

Fine tune large language models (LLM's)

- Ενα μεγάλο μοντέλο γλωσσικής πρόβλεψης που έχει εκπαιδευτεί σε τεράστιο όγκο δεδομένων κειμένου.
- Κατανοεί την ανθρώπινη γλώσσα και παρέχει κείμενο.
- Χρησιμοποιείται για περήφλυτη υερμένου
- large ~ προτίμεται αυτό τις πολλές παραμέτρους που έχει το μοντέλο.
- Εκπαιδεύονται πάνω σε πλούσια κειμένων που προέρχονται από τον κόσμο, που πελάγουν διάφορα θέματα ή γλωσσικά γεγονότα. Με αποτέλεσμα να κατανοούν ή να παράγουν κείμενο σε διάφορες γλώσσες ή σε διαφορετικές ενότητες.

Πρέπει να ωριμάσει πρώτα τι θα κάνει

τα LLM's γνωστούνται από περιηγουμεί αρχιτεκτονικές.

Πλήρουν χρόνο ή κρίματα για να ευπαιδεύτονται κατά στα αυτό που τα θέλουμε.

Στα αυτό όμοιό είναι να χρησιμοποιούνται pre-trained models που παρέχονται από πολλούς οργανισμούς ανοικτού κώδικα.

Άλλο ότι από όταν γίνεται η ΧΥΓΙΑ τα

Pre-trained models οι οποίες δουλεύουν το iΟΙΟ.

Μοντέλα που έχουν ευπαιδευτεί επί πολλές ώρες σε τεράστιο όγκο δεδομένων υερμένων, από ανοικτές πηγές, όπως διαδίκτυο.

Τα μοντέλα αυτά πρέπει να προβαρμόνιστούν για να επιστρέψουν το επιευπτικό αποτελέσματα. (Προβαρμόνιστας είναι αναγρέσεις)

Κύρια διαφορά είναι η ευπαιδεύση που δεν χρειάζεται

API's Based LLMs

Υπάρχουν ανοικτού κώδικα οργανισμοί ή πλατφόρμες που παρέχουν αυτήν πρόσθιαν σε προετοιμασμένα μοντέλα μέσω διεπαφών προγραμματικούς εφαρμογών (API's). Εάν διαθέρουμε να χρησιμοποιήσουμε αυτά τα μοντέλα χωρίς να χρειάζεται να τα εκπαιδεύουμε, από την αρχή.

Παραδειγματα Model ΛLN's

GPT 2

BERT sequence classification

* BLOOM

ROBERTA

TS (text - text transfer transformer)

DistilBERT

Electra

XLNet

Ilua ?

Questions

- Πρέπει να διατίθεται ένα NLP model για επειρατώσεις
στον διάλογο που θα γενικώνει;
- Είναι εύκολο να που δημιουργήσεις για
την εγγύηση των αποτελεσμάτων;
- Data πρέπει να είναι της ίδιας πορείας
- Κοινωνία ου σε είναι για ωρες υπεργόραφο ένα με
διάρκεια ιωρεί ήτε τα κατανοητές μέσα για
τα αρχιπέλαγα.
- Χρησιμοποιούμε τα pattern που έχουν, γιόν, γενναντεί
στοιχεία που θα συνδέονται με την επειρατώση
- Υπολογιστή; Τι θα χρειαστεί;

Eίναι η διαδικασία.

Finetuning Large Language Models

What is finetuning?

fine tuning ειναι η προσέγγισης στο fine tuning προσβλησηών, με τα να επιταχύνεται αλληλεπίδραση μεταξύ αντικειμένων που διορθώνονται μεταξύ τους.

Είναι τύπος τεχνικής, που χρηγάπτοσθηκε σε διάφορα
εκπαιδευτέο ποντέλο ή το ειναιδεράδηκοια! δεσμεύει
κι αυτό συγχρόνως εφαπτίζει την ποντίνα.

Tl-x GPTA - Github Copilot / Primary care physician - Cardiologist

What does finetuning do for the model?

