Generisanje teza iz video sadržaja koristeći veštačku inteligenciju

Definicija problema

Cilj projekta je razvoj aplikacije koja na osnovu unetog videa automatski generiše teze koje ne prenose samo osnovno značenje već i dublju strukturu sadržaja. Uz svaku tezu, sistem bi generisao i vremenski trenutak u kojem je ona izrečena.

Motivacija

Količina video sadržaja eksponencijalno raste. Takođe, zbog trenutne situacije u državi, nastava na fakultetima se održava onlajn tako da bi ova aplikacija bila veoma korisna studentima. Osim toga, svako se bar jednom našao u situaciji da treba da pogleda video, a da nema dovoljno vremena ili volje da to učini.

Skup podataka

Za treniranje I evaluacijumodela koriste se javno dostupni skupovi podataka:

1. TvSum I SumMe – sadrže videe koji su anotirani od strane ljudi – naznačeni su delovi videa koji su bitni i koji nisu što značajno pojednostavljuje treniranje modela
2. Video snimci predavanja iz tekuće fakultetske godine

Pretprocesiranje podataka

1. Automatska transkripcija govora (ASR)
2. Čišćenje i normalizacija teksta
3. Detekcija scena i segmentacija videa
4. Poravnanje transkripta sa vremenskim intervalima

Metodologija

1. **Transkripcija videa** - Video materijal se pretvara u tekstualni transkript pomoću gotovog ASR sistema (npr. Whisper). Dobija se tekst sa vremenskim oznakama.
2. **Detekcija ključnih segmenata** - Kombinuju se vizuelni signali (npr. promena kadra, analiza scena) i tekstualne informacije (važnost rečenica, semantička koherentnost) kako bi se izdvojili najznačajniji delovi videa.
3. **Ekstrakcija I generisanje kandidata za teze** - Iz transkripta se koriste tehnike poput TextRank-a, TF-IDF skora i word embeddinga za ekstrakciju važnih rečenica i izraza. Na ovim kandidatima se dalje gradi model koji generiše kratke teze i povezuje ih sa prethodno detektovanim važnim segmentima.
4. **Izlaz sistema** - koherentan sažetak koji se sastoji iz liste teza vremenskih trenutaka relevantnih za svaku tezu.

Evaluacija

Radi evaluacije kvaliteta modela koristiće se sledeće metrike:

1. F1 – meri uspešnost detekcije važnih delova
2. ROUGE – meri Koliko se reči I fraze poklapaju sa ljudskim sažecima
3. BERTScore – meri Koliko se značenje poklapa – čak iako je fomrulisano drugim rečima

Train/Val/Test podela dataset-a u odnosu 70/15/15.

Tehnologije

1. Automatska transkripcija: OpenAI Whisper
2. Programski Jezik: Python(PyTorch, za obradu modela, OpenCV za rad sa videom)
3. Algoritmi za tekstualnu obradu:
    A. TF-IDF, TextRank – za identifikaciju i rangiranje važnih rečenica
    B. Word Embeddings – za merenje sličnosti rečenica i eliminacije duplikata

Literatura i resursi

1. TvSum Dataset
2. SumMe Dataset
3. OpenAI Whisper
4. Word Embeddings
5. Riverside Implemetacija

Mihajlo Orlović SV13/2022

Predlog projekta - Osnove Računraske Inteligencije