



Semestrální práce

SoPa: Bridging CNNs, RNNs, and Weighted Finite-State Machines

Studijní program:

N0613A140028 – Informační technologie

Studijní obor:

N0613A140028IS – Inteligentní systémy

Autor práce:

Bc. David Šafařík

Vedoucí práce:

doc. Mgr. Ing. Václav Záda, CSc.

Liberec 2022

Anotace

V tomto článku je představen model SoPa, který se snaží propojit rekurentní a konvoluční neuronové sítě jakožto odlišné přístupy pro kódování výroků v přirozeném jazyce. SoPa kombinuje učení neuronové reprezentace s váženými konečnými automaty, aby se naučil měkkou reprezentaci povrchových vzorů. Na třech úlohách klasifikace textu je ukázáno, že SoPa je srovnatelný nebo lepší než oba referenční modely rekurentní a konvoluční neuronové sítě, a je zvláště užitečný při malých datových sadách.

1 Úvod

V úvodu článku jsou rekurentní neuronové sítě (RNN) a konvoluční neuronové sítě (CNN) představeny jako jedny z nejužitečnějších nástrojů pro učení se textové reprezentace v oboru zpracování přirozeného jazyka (NLP). RNN jsou schopné zakódovat libovolně dlouhé posloupnosti textu s vysokou mírou expresivity. CNN jsou, na druhou stranu, mnohem více lokální ve smyslu toho, že jsou schopné zakódovat pouze text v okýnku pevné délky. Tento článek nabízí řešení pro spojení obou metod v podobě modelu **SoPa** (zkratka pro **Soft Patterns** - jemné vzory), který leží na hranici mezi RNN a CNN.

SoPa je neuronová verze vážených konečných automatů (WFSA) s omezenou množinou přechodů. SoPa je zajímavý z hlediska lingvistiky díky své schopnosti zachycovat jemné pojetí povrchových vzorů. Definuje si sadu jemných vzorů různých délek s tím, že každý vzor je reprezentován pomocí WFSA. Zatímco počet a délka těchto vzorů jsou hyperparametry, samotné vzory jsou parametry, které se učí pomocí principu konec-konec.

2 Pozadí

Autoři popisují povrchové vzory jako posloupnost slov a zástupných slov, které mohou být definovány manuálně, nebo mohou být extrahovány z korpusu za použití statistiky kookurence. Článek pokračuje definicí váženého konečného automatu s ϵ přechody.

3 SoPa: RNN založená na vážených konečných automatech

V této části článku je představen model SoPa jako RNN založená na WFSA, která je navržena pro reprezentaci textu jako soubor výskytů povrchových vzorů.

Článek navazuje popisem, jakým způsobem vážené konečné automaty reprezentují vzory. Každý vzor je představen samostatným automatem. Model byl navržen tak, aby se mohl učit přímo metodou konec-konec pomocí zpětné propagace.

Každý vzor tak má mezi 2 až 7 stavy. Každý stav má tři možné přechody: sám na sebe se spotřebováním slova, do dalšího stavu se spotřebováním slova, do dalšího stavu bez spotřebování slova. Vzory s d stavy se snaží srovnat s rozsahy délky $d - 1$.

4 SoPa jako rozšíření konvolunční neuronové sítě

V této části autoři ukazují, že SoPa je rozšířením jednovrstvé CNN s max-pool vrstvou. Docházejí k závěru, že jednovrstvé CNN se učí ještě omezenější třídu vážených konečných automatů, které zachycují pouze vzory fixní délky. Dále poukazují na rozdíly mezi SoPa a konvolučními neuronovými sítěmi. Uvádějí příklad, že CNN mohou používat jakýkoliv filtr.

5 Experimentální část

Pro vyhodnocení účinnosti SoPa byli v článku využity následující datasety:

- **SST.** The Stanford Sentiment Treebank obsahuje přibližně 10 tisíc filmových recenzí označovaných celým číslem na stupnici od 1 do 5. Autoři článku označili, pro účely úlohy binárního klasifikace, 1 a 2 jako negativní, 4 a 5 jako pozitivní, a recenze s hodnocením 3 ignorovali.

