Rozšířený Abstrakt

Využití hlubokých neuronových sítí v systémech rozpoznávání řeči

Martin Paroubek, Ing. Petr Červa, Ph.D.

Abstrakt

V příspěvku proběhly experimenty s neuronovými sítěmi, které představovaly akustické modely v hybridním systému DNN-HMM. Ten se zdá být výhodnější než současný přístup. Zjišťovala se optimální topologie neuronové sítě, vliv množství dat na přesnost rozpoznávání a vliv předtrénování.

Úvod

Hluboké neuronové sítě byly označeny v článku NY Times [1] za největší změnu v přesnosti od roku 1979 a vzhledem k množství článků s překonáním současných systémů založených na GMM - HMM bylo rozhodnuto, že je nutné tento přístup otestovat i v laboratoři počítačového zpracování řeči na TUL.

Neuronové sítě byly v příspěvku použity pro vytvoření akustických modelů, které nahradily GMM část současného systému rozpoznávání. Experimentovalo se s různou topologií neuronové sítě, použitým postupem pro trénování a s různým množstvím dat v trénovacím korpusu. Vzhledem k množství různých parametrů a jejich vzájemné závislosti se vycházelo z již úspěšných pokusů publikovaných pro AJ.[2]

Experiment a metody

Veškeré experimenty sdílely nastavení těchto parametrů: $\eta = 0{,}08$ (parametr učení), $\alpha = 1$ (momentum), $i = 429$ (počet vstupních neuronů resp. Příznaků) a $o = 3180$ (počet výstupních neuronů resp. stavů HMM). Trénování akustických modelů probíhalo na GPU pomocí algoritmů největšího spádu a zpětného šíření chyby. Trénovací korpus obsahoval 56 hodin spojité polské řeči. U výsledných modelů byla experimentálně ověřena přesnost rozpoznávání na třech testovacích sadách a byla porovnána s přesností při použití současného systému (viz tabulka 1).

Tabulka 1. Přesnost rozpoznávání testovacích sad současným systémem [%]

Broadcast	Justice	Pomocne
73,14	88,09	85,16

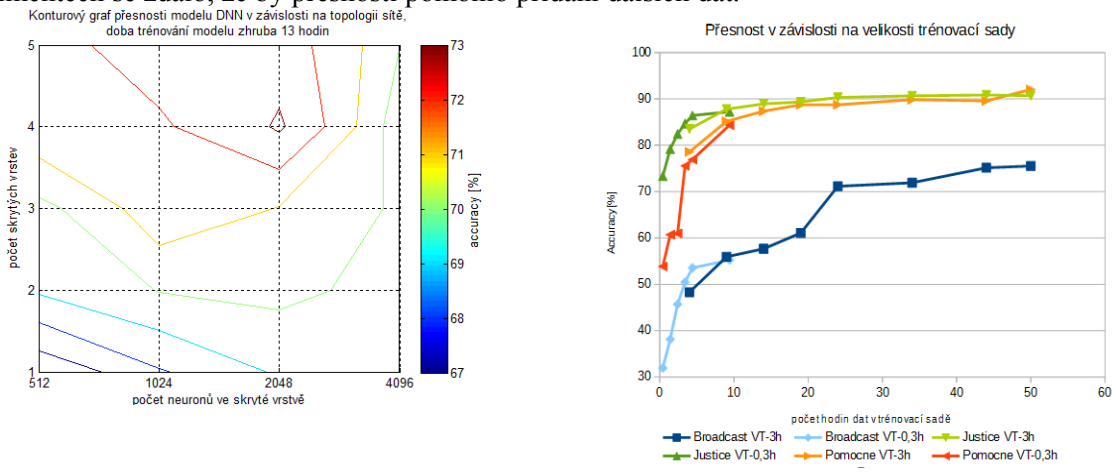
Výsledky a diskuze

Nejdříve byly odzkoušeny různé topologie neuronových sítí. Doba trénování těchto modelů se pohybovala od 12 hodin až po 7 dní. Bylo zjištěno, že příliš velké sítě (4096 neuronů ve skryté vrstvě) je náročnější natrénovat a nedosahují přesností, které byly natrénovány pro sítě o 1024 nebo 2048 neuronech ve skrytých vrstvách (viz obrázek 1 vlevo). Zároveň bylo zjištěno, že vyšší počet vrstev přináší užitek do 5 až 6 skrytých vrstev. Nejlepší modely byly dotrénovány a bylo zjištěno, že jako nejpřesnější se zdá topologie o šířce 1024 neuronů a 5 skrytých vrstvách po 80 hodinách trénování. Na testovacích sadách došlo k absolutnímu zlepšení přesnosti o 3 - 7% oproti současně používanému systému.

V další fázích experimentů byla testována metodu předtrénování po jednotlivých vrstvách. Byly odzkoušeny dva možné postupy. V prvním byla každá nová vrstva natrénována do co nejlepší

přesnosti na testovací části trénovacího korpusu a v závěru proběhlo diskriminativní trénování celé sítě. V druhém pokusu byly jednotlivé vrstvy trénovány po 5 epochách. Ani v jednom případě nedošlo ke zlepšení přesnosti oproti metodě bez předtrénování.

V posledním experimentu bylo úkolem zjistit, o kolik lze přesnost modelu zlepšit přidáním dalších dat. Tento průběh byl sledován na trénovacích korpusech o velikosti 1 hodiny až 56 hodin (viz obrázek 1 vpravo). Pro testovací sady Justice a Pomocne bylo zjištěno, že dostačující velikost je okolo 30 hodin. Pro testovací sadu Broadcast, která měla celkově nižší hodnoty přesnosti ve všech experimentech se zdálo, že by přesnosti pomohlo přidání dalších dat.



Obrázek 1. Vlevo:Konturový graf přesnosti modelu DNN v závislosti na topologii sítě. Vpravo: Graf přesnosti rozpoznávání v závislosti na velikosti trénovací sady

Závěr

Výsledný model zaznamenal zlepšení o 3 - 7% oproti současným systémům. Přestože metoda předtrénování se nezlepšila oproti běžnému postupu trénování, mohla by být užitečná v případě menších trénovacích korpusů. Množství dat nutné pro trénování neuronových sítí se zdá být závislé na kontextu použití. Další pokusy by se mohly odvíjet směrem k použití jiných aktivačních funkci (např. ReLU) anebo k natrénování akustického modelu jiného jazyka při použití již vytvořeného modelu jako iniciálního stavu.

Poděkování

Rád bych na tomto místě poděkoval panu Ing. Petru Červovi, Ph.D. za cenné rady, připomínky, ochotu a čas věnovaný konzultacím mé diplomové práce a také panu Ing. Ladislavu Šepsovi za uvedení do problematiky a četné konzultace.

Reference

[1] Scientists See Promise in Deep-Learning Programs. *The New York Times* [online]. 2012 [cit. 2014-05-10]. Dostupné z: http://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html

[2] HINTON, Geoffrey, Li DENG, Dong YU, George DAHL, Abdel-rahman MOHAMED, Navdeep JAITLY, Andrew SENIOR, Vincent VANHOUCKE, Patrick NGUYEN, Tara SAINATH a Brian KINGSBURY. Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine.* 2012, s. 82-97. Dostupné z: https://www.cs.toronto.edu/~hinton/absps/DNN-2012-proof.pdf