

Transzfer learning alapok kísérletezőknek és menedzsereknek

Molnár Tamás

2022. 05. 04.

- Néhány bevezető szó

- Mi a transzfer learning?
 - Lényege
 - Hasznai
 - Az újrahasznosítás, mint egy általános eszköz
- Aláírás azonosítás esetpélda
 - A megoldandó feladat
 - A megoldás szakaszai
 - A koncepcionális fejlődés szakaszai, beleértve a transzferek alkalmazását

Mi a transzfer learning?

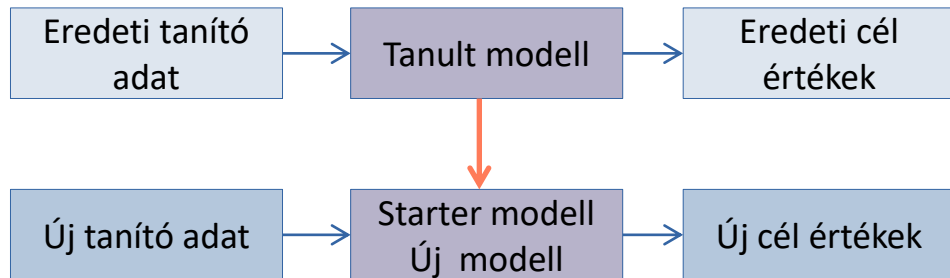
- A transzfer learning egy olyan gépi tanulási technika, ahol egy bizonyos probléma megoldására betanított modellt alkalmassá teszünk egy másik, hasonló probléma megoldására

Az “alkalmassá tétel” a kész modell

- Továbbtanítását új feladat adatokon

és esetleg

- Kismértékű szerkezeti módosítását jelenti



Az alkalmazásától várható előnyök:

- Lerövidült fejlesztési idő, mert nem nulláról indulunk, vagy
- Közvetlenül nehezen támadható feladat megoldása vagy
- Pontosabb előrejelző képességű végső modell építése

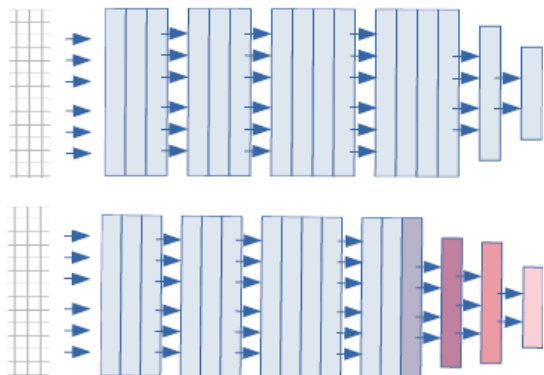
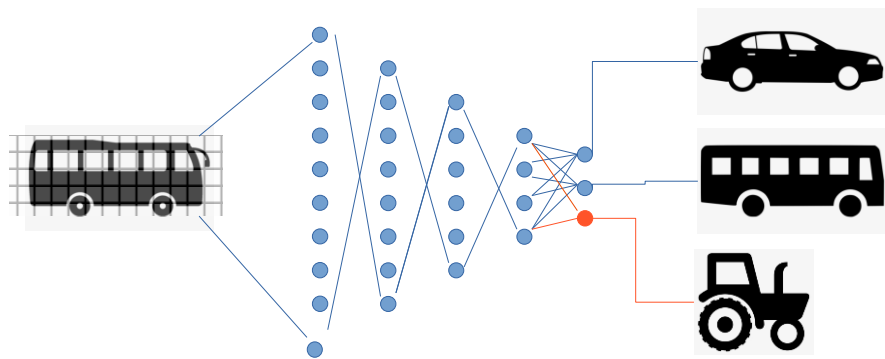
Szerencsés esetben külső forrásból elérhetünk releváns előtanított modelleket. Jól jöhetnek, ha

- Kevés adatunk van
- Nincs erőforrásunk sok adat használatához

Mi a transzfer learning?

