

Pletykaalapú gépi tanulás teljesen elosztott környezetben

Hegedűs István

Témavezető

Dr. Jelasity Márk

MTA-SZTE Mesterséges Intelligencia Kutatócsoport

Informatika Doktori Iskola

Szegedi Tudományegyetem



Doktori értekezés összefoglaló

Szeged

2016

Bevezetés

Az adatbányászati és gépi tanuló algoritmusok jelen vannak a digitális életünkben, még ha ez nem is tűnik fel nekünk. Ezen algoritmusok szolgálnak a megtévesztő levelek kiszűrésére a levelezőprogramunkban, hozzászólások automatikus moderálására a közösségi oldalakon és blogokban, automatikusan kiegészítik a keresőkifejezéseinket vagy a szavakat a szövegszerkesztőkben, filmeket vagy termékeket ajánlanak egy webshop-ban. Más algoritmusok segítenek megóvni az egészségünket egy orvosi alkalmazásban, vagy csak lenyűgöznek minket egy videojátékban. Felismerik a kiejtett szavakat és a kézzel írt szöveget. Részletekbe menő leírás nélkül had idézzük Tom M. Mitchell definícióját a gépi tanulásra [13]:

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ,”

E definíció alapján a cél a példákából való tanulás. Ezen példákat általában mi állítjuk elő, amikor egy levelet bedobunk a spam mappába, vagy amikor rákattintunk a keresésünk eredményére vagy egy reklámra egy weboldalon, mobilalkalmazásban. Ezekből a példákából adatbázisok építhetők, melyeket a gépi tanuló algoritmusok tanítására használhatunk, ezáltal kinyerjük az adatbázisból a benne lévő információt és azonosítjuk a benne rejlő mintázatokat. Ezek a minták és az információ segíthet megérteni, hogyan működik a minket körülvevő világ, azonosítani vagy előre jelezni a trendeket, optimalizálni a városok forgalmát.

A jelentős mennyiségű adat általában cégek tulajdonában lévő szervereken, klasztereken vagy a felhőben van eltárolva, viszont a mi eszközeinken (PC-ken, laptopokon, okos telefonokon, hordozható eszközökön, okos házak érzékelőin, ...) lett előállítva. Ezen adat feldolgozása egyre nagyobb kihívás többek között annak mérete és a korlátozott hozzáférés miatt. A párhuzamos számítási módszerek még képesek kezelni a méret problémáját, de ahogy az idő telik, az adat egyre csak gyűlik, és a cégeknek egyre több erőforrásra lesz ehhez szükségük. Egyre nagyobb szerverfarmokra, amelyek képesek tárolni az összegyűlt adatokat, egyre nagyobb számítási kapacitásra, hogy kiszolgálják a kéréseinket, és egyre több memóriára az adatbányászati feladatok lefuttatására. Továbbá, a hozzáférés ezekhez az adatokhoz általában nem lehetséges még a kutatók számára sem. A teljesen elosztott (pl. peer-to-peer (P2P)) gépi tanuló algoritmusok használatát leginkább a fent említett okok motiválják. Mivel az adat az általunk minden nap használt eszközökön állítódik elő, ezek az eszközök végezhetnék közösen az adat feldolgozását is, valamint a gépi tanuló algoritmusok tanítását is az eszközöknek kellene elosztottan végrehajtaniuk.

1. táblázat. Az értekezés fejezetei és a hozzá kapcsolódó publikációk (ahol a ● jelöli a fő, míg a ○ jelöli a kapcsolódó publikációkat).

