Gaussian Mixture Modell alapú Fisher vektor számolás GPGPU-n

Daróczy Bálint és

András Benczúr, Bodzsár Erik, Petrás István, Siklósi Dávid, Nikházy László, Pethes Róbert

Adatbányászat és Webes Keresés Kutatócsoport

MTA SZTAKI

A csoport elsősorban egyedi, nagy méretű, különösen összetett rendszerek (nagy intranetek, nagy forgalmú tartalomszolgáltatók, publikus Web, tudományos adatbázisok) számára kínál webes keresési, ill. adatbányászati megoldásokat. https://dms.sztaki.hu

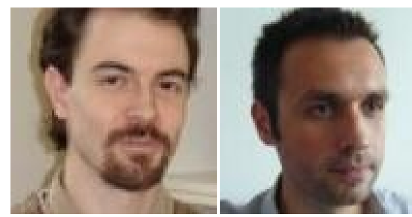
Kereső motorok

„Big Data” üzleti intelligencia

Laborvezető: Benczúr András

Gépi tanulás

Ajánló rendszerek

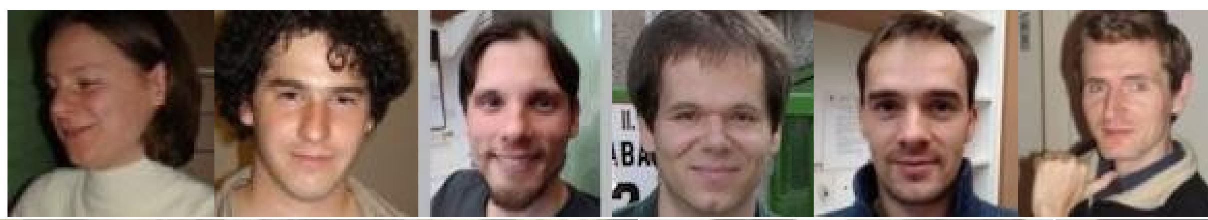


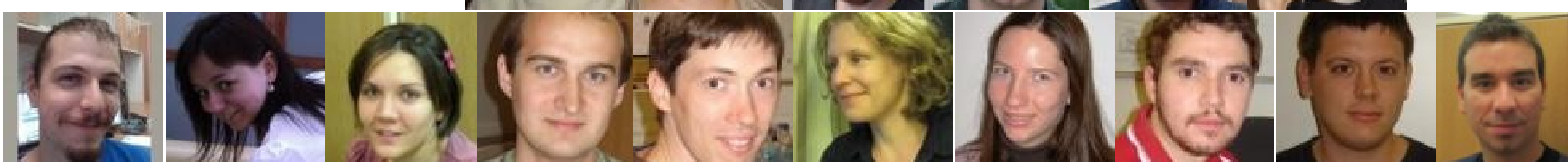
Gráf vizualizáció

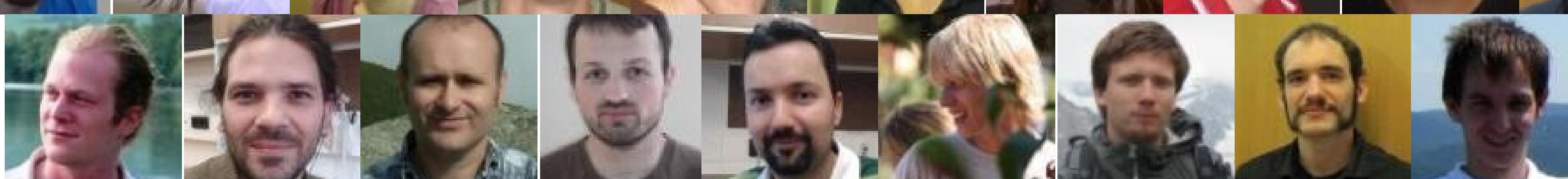
Azonosság feloldás

Bioinformatika

Web spam szűrés







GPGPU: General-purpose computing on graphics processing units

Miért jó számunkra:

- sok egyszerű, egyben lassú számoló egység (>100) → gyakorlatban fp32....
- kicsi cserébe nagyon gyors memória (>100GB/sec)

vs. CPU:

- kevés (<20), de komplex és gyors számoló egység
- arányaiban lassú memória (<25GB/sec)

Pl.: Nvidia Tesla c2075 : 448 CUDA cores with 6GB (144GB/sec)

Hatékony architektúra erősen párhuzamosítható algoritmusokhoz kis memória igénnyel (Nvidia Denver?):

- matrix műveletek, DFT, wavelet, rendezés, gráf alg. stb.
- titkosítás, tömörítés, stb.