- Βοηθεί στο να αντέι το μοντέρνο τα δεδομένα, και να πλαιρεί πληροφορία χρήσιμη, στις ώστε να παράγει αποτίσεις ή προβλέψεις σχετικά με αυτά. Αυτή να αναγράφει τι αναπαράγει τα δεδομένα
 - Ελατερίνει το μοντέρνο σε πιο εύκολης αποτελέσματα
 - Προσαρμόζει το μοντέρνο σε μια ευξειρήμενη περιπτώση
 - Μειώνει τις γενευστίσεις

Prompt Engineering Finetuning Engineering

Ειναι τεχνιτοι Αναρρηπεται στην Σιδηρατοια 6xσιδηροι ④

ΤΗΝ ΜΟΝΤΕΔΙΑΝΗ ΔΙΑΓΩΡΦΩΣΗΝ ΤΟΥ ΣΤΑΔΙΟΥ ΦΡΑΓΜΑΤΟΣ ΤΟΥ

in epwings fine-tuning για να ΕΠΙΤΕΥΧΕΙ καλύτερη

Η ιδέα είναι αιδούση + επίτευξη συγκριτικών δρόσων

μουτέρο έργοι Διυλασθή αφορά το σχεδιασμό Θ την Σιαχεί-

outlawing is part of this dialectic's finetuning goal

Autor proprio

Evtodias ne Če

Kadugyanan na paulosan ang iba't katuwang

8 Η διαμόρφωση της ειδούσας βονέται για να
9 τις στηριζεται στην ισχύ της από τη λογική της

Finetuning Large Language Models

What is finetuning?

Το fine tuning είναι η προβεβλεπόμενη διαδικασία γιών, με το να ευπλαδεύεται σταθερά μεταξύ των αποδεσμένων μοντέλων. Η απόδοση πολύ μεγάλη είναι μια τεχνική, που χρηγάπτεται σε πολλούς επαγγελματικούς κλάδους. Το πρώτο μοντέλο ή το είναι διαδικασία διαδέχεται με την αντικατίσταντη εργασία στην πράξη.

TL;DR: GPT-3 - Github Copilot / Primary care physician - Cardiologist

What does finetuning do for the model?

- Βοηθεί στην ανάπτυξη της απόδοσης του μοντέλου
- Παρέχει απλοπίσεις ανταρρητικού πολλών προβλημάτων
- Μετατίνει την απόδοση σε επαγγελματικούς κλάδους
- Επεκτείνει την απόδοση σε νέα περιοχές
- Μειώνει την απόδοση σε προστατευόμενα περιοχές

Prompt Engineering

Είναι τεχνική που χρηματοποιείται για την απόδοση των μοντέλων, ανάλογα με το πώς αντέρχονται εντονές σε ερωτήσεις που τας δίνονται.

Η ιδέα είναι να διαφοροποιηθεί η είδοση που διετέλεσε στο μοντέλο έτσι ώστε το μοντέλο να παράγει επιευγές στα λαντίνες.

Αυτό μπορεί να κυριεύει λαμβάνοντας διαφορετικές εντονές με λέξεις - λατίνια + άλλες τεχνικές για την κατεύργανση των λαντίνων στην ανθρώπινη γλώσσα.

Η διαφορετική είδοση που διετέλεσε στο μοντέλο για να πετύχαινει

- τις επιευγές λαντίνες από το μοντέλο

(Draupl Eugineering)
Drauplting

Prompting

Είναι η δράση που
χρησιμοποιούμενη τις
κνήσεις (prompts) κατατίνει
διαδικασία εκτίθεσης
(prompt engineering),
για να δημιουργήσεις απόφε-
νεις από ένα πούτερο.

Finetuning

- Αγνοείται με την προσαρμοξή
του λογισμού για να είναι
τιο αποτελεσματικό σε
εγκερωτικά σεμεία

Positives

- Dev xperialefou data ja va ferivnders. (You can start chatting with the model)
 - Murop kōbros
 - Dev xperialefou iðicitepes TEXVILIES yvñðeis (attai pñws va biññers yinuñda)
 - Quñðeis ta ðeðofjera þef as avautungs (RAG^④ ret augmented generation) Eta pñopeis va ðiañdefers ti ðeðofjera xperialefou m ea nñave gto eñwtepiñd t

Positives

- Ηπορεις να αποθηκεύεται
μερίδια όχι σε διδύμενων
 - Ηπορεις να πάρει νέες πλανοδοτήσεις
από τα πλανή σε διάφευκα
 - Ηπορεις να τα πάρει διατα
διορθώσεις τα ήδην της ηπορεις να
εξαγοράσει
 - Ημιφόρερο κόστος αν σο
κοντέρα είναι μηρό
 - Ηπορεις να χρησιμοποιούνται
ιεραλ το RAG και εδώ, ωστε να μπορεις
να το βυθίσεις Θ) να μερισθούτερα.
κοντέρα αριθμα Θ) αν εξει πάστει
όδια αυτη την πλανοδοτηση.