	3. fejezet	4. fejezet	5. fejezet	6. fejezet
CCPE 2013 [6]	●	○	○	○
EUROPAR 2012 [3]	○	●		
SASO 2012 [4]	○		●	
SISY 2012 [2]	○		●	
ACS 2013 [5]	○		●	
P2P 2014 [9]	○			●
EUROPAR 2011 [1]	○			
ICML 2013 [7]	○			
ESANN 2014 [8]	○			
TIST 2016 [11]	○			○
PDP 2016 [12]	○			
PDP 2016 [10]	○			○

További előnye a teljesen elosztott módszereknek a személyes adatok védelme. Senki sem szeretné, hogy a keresési előzménye, a személyes fényképei vagy az általa megnézett filmek listája kiszivároгjon a felhőből, csak mert egy hacker talált egy hátsó kaput egy szerveren. Ennek érdekében ezen elosztott módszerek a személyes adatainkat a saját eszközünkön tartják, ahelyett hogy feltöltenék egy adatközpontba, és mégis képesek gépi tanuló modellt építeni az adatokon. Ebben az értekezésben bemutatjuk egy lehetséges módját a teljesen elosztott gépi tanulásnak.

A dolgozat második fejezetében elsőként adunk egy összefoglalást a kapcsolódó háttértudásról, melyben bevezetjük a felügyelt tanulást, adunk egy áttekintést az alkalmazott rendszermodellről és az adatok elosztásáról; valamint bevezetjük a teljesen elosztott algoritmusokat pár példa segítségével. A továbbiakban, a 3 – 6. fejezetekben bemutatjuk a tézis fő pontjait, ahol bevezetünk egy teljesen elosztott tanulási rendszert, melyben az adat modellező algoritmusok online módon taníthatóak. Ehhez megadunk számos algoritmust, köztük fejlettebbeket is, melyek online módon taníthatóak. A későbbi fejezetekben leírunk korszerű gépi tanulási módszereket is, mint például a boosting és a mátrixfelbontás. Valamint bemutatjuk a keretrendszerünk egy továbbfejlesztett változatát, mely hatékonyan képes kezelni a fogalomsodródás problémáját. Az 1. táblázat mutatja a kapcsolódó publikációkat ¹.

¹A tézisben bemutatott algoritmusok megvalósítása online elérhető: <https://github.com/isthededus/Gossip-Learning-Framework>

A tézis eredményeinek összefoglalása

Ahogy azt a bevezetőben említettük, az érdekes információt tartalmazó adat a napi használatban lévő eszközeinkről származik. Az összegyűjtött adathalmazhoz sajnos nincs hozzáférésünk, mivel az magánkézben lévő szervereken tárolódik. Leginkább ezek a tényezők motiváltak minket arra, hogy olyan módszereket dolgozzunk ki, amelyek teljesen elosztott módon képesek feldolgozni az adatot.

A tézis fő célkitűzése egy olyan módszer bemutatása, amely segítségével teljesen elosztott adatbázisokon alkalmazhatunk gépi tanuló módszereket anélkül, hogy az adatokat egy központi helyre összegyűjtenénk. Az elvárásunk az, hogy a nagyszámú számítástechnikai eszköz együttműködve oldja meg a gépi tanulási problémákat kizárólag üzenetküldések segítségével. Bemutatunk egy pletykaalapú keretrendszert, amely képes megoldani a fent említett problémát. Ebben a keretrendszerben különféle algoritmusok használhatók. A harmadik fejezetben részletesen bemutatjuk e keretrendszert, valamint számos, a keretrendszerben alkalmazható tanuló algoritmust. Itt a felügyelt tanuló algoritmusokra koncentrálunk, mint például a logisztikus regresszió, Support Vector Machines és a mesterséges neuronháló. A későbbi fejezetekben pedig fejlettebb módszereket és a keretrendszer alkalmazhatóságát vizsgáljuk. Bemutatunk egy boosting módszert, amellyel az algoritmusok osztályozási teljesítménye javítható, valamint egy módszert, amellyel kezelhetővé válik a fogalomsodródás problémája. Ezentúl egy felügyelet nélküli elosztott mátrix felbontó algoritmust is ismertetünk. Végül fontos megjegyezni, hogy a fent említett keretrendszer támogatja a személyes adatok védelmét, mivel az adat feldolgozása során a lokális adat sosem hagyja el az eszközt, amely tárolja azt.