Eddig adoptált Adatbányászati és képi algoritmusok:

- SVD számolás ajánló rendszerekhez: túl sok a memória mozgás!
- Képszegmentálás: optimális vágás megvalósítható nagyságrendben gyorsabban
- Képleírók készítése: Fisher vektor és GMM

Terveinkben:

- alacsony szintű leírók számolása: jelenleg ~10-15 sec képenként egy szálon (videó?)
- gépi tanulás: regresszió és Support Vector Machine, MKL
- Hasonlóság számolás: memória korlátok?

Képi annotáció feladata egy ismeretlen képről eldönteni, hogy melyek a rá jellemző fogalmak.

Pl. nappal, ég, hajó, tó (tenger?) stb.

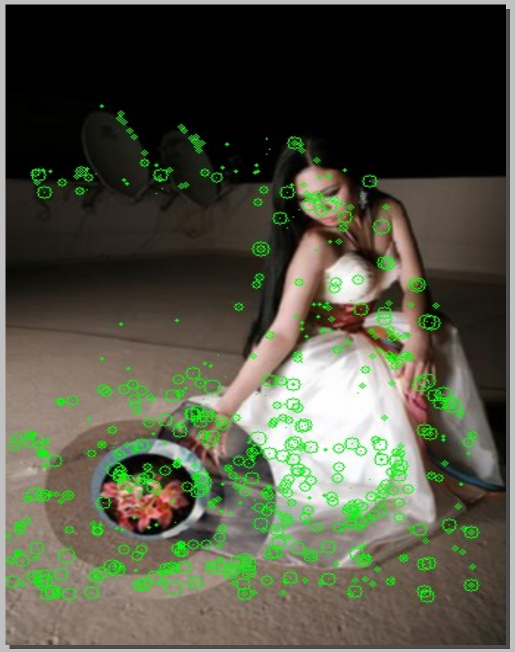
Adott:

- maga a kép
- a képhez szorosan kapcsolódó szöveges információk

  pl. Flickr, honlap, kép neve, exif stb.