Negatives

- Οι νομοί που χρησιμοποιούνται για την απόδειξη μονάδας διαχείρισης είναι:
 - Οι αριθμοί που παραπέμπουν στην ποσότητα της μονάδας.
 - Οι λέξεις που παραπέμπουν στην ποσότητα της μονάδας.
 - Το πιο σημαντικό στοιχείο που παραπέμπεται στην μονάδα είναι η μονάδα μέτρησης.
 - Το πιο σημαντικό στοιχείο που παραπέμπεται στην μονάδα είναι η μονάδα μέτρησης.
 - Το πιο σημαντικό στοιχείο που παραπέμπεται στην μονάδα είναι η μονάδα μέτρησης.

- όπισθις θίασοις τα λάθη του υπορει για
ταχία ° Μηδότερο κόστος αν συ-
μψυκτική ισοτιμία είναι μηδέ
ούτε ποτέ ° Η πορεία να χρησιμοποιηθεί
για την επιστροφή των λαθών, όπου να πορεια-
στην) να το ευνόεις Θ ήτε περισσότερο.
Δείξεις Βούτηλα αρότρα Θ αν εξει πάει
και που ολόκληρη αυτή την παραδοσία.

μέρος της Negatives

- ° Χρησιμότερη περιβόλτηρα διδαχένα,
που να είναι high quality
προϊόντα ° Εξει τα αρχικά κόστοι
απα. ° Χρησιμά τεχνικές γνώσεις
προτιμώτες Ιδιαίτερα για τα δεδομένα.

ie1

1. Prompt engineering

Παράγει ευτολής καταλληλες
Θ αποτελεσματικές

ie2

2. Prompting

Χρησιμοποιει τις ευτολής
για να πάρει απαντήσεις

Benefits of finetuning your own LLM

Performance

Ως καπαζίνει το λογότυπο σου να γίνεται δικό του οξεία

Απούρια περισσότερη εμπειρία στον τομέα του σεβετες να το ελιμεινήσεται

Είναι πιο δυνητικός. (Κάποιες δομές τα ποντέα δεν είναι δυνητικές. Τι μια πέρα δίνουν τέλειο output την άλλη όχι)

Μπορείς να λύνεις το ποντέα σου πιο περιπολατής, ανεξάρτητα από την αρχική

Privacy

Μπορείς να προστατεύεις τα δεδομένα σου γιατί γίνεται δεν διύλλει την VPC ή on-premises. No data breaches

Cost

Μπορείς να διαχειρίζεσαι το κόστος, με διαδίκτυο

Reliability

Μπορεί να διαχειρίζεσαι το χρόνο αντανακλήσης του ποντέα

Μπορείς να διαχειρίζεσαι το ποντέα σου όπως θελεις. Να λύνεις συγκεκριμένες αιτησίες

Python libraries for finetune

- Pytorch (καμποτέρο επιτέλους διαθέσιμον)
- Huggingface
- Llava library (train models with 3 liner code)

Where does finetuning fit in?

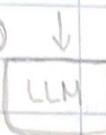
Pretraining: Ήπιως σημα πριν το finetuning



- ② Ηπιως έχει τυχαιό πουλέσθηκε στην αρχή που
- δεν είναι ευνοϊκότερο.

Once upon a
midnight...

- Οδα τα weights είναι random
- Δεν έχει καμιά language skills
- Ο μαθησιακός γρόξος που έχει είναι η ετοιμεύνη πρόβλεψης



- ① Ετοιμεύεται να predict το αρχικό γράμμα είναι sd!@ (λέξος να προβλέψει την γένη μετά το Once, από το upon) μαρκάρισε από αυτό ολό το δικτυο. Συνάντηση τη δεδομένη τα γένη unlabeled επειδή δεν είναι κάτι που έχει το ίδιον, αλλαί αναπαγεί από το internet.

- Το πουλέσθηκε από είναι supervised learning καθώς εκπαιδεύεται το κατό του με την ετοιμεύνη πρόβλεψης. Αλλιώς δεν θα έχει labels
- Μετά το training μπορεί να προβλέψει την γένη upon τα και να μάθει γάνωσα, είδος γύρων

Data scraped from internet

Refers to information that has been collected or extracted from world wide web.

Scraping is a technique where automated tools or scripts are used to extract data from websites. Scrap from internet is a vast and complex task.