1. tézis: Pletykaalapú gépi tanulás

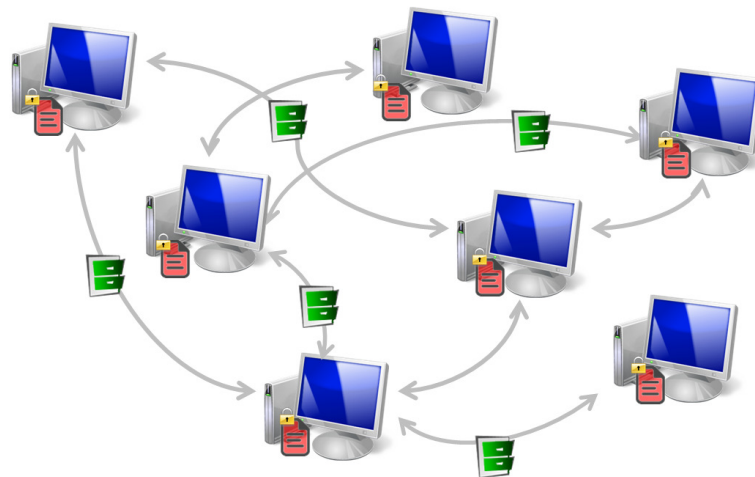
Algoritmus 1 Pletykaalapú tanulás (GOLF)

```

1:  $x, y$  ▷ local data
2:  $currentModel \leftarrow initModel()$ 
3: loop
4:    $wait(\Delta)$ 
5:    $p \leftarrow selectPeer()$ 
6:   send  $currentModel$  to  $p$ 
7: end loop
8: procedure ONRECEIVEMODEL( $m$ )
9:    $m.updateModel(x, y)$ 
10:   $currentModel \leftarrow m$ 
11: end procedure

```

Bemutatunk egy pletykaalapú általános keretrendszert, amely gépi tanuló modellek sztochasztikus gradiens alapú tanítását teszi lehetővé P2P rendszerekben, teljesen elosztott adatok felett. A pletykaalapú tanulás alapötlete, hogy modellek vándorolnak a hálózat elemein véletlen sétát leírva, miközben az eszközök frissítik a fogadott modelleket. Ebben a keretrendszerben minden sztochasztikus gradiens alapú módszer megvalósítható. A hálózatban jelen lévő csomópontok ugyanazt az algoritmust hajtják végre (amelyet az 1. algoritmus reprezentál), ezáltal megvalósítva a együttes modell tanítást. Amikor egy eszköz csatlakozik a hálózathoz, inicializál egy úgynevezett lokális modellt. Ezután rendszeresen (Δ időközönként) elküldi a lokális modelljét egy szomszédjának. Amikor egy eszköz fogad egy modellt, frissíti és eltárolja azt. A hálózati csomópontok szomszédjainak karbantartását a NEWSCAST peer mintavételező szolgáltatás végzi, amely képes egyenletesen véletlenszerűen mintázni a hálózatot. Ameddig a peer mintavételező szolgáltatás megfelelően működik, a modellek véletlen sétát írnak le a hálózatban. Az 1. ábrán



1. ábra. Pletykaalapú tanulás

látható az általunk használt rendszermodell és az elosztott tanulás, ahol a zölddel jelzett model-

lek vándorolnak a kommunikációs csatornán, a piros színnel jelölt adat pedig nem hagyja el az eszközt.

Ebben a keretrendszerben minden algoritmus használható, amely képes az adat streamként való feldolgozására. Továbbá, az algoritmusok, amelyek sztochasztikus gradiens (SGD) alapú módszert használnak és konvex célfüggvényt optimalizálnak, a probléma optimumához fognak konvergálni, amíg a peer mintavételező szolgáltatás megfelelően működik. Ezáltal a modellek a hálózatban véletlen sétát írnak le és egyenletesen véletlenszerű példával lesznek frissítve. Bemutatunk számos gépi tanuló algoritmust, amelyek beilleszthetők ebbe a keretbe, mint például: Pegasos SVM, Logistic Regression, ANN.