Low-level feature-ök

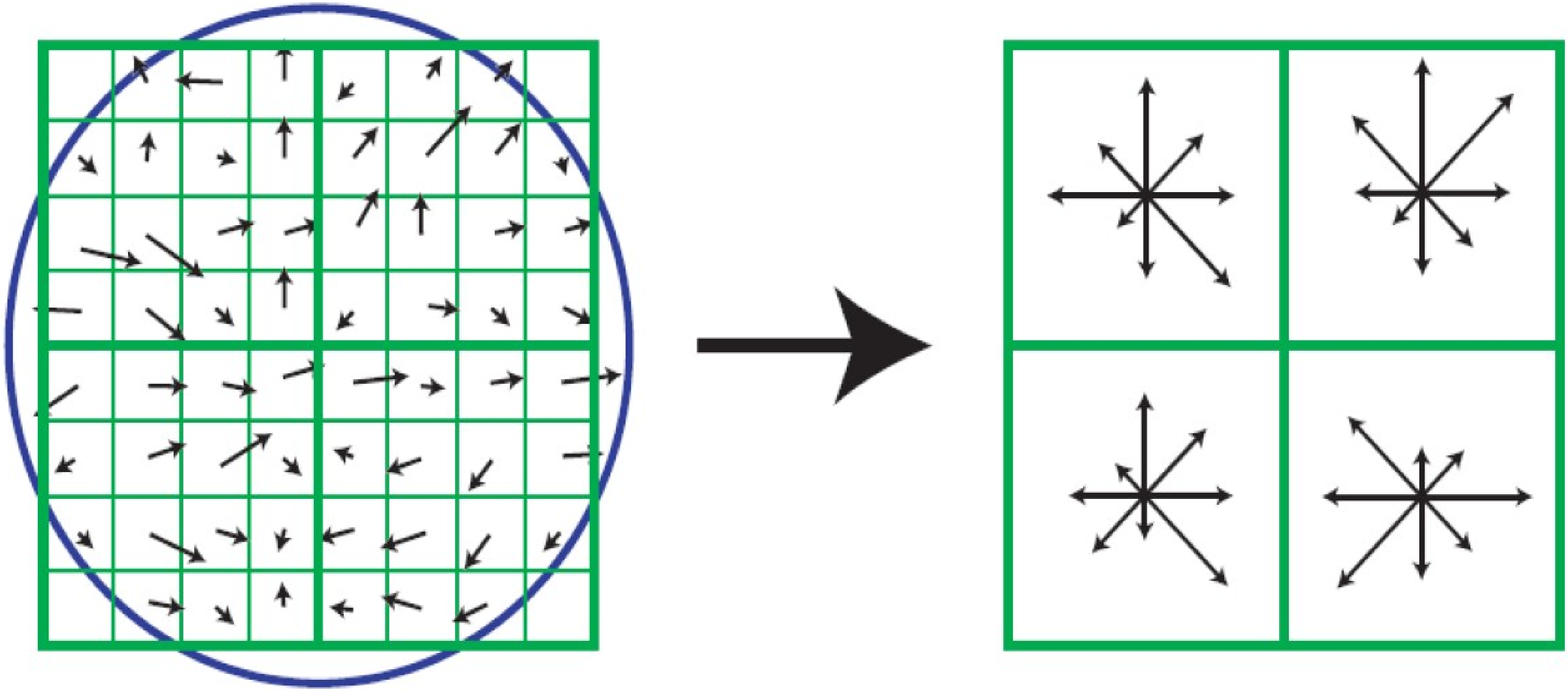
Szegmens alapú
~10..100



ROI alapú
~600..2000

Grid alapú
~5..30 ezer

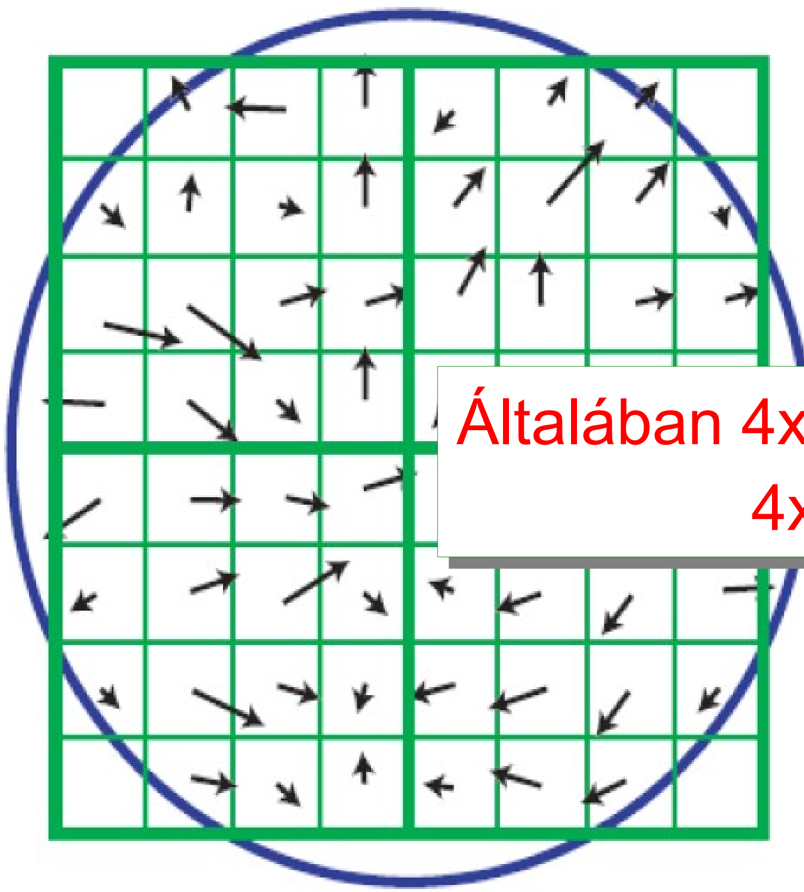
Pl. Scale-Invariant Feature Transform – SIFT [1]

