

A SUPPLEMENT

Algorithm 3 shows the details of how to sample a mini-batch using k -DPP [19] which is used for the DM-SGD and DM-SVI algorithm in the paper.

Algorithm 3 Mini-batch Sampling

Input: Mini-batch size k , eigendecomposition $\{(v_n, \lambda_n)\}_{n=1}^N$ of similarity matrix L .

Compute the elementary symmetric polynomials

$$e_0^n \leftarrow 1 \forall n \in \{0, 1, 2, \dots, N\}$$

$$e_l^n \leftarrow 1 \forall l \in \{1, 2, \dots, k\}$$

for $l = 1, 2, \dots, k$ **do**

for $n = 1, 2, \dots, N$ **do**

$$| \quad e_l^n \leftarrow e_l^{n-1} + \lambda_n e_{l-1}^{n-1}$$

end

end

for $t=1$ to Number of subset samples to generate **do**

Sampling k eigenvectors V with indices J

$$J \leftarrow \emptyset$$

$$l \leftarrow k$$

for $n = N, \dots, 2, 1$ **do**

if $l = 0$ **then**

 | break;

end

if $u \sim U[0, 1] \leq \lambda_n \frac{e_{l-1}^{n-1}}{e_l^n}$ **then**

$$| \quad J \leftarrow J \cup \{n\}$$

$$| \quad l \leftarrow l - 1$$

end

end

Sample k data points indexed by J using V .

$$V \leftarrow \{v_i\}_{i \in J}$$

$$Y \leftarrow \emptyset$$

while $|V| > 0$ **do**

 Select i with $Pr(i) = \frac{1}{|V|} \sum_{v \in V} (v^T e_i)^2$

$Y \leftarrow Y \cup i$ $V \leftarrow V_{\perp}$, an orthonormal basis for the subspace of V orthogonal to e_i

end

Output: Y

end

Table 2 and 3 show the top words using $K = 30$ for LDA using traditional SVI and our proposed DM-SVI respectively. We can see that the topics that are learned by DM-SVI are more diverse and rare topics such as grain (colored in blue) are captured.

Figure 10 shows the synthetic data that are used in the LDA experiment. Each row represents a document and each column represents a word.

The sampling time in seconds for the R8 dataset is listed in Table 4. There are 5485 training documents. The first row in the table shows the sampling time for different mini-batch sizes k and different versions of k -DPP sampling. In

Topic 1	pct shares stake and group investment securities stock commission firm
Topic 2	year pct and for last lower growth debt profits company
Topic 3	and merger for will approval companies corp acquire into letter
Topic 4	and for canadian company management pacific bid southern court units
Topic 5	baker official and that treasury western policy administration study budget
Topic 6	and president for executive chief shares plc company chairman cyclops
Topic 7	bank pct banks rate rates money interest and reuter today
Topic 8	and unit inc sale sell reuter company systems corp terms
Topic 9	mln stg and reuter months year for plc market pretax
Topic 10	and national loan federal savings reuter association insurance estate real
Topic 11	trade and for bill not united imports that surplus south
Topic 12	and february for china january gulf issue month that last
Topic 13	market dollar that had and will exchange system currency west
Topic 14	dhrs quarter share for company earnings year per and fiscal
Topic 15	billion mln tax year profit credit marks francs net pct
Topic 16	usair inc twa reuter trust air department chemical diluted piedmont
Topic 17	and will union spokesman not two that reuter security port
Topic 18	offer share tender shares that general and gen-corp dhrs not
Topic 19	and company for that board proposal group made directors proposed
Topic 20	that japan japanese and world industry government for told officials
Topic 21	american analysts and that analyst chrysler shear-son express stock not
Topic 22	loss profit mln reuter cts net shr dhrs qtr year
Topic 23	mln dhrs and assets for dlr operations year charge reuter
Topic 24	mln net cts shr revs dhrs qtr year oper reuter
Topic 25	cts april reuter div pay prior record qtrly march sets
Topic 26	dividend stock split for two reuter march payable record april
Topic 27	oil and prices crude for energy opec petroleum production bpd
Topic 28	agreement for development and years program technology reuter conditions agreed
Topic 29	and foreign that talks for international industrial exchange not since
Topic 30	corp inc acquisition will company common shares reuter stock purchase

Table 2: Top 10 words for each topics learned from LDA with traditional SVI.