A keretrendszer lehetővé teszi a hálózati csomópontok számára a predikció lokális kiszámítását, bármilyen üzenet küldése nélkül. Továbbá a kommunikációs költsége is kielégítő, csomópontonként egy-egy üzenet elküldés történik minden ciklusban.

Azáltal, hogy a lokális adat sosem hagyja el az eszközt, amely tárolja azt, a keretrendszer támogatja az érzékeny adatok védelmét. Ezen adatok csak speciálisan kialakított modellek segítségével figyelhetők meg.

Fő eredmények:

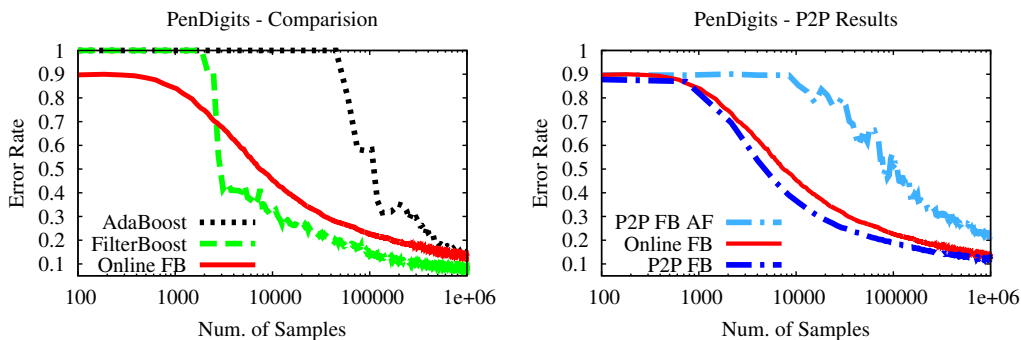
- Teljesen elosztott tanuló keretrendszer (GOLF);
- Számos megvalósított tanuló algoritmus;
- Lokálisan használható modellek;
- Érzékeny adatok védelmének támogatása;
- Kapcsolódó publikáció: [6]

Róbert Ormándi, István Hegedűs, and Márk Jelasity. Gossip learning with linear models on fully distributed data. *Concurrency and Computation: Practice and Experience*, 25(4):556–571, 2013

2. tézis: Teljesen elosztott turbózás

Bemutattuk, hogy a fenti keretrendszer alkalmas a többsztályos turbózás megvalósítására is. A turbózás módszere alkalmassá teszi a gépi tanuló módszereket, hogy azok nagyobb reprezentációs erővel bírjanak egy megfelelően súlyozott tanítás és szavaztatás segítségével.

Ennek elérése érdekében bemutattunk egy módosított FILTERBOOST algoritmust, amely így teljesen online módon képes többsztályos modellek tanítására. Elméletileg igazoltuk, hogy a definiált módszer optimalizálja a megadott negatív log-likelihood mértéket. Az eredmények jelentősége, hogy a gépi tanuló modell minősége szempontjából fejlett algoritmusok is alkalmazhatók teljesen elosztott rendszerekben. A 2. ábra szemlélteti, hogy az elosztott algoritmus versenyképes más megvalósításokkal.



2. ábra. Boosting algoritmusok összehasonlítása (bal oldal) és P2P kiértékelések (jobb oldal). FB és AF jelölik a FilterBoost és a „hibával terhelt” eseteket.

Továbbá rámutattunk, hogy a modellek diverzitásának csökkenése lehetséges probléma a GOLF keretrendszerben. Adtunk egy megoldást, amely segítségével fenntartható a hálózatban jelen lévő modellek változatossága. Ez lehetővé teszi a legjobb modellek elterjesztését az eszközön, kihasználva a hálózatban rendelkezésre álló modellek számát.