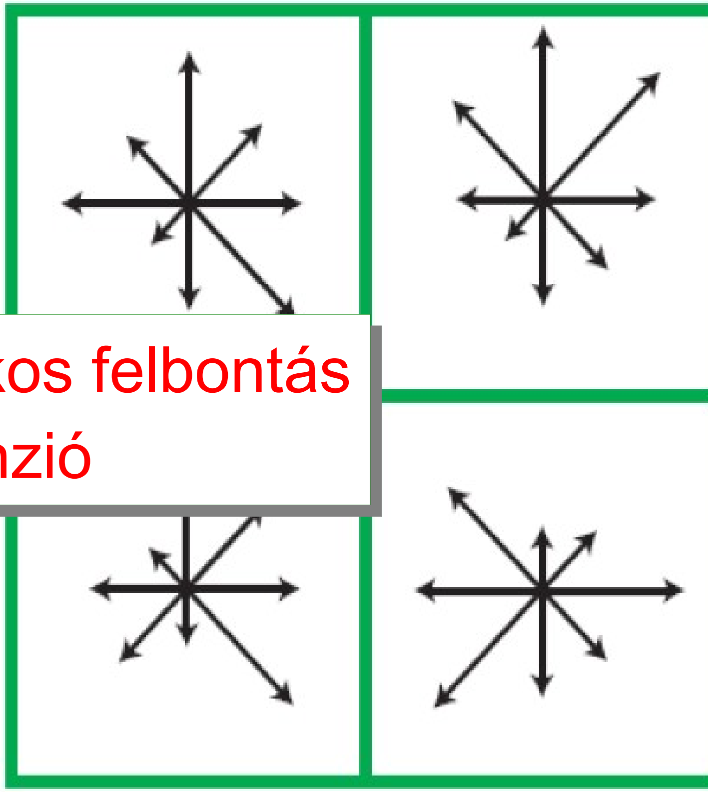


Kép orientációk (n=8 → 8x8)

m=2, 2x2-es sub-block orientáció hisztogram

Pl. Scale-Invariant Feature Transform – SIFT [1]





Általában 4x4 block és 45 fokos felbontás
4x4x8 = 128 dimenzió

Kép orientációk (n=8 → 8x8)

m=2, 2x2-es sub-block orientáció hisztogram

Gaussian Mixture Model

EM algoritmus

- nem hierarchikus ellentétben a gyakori megoldásokkal [2,3,6]
- nagy dimenziós térben könnyen alulcsordul
- nem megfelelő előkészítés esetén kiürülnek a klaszterek

Algorithm 1 The GPU GMM algorithm

Input: data points $\{x_i\}_{i=1}^{N}$, dimension D, mixture number K.

Output: $\{P_i\}_{i=1}^{K}$, $\{\mu_i\}_{i=1}^{K}$, $\{\sigma_i\}_{i=1}^{K}$, where $\mathcal{N}_i$ is normally distributed with parameters (μ_i, σ_i) and σ_i is assumed to be diagonal.