Jellemzően

- Deep learning környezetben merül fel



Nevezetes példák:

Képfeldolgozás (kép osztályozás)

- ImageNet alapú osztályozó továbbtanítása

Natural language processing

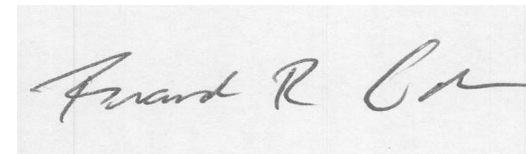
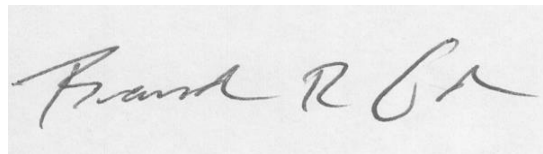
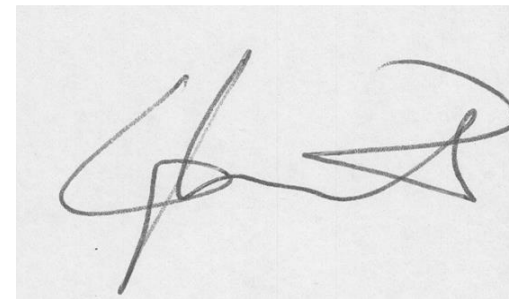
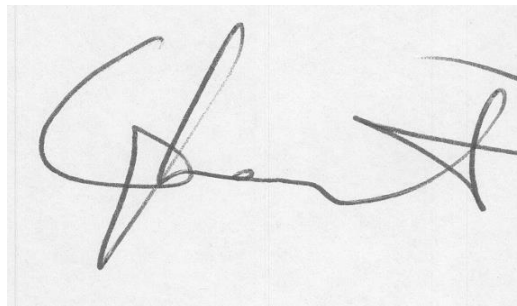
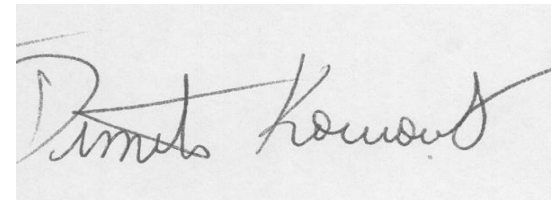
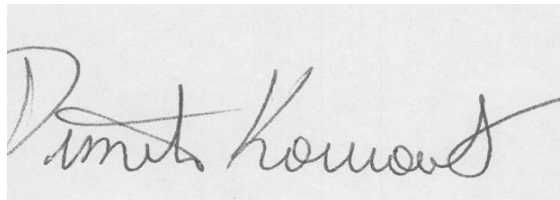
- GloVe szó reprezentáció használata saját célra

Félig jó (nem kanonikus) példák:

- Módszertan átvitele egyik domainből a másikba
- Célfüggvény csere és újratanítás
- Modell stacking
- Modell update új adatokkal újraépítés nélkül

A megoldandó feladat

- Jegyzőkönyv aláírók beazonosítása
 - tableten vagy mobilon felvett
 - aláírás (szignó) alapján
 - utólag, audit támogatási célból
 - „gyorsan”
-
- Hamisítást nem tételezünk fel.
 - Az aláíró jogosan jár el, csak nincs letárolva a kiléte; 1500+ fő.
-
- Rendelkezésre áll több darab aláírás minta ugyanattól a személytől
- POC (proof of concept) cél:
- Mutassuk meg, hogy bármilyen szinten a probléma kezelhető
- MVP (minimum viable product) cél:
- Legyen egy alkalmazásunk, ami az aláírások 80%-át „gyorsan” beazonosíthatóvá teszi



A megoldás szakaszai – inkább modellezői szemmel

Irodalmazás

POC szakasz

MVP szakasz

Első
implementáció

Továbbfejlesztés

- 11+ cikk
- Gyűjtés: 2 hétvége

- CEDAR aláírás adatbázis (24 aláírás / fő, 55 személy)



- Valós aláírás minták (2*10 darab / fő, 303 személy)