Topic 1	oil and that prices for petroleum dlrs energy crude field
Topic 2	pct and that rate market banks term rates this will
Topic 3	billion and pct mln group marks sales year capital rose
Topic 4	and saudi oil gulf that arabia december minister prices for
Topic 5	and dlrs debt for brazil southern mln will medical had
Topic 6	and grain that will futures for program farm certificates agriculture
Topic 7	bank banks rate and pct interest rates for foreign banking
Topic 8	and union for national seamen california port security that strike
Topic 9	and trade that for dollar deficit gatt not exports economic
Topic 10	and financial for sale inc services reuter systems agreement assets
Topic 11	dollar and for yen mark march that dealers sterling market
Topic 12	and for south unit equipment reuter two will state corp
Topic 13	and firm stock company will for pct not share that
Topic 14	and world that talks economic official for countries system monetary
Topic 15	and gencorp for offer general company partners that dlrs share
Topic 16	mln canada canadian stg and pct will air that royal
Topic 17	usair and twa that analysts not for pct analyst piedmont
Topic 18	and that for companies not years study this areas overseas
Topic 19	trade and bill for house that reagan foreign states committee
Topic 20	company dlrs offer stock and for corp share shares mln
Topic 21	dlrs year and quarter company for earnings will tax share
Topic 22	mln cts net loss dlrs profit reuter shr year qtr
Topic 23	exchange paris and rates that treasury baker allied for western
Topic 24	and shares inc for group dlrs pct offer reuter share
Topic 25	merger and that pacific texas hughes baker commerce for company
Topic 26	and american company subsidiary china french reuter pct for owned
Topic 27	japan japanese and that trade officials for government industry pact
Topic 28	oil opec mln bpd prices production ecuador and output crude
Topic 29	and that had shares block for mln government not san
Topic 30	mln pct and profits dlrs year for billion company will

Table 3: Top 10 words for each topics learned from LDA with DM-SVI.

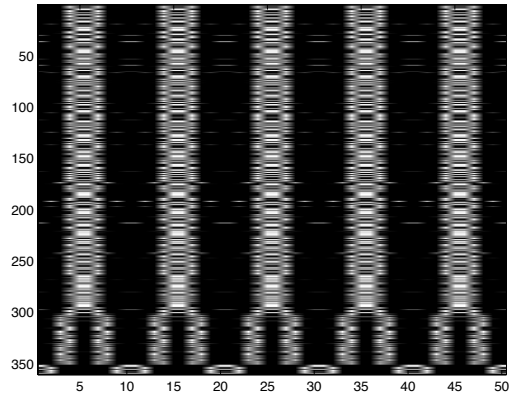


Figure 10: Synthetic data used to train the LDA model in the main paper. Each row presents a document and each column represents a word. Documents share topics with highly imbalanced proportions.

Size	k = 10	k =30	k=50	k=80
Fast k-DPP	0.001	0.0139	0.0541	0.2199
k-DPP	0.0098	0.1468	0.6438	2.6698
LDA	0.8777	1.2530	1.6414	2.2312

Table 4: Sampling time (in sec) for LDA on the R8 dataset with different mini-batch sizes.

practice, we use the original implementation from [23] with $M = 100$. To compare with the traditional k-DPP, we listed the elapsed time with [19]. The last row shows the running time per local LDA update, excluding sampling.

The computational time for training a neural network highly depends on the network structure and implementation details. For example, when using only one softmax layer as in the flower experiment, the cost per gradient step is in the milliseconds. In this setup, k-DPP is not effective from a runtime perspective, but still results in better final classification accuracies. However, the cost for each gradient step for a simple 5 layer NN as in the MNIST experiment with $K = 100$ is 1.294 seconds. In the latter case, this time is comparable to k-DPP sampling (0.7941 sec) see Table 5. We thus expect our methods to benefit expensive models and imbalanced training datasets more.

Figure 11 shows the bar plots of the frequency of images in each class for Oxford Flower dataset using the number of classes as the mini-batch size. With this setting, we can see that when $w = 1$, DM-SGD is equivalent to StS.

Size	k = 10	k =100	k=200
Fast k-DPP	0.0012	0.7941	5.4216
NN cost	0.166948	1.29452	2.64811

Table 5: Five Layer NN trained on MNIST with different mini-batch sizes. Top row: sampling time (in sec) using the fast k-DPP approach. Bottom row: run time for each update step (excluding mini-batch sampling).

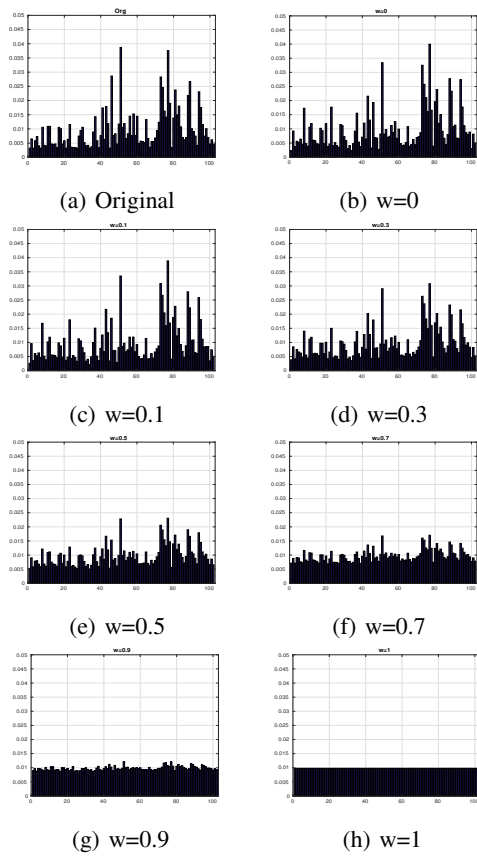


Figure 11: The frequency of images in each class for Oxford Flower dataset, with $k = 102$.