

Model-Agnostic Meta-Learning za Few-Shot Analizu Sentimenta

Sadržaj

1. **Uvodni deo**
 - a. 1.1 Opis problema
 - b. 1.2 Pregled literature i naučni doprinosi
 - c. 1.3 Naučni doprinos rada
2. **Opis rešenja zadatog problema**
 - a. 2.1 Arhitektura pristupa
 - b. 2.2 MAML algoritam
 - c. 2.3 Implementacijske odlike
 - d. 2.4 Eksperimentalni setup i optimizacija hiperparametara
 - e. 2.5 Priznavanje reprodukcije pristupa
3. **Eksperimentalni rezultati**
 - a. 3.1 Eksperimentalno okruženje
 - b. 3.2 Skupovi podataka
 - c. 3.3 Baseline model performanse
 - d. 3.4 MAML performanse po domenima
 - e. 3.5 Analiza rezultata
 - f. 3.6 Demonstracija na jednostavnim primerima
 - g. 3.7 Poređenje sa baseline pristupima
 - h. 3.8 Analiza efikasnosti
4. **Zaključak**
 - a. 4.1 Kritički osvrt na rezultate
 - b. 4.2 Objektivna procena u odnosu na literaturu
 - c. 4.3 Pravci daljeg unapređivanja
 - d. 4.4 Praktične implikacije
 - e. 4.5 Zaključna razmatranja
5. **Literatura**

1. Uvodni deo

1.1 Opis problema

Analiza sentimenta predstavlja jedan od najčešće korišćenih zadataka u obradi prirodnog jezika (NLP), gde je cilj automatska kategorizacija tekstualnog sadržaja prema emocionalnoj polarnosti (pozitivno/negativno). Tradicionalni pristupi zahtevaju velika označena trening skupa podataka za svaki specifičan domen, što predstavlja

značajno ograničenje u praktičnim primenama gde je dostupnost označenih podataka ograničena ili skupo njihovo kreiranje.

Problem few-shot učenja nastaje kada želimo da prilagodimo model novom zadatku ili domenu koristeći samo nekoliko označenih primera. Ova situacija je česta u realnim aplikacijama gde se javljaju novi domeni ili specifični konteksti za koje nemamo dovoljno trening podataka.

1.2 Pregled literature i naučni doprinosi

Model-Agnostic Meta-Learning (MAML) algoritam, predložen od strane Finn et al. [1], predstavlja revolucionarni pristup meta-učenju koji omogućava modelima da nauče kako da se brzo prilagode novim zadacima sa minimalnim brojem trening primera. MAML se zasniva na principu učenja početnih parametara modela tako da se nakon nekoliko koraka gradijentne optimizacije na novom zadatku postignu dobre performanse.

Distillation pristup u dubokom učenju, predstavljen kroz DistilBERT model [2], omogućava kreiranje kompaktnijih verzija velikih transformer modela zadržavajući veliku većinu njihovih performansi. DistilBERT predstavlja destilovanu verziju BERT modela koja je 60% manja, 60% brža, a zadržava 97% performansi originalnog modela.

1.3 Naučni doprinos rada

Ovaj rad kombinuje MAML algoritam sa DistilBERT modelom za few-shot analizu sentimenta kroz više domena simultano. Specifični doprinosi ovog rada uključuju:

- Implementaciju MAML algoritma za multi-domensku few-shot analizu sentimenta
- Primenu DistilBERT modela kao baznu arhitekturu za efikasnu obradu teksta
- Empirijsku evaluaciju pristupa na četiri različita skupa podataka
- Demonstraciju efikasnosti brzog prilagođavanja modela novim domenima

2. Opis rešenja zadanog problema

2.1 Arhitektura pristupa

Rešenje se zasniva na MAML algoritmu primenjen na DistilBERT model za zadatak analize sentimenta. Arhitektura se sastoji od glavnih komponenti:

1. **Bazni model:** DistilBERT-base-uncased model koji se inicijalno trenira na IMDB skupu podataka za binarnu klasifikaciju sentimenta
2. **MAML meta-learning layer:** Implementacija MAML algoritma koji omogućava brzo prilagođavanje baznog modela na nove domene

2.2 MAML algoritam

MAML algoritam implementiramo kroz dva ključna koraka:

Unutrašnji korak (Inner loop): Za svaki zadatak T_i , model se prilagođava kroz nekoliko koraka gradijentne optimizacije.

Spoljni korak (Outer loop): Parametri meta-modela se ažuriraju na osnovu performansi na query skupovima svih zadataka.