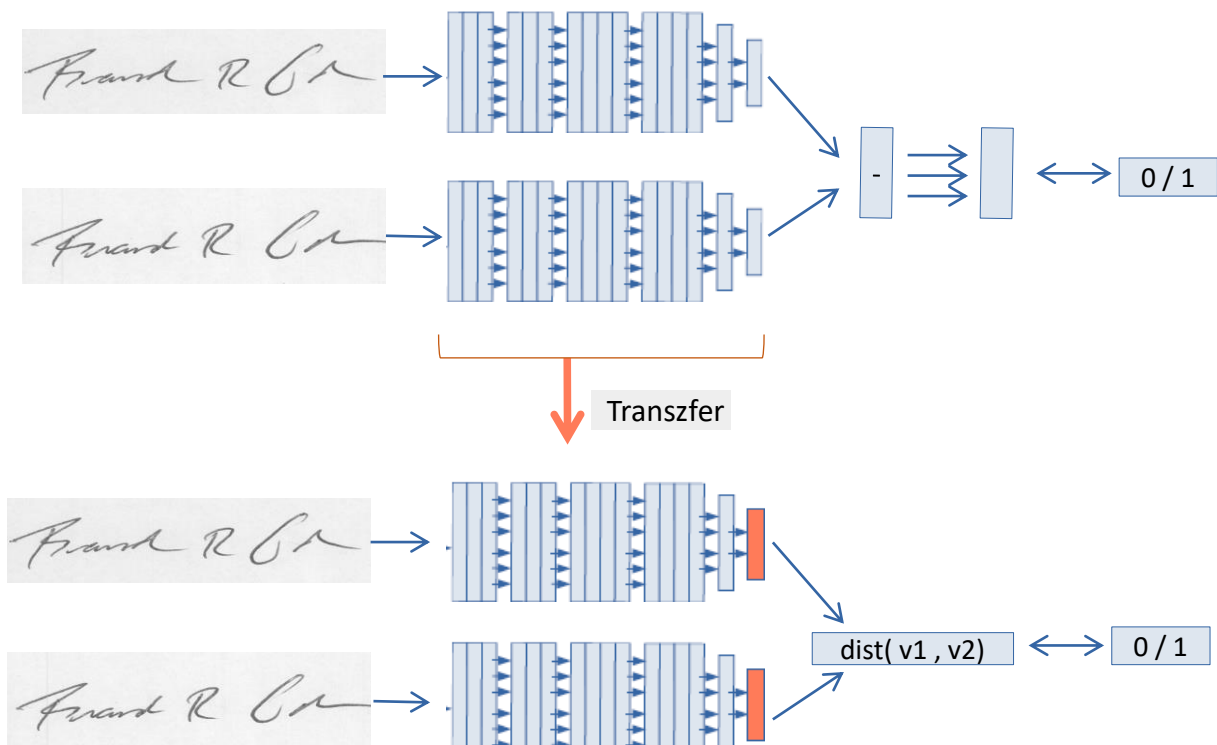
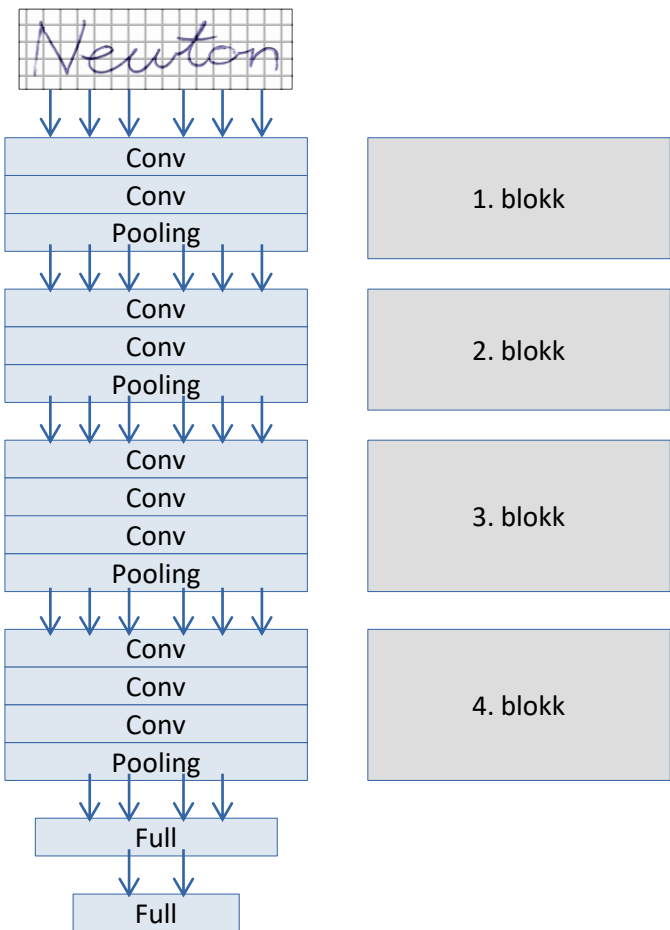
- Dokumentum és aláírás minta törzs menedzsment rendszer