!!! in?

weights: Παραμέτρους ενός μοντέλου, εδινότε-

ρα το πλασίο των νευρωνικών σιτών. Το μοντέλο τα χρησιμοποιεί για να πραγματοποιεί προβλέψεις ή αποφάσεις

ja τυχαιο μοντέλο * see more at page

write
it from
class opt

ευπαιδευμένο.

weights είναι random

αιόμα, language skills

akόs, erόxos, που έχει είναι αβούρενα

to predict τι αρχικό στάδιο είναι ιόχος να προβλέψει την γένη μετά από το upon) μακριά από αυτό

θιομένως παίρνει δεδομένα από

2 έννοια τα δεδομένα τα οποία είναι unlabeled
πι πάτη που έχει το σύνον, αλλά co

convised learning ratios επαιδεύτη

Dataset Pile from EleutherAI

Είναι μεγάλο δίκτυο σε δεδομένων χιλιάδων της επαγγελματικής
τεχνητής νευρωνικής δικτύων που αναπτύχθηκε
από την EleutherAI.

Αποτελείται από 825 εβ διάφορων τύπων κεφένων, και
περιλαμβάνει πληροφορίες από πολλές διαφορετικές
πηγές. Το pile δημιουργήθηκε συνδιάσοντας οι
διάφορα, υγιάς ποιότητας δίκτυα δεδομένων, που είναι
προστιά των ανοικτού κώδικα.

Διατί: Επιταχεινούν μοντέλων για την επεξεργασία
φυσικής γλώσσας όπως τα νευρωνικά δίκτυα.
Μπορεί να επιταχεινεί μοντέλα που αποτινούν
πολυμορφία στα δεδομένα τους και μπορεί
να βοηθήσει στην ανάπτυξη προηγμένων
λύσεων στο τομέα της επεξεργασίας φύσικης
γλώσσας.

Αυτό το είδος μεγάλων δεδομένων είναι σημαντικό χιλιάδων
την επίτευξη υγιάς αρκιβεΐδας & επιδόσης στην
επιπλόσια τεχνητής νευρωνικής δικτύων.

is in jsonlines data

Limitations of pre-trained base models

Επομένως pre-trained models από το διαδικτυού
όπως το Pile, είναι χρήσιμα επειδή έχουν μέχριο
όχρη δεδομένων αλλά οι απαντήσεις που δουν
επιβεβαιώνονται δεν θα μπορούν να επιβεβαιωθούν
που τοις λειτιδικές αλλαγές που θα γίνονται, οχι αυτό μας λειτιδικές.

Input
↓
LLM
↓
Output
→
Π.χ. Η ιδέα είναι να προτείνουμε τις λογικές,
ανθρώπινες ή ανθρώπινες λογικές λογικές,
πληρεσμένες με πληροφορίες για το LLM, σαν
να μπορεί να λειτουργεί σα chatbot interface;

Finetuning after pretraining

To finetuning σε βασικές ψηφιακές πλατφόρμες ή σε pre-trained model, για μοντέλο chatbot interface, και μέσω της αλλαγής προσθήσεων αυτής να εκπαιδεύεται περισσότερο το μοντέλο καί το σκοπό του το ίσιο.

Επομένως: - Pretraining (1^ο βήμα): Για να πάρεις το βασικό δαχτυλίδι μοντέλο

- Add more data
- Finetuning : Για να πάρεις fine-tuned model
- Μπορείς μετά να προστέθεις finetuning steps

Από:

To Finetuning χρησιμοποιείται για νετέλπιστα εκπαίδευση

Why finetuning refers to training further?

- Μπορείς να πάρεις τους ίδιους τύπους δεδομένων ή να συγχεινείς δεδομένα από διάφορες πηγές ή να τα ευθύνεστε (self-supervised unlabeled data)
- Μπορείς να δομίσεις τα δεδομένα και πάνω σαν "labeled"
- Επειδή σαν δικτυακά δεδομένα
- Tool in your toolbox

Actually you take a basic pretrained model with much knowledge and basic language skills and you taking it to the next level.

Finetuning for generative tasks / διαπίπτεις εργασίες παραγωγής

It is not well-defined! Why?

Because: - Ο αριθμός του finetuning είναι δικό της αριθμός πλειόντων πραγματοποιούντων ενημερώσεις στις παραγετέρους (weights) οδουδίπου του μοντέλου, οχι μέρος του

- Εκτινάζει τον ίδιο στόχο εκπαίδευσης (την επίλευγη λέξη) και το πάνω που μπορεί να γίνεται είναι να αλλάξουν τα δεδομένα να είναι πιο διατυπωμένα.

• Περισσότεροι εφαγμένα τρόποι να περιορίσεται
μουτέδα.

What is finetuning doing for you?

1. Behavior change

- Αναδειχθεί την συμπεριφορά του μουτέδα, το κάνει
chatbot. Το αποτέλεσμα σήμερι το μουτέδα να
απαντά με βιντέο.