Fő eredmények:

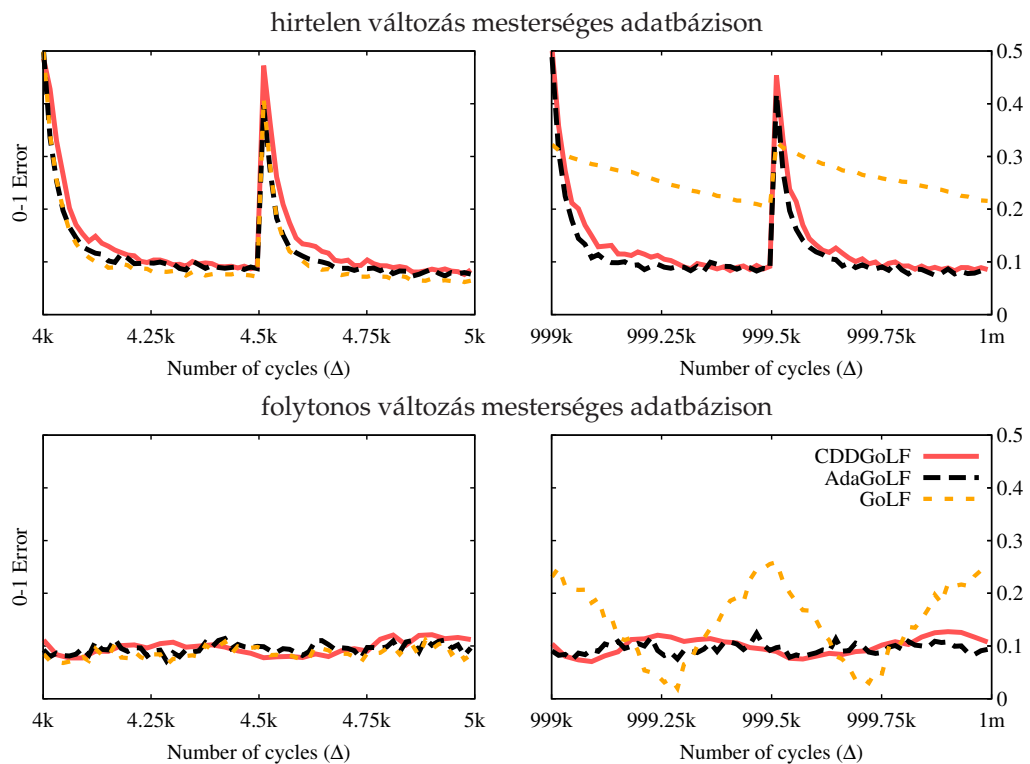
- Elosztott többsztályos turbózás megvalósítása;
- Online FILTERBOOST algoritmus és elméleti levezetése;
- A modellek változatosságának fenntartása a GOLF keretrendszerben;
- Kapcsolódó publikáció: [3]

István Hegedűs, Busa-Fekete Róbert, Ormándi Róbert, Jelasity Márk, and Kégl Balázs. Peer-to-peer multi-class boosting. In Christos Kaklamanis, Theodore Papatheodorou, and Paul Spirakis, editors, *Euro-Par 2012 Parallel Processing*, volume 7484 of *Lecture Notes in Computer Science*, pages 389–400. Springer Berlin / Heidelberg, 2012

3. tézis: Fogalomsodródás kezelése

Bemutatjuk a GOLF keretrendszer két adaptív változatát, melyek az ADAGOLF és a CDDGOLF. Ezekkel a módszerekkel a keretrendszer alkalmas az adatokban rejlő azon minták változásának kezelésére, amelyeket meg akarunk tanulni. A változás különféle okok miatt következhet be. Ezek például: a felhasználók cserélődhetnek egy alkalmazásban, vagy külső tényezők megváltozhatnak, mint például az időjárás. Ez hatással lehet arra, ahogy az emberek reagálnak dolgokra, milyen filmet néznek, hogyan közlekednek. Valamint az emberek különbözően viselkednek napi tevékenységeik folyamán is.