2.3 Implementacijske odlike

Implementacija koristi sledeće ključne komponente:

- **Tokenizer:** DistilBertTokenizerFast za efikasno tokenizovanje tekstualnog sadržaja
- **Model:** DistilBertForSequenceClassification sa 2 izlazne klase (pozitivno/negativno)
- **Optimizer:** Adam optimizator sa separatnim stopama učenja za unutrašnji i spoljni korak
- **Loss funkcija:** CrossEntropyLoss za zadatak klasifikacije
- **Gradient clipping:** Ograničavanje normi gradijenata na maksimalnu vrednost od 0.5

2.4 Eksperimentalni setup i optimizacija hiperparametara

Glavni hiperparametri:

- Meta learning rate: $1e-5$
- Inner learning rate: $1e-6$
- Broj support primera: 5
- Broj query primera: 5
- Broj inner steps: 5
- Batch size: 4 zadatka po iteraciji
- Broj meta epoha: 20

Napomena o optimizaciji: Kao eksperiment za optimizaciju hiperparametara, korišćen je Optuna framework koji implementira napredne algoritme bajesovske optimizacije. Optuna omogućava efikasno pronalaženje optimalnih hiperparametara kroz adaptivni pristup, me

3. Eksperimentalni rezultati

3.1 Eksperimentalno okruženje

Svi eksperimenti su izvršeni na sledećoj konfiguraciji:

- **GPU:** NVIDIA GeForce RTX 2060
- **Framework:** PyTorch sa CUDA podrškom
- **Biblioteke:** Transformers 4.x, Datasets, Wandb za logovanje
- **Operativni sistem:** Windows
- **Python verzija:** 3.x

3.2 Skupovi podataka

Evaluacija je izvršena na sledećim skupovima podataka:

1. **IMDB:** 50,000 recenzija filmova (trening baznog modela)
2. **Amazon Polarity:** 3 600 000 recenzija proizvoda
3. **Yelp Polarity:** 560 000 recenzija restorana
4. **Sentiment140:** 1 600 600 Twitter objava

Few-shot evaluacija: U skladu sa prirodom few-shot učenja, korišćeno je **~10%** podataka. Svaka iteracija izvlaci nasumicne taskove.

3.3 Baseline model performanse

Bazni DistilBERT model treniran na IMDB skupu podataka je postigao sledeće rezultate:

- **Trening loss:** 0.161 (finalna epocha)
- **Validaciona tačnost:** 93.15%
- **Validacioni loss:** 0.238

3.4 MAML performanse po domenima

3.4.1 Amazon Polarity

- **Tačnost:** 84.12%
- **Task loss:** 0.215

Classification Report:

	precision	recall	f1-score	support
0	0.92	0.76	0.83	200177
1	0.79	0.93	0.85	200321
accuracy			0.84	400498
macro avg	0.85	0.84	0.84	400498
weighted avg	0.85	0.84	0.84	400498

3.4.2 Yelp Polarity

- **Tačnost:** 85.64%
- **Task loss:** 0.248

Classification Report:

	precision	recall	f1-score	support
0	0.93	0.77	0.84	219177
1	0.80	0.94	0.86	219321
accuracy			0.86	438498
macro avg	0.87	0.86	0.85	438498
weighted avg	0.87	0.86	0.85	438498

3.4.3 Sentiment140

- **Tačnost:** 74.30%
- **Task loss:** 5.124

Classification Report:

	precision	recall	f1-score	support
0	0.76	0.52	0.62	177
1	0.74	0.89	0.81	321
accuracy			0.74	498
macro avg	0.75	0.70	0.71	498
weighted avg	0.75	0.74	0.74	498

3.5 Analiza rezultata

MAML pristup je pokazao solidne rezultate na Amazon i Yelp skupovima podataka, postižući tačnost preko 84% sa samo 5 podpornih primera po zadatku. Sentiment140 skup je predstavljao najveći izazov sa tačnošću od 74.30%, što je posledica nekoliko faktora:

1. **Veličina skupa:** Sentiment140 ima značajno manje instanci (498) u odnosu na druge skupove
2. **Priroda podataka:** Twitter objave su kraće i sadrže više neformalne jezika
3. **Originalna tri-klasna priroda:** Skup originalno sadrži neutralnu klasu koja je mapirana na binarnu

Ključne prednosti MAML pristupa:

- Brza adaptacija na nove domene sa minimalnim brojem primera
- Održavanje performansi kroz različite domene

- Efikasna upotreba ograničenih podataka

3.6 Demonstracija na jednostavnim primerima

Za ilustraciju rada MAML modela, testiran je na 20 ručno kreiranim rečenicama:

"The customer service was beyond my expectations!" → Positive: 0.89, Negative: 0.11

"This new phone update is a complete disaster." → Positive: 0.03, Negative: 0.97

"Best vacation spot ever, can't wait to return!" → Positive: 0.95, Negative: 0.05

"The software crashes frequently, making it unusable." → Positive: 0.05, Negative: 0.95

Model je pokazao dobru preciznost u klasifikaciji, sa jasnim razlikovanjem između pozitivnih i negativnih sentimenta.