Initialize the Gaussian distributions with random parameters

repeat

for all n and k **do**

Expectation: Compute likelihood $p_{nk} = \frac{f_k(x_n)P_k}{\sum_{i=1}^{K} f_i(x_n)P_i}$ where f_i is the density of $\mathcal{N}_i$

for all k and d **do**

Maximization: compute $P_k = \frac{\sum_{n=1}^{N} p_{nk}}{N}$, $\mu_{kd} = \frac{\sum_{n=1}^{N} p_{nk} x_{nd}}{\sum_{n=1}^{N} p_{nk}}$ and $\sigma_{kd}^2 = \frac{\sum_{n=1}^{N} p_{nk}(x_{nd} - \mu_{kd})^2}{\sum_{n=1}^{N} p_{nk}}$

until until converge

GPU implementáció [6]

Numerikus hibák:

- Egy naiv megvalósítás gyakran alulcsordul a sűrűség meghatározásakor σ^2_{kd}
- Feloldható logaritmusok és úgynevezett „bucket”-ek használatával

$$\log(p_{nk}) = \log(f_k(x_n)) + \log(P_k) - \log \sum_{i=1}^{K} \exp(\log(f_i(x_n)) + \log(P_i))$$

$$\log(P_k) = \log \sum_{n=1}^{N} \exp\log(p_{nk}) - \log N$$

$$\mu_{kd} = \frac{\sum_{n=1}^{N} \exp(\log(p_{nk}) + \log(x_{nd}))}{\sum_{n=1}^{N} \exp\log(p_{nk})}$$

$$\sigma^2_{kd} = \frac{\sum_{n=1}^{N} \exp(\log(p_{nk}) + 2\log(x_{nd} - \mu_{kd}))}{\sum_{n=1}^{N} \exp\log(p_{nk})}$$

http://datamining.sztaki.hu/?q=en/GPU-GMM

Fisher vektor modellek

Legyen λ egy Gaussian Mixture Modell (k eloszlás):

$$\lambda = [p_{1,..}, p_k, \mu_{1,..}, \mu_k, \Sigma_{1,..}, \Sigma_k]$$

The idea is to describe the low-level features with the gradient of the log-likelihood of their conditional probability [2,3]:

$$F_X(\lambda) = \nabla_\lambda \log P(X|\lambda) = \frac{\partial L(X|\lambda)}{\partial \lambda}$$ ahol $$p(\boldsymbol{x_n}|\lambda) = \sum_{k=1}^{K} p_k g(\boldsymbol{x_n}; \boldsymbol{\mu_k}, \boldsymbol{\sigma_k})$$

$$L(X|\lambda) = \sum_{n=1}^{N} \log p(\boldsymbol{x_n}|\lambda)$$

$$\frac{\partial L(X|\lambda)}{\partial p_k} = \sum_{n=1}^{N} \left(\frac{p(k|\boldsymbol{x_n})}{p_k} - \frac{p(1|\boldsymbol{x_n})}{p_1} \right), \quad k = 2 \ldots K$$

$$\frac{\partial L(X|\lambda)}{\partial \mu_k^d} = \sum_{n=1}^{N} p(k|\boldsymbol{x_n}) \left(\frac{x_n^d - \mu_k^d}{(\sigma_k^d)^2} \right), \quad k = 1 \ldots K, d = 1 \ldots D$$

$$\frac{\partial L(X|\lambda)}{\partial \sigma_k^d} = \sum_{n=1}^{N} p(k|\boldsymbol{x_n}) \left(\frac{(x_n^d - \mu_k^d)^2}{(\sigma_k^d)^3} - \frac{1}{\sigma_k^d} \right), \quad k = 1 \ldots K, d = 1 \ldots D$$

Fisher vektor modellek

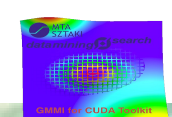
Probléma:

- Igazán hatékony GMM: legalább 256 eloszlás
- Túl sok számítás:
  - független számítások száma nagy
    - GPU
    - Kis adatmennyiség sok párhuzamos számítás (egyetlen kép alacsony szintű leírói: ~30MB)
  - egyetlen gradiensnél a Fisher vektor dimenziója N-1+2*D*N, ahol N az eloszlások száma, D az alacsony szintű leíró dimenziója:

    SIFT: 127+2*128*256 = 65,663 dimenzió

Ötletek Fisher vektor közelítésére[2,3]:

- diagonális kovariancia mátrix
- csak a legjobban illeszkedő eloszlásokat vesszük figyelembe, ebben az esetben a dimenzió K+2*K*D



http://datamining.sztaki.hu/?q=en/GPU-GMM

Fisher vektor modellek

Három példa:

→ ritka, gyors: approxFV csak a top x illeszkedő eloszlást vesszük figyelembe, cserébe túl ritka fp32 mellett [2,3]

→ Módosított Fisher információ mátrix a ritkaság feloldására: az eloszlások előfordulásával arányosan normálunk, nem a képen található elemek számával[2,3], „ha előfordult mennyire biztosan fordult elő „ locFV

$$f_{\omega_i} = T_i \frac{1}{\omega_i},$$

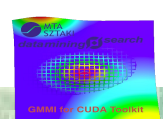
$$f_{\mu_i^d} = T_i \frac{\omega_i}{(\sigma_i^d)^2},$$

$$f_{\sigma_i^d} = 2T_i \frac{\omega_i}{(\sigma_i^d)^2},$$

→ Használjuk mindegyik eloszlást: sűrű Fisher vektor dFV (CUDA) ~44..70x gyorsulás GTX285 vs. Intel E5620 egy szálon

A GMM modell tapasztalatati alapján tudjuk számolni a feltételes valószínűségeket

- D*N független számítás
- Majd egy összegzés: már csak N elem



http://datamining.sztaki.hu/?q=en/GPU-GMM

Gépi tanulás

SVM[4], Fisher vektorokon értelmezett „természetes magfüggvény”[2]:

$$K_{Fisher}(X,X)=F_X F_{Information}^{-1} F_X = F_{Information}^{-1/2} F_X F_{Information}^{-1/2} F_X(\lambda) = F_{norm}(X,\lambda) F_{norm}(X,\lambda)$$

ahol $F_{Information}(\lambda)=E[\nabla_\lambda \log P(X|\lambda) \nabla_\lambda \log P(X|\lambda)^T]$

Multiple-Kernel probléma[4,5]:

Maximalizáljuk $L(\alpha_c,\beta_c)=\sum_{i=1}^{n}\alpha_{ci}-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ci}\alpha_{cj}y_{ci}y_{cj}\sum_{n=1}^{N}\beta_{cn}K_n(x_j,x_i)$

ahol $\sum_{i=1}^{n}\alpha_i y_i=0$ és $\forall\alpha_i\geq 0$

Unified representation[7,8]:

Kombináljuk a különböző modalitásokból származó információkat még a klasszifikáció előtt (pl.: szín, textúra, szöveges környezet, link adatok stb.):

- adott egy B referencia képhalmaz
- minden elemre értelmezzünk R reprezentációt

$$L_R(X,B)=[\sum_{r=1}^{R}\beta_r dist_r(X_r,B_{r1}),..,\sum_{r=1}^{R}\beta_r dist_r(X_r,B_{rb})]$$

Gépi tanulás

Számítási problémák:

- Képenként |B|*|R| reprezentáció adaptív hasonlóság számítás: túl sok számítás, túl nagy memória igénnyel

  Pl. szöveges hasonlóság

$$K_{k_{tag}}(X, I_t) = dist_{Jensen-Shannon}(X_{tags}, I_{t_{tags}}) = \frac{1}{2} D(P|M) + \frac{1}{2} D(Q|M)$$

- SVM modell kiértékelése: függ a tanulás után megmaradt support vektorok számától, sok esetben több ezer (memória)

- Bi-klaszterezés[8]: dokumentumok együttes csoportosítása többféle szöveges és /vagy képi hasonlóságok alapján

- Osztályonként különböző optimális arányok:
  - Súlyok meghatározása
  - Minden válaszhoz külön unified vektor?