Az ADAGOLF esetén az adaptivitást a hálózatban jelen lévő modellek életkoreloszlásának fenntartásával értük el, figyelve a megfelelő változatosság fenntartására az életkorokban. Így a hálózatban lesznek fiatal (adaptív) és idős (jó teljesítményű) modellek. A CDDGOLF szintén a modellek újraindításával éri el az adaptivitást, viszont a döntés itt a modell teljesítményének az előzményétől függ. Ezen módszer képes jelezni az eloszlás változását is.



3. ábra. A "beragadás,, jelensége, mint motiváció az adaptivitásra.

A 3. ábra szemlélteti a fogalomsodródás jelenségét. Ez azt jelenti, hogy az alap gépi tanuló módszerek nem tudják kezelni az adat változását, egy idő (megfelelő számú frissítés) után nem

tudnak igazodni az adat eloszlásának változásához. Az általunk bemutatott módszerek segítenek az algoritmusoknak elkerülni ezt a problémát.

A legfőbb eredményünk azon esetekhez köthető, amikor az eloszlás mintázásának a sebessége viszonylag alacsony az eloszlás változásának a sebességéhez képest. Ebben az esetben a módszerünk teljesítménye jelentősen felülmúlja az alap algoritmusok eredményeit és megközelíti az elérhető legjobb teljesítményt.

Továbbá bemutattuk, hogy nagyobb mintavételezési sebesség esetén is alkalmazható a módszerünk, habár ebben az esetben a lokális adaton tanított modell is kielégítő eredményt ér el.

Fő eredmények:

- Kétféle adaptív módszer a GOLF keretrendszerhez, melyek
 - Egyike fenntartja a modellek életkorának eloszlását
 - A másik pedig újraindítja az alacsony teljesítményű modelleket;
- Eloszlás változásának követése, valamint detektálása;
- Kiemelkedő teljesítmény az eloszlás ritka mintavételezése esetén;
- Kapcsolódó publikációk: [2, 4, 5]

István Hegedűs, Ormándi Róbert, and Jelasity Márk. Gossip-based learning under drifting concepts in fully distributed networks. In *2012 IEEE Sixth International Conference on Self-Adaptive and Self-Organizing Systems, SASO'12*, pages 79–88. IEEE, 2012

István Hegedűs, Lehel Nyers, and Róbert Ormándi. Detecting concept drift in fully distributed environments. In *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics, SISY'12*, pages 183–188. IEEE, 2012

István Hegedűs, Róbert Ormándi, and Márk Jelasity. Massively distributed concept drift handling in large networks. *Advances in Complex Systems*, 16(4&5):1350021, 2013

4. tézis: Szinguláris felbontás

Bemutatunk egy SGD alapú módszert, amely segítségével alacsony rangú mátrixfaktorizáció feladata oldható meg, teljesen elosztott környezetben, továbbá a módszer egy módosítását, amely egy A mátrix SVD felbontásához konvergál. Az algoritmus eredménye k rang esetén azon X és Y mátrixok, amelyek az A mátrix első k jobb és bal szinguláris vektorainak skálázott változatát tartalmazzák. Az X és Y mátrixok egyediek, eltekintve az oszlopaik hosszától. A feladat az alábbi célfüggvény optimalizálásával oldható meg

$$J(X, Y) = \frac{1}{2} \|A - XY^T\|_F^2 = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (a_{ij} - \sum_{l=1}^k x_{il}y_{jl})^2, \quad (1)$$

ahol $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{m \times k}$, $Y \in \mathbb{R}^{n \times k}$, és a gradiensek (a modell frissítéshez) a fenti függvény parciális deriváltjai

$$\frac{\partial J}{\partial X} = (XY^T - A)Y, \quad \frac{\partial J}{\partial Y} = (YX^T - A^T)X. \quad (2)$$