- Lényegesen több aláírás minta
- De nem egységes számban

Newton

Newton

A POC szakasz indítása: az analitikus ötlet

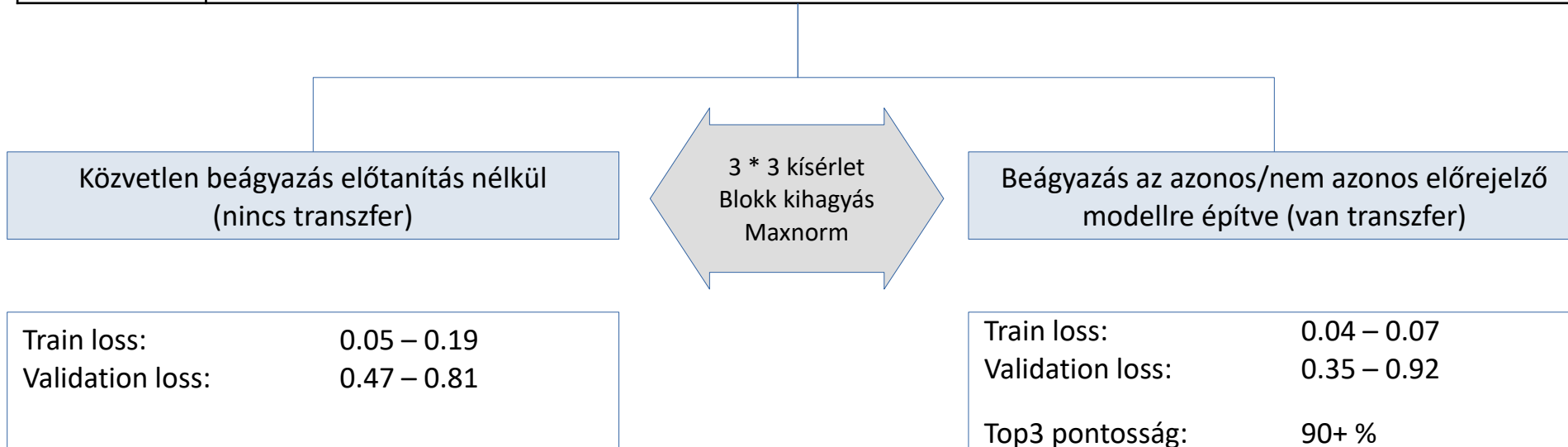


A POC szakasz gondolati fejlődése -1

Szakasz	Kísérleti csapás	Megfigyelések
Tú a kazalban	Változó számú végső dense layer Más méretű dense layer	Rossz train loss. Nagyon rossz és sokszor instabil validation loss.
	Változó strides – mennyire gyorsan bontsa le a képet	
	Dropout a dense layerek után Különböző aktivációs függvények a dense layerekben Dropout a konvolúciós rétegek között	
	Különböző aktivációs függvények a final node-ban	~30-40 kísérlet után: legalább már a train loss jó, bár a validation még rossz
	Maxnorm bevezetése	
	Más pozitív-negatív részarányú tanító adatbázisok	
Homály	Kísérletek az eddigi paraméterek mentén	~60 kísérlet után Kevés hatás, legfeljebb extrém beállítás esetén és akkor is negatív.
	Játék a vektorok összevezetésének módjával	
	Batch size változtatása	

A POC szakasz gondolati fejlődése - 2

Szakasz	Kísérleti csapás	Megfigyelések
Remény	Konvolúciós rétegek és teljes blokkok elhagyása	Lényegesen jobb validációs hiba ~100 kísérlet után: még nem túl jó, de reményteli teljesítmény
	Váltás beágyazási célfüggvényre	