- **Amazon.** The Amazon Review Corpus obsahuje recenze elektronických zařízení. Každý dokument v tomto datasetu obsahuje recenzi a shrnutí. Autoři použili pouze recenze, a zaměřili se pouze na pozitivní a negativní recenze.
- **ROC.** The ROC story cloze task je úloha porozumění příběhu. Úloha je složena z prefixu, který se skládá ze čtyř vět, a dvou soutěžících závěrů příběhu. Jeden ze závěrů je příběhově konzistentní s prefixem, druhý je příběhově nesouvislý s prefixem. Autoři tento dataset využili pro úlohu detekce stylu. Z toho důvodu všechny konzistentní závěry považují za pozitivní vzorky, a naopak nekonzistentní závěry za negativní vzorky.

Autoři dále rozdělili trénovací množiny jednotlivých datasetů na různě velké podmnožiny za účelem zjištění toho, jak je model SoPa schopný učit se z malých datasetů.

K porovnání výsledků SoPa autoři využili čtyři referenční modely: BiLSTM (Bi-directional LSTM), jednovrstvá CNN, DAN (jednoduchá alternativa k RNN) a feature-based classifier trained with hard-pattern features.

- **BiLSTM.** Autoři nechali jednovrstvou BiLSTM naučit se reprezentaci dokumentu, následně předávali průměr všech skrytých stavů neuronové síti typu MLP.
- **CNN.** Autoři natrénovali jednovrstvou CNN s max-poolingem, následně předávali výslednou reprezentaci neuronové síti typu MLP.
- **DAN.** Deep averaging network s word dropout. Jednoduchý, ale silný model pro klasifikaci textu.
- **Hard.** Autoři natrénovali klasifikátor logistické regrese s více třídami. Slova s nízkým výskytem nahradili speciálními zástupnými slovy. Model se učí posloupnost 1-6 konkrétních slov s tím, že mezi každá dvě slova může být vložen libovolný počet zástupných slov. Nakonec slova s pravděpodobností výskytu alespoň 0,01% z trénovací množiny považují za konkrétní slova, a slova s pravděpodobností výskytu 1% a méně považují za zástupná slova.

6 Výsledky

Výsledky experimentů ukazují, že SoPa má výsledky srovnatelné, nebo lepší než referenční modely. Na datasetech ROC a SST má SoPa nejlepší výsledky ze všech modelů. Na datasetu Amazon model BiLSTM překonává všechny ostatní modely, ale výsledky ze SoPa jsou téměř srovnatelné s výsledky BiLSTM. Nicméně SoPa má třikrát až šestkrát méně parametrů než BiLSTM v závislosti na použitém datasetu.

Autoři dále zjistili, že SoPa výrazně vyniká při malých datasetech. Tento rozdíl se ale uzavírá se zvětšující se trénovací množinou.

7 Interpretabilita modelu

Interpretace samostatných vzorů. Autoři za účelem vizualizace vzoru spočítali pro každou frázi v trénovací množině skóre shody se vzorem, a následně vybrali k frází se nejvyšším skóre. Tímto způsobem zjistili, že vzory enkodují významově souvislé výrazy. Velká část tak odpovídá sentimentu, ale zachycují i například vyjádření času.

Dále autoři poukazují na fakt, že vzory modelu SoPa jsou relativně měkké, a umožňují lexikální ohebnost. Některé vzory však určitá slova fixují a neumožňují tak velkou ohebnost. Autoři navrhuji možnost využití SoPa jako dekodéru v budoucích pracích na základě toho, že by mohlo být možné podpořit model učit se konkrétnější slova.

Interpretace dokumentu. Interpretovatelná reprezentace dokumentu je vektor maximálně shodného skóre každého ze vzorů s jakýmkoliv rozsahem v dokumentu. Aby autoři zjistili, který z k vzorů nejvíce přispívá k rozhodnutí výsledné klasifikace, aplikovali metodu leave-one-out. Výsledkem je označení částí textu, které nejvíce přispívají buď k pozitivnímu označení textu, nebo k negativnímu označení.

8 Závěr

Autoři představili model SoPa, který kombinuje učení neuronové reprezentaci s váženými konečnými automaty. Pomocí experimentů dokázali, že SoPa výsledky předčí nebo je alespoň srovnatelný se čtyřmi referenčními modely, přestože obsahuje mnohem méně parametrů než modely s lepšími výsledky. S malou trénovací množinou ale SoPa vždy předčil referenční modely. Implementaci SoPa autoři zveřejnili v repozitáři na stránce github.com.