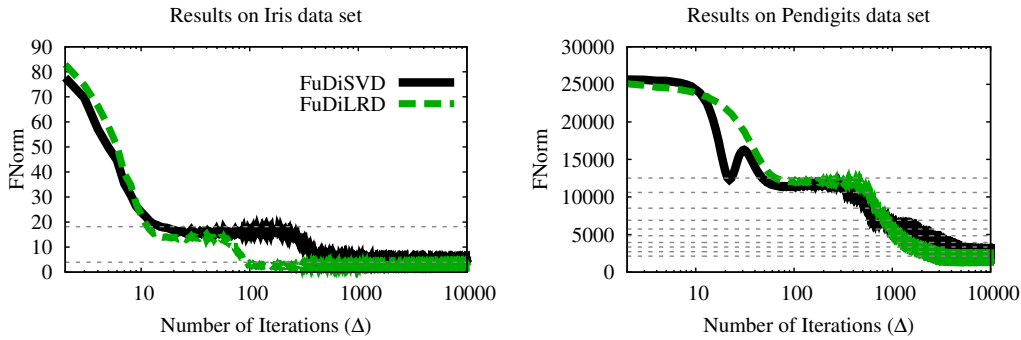
Ezen eredmény fontossága számos problémával igazolható, amelyek mátrixfelbontással oldhatók meg. Ezek például az ajánlórendszerek, ahol az A mátrix tartalmazza a felhasználók termékekre tett értékeléseiket, a feladat pedig a hiányzó értékek közelítése; dimenziócsökkentés, ahol az A sorai a tanítópéldák, az eredmény pedig a példák egy tömör reprezentációja; gráfklaszterezés, ahol az A a normalizált szomszédsági mátrix; és a Latent semantic indexing (LSI) módszer, amely segítségével kulcsszavak nyerhetők ki szöveges dokumentumokból.

Algoritmus 2 P2P alacsony rangú felbontás az i eszközön

<pre> 1: a_i 2: initialize Y 3: initialize x_i 4: loop 5: wait(Δ) 6: $p \leftarrow \text{selectPeer}()$ 7: send Y to p 8: end loop 9: procedure ONRECEIVE$Y(\tilde{Y})$ 10: $Y \leftarrow \tilde{Y}$ 11: $(Y, x_i) \leftarrow \text{update}(Y, x_i, a_i)$ 12: end procedure </pre>	<pre> 13: η 14: procedure UPDATE(Y, x_i, a_i) 15: $\text{err} \leftarrow a_i - x_i Y^T$ 16: $x'_i \leftarrow x_i + \eta \cdot \text{err} \cdot Y$ 17: $Y' \leftarrow Y + \eta \cdot \text{err}^T \cdot x_i$ 18: return (Y', x'_i) 19: end procedure </pre>
---	--

A 2. algoritmus mutatja a GOLF egy módosított változatát, amely megoldja az elosztott mátrixfaktorizációt. Az A és X mátrixok sorait csak azok az eszközök érik el, amelyek tárolják azt. Az Y mátrix pedig véletlen sétát ír le a hálózatban, és frissül a csomópontok által az X mátrix lokális

sorával együtt. Továbbá az Y mátrix egy-egy példánya lokálisan elérhető a hálózat összes eleme számára.



4. ábra. Konvergencia valós adatbázisokon. A hiba mértéke Frobenius normával van szemléltetve. A vízszintes szaggatott vonalak mutatják fentről lefelé haladva az FNORM értékét az optimális i rangú felbontásnak, $i = 1, \dots, k$

A tapasztalati kiértékelések alapján vizsgáltuk az algoritmus konvergenciájának sebességét (4. ábra), és bemutattuk, hogy versenyképes más gradiens alapú módszerekhez képest, amelyek ráadásul extra adathozzáférést igényelnek. Továbbá a módszer hálózati hibatűrő képessége is figyelemre méltó.

