

# Mély neuronhálók az akusztikus modellezésben

Grósz Tamás

II. évf. programtervező informatikus MSc

Témavezető: Dr. Tóth László

MTA-SZTE Mesterséges Intelligencia Kutatócsoport

Az elmúlt néhány évtizedben a mesterséges neuronhálók számos változatát kipróbálták a beszédfelismerésben, általános elismertséget azonban csak a többrétegű perceptron-hálózatokra (MLP) épülő ún. hibrid HMM/ANN modellnek sikerült elérnie. Mindez megváltozni látszik azonban az ún. mély neuronhálók (deep neural nets) megjelenésével. A mély neuronhálót (és tanítási algoritmusát) 2006-ban publikálták először [1], a kezdeti cikkben képi alakfelismerési tesztekkel használták demonstrációként. Az első cikk a mély neuronhálók beszédfelismerési alkalmazásáról 2009-ben jelent meg, melyben sikerült megdönteni az összes korábban elért felismerési pontosságot [2] egy közismert tesztfeladaton, a TIMIT adatbázison.

Dolgozatomban a mély neuronhálókra épülő akusztikus modelleket beszédhang-felismerési teszteken értékelem ki és vetem össze eddigi eredményekkel. Részletesen bemutatom a mély neuronhálók tanítási algoritmusát, amely két részből áll: egy felügyelet nélküli előtanításból és egy finomhangolásból. A tanuló adatbázisok nagysága és a tanítási algoritmus nagy időigénye miatt a saját implementációt GPU-n futtathatóvá tettem, ehhez CUDA-t használtam. Elsősorban a TIMIT adatbázison elért eredményeimet közlöm, mert a legtöbb cikk is ezt az adatbázist használja, így összevethetőek az eredményeimel. Továbbá az akusztikus modelleket kiértékelem még két magyar nyelvű adatbázison (egy hangoskönyv és egy híradós adatbázis) is. A TIMIT-en elért eredményemet hasonlítja össze az alábbi táblázat az irodalomból ismert értékekkel (beszédhang-felismerési hiba a „core” tesztalmanachon):

Trifón HMM [2]	27,3%
Rekurrens Neuron Háló [2]	26,1%
Monofón HTM [2]	25,5%
Monofón Deep Net [2]	23%
Monofón Deep Net, saját implementáció	22,79%

Kulcsszavak: Beszédfelismerés, akusztikus modell, CUDA.

[1] Hinton, G. E., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets, Neural Computation, Vol. 18, pp. 1527-1554, 2006.

[2] Mohamed, A., Dahl, G., Hinton, G.: Deep belief networks for phone recognition, Proc. NIPS Workshop on Deep Learning for Speech Recognition and Related Applications, 2009.