- Το μουτέδα μπορεί να επιτεντώσει

- Μπορεί να περαπλανηθεί μαζί του για να δεις
που είναι καλύτερα, χωρίς να κάνεις prompt
engineering για να εξετάσεις που σου δίνει
τις καλύτερες απαντήσεις

2. Gain knowledge

- Αναπτύσσει την γνώση του για συγκεκριμένα είδα.
Του δεν είναι στο pretrained model

- Μπορεί επίσης να διορθώσει incorrect γνώσεις, εάν
μπορεί με την σύλληξη να το ενημέρωσεις
και να το διορθώσεις

3. Both

Αναδειχθεί την συμπεριφορά + το εμπλοκής
με γνώσεις

Tasks to finetuning

Τα finetuning μουτέδα χρησιμοποιούνται για να
δινουν + να μαρτυρούν text από τα LLM's

Τα task των finetuning μπορούν να τοιχωθούν σε
2 κατηγορίες:

reading
key
key

← • Μπορείτε να παρέστε κάποιο κείμενο ως εισάδο +
να παράγετε λιγότερο κείμενο ως αποτέλεσμα.

Π.χ. μπορείς να ελέγξεις γεγενικά κείμενα από ένα
κείμενο ή να αποφασίσεις πώς να υπενθύμισεται
μια συνομιλία προς ένα API.

Ότε αυτή την περίπτωση δινες κείμενο ως εισόδο Φ
λαμβάνεις περισσότερο κείμενο ως αποτέλεσμα. Για
παραδειγματικά μπορεις να χρησιμοποιεις το ποντέλο
για να δινούμεις, να γραφεις emails ή κώδικα.
Το πιο βιβραντικό είναι ειναι να λέπεις το αποτέλεσμα
σου. Να λέπεις πως μπορει ενα καλό αποτέλεσμα,
πως μπορει ενα κακό αποτέλεσμα και να λέπεις
τι να κάνετε για να βελτιώσετε την απόδοση του
μοντέλου σας. ! Task clarity is the key

Steps for your first time finetuning

1. Identify task(s) by prompt-engineering a large LLM

Πλαστε με ενα LLM όπως το chat Gpt. Δινοριανοείς
να δινεις του ραν δρόμος κάποια αδιάνευτη task, κατα
που δεν ειναι τέλειο, αλλα μπορει να τελειοποιηθει.

2. Finds task that you see an LLM doing ~OK at

3. Choose one task

4. Get ~1000 inputs and outputs for the task

(~better than the OK from the LLM)

Δώσε text σαν input και πλαστε text σαν output.

Πρέπει να έχει ~1000 σειράρια απο αυτά για
ονοια να λέπεις το αποτέλεσμα που θα δημιουργησες.

Τα 1000 pairs of data ειναι ενας καλος αριθμος
για να μπορει το μοντέλο να μάθει να αντιστοιχει
την εισόδο Φ ελόδο.

Τα σειράρια αυτά μπορει να ειναι καλύτερα απο
αυτά που ξεκινησαν το LLM πριν.

5. Finetune a small LLM on this data

Μετα μπορεις να πεισματισεις ενα μικρο χωροστικο
μοντέλο να βάσην αυτά τα δεδομένα απλώς για
να απορριψεις μια ειδονα της βελτιωσης της
απόδοσης.

Steps to prepare your data

1. Collect instruction-response pairs

Διαχειρίσιμες οδη τα δεδομένα που χρειάζεται, σε μορφή λεξαριών ερώτηση-απάντησης.

2. Concatenate pairs (add prompt template, if applicable)

Στην συνέχεια πρέπει να συνδιάσει τα data, με βάση κάποιο template

3. Tokenize: Pad, Truncate

Tokenize είναι η διαδικασία διάτασης λειμένων σε μικρότερες μονάδες που αναγνωρίζονται tokens. Στην επεξεργασία διαχειρίζονται για παραδειγματικός λόγων, τα token μπορεί να είναι λέξεις, γράμματα, αριθμοί, σύμβολα κ.λπ. Η διαδικασία διατάσσει στην προετοιμασία του λειμένου για περισσότερων ανάλυσης από αλγόριθμους ή μοντέλα μηχανικής μάθησης.

4. Split into train/test

Padding:

Χωρίζονται τα δεδομένα σε train & test

Padding ?? ~what is it
check AutoTokenizer from transformers

Tokenizing

Fine Tuning is fun for all! ↴ Truncation:

[34389 13932 378 318 1257 329 477 0] ↴ decoding
Fine Tuning is fun for all!

Η διαδικασία αυτή παίρνει το λειμένο και το μετατρέπει σε αριθμούς. Κάθε κομμάτι λειμένου αυτοματοποιείται είναι αριθμό.

