

Muturretik muturrerako solasaldi sistema

1. Proposatzailea: Jon Ander Campos

2. Deskribapena

Proiektu honetan ikasketa sakonean oinarritutako muturretik muturrerako solasaldi sistema bat garatuko duzu Bahdanau et al. (2014) lanean oinarritua eta filmetako azpigituluak erabiliz (Lison et al. (2016)). Honetarako, dialogoa itzulpen ataza bat bezala proposatuko dugu, ikusi hurrengo adibidea:

- Itzulpen automatikoa :
 - Sistemaren sarrera -> esaldia jatorri hizkuntza batean: "Egun on guztioi."
 - Sistemaren irteera -> esaldia helburu hizkuntzan: "Buenos días a todos."
- Dialogoa:
 - Sistemaren sarrera -> dialogoko partaide baten esaldia: "Egun on guztioi."
 - Sistemaren irteera -> sarrerako esaldiari erantzuna: "Baita zuri ere."

Proiektuan muturretik muturrerako sare errekorrenteak erabiliko dituzu eta zure sistema Telegramen bot bezala egokitze aukera izango duzu.

3. Helburuak

Helburuak zailtasun mailaren arabera izango dira:

- Z1: Deskargatu ingelesezko entrenatua izan den muturretik muturrerako solasaldi sistema eta probatu ezazu (inferentzia garaian CPUan exekutatzeko gai izan beharko zinateke). Aztertu itzazu ere sistemaren arkitektura eta entrenamendurako erabili diren datuak.
- Z2: Orain duzun sistemak ingelesez bakarrik ulertzen du, zergatik ez hau euskarara moldatu? Deskargatu itzazu euskarazko filmetako azpigituluak eta entrenatu ezazu sistema berri bat. Sistemaren entrenamendua Google Colaboratory erabiliz egin behar baduzu sarearen tamaina txikitu beharko duzu. Kodean bertan topatuko dituzu parametro gomendagarrienak.
- Z3 (1. aukera): Esku artean dituzun sistemekin solasteko modu oso interesgarria eskaintzen du Telegramek. Aukeratu ezazu bi sistemetako bat eta moldatu Telegrameko bot bezala funtziona dezan. Kasu honetan sistema inferentziarako erabili behar denez, CPUan exekutatzeko gai izan beharko zinateke.
- Z3 (2. aukera): Orain arte erabili dituzun sistema guztiak testuingurua kontutan hartu gabe funtzionatzen dute eta hau oso hurbilpen kaxkarra da dialogorako. Hortaz, saiatu zaitez dialogoaren testuingurua kontuan hartzen duen sistema berri bat garatzen. Nahi bezain besteko konplexutasuna gehitu daiteke atal

honetan baina aurreko txandako galdera kontutan hartzea nahikoa izango litzateke testuinguruaren ezagutza minimo bat sistemari emateko.

4. Materialak

Proiektu honetarako materialak hurrengoak dira:

- Z1: Bertan topatuko dituzu sistema eta hau exekutatzeko kodea. Ezertan hasi aurretik irakurri ezazu “IRAKURRI.txt” fitxategia.
- Z2: Hurrengo helbidean hizkuntza askotako azpтитulu fitxategiak dituzu: <http://opus.nlpl.eu/OpenSubtitles-v2018.php>. Euskarakoak [bertan](#) daude.
- Z3 (1.aukera): Hurrengo helbidean topatuko duzu Telegrameko APIa Pythonerako: <https://github.com/python-telegram-bot/python-telegram-bot>. Lehen bota sortzeko jarrabideak [bertan](#) topa ditzakezu.

[1] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*

[2] Lison, P., & Tiedemann, J. (2016). Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles.