CSE291 Report: Topic Discovery using Nonnegative Matrix Factorization under Separability assumption

Chicheng Zhang

June 14, 2013

1 Problem Setting

1.1 Topic Discovery: Generative Model

Throughout the report, we assume the following generative model of corpus: Suppose the topic matrix A is fixed, each column of A is a topic(sports, technology, etc), i.e. a discrete distribution over a n-word vocabulary. for each document $D_i, i = 1, \ldots, m$, the topic proportion $\theta_i \sim T$, T is a distribution over Δ^{k-1}, e.g. $\mathbf{Dir}(\alpha)$. Then the j^{th} word w_{ij} are generated by a categorical ditribution $\mathbf{Cat}(A\theta_i)$, $j = 1, \ldots, N_i$, where N_i is the number of words in document i, i.e. $w_{ij} = e_k$ if w_{ij} is the k^{th} word in vocabulary. Our goal is:

Given a corpus $C = \{D_i, i = 1, \ldots, m\}$, recover the topic matrix A.

1.2 Tool: (Approximate)Nonnegative Matrix Factorization

Suppose a matrix $M \in \mathbb{R}^{a\times b}$, $M \geq 0$ entrywise, assume it can be factorized as $M = FW$, $F \in \mathbb{R}^{a\times r}$, $W \in R^{r\times b}$, $F, W \geq 0$ entrywise. Given a corrupted version of M, i.e. $\hat{M}$, each row $|\hat{M}^i - M^i| \leq \epsilon$, our goal is to find $\hat{F}, \hat{W} \geq 0$ such that $\forall i, j, (\hat{F} - F)_{ij}, (\hat{W} - W)_{ij} \leq f(\epsilon)$ for some function f. As we shall see next, the technique of solving nonnegative matrix factorization can be used in topic discovery problem.

1.3 Separability Assumption

Given $M = FW$, separability assumption says for each column i of F, there is a row $F^{f(i)}$ such that it is nonzero only in column i. Moreover we call it p-separable if the entry $F_{f(i),i}$ is greater than p. From NMF's perspective, if separability is not assumed, the factorization solution can be not unique. Further discussion of the assumption is provided in subsection 5.1.

1.4 Which Matrix to Factorize?

Given the corpus C, there are two natural choice of factorization:

The first choice is $\hat{M}_1 \in \mathbb{R}^{n\times m}$, whose row i containing the average word frequency of document i, i.e. $\sum_j w_{ij}/N_i$. It can be seen that $\mathbb{E}(M_1|\theta_1, \ldots, \theta_m) = A(\theta_1, \ldots, \theta_m)$. we can recover A based on $\hat{M}_1$. Unfortunately this direct approach suffers from noise too much, since if N_i are small, each entry of $\hat{M}_1$ is also noisy. As $m \to \infty$, it cannot guarantee consistency of estimation of A.

The second choice is the Gram matrix $\hat{M}_2 = \hat{\mathbb{E}}(w_1 w_2^T) \in \mathbb{R}^{n\times n}$, which characterizes word-word correlation within a document. As the number of document increases, the empirical estimate $\hat{\mathbb{E}}(w_1 w_2^T) \to \mathbb{E}(w_1 w_2^T)$. Note in this approach as long as $N_i \geq 2$, if $m \to \infty$, we are still able to recover A. Under the generative model defined above, $\mathbb{E}(w_1 w_2^T) = ARA^T$, where $R = \mathbb{E}(\theta_i \theta_i^T)$, Taking $M = ARA^T$, $F = A \in \mathbb{R}^{n\times k}$, $W = RA^T \in \mathbb{R}^{k\times n}$ in subsection 1.2, the topic discovery problem reduces to recover A with access to a noisy version of M, assuming A is separable. It is argued that the separabilty assumption holds in real corpus:

for example, "kobe_bryant" is likely to appear only in topic of basketball, "bunt" is likely to appear only in topic of baseball, etc, and these are often called "anchor" words.

2 Literature Overview

2.1 NMF under Separability Assumption

Recently [3] open a new avenue of NMF, relying on separability assumption. The intuition is described as follows. Observe that under separability assumption, after normalizing each row of M, the convex hull of rows is simple – a r-dimensional simplex. Then we can recover F and W approximately using a two-stage algorithm: first find the vertices of the convex hull of rows based on linear programming(finding robust loners), then they are the perturbed version of a permutation of rows of W. Then solve the constrained optimization problem $\min_{\hat{F}} ||\hat{M} - \hat{F}\hat{W}||_1$ to obtain $\hat{F}$. [9] solve the problem based on localizing factorization, using a single linear program involving $n \times n$ variables, which they argue can be solve fastly by incremental gradient descent. [8] proposed some heuristic to find the vertices, based on unnormalized M, and found extreme rays of M based on repeat detection and projection.

2.2 Application to Topic discovery

[4, 2] tailored the NMF algorithm to topic discovery problem. [4] improved upon the previous work using more delicate definition of robust loner, and used matrix inversion to recover A and R. [2] presents a combinatorial algorithm to find vertices, hence finding W, then use Bayes' Theorem to recover A, which is more robust than the inversion approach. Our implementation is mainly based on this line of work. [6] proposed to find anchor rows using data dependent projection/random projection, but relies on an unnatural quantity $\beta_\wedge$ (the lower bounds of non-zero elements of A).