Βασίζεται στην συχνιτικά εμφάνισης & όχι απαραίτητη στην εύρους λέξης.

Πολύ συνήθες είναι να εμφανίζουν την κατάλληλη padding για την λεξική του λειμένου.

Κάθε μοντέλο ευποιεύεται με ένα συχνευρίσκοντας tokenizer που ταυτίζει με την σούπερ & την χάρηση του λειμένου.

Αν χρησιμοποιούσουμε σιδηροτελές tokenizer το model θα μπορεύει και σεν θα λειτουργήσει καλά.

① Data Preparation

What kind of data we need?

Χριστούντες data:

• High Quality: Αν τώστες lower quality data, εσείς πάπεις και lower quality output.

• Diversity: Πρέπει να έχεις σιδηρότερα data για να υλοποιήσεις πολλές αντικείμενα μεσανθρώπους είναι χρήσιμο. Αν τα input και output είναι τα ίδια το ποντέλο τα απομακρύνεται και δεν θα έχει συνέχεια τα ίδια και τα ίδια.

• Real / generated: Μπορούμε να δημιουργήσουμε ναι δική μας data, οράσα αν χρησιμό. Ποιούμε generate data υπάρχουν συχνευτικά μοτίβα που είναι αποτέλεσμα της φύσης ή χρήσης για ελάσσον.

• More data: Είναι σημαντικό να έχεις πολλά data, βέβαια τε ποντέλο που έχουν pre-trained με data από το σιδηρότερο. Το να έχεις πολλά data δεν είναι η τόσο σημαντικό γιατί το ποντέλο δεν φεύγει από το μπόεν.

8.

Steps to prepare your data

1. Collected instruction-response pairs

Συγκέντρωσε όλα τα δεδομένα που χρειάζεται, γε
μαρτίν Σεκαρίου ερώτο-απαντήσεων.

2. Concatenate pairs (add prompt template, if applicable)

Στην ουνέξεια πρέπει να ουντάσθε τα data, όπως βλέπεται στο template.

3. Tokenize: Pad, Truncate

Tokenize είναι η διαδικασία διάτυπου κειμένου σε μικρότερες μονάδες που αναφέρονται tokens. Στην επεξεργασία δυοινιών γαϊδουρών, τα token μπορεί να είναι λέξεις, γρήγορια, σύνθετα γρήγορια ή άλλες χραστικές μονάδες. Η διαδικασία tokenization στην προετοιμασία του κειμένου για περισσότερων ανάλυση από αλγόριθμους θα ποντέζει μηχανική ονόματα.

A. Split into train/test

Padding:

Χωρίσου τα δεδομένα σε train + test

Padding ?? ~what is it

Padding : with N/A
check Autotokenizer from transformers

Tokeuiting

Fine Tuning is fun for all! Truncation:

[34389 13929 248 318 1251 329 477 0] *decoding*

Five Tuning is fun for all! - decoding

Η διαδικασία αυτή πλαισίει το γείφενο και το μετατρέπει σε αριθμούς. Καθε κομμάτι γείφενου αντιπροσωπεύει έναν αριθμό.

Βασιζεται στην συχνωτικη εμφανισης η οχι απαραιτητη
ητην ειγραση τηλε.

Πολιύ συνήθες είναι να εργάζονται την κοπιάκην -ing σαν γεωργικό τόκον.

Καθε ποτέδη ευμάρτυρας ήταν οικείος τοποθεσία
που ταπιάζει ότι τις άστρους ή τις γλώσσας του ποτέδου

Αν χρησιμοποιούμε διαφορετικό τοποτείχο το model θα μπορεύει και σεν θα λειτουργήσει όμως τα.

② Training Process

Same as other neural networks.

- Τρίτα βάσεις σαν input τα δεδομένα εκπαίδευσης
- Υπολογίζεις την απώλεια σε σύγκριση με τις πραγματικές απάντησης που έχει υπάρξει (pawn)
- Ενημερώνεις τα weight
- Back prompt για να γίνεται update για να το βελτιωθεί, εάν ώστε το μοντέλο να μάθει + να βγάλει σαν αποτέλεσμα κάτια κοντά στο pawn να

Υπερπαραγέτροι που συμμετέχουν στην εκπαίδευση των LLM, μερικές είναι:

- Ταχύτητα μάθησης (πως ενημερώνονται τα weights)
- Προγραμματιστής μάθησης (πως αλλάζει η ταχύτητα)
- Υπερπαραγέτροι βελτιστοποίησης (μέσοι ελαστικότητας)

Οι υπερπαραγέτροι αριθμούν αυτό την έναρξη της διαδικασίας της εκπαίδευσης του μοντέλου.