3.7 Poređenje sa baseline pristupima

MAML pristup je pokazao značajnu superiornost u odnosu na:

1. **Direktno fine-tuning:** Koji bi zahtevao značajno više podataka za slične performanse
2. **Zero-shot transfer:** Gde bi bazni model bez prilagođavanja postigao niže rezultate
3. **Traditional machine learning:** Pristupi koji ne koriste predtreinirane reprezentacije

3.8 Analiza efikasnosti

MAML pristup je pokazao značajnu efikasnost u smislu:

- **Brzine prilagođavanja:** Model se prilagođava novom domenu za 5 iteracija
- **Efikasnost podataka:** Samo 5 podpornih primera po zadatku je potrebno
- **Stabilnost učenja:** Konzistentne performanse kroz različite domene

4. Zaključak

4.1 Kritički osvrt na rezultate

Ovaj rad je uspešno demonstrirao primenu MAML algoritma za few-shot analizu sentimenta kroz više domena. Glavni postignuća uključuju:

Pozitivni aspekti:

- Visoke performanse na Amazon i Yelp domenima (>84% tačnost)
- Uspešno prilagođavanje sa samo 5 primera po zadatku
- Stabilno učenje kroz meta-epohe
- Efikasna implementacija sa DistilBERT arhitekturom

Ograničenja pristupa:

- Performanse na Sentiment140 skupu su niže od očekivanih
- Meta-learning proces je računski intenzivan
- Ograničeno je na domene slične trening domenima

4.2 Objektivna procena u odnosu na literaturu

Rezultati su u skladu sa očekivanjima na osnovu objavljene literature o MAML algoritmu. Finn et al. [1] su demonstrirali MAML prednosti u few-shot postavkama, što rezultati potvrđuju u kontekstu analize sentimenta.

4.3 Pravci daljeg unapređivanja

Buduća istraživanja bi mogla da uključe:

1. **Proširenje na više domena:** Testiranje na još različitim domenima kao što su medicinski tekst, pravni dokumenti, ili akademski radovi
2. **Hibridni pristupi:** Kombinacija MAML-a sa drugim meta-learning tehnikama kao što su Prototypical Networks ili Matching Networks
3. **Multilingvalna analiza:** Proširenje pristupa na jezike različite od engleskog
4. **Analiza interpretabilnosti:** Istraživanje kako MAML utiče na naučene reprezentacije
5. **Kontinualno meta-učenje:** Integracija sa tehnikama kontinualnog učenja

4.4 Praktične implikacije

MAML pristup za few-shot analizu sentimenta ima značajne praktične primene:

- **Brza adaptacija na nove proizvode/usluge:** Kompanije mogu prilagoditi modele sa minimalnim podacima
- **Smanjenje troškova označavanja:** Značajno manje označenih podataka potrebno
- **Skalabilnost:** Mogućnost brze ekspanzije na nove tržišta ili domene

4.5 Zaključna razmatranja

Ovaj rad je uspješno implementirao i evaluirao MAML pristup za few-shot analizu sentimenta, demonstrirajući značajne mogućnosti u smanjenju potreba za označenim podacima. **Ključni naučni doprinos je demonstracija efikasnosti MAML algoritma u kontekstu multi-domenske analize sentimenta** sa DistilBERT arhitekturom.

MAML predstavlja moćan pristup koji može značajno poboljšati praktične primene few-shot učenja kroz dramatično smanjenje potreba za podacima. Uspješna primena zahteva pažljivo balansiranje meta i inner parametara učenja.

Literatura

- [1] Chelsea Finn, Pieter Abbeel, Sergey Levine. *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*. ICML 2017.
- [2] Victor Sanh, Lysandre Debut, Julien Chaumond, Thomas Wolf. *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2019.
- [3] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, Masanori Koyama. *Optuna: A Next-generation Hyperparameter Optimization Framework*. KDD 2019.