok su key i value atributi drugih riječi u rečenici koje bi mogle biti relevantne za razumijevanje konteksta queryja.
2. **Izračun bodova**: Model izračunava bodove pažnje tako što uspoređuje query s svakim key-em u podacima. Ovi bodovi odražavaju koliko je svaki dio podataka relevantan za trenutno obrađivani query.
3. **Softmax normalizacija**: Bodovi se normaliziraju koristeći softmax funkciju, koja pretvara izračunate bodove u vjerojatnosne vrijednosti koje zbroje do jedan. Ovo omogućuje modelu da distribuira svoj "fokus" ili "pažnju" na način koji favorizira relevantnije podatke.
4. **Težinska suma**: Nakon toga, svaki value se množi s odgovarajućom normaliziranom pažnjom, te se zbroje svi takvi proizvodi. Rezultat je agregirani izlaz koji sintetizira informacije iz ulaznih podataka na način koji uzima u obzir kontekstualnu važnost svakog dijela.

**Primjeri primjene**

Mehanizmi pažnje koriste se u raznim aplikacijama za:

* **Prevođenje jezika**: Omogućuje modelima da se fokusiraju na relevantne riječi u izvornom jeziku kada generiraju prijevod na ciljnom jeziku.
* **Generiranje teksta**: Pomaže u održavanju konteksta i koherencije generiranog teksta, usmjeravajući pažnju na relevantne dijelove prethodnog teksta.
* **Analiza slika**: Primjena pažnje u računalnom vidu omogućuje modelima da identificiraju i fokusiraju se na važne regije unutar slike za zadatke poput prepoznavanja objekata ili segmentacije.

**Izazovi i ograničenja**

Unatoč svojoj svestranosti i učinkovitosti, mehanizmi pažnje donose izazove:

* **Složenost računanja**: Pažnja može biti računski zahtjevna, posebno za velike skupove podataka ili duboke modele.
* **Interpretacija**: Iako mehanizmi pažnje omogućuju modelima da se "fokusiraju", ponekad je teško interpretirati na što točno modeli obraćaju pažnju, što može otežati razumijevanje i poboljšanje modela.

Ukratko, mehanizam pažnje transformirao je pristup rješavanju složenih zadataka u dubokom učenju, pružajući modelima moć da razumiju i interpretiraju velike količine podataka na način koji imitira ljudsku sposobnost fokusiranja na relevantne informacije

**Zaključak**

Vizualni transformatori predstavljaju značajan napredak u računalnom vidu, nudeći novi pristup analizi i obradi slika. Njihova sposobnost da efikasno rješavaju složene zadatke učinila ih je važnim alatom u arsenalu modernih tehnologija računalnog vida. Nastavak razvoja i poboljšanja ove tehnologije vjerojatno će rezultirati još širom primjenom i boljim razumijevanjem vizualnih podataka.