Ο αντιός γνωρίζει πώς μπορει να διαφοροποιήσει συγκατα το αποτέλεσμα της εκπαίδευσης.

```
chunk      Run through general chunks of training process in code
μέρος     chunk of training code in PyTorch
νομίσει   for epoch in range(num_epochs): (1)
δεξορεύω
οπτικό
κεντρικό
έχει
δεσμένη
    for batch in train_dataloader: (2)
        outputs = models(**batch) (3)
        loss = outputs.loss (4)
        loss.backward() (5)
        optimizer.step() (6)
```

(1) Μόριες φορές το μοντέλο θα περάσει από το στάδιο της εκπαίδευσης

Η διεργασία για κάθε μία epoch απαιτεί την επανάληψη της εκπαίδευσης με τα ίδια data πολλές φορές. Αυτό βοηθάει στην βελτιστοποίηση των weights του να προβλεπει καλύτερα το train

(2) Προτίναυρε τα σεδόφινα σε batches. Η την όποια batches ενσωματίζεται δεδομένων (tokenizing). Μετά την διάρκεια της επεξεργασίας το μοντέλο επιταχύνεται σε νέα batch λεχωριστικά. Αυτό επιτρέπει στο μοντέλο να επιταχύνεται σε batches με μηροτερη πολυπλοκότητα ή επιτρέπει την διεξαγωγή μεγαλύτερου αριθμού epochs, χωρίς να αυξανεται διμερώς ο απαιτούμενος χρόνος επεξεργασίας.

(3) Περινόρει το batch μέσα από το model για να πλήρωνε αποτελεσματικά

(4) Υπολογίζουμε την απώλεια του model. Έποχας είναι να μειωθει η απώλεια. Ωδαδόν να μπορεί να προβλέψουμε τα σεδόφινα όποιο το διαναύσματος θέλουμε στην πραγματικότητα.

(5) Η διαδικασία προς τα πίσω είναι η διαδικασία όπου τα gradients της απώλειας, ενημερώνουν τα weights ώστε να προσπαθήσουν να μειωθεί την απώλεια. Αυχεντικό μένα χρησιμοποιεί την προπαραγωγή για να υπολογίσει τις αλλαγές που απαιτούνται για την ↓ απώλεια.

(6) Update του optimiser, ο οποίος είναι υπεύθυνος για την ενημέρωση των weights ή βάση των gradients που υπολογίζεται στο πίσω φίρμα.

Lawini library train model
Linux in 3 lines
and can run on
open source model
This code is hosted on
an external GPU.

GPU / PyTorch
or
CPU

③ Evaluation and iteration

Δεν μπορεί να έχει παθητές μετρήσεις σε αυτά τα μοντέλα, και η αποτελεσματικότητα του εφεγγέρου πέρα με την πέρα, με αποτελεσματικές οι μετρήσεις αυτές να γίνουν μπορούν να συμβαίσουν.

• Είναι καθός τρόπος να αφιοδοξήσεις το μοντέλο είναι να χρησιμοποιήσεις των ανθρώπινων αφιοδόγνων, να δρεις ανθρώπινης είδησης πάνω στην εφιδικευμένη μοντέλου και να αφιοδοξήσουν τα αποτελέσματα.

• Επίσης σιασφαλίζεις την ποιοτότητα των σκέψεων να είναι υγιείς, αυριβίς, χειρισμένο για να καλύπτει διάφορες περιπτώσεις ελέγχου ή φυσικά να μην έχει δει τα training data

• Σύγχρονη ELO μοίραση σχεδίου με ένα A-B test μεταξύ πολλαπλών μοντέλων. Οι βαθμολογίες είναι χρησιμοποιούνται στο διάτη ή αυτός ο τρόπος βοηθεί να καταγάρωσε το ποια μοντέλα αποδίδουν καλά ή όχι.

LM Benchmarks: Suite of Evaluation Methods

Για την αφιοδόγνωση των μοντέλων LM χρησιμοποιούνται ένα σύνολο αφιοδόγνωσης για να δοκιμάσουν την απόδοση των μοντέλων δύποτα από αυτά είναι:

ABC: Αφιοδοξεί τις μανόντες ενός μοντέλου να απαντά σε ερωτήσεις που απαιτούν λογική συγένη ή ματρικόντων βασικών ευπλανητικών οφιδών

HellaSwag: Λογικάδουν την τοινή λογική ή την μανότητα ενός LM να προβλέπει λογικές αριθμητικές συμβάσεις

NMLU: Αφιοδοξεί την χωνέτην ενός LM GE γεγαρά όπως γραμματική, μαθηματικές ελισσώσεις χειρόγραφία κλπ.

