საღი აზრით მსჯელობა

(Common Sense Reasoning)

ამოცანა

მას შემდეგ, რაც ChatGPT ხელმისაწვდომი გახდა ფართო პოპულაციისთვის, ცხადი გახდა, რომ ხშირად დიდ ენობრივ მოდელებს კარგად არ ესმით საღი აზრი. საღი აზრის მსჯელობის პრობლემა LLM-ში საკმაოდ რთული ამოცანაა, რომელსაც მარტივი გამოსავალი არ აქვს.

ვიკიპედია ამას გვეუბნება საღ აზრზე: „საღი აზრი არის „ცოდნა, განსჯა და გემოვნება, რომელიც მეტ-ნაკლებად უნივერსალურია და რომელიც მეტ-ნაკლებად იმართება არგუმენტის გარეშე. როგორც ასეთი, ის. ხშირად განიხილება, რომ წარმოადგენს ჯანსაღი პრაქტიკული განსჯის ან ძირითადი ფაქტების ცოდნის საბაზისო დონეს, რომელიც ნებისმიერ ზრდასრულ ადამიანს უნდა ჰქონდეს.”

სრულყოფილად რომ ჩამოვაყალიბოთ ამოცანა, საღი აზრით მსჯელობა გულისხმობს რომ მოდელს უნდა შეეძლოს ისეთ კითხვებზე პასუხის გაცემა რაც აშკარა არის ადამიანისთვის. მაგალითად რომ ავიღოთ:

“თასი ზედმეტად დიდია რომ ჩაეტიოს ჩემოდანში. რა არის ზედმეტად დიდი?” ამ კითხვაზე პასუხის გაცემა უნდა შეეძლოს მოდელზე.

არსებული Benchmark Dataset-ები

ამ პრობლემისთვის

Event2Mind - ამ ბენჩმარკ დატასეტით გატესტვა შეიძლება თუ მოდელებს რამდენად კარგად შეუძლიათ მიიღონ დასკვნა თუ რატომ უნდოდა ვინმეს ამ მოვლენის გამოწვევა და თუ რა რეაქციები ჰქონდათ ამ მოვლენაში მონაწილე პირებს. კონკრეტულად, ეს ბენჩმარკი ამოწმებს ორი სახის დასკვნას:

1. Intent (სურვილი) - ეს ანოტაცია აღწერს თუ რა იყო X ადამიანის Event (მოვლენის) განხორციელების სურვილი, მაგალითად “X gives Y a gift” აქ Intent იქნება “To make Y happy”.
2. Reactions (რეაქციები) - ესენი აღწერენ როგორც X, ასევე Y პირების რეაქციებს. წინა მაგალითიდან გამომდინარე Y-ის რეაქცია იქნება “Y feels grateful”.

ამ დატასეტში სწორად ეს ველები არის ანოტირებული, რამდენიმე სხვა და სხვა ვარიანტით. Intent - Event - X Reaction - Y Reaction

ნაშრომში განხილული აქვთ მოდელი რომელიც აგენერირებს ამ სამივე პრაგმატულ დასკვნას Event-იდან გამომდინარე. თითოეული მოვლენა არის ენკოდირებული Word Embedding-ების მიმდევრობად და ტრანსფორმირებულია ფიქსირებული ზომის ვექტორში. ამ ვექტორს გადასცემენ თავიანთ მოდელს Input-ად.

მოდელს გააჩნია სამი დეკოდირების მოდული, თითო მათგანს გადაეცემა დაენკოდირებული Phrase (ფრაზა) და პასუხად იძლევიან სამ დისტრიბუციას: Vi, Vx, Vo (Intention, X Reaction, Y/Others Reaction). იყენებენ ორ სხვა და სხვა დეკოდერის სეტაპს:

1. N-gram Re-ranking - ფრაზის ემბედინგს აპროექტებს (projects) დაბალ განზომილებიან ვექტორში, რომელსაც softmax-ს მოსზდებს ზემოდან. უშუალოდ ეს არის ფორმულა Vi = softmax(Wi\*phrase + bi) სადაც Vi არის X-ის Intent-ის დისტრიბუცია.
2. Sequence Generation - RNN-ის გამოყენებით (GRU) დეკოდერის საწყის state-ს ანიჭებენ ფრაზის ემბედინგს და შემდგომ სიტყვა სიტყვით აგენერირებენ.

არსებული პრობლემები

რამდენიმე არსებული პრობლემა გააჩნია ამ განხრით პრობლემის გადაჭრას, რაც საკმაოდ მნიშვნელოვანია რომ გადაიჭრას

1. პირველ რიგში, ბაიასების არსებობა. ასეთი საღი აზრის მოდელები უფრო მეტად არიან მსხვერპლი ბაიასის სწავლის, რის შედეგადაც ასეთი მოდელების წინასწარმეტყველებები კარგავენ სიზუსტეს.
2. ყველაზე დიდი პრობლემა ამ განხრით არის დატასეტების შექმნა. ესეთი კონკრეტული პრობლემა უნდა იყოს სრულიად ანოტირებული, ავტომატურად ამის გაკეთება შეუძლებელი ან რთულია, შესაბამისად საკმაოდ დიდი რესურსი იხარჯება მაღალი ხარისხის დატასეტების შექმნაში.

ამ კონკრეტულ ამოცანაზე საკმაოდ ნაკლები კვლევები არის ჩატარებული, ჩემი ვარაუდით ეს ასე იმიტომ არის რომ უფრო მეტად პოპულარული გახდა LLM მოდელები რომლებიც გიგანტურ დატაზე არიან დატრენინგებულები და Good Enough პერფორმანსს დებენ საღი აზრის ბენჩმარკებზე, თუმცა, ვფიქრობ რომ საჭირო არის მაინც ამ განხრით უფრო მაღალი ხარისხის და შედარებით დიდი დატასეტების შექმნა რადგან საღი აზროვნების მოდელების განსავითარებლად და შესასწავლად გამოვიყენოთ. ვფიქრობ რომ მარტო მეტი დატათი არ გადაიჭრება LLM-ების ეს პრობლემა და საჭიროა უშუალოდ უკეთესი მეთოდების პოვნა.

სამომავლო სამუშაო

საკმაოდ ბევრი სამუშაო არის ამ პრობლემის განხრით რაც ბევრ შედეგს მოგვცემს

1. კარგი იქნება თუ ისეთი მოდელები შეიქმნება რომლებიც შეძლებენ ისეთი წინასწარმეტყვლებების დაგენერირებას რომლებიც დასატრენინგებელ დატაში არ ჰქონდათ, Event2Mind მაგალითად ასეთ წინასწარმეტყველებებს ვერ იძლევა. ამისათვის შეიძლება უფრო მრავალფეროვანი დატასეტები, unsupervised learning ან online learning დაგვეხმაროს.
2. დამატებით შეიძლება მულტიმოდალური დატა გამოიყენებოდეს უკეთესი მოდელებისთვის, მაგალითად ვიზუალური სურათებით + ტექსტური კონტექსთით რომ შეეძლოს მოდელს წინასწარმეტყველება.

წყაროები

*Event2Mind: Commonsense Inference on Events, Intents, and Reactions*. https://arxiv.org/pdf/1805.06939.

Grosseteste, Robert, and Roger Bacon. “Common sense.” *Wikipedia*, https://en.wikipedia.org/wiki/Common\_sense.

*Nlp-Progress: Common Sense*. https://nlpprogress.com/english/common\_sense.html.