Fő eredmények:

- SGD alapú alacsony rangú matrixfelbontó módszer a GOLF keretben;
- Egy módszer, amely az SVD eredményéhez konvergál;
- Az érzékeny adatok nem hagyják el az eszközöket;
- Kapcsolódó publikáció: [9]

István Hegedűs, Márk Jelasity, Levente Kocsis, and András A. Benczúr. Fully distributed robust singular value decomposition. In *Proceedings of the 14th IEEE Fourteenth International Conference on Peer-to-Peer Computing (P2P)*, P2P'14. IEEE, 2014

Hivatkozások

- [1] Róbert Ormándi, István Hegedűs, and Márk Jelasity. Asynchronous peer-to-peer data mining with stochastic gradient descent. In *Proceedings of the 17th international conference on Parallel processing - Volume Part I, Euro-Par'11*, pages 528–540, Berlin, Heidelberg, 2011. Springer-Verlag.
- [2] István Hegedűs, Lehel Nyers, and Róbert Ormándi. Detecting concept drift in fully distributed environments. In *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics, SISY'12*, pages 183–188. IEEE, 2012.
- [3] István Hegedűs, Busa-Fekete Róbert, Ormándi Róbert, Jelasity Márk, and Kégl Balázs. Peer-to-peer multi-class boosting. In Christos Kaklamanis, Theodore Papatheodorou, and Paul Spirakis, editors, *Euro-Par 2012 Parallel Processing*, volume 7484 of *Lecture Notes in Computer Science*, pages 389–400. Springer Berlin / Heidelberg, 2012.
- [4] István Hegedűs, Ormándi Róbert, and Jelasity Márk. Gossip-based learning under drifting concepts in fully distributed networks. In *2012 IEEE Sixth International Conference on Self-Adaptive and Self-Organizing Systems, SASO'12*, pages 79–88. IEEE, 2012.
- [5] István Hegedűs, Róbert Ormándi, and Márk Jelasity. Massively distributed concept drift handling in large networks. *Advances in Complex Systems*, 16(4&5):1350021, 2013.

-
- [6] Róbert Ormándi, István Hegedűs, and Márk Jelasity. Gossip learning with linear models on fully distributed data. *Concurrency and Computation: Practice and Experience*, 25(4):556–571, 2013.
- [7] Balázs Szörényi, Róbert Busa-Fekete, István Hegedűs, Ormándi Róbert, Jelasity Márk, and Kégl Balázs. Gossip-based distributed stochastic bandit algorithms. In *Proceedings of The 30th International Conference on Machine Learning*, volume 28(3) of *ICML'13*, page 19–27. JMLR Workshop and Conference Proceedings, 2013.
- [8] Árpád Berta, István Hegedűs, and Róbert Ormándi. Lightning fast asynchronous distributed k-means clustering. In *22th European Symposium on Artificial Neural Networks, ESANN 2014*, pages 99–104, 2014.
- [9] István Hegedűs, Márk Jelasity, Levente Kocsis, and András A. Benczúr. Fully distributed robust singular value decomposition. In *Proceedings of the 14th IEEE Fourteenth International Conference on Peer-to-Peer Computing (P2P)*, P2P'14. IEEE, 2014.
- [10] Árpád Berta, István Hegedűs, and Márk Jelasity. Dimension reduction methods for collaborative mobile gossip learning. In *2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*, pages 393–397, Feb 2016.
- [11] István Hegedűs, Árpád Berta, Levente Kocsis, András A. Benczúr, and Márk Jelasity. Robust decentralized low-rank matrix decomposition. *ACM Trans. Intell. Syst. Technol.*, 7(4):62:1–62:24, May 2016.
- [12] István Hegedűs and Márk Jelasity. Distributed differentially private stochastic gradient descent: An empirical study. In *2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*, pages 566–573, Feb 2016.
- [13] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 2 edition, 1997.