Benchmarks: Συστήματα που διαδίδουν την χρήση μονοποιείτου για την ασφαλή λειτουργία την απόδειξη ενός
ευενθύμιατος

TruthfulQA: Μετρά τις δυνατότητες του LLM να αντιταράξει γενδή σεσοφένα που μπορούν να ερεθίσουν συνήθως στο διαδίκτυο χρησιμοποιούται πληρώς για να δει αν το μοντέλο διαπίνει την αλήθευτη από το γέρνα.

Τα παραπάνω μοντέλα είναι χρήσιμα για την σύγκριση και την βελτίωση της απόδοσης των μοντέλων.

Error Analysis

- Ηια σάλλη μεθόδος αισιοδόχησης είναι η Error Analysis
- Αυτή η μέθοδος καταχρεωποίει error's για να ανιανθείσι ποιός τύπος σεσοφένων είναι πιο σύνηθες, ώστε να μπορέσεις να τα μείνεις.
- Πολλές φορές αυτή η ανάλυση προϋποθέτει να έχει επιταχύνει το μοντέλο σου, άλλο εφόσον για fine-tuning models παν ήτη το base model τους έχει ήδη επιταχύνει από πριν, μπορεις να εφαρμόσεις την error analysis πριν fine-tune το μοντέλο.
- Αυτό θε βοηθεί να ανιανθείς τα χαρακτηριστικά του basemodel ώστε να λέπεις τι data περιέχει & να μπορέσεις να έχεις μεγαλύτερη απόδοση στο finetuning.
- Μερικά errors έχουν να ισχουν όπως:
 - Misspellings
 - Too long output
 - Repetitive output's

Consideration on getting started now

Practical approach to finetuning

- 1 Figure out the task
- 2 Collect data related to your task's inputs/outputs
- 3 Generate data if you don't have enough data

Av dev exis aplwta data, ta duvouzhoi

Diua oas i xpoliopoiise prompt template

gla va spidais iepiuia.

- 4 Finetune a small model ($100 \approx$ parameter model)

400 gla va aploptis tnv aibangn tns
aridogns tou poutelou

- 5 Vary the amount of data you give the model

Diafporoiointe zo nlobi twv deobekov tlv diwe
6to pouteho gla va afiologhies tlv ofei to
model 600, avdiyga fe to nroga data to empelou

- 6 Evaluate your LLM to know what's going well vs not

- 7 Collect more data to improve

Tia va avlinber tnv aridogn piew tou evaluation

- 8 Increase task complexity

Tia va to uavers tio duvoulo

- 9 Increase model size for performance

Tia va ↑ tnv aridogn be tio duvoulo task's

Writing tasks are a lot harder than the reading task's

More expansive tasks are chatting, emails, code
and this because they are more tokens that
are produced by the model. This hard task need
larger models to handle them.

Another way to have harder tasks is to have a
combination of task. Ask the model to do a lot
of things instead of one.

In order to create an application where you can chat with your data, you first have to load your data into a format you can work with (LangChain document loaders)

Document Loaders

- Αρχικούνται με τα νέα δώσουν προτύπων ④ να μετασχηματίσουν δεδομένα, από διάφορες πηγές ④ πηγές γενικά συγκεντριμένο τύπο.
- Μπορούμε να φορτώσουμε δεδομένα από διάφορες βάσεις δεδομένων ή το internet (π.χ. YouTube).
- Τα δεδομένα μπορούν να είναι διάφορους τύπους, όπως PDF's, HTML, JSON.
- Επομένως τα document loaders, μαρτυρούν δεδομένα από διάφορα data sources και να τα ανεβάσουν γενικά συγκεντριμένο τύπο δεδομένων, το οποίο θα αποτελείται από το μερικόνευρο και τα διεταρά μεταδεδομένα.
- Η εφαρμογή μπορεί να φορτώσει unstructured data, όπως κείμενα, από public data sources, όπως YouTube, wikipedia, twitter κ.λπ. Η από proprietary data sources (personal / company data όπως messenger, git hub, powerpoint, jupyter κ.λπ.
- Μπορεί επίσης να φορτώσει structured data, data που είναι σαν πίνακες, που μπορεί να είχουν δείγματα σα κελιά γραμμών ή στηλών, που μπορείς να έχεις να ρυθμίζεις ④ να αυτονομείς. Τα δεδομένα αυτά μπορείς να έχεις να τα ρίξεις από public sources, όπως hugging faces ή από proprietary sources, όπως excel, pandas κ.λπ.