Az MVP szakasz fő alverziói

Papír + Papír

Newton Newton

Tablet + Tablet

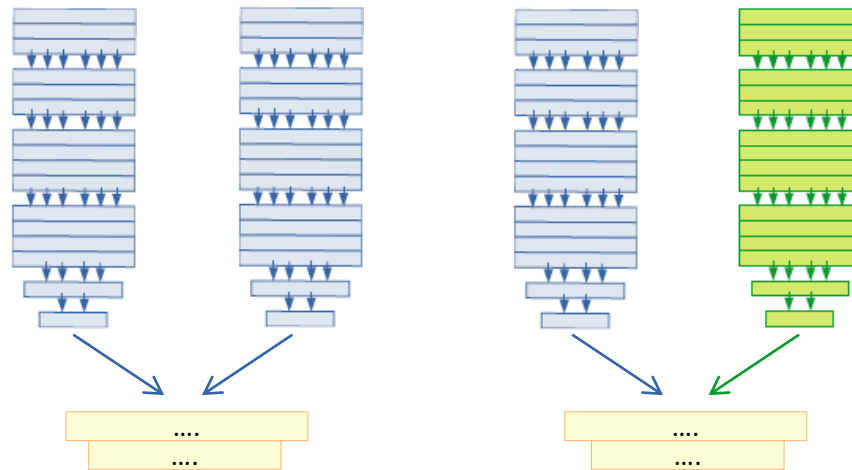
Newton Newton

Papír + Tablet

Newton Newton

Papír + Tablet

Newton Newton



8 kísérlet
0-ról indítva 160-260 epoch
Egyszerűen jó
teljesítmény

35 kísérlet
0-ról indítva 80-180 epoch
TL: + 150-160 epoch
Egyszerűen jó
teljesítmény

110 kísérlet – 1 kísérlet ~ 1 nap időtartam
Jobb megoldások transzfer nélkül 0-ról indítva ~90 epoch
(71% - 80% - 85% - top2 – top3 - top4)
Jobb megoldások transzferrel +200-300 epoch
(70% - 79% - 85% top2 – top3 - top4)

Az MVP gondolati irányjai

Ami nem jött be

Batch normalization trainable

Embedding layer trainable

Add extra dense layer

Batch size

Learning rate decay

Ami kellett valamennyire

Azonos vagy különböző tablet / papír embedder

Maxnorm

Ami hasznosnak bizonyult

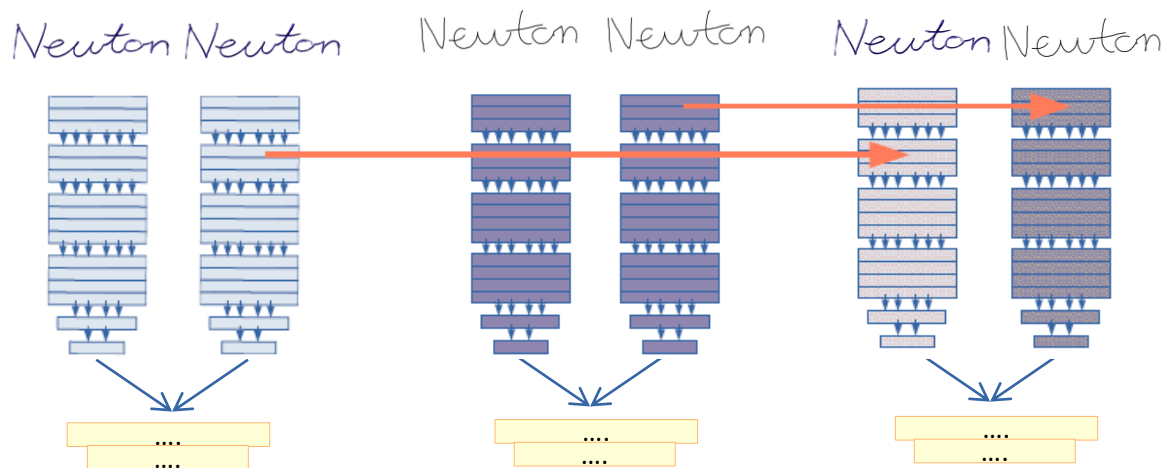
Direkt struktúra egyszerűsítés

Learning rate beállítás

Nem direkt, hanem transzfer tanítás

A transzfer alap megválasztása

Türelem az epoch számokkal



A megoldás szakaszai – inkább menedzseri szemszögből

Irodalmazás

- 11+ cikk
- Gyűjtés: 2 hétvége

POC szakasz

- CEDAR aláírás adatbázis (24 aláírás / fő, 55 személy)



MVP szakasz

- Valós aláírás minták (2*10 darab / fő, 303 személy)

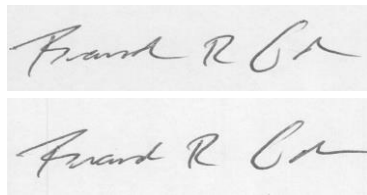


Első implementáció

- Dokumentum és aláírás minta törzs menedzser rendszer

Továbbfejlesztés

- Lényegesen több aláírás minta
- De nem egységes számban



Newton

Newton

Választott alaplómódszertani elvek

Kidolgozott alaplómódszertan (kód)

Valódi feladatra működőképes módszertan (kód)

Felhasználói alkalmazás + adatbázis

Hatékony alkalmazás

3 hét

10 hét

4 hónap / 3 hét

5 hónap

Köszönöm a figyelmet!