Összefoglalás

Váltás:

- GPU vs. CPU : memória , PCIe sávszélesség, virtuális memória
- Nvida vs. AMD/Ati: OpenCL-re váltás a korábban használt GPU-khoz képest nem tűnik veszteségnek
- CPU vs. OpenCL: POSIX thread vs. OpenCL
- Cloud: új api-k?
- Mobil applikációk: kamerával felvett képinformációk előfeldolgozása vs. nyers adat átküldése

Megoldandó problémák:

- Alacsony szintű képfeldolgozás: alacsony memória igény, bizonyítottan működik pl. szegmentálás
- Koherencia miatt OpenCL
- Gépi tanulás gyorsítása (ritka adat?), vagy legalább a modell kiértékelése
- Mozgóképekre 40-50x-es gyorsulás lenne legalább szükséges
- Egyéb adatbányászati problémák: bi-klaszterezés [8], bioinformatika

Eredmények1: Yahoo MIRFLickr dataset: ImageCLEF 2011 Photo Annotation Task

			MAP	EER	AUC
	Egységes súly: dFV + Jensen-Shannon	IMG+TXT	**0.4512**	0.233511	0.8378
1.	TUBFI - Technische Universität Berlin + Frauenhofer	IMG+TXT	**0.443449**	0.233690	0.835753
2.	LIRIS - Universite de Lyon, CNRS, Ecole Centrale de Lyon	IMG+TXT	0.436968	0.232860	0.836570
3.	**BPACAD - SZTAKI**	IMG+TXT	**0.436294**	0.241691	0.827747
4.	ISIS - University of Amsterdam	IMG+TXT	0.432758	0.245721	0.821462
5.	MLKD - Department of Informatics, Aristotle University of Thessaloniki	IMG+TXT	0.401642	0.253232	0.817084
6.	CEALIST - Laboratoire d' Intégration des Systèmes et des Technologies	IMG+TXT	0.383528	0.250714	0.819797
7.	CAEN – University of Caen	IMG	0.382403	0.262823	0.805420
8.	MRIM - Laboratoire d'Informatique de Grenoble	IMG+TXT	0.377179	0.260416	0.809472
9.	IDMT – Fraunhofer	IMG+TXT	0.370975	0.303458	0.751788
...					

Eredmények2: Pascal VOC 2011 Classification

		Előre adott boxok használata nélkül	Előre megadott mintaboxok használata mellett
1.	NUS - National University of Singapore	-	0,7856
2.	NLPR - National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences	0,6108	0,7823
3.	ISIS - University of Amsterdam, University of Trento	-	0,7335
4.	MSR - Microsoft Research Asia & University of Science and Technology of China	-	0,7049
5.	LIRIS - LIRIS, Ecole Centrale de Lyon, CNRS, UMR5205, France	0,6094	0,6681
6.	**DMWS, MTA SZTAKI**	**0,6139**	**-**
7.	SJT - Shanghai Jiao Tong Univeristy	0,5485	0,604
8.	JDL - Institute of Computing Technology, Chinese Academy of Sciences	0,5478	-
9.	KUL - University of Leuven	0,4512	-
...			

MTA SZTAKI

ERCIM

Referenciák

1. D. Lowe: “Object recognition from local scale-invariant features”, in: International Conference on Computer Vision, volume 2, pp. 1150–1157.

2. F. Perronnin, C. Dance: “Fisher kernels on visual vocabularies for image categorization”, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR’07, pp. 1–8.

3. F. Perronnin, J. Sanchez, T. Mensink, “Improving the fisher kernel for large-scale image classification”, in: 11th ECCV,2010.

4. C. Cortes, V. Vapnik: “Support-vector networks”, Machine Learning 20 (1995) 273–297.

5. M. Kloft, U. Brefeld, S. Sonnenburg, A. Zien: “lp-norm multiple kernel learning”, Journal of Machine Learning Research 12 (2011) 953–997.

6. E. Bodzsár, B. Daróczy, I. Petrás, A. A. Benczúr: “GMM based fisher vector calculation on GPGPU” (2011). http://datamining.sztaki.hu/?q=en/GPU-GMM.

7. B. Daróczy, R. Pethes, A. Benczúr: „SZTAKI @ ImageCLEF 2011”, In Working Notes for ImageCLEF Workshop at CLEF Amsterdam, Nederland, 2011

8. Dávid Siklósi, Bálint Daróczy, András A. Benczúr: “Content-Based Trust and Bias Classification via Biclustering”, WWW'12 Lyon, France, Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality , Pages 41-47 ACM, New York, NY, USA ©2012 ISBN: 978-1-4503-1237-0

Köszönöm a figyelmet!