3 Implementation

In our implemetation we only conducted experiments on [2], since it has both solid theoretical guarantees and practical. To construct the empirical cross word moment matrix we use:

$$\hat{M} = \hat{\mathbb{E}}(w_1 w_2^T) = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{|N_i|(|N_i|-1)}(c_i c_i^T - \mathbf{diag}(c_i))$$

where $c_i = \sum_{j=1}^{N_i} w_{i,j}$ is the count of words in document i. It can be seen that

$$\hat{M} \xrightarrow{p} \mathbb{E}(w_1 w_2^T) = \mathbb{E}(\mathbb{E}(w_1 w_2^T|\theta)) = A\mathbb{E}(\theta\theta^T)A^T = A \cdot RA^T = A \cdot W$$

3.1 Finding Anchor Rows

Suppose we get the true $M = AW$, we normalize each row of M,A,W; denote $\bar{M} = \mathbf{diag}(M\mathbf{1})^{-1}M$, $\bar{A} = \mathbf{diag}(M\mathbf{1})^{-1}A\mathbf{diag}(W\mathbf{1})$, $\bar{W} = \mathbf{diag}(W\mathbf{1})^{-1}W$ as the row-normalized version respectively. A simple observation yields that $\bar{M} = \bar{A}\bar{W}$. The row $f(i)$ corresponding anchor words for topic i, of $\bar{A}$ is (0,...,0,1,0,...,0), so the row $f(i)$ of $\bar{M}$ stores exactly a copy of $\bar{W}^i$. Moreover, non-anchor rows are within the convex hull of $\bar{W}^i$, since $\forall j$, $\bar{M}^j = \sum_{i=1}^{k} \bar{a}_{ij}\bar{W}^j$. In other words, the rows of $\bar{M}$ lies within a simplex with vertices $\bar{M}^{f(i)}, i = 1, \ldots, k$. Moreover if the A is p-separable and R's ℓ_1-condition number is lower bounded by γ, then $W = RA^T$ is $O(\gamma p)$ robustly simplicial, i.e. the ℓ_2 distance of any $\bar{W}^i$ to the convex hull of $\{\bar{W}^j : j \neq i\}$ is at least γ. This leads to the robust anchor row finding algorithm [2]:

Algorithm 1 Robust Anchor Row Finding

Inputs: Gram matrix $\bar{M}$.
Outputs: Approximate anchor rows $\{v'_1, \ldots, v'_k\}$.
$S = \{M^i\}$ such that $\bar{M}^i$ is the farthest point from the origin
for $i = 1$ to $k - 1$ **do**
  Let M^i be the point in $\{\bar{M}^1, \ldots, \bar{M}^n\}$ that has the largest distance to **span**(S)
  $S = S \cup \{\bar{M}^i\}$
end for
$S = \{v'_1, \ldots, v'_k\}$
for $i = 1$ to k **do**
  Let $\bar{M}^j$ be the point that has the largest distance to **span**$(\{v'_1, \ldots, v'_k\} \backslash \{v'_i\})$
  update v'_i to $\bar{M}^j$
end for
Return $\{v'_1, \ldots, v'_k\}$

3.2 Recover Topic Matrix Using Bayes' Theorem

The row normalization has a probablistic interpretation: suppose originally $M_{ij} = \Pr(w_1 = i, w_2 = j)$, then

$$\bar{M}_{ij} = \Pr(w_2 = j | w_1 = i) = \sum_k \Pr(w_2 = j | z_1 = k) \Pr(z_1 = k | w_1 = i)$$

If row $f(k)$ is an anchor row for topic k, which means $\Pr(w_1 = f(k)|z_1 = k) > 0$, while $\Pr(w_1 = f(k)|z_1 = l) = 0, l \neq k$, then it implies $\Pr(z_1 = k|w_1 = f(k)) = 1$. So picking the rows $f(k)$, $\bar{M}_{f(k)j} = \Pr(w_2 = j|w_1 = f(k)) = \Pr(w_2 = j|z_1 = k)$. Hence we can get $\Pr(z_1 = k|w_1 = i)$ for all i, since we have an overdetermined equation:

$$\bar{M}_i = \sum_k \Pr(z_1 = k|w_1 = i) \bar{M}_{f(k)}$$

which has n variables and k equations. In implementation we solve the least squares problem, as well as ensuring positiveness and normalization of $\Pr(z_1 = k|w_1 = i)$(denote it as C_{ik}), i.e.

$$\begin{aligned} &\min ||\bar{M}_i - C_i^T (\bar{M}_{f(1)}, \ldots, \bar{M}_{f(k)})^T||^2 \\ \text{subject to} &C_i \geq 0 \\ &\sum_{i=1}^{k} C_i = 1 \end{aligned}$$

Finally using Bayes' theorem, the topic matrix A can be recovered:

$$A_{ik} = \Pr(w_1 = i|z_1 = k) = \frac{\Pr(z_1 = k|w_1 = i) \Pr(w_1 = i)}{\sum_k \Pr(z_1 = k|w_1 = i) \Pr(w_1 = i)}$$

where $\Pr(w_1 = i) = \sum_j \Pr(w_1 = i, w_2 = j)$ can be directly obtained in M.

4 Experiments

4.1 Real Coupus

We conducted experiments on New York Times as well as NIPS abstracts corpus. The illustrative results(top words from extracted topics) is provided in appendix A.

4.2 Semi-Synthetic Datasets: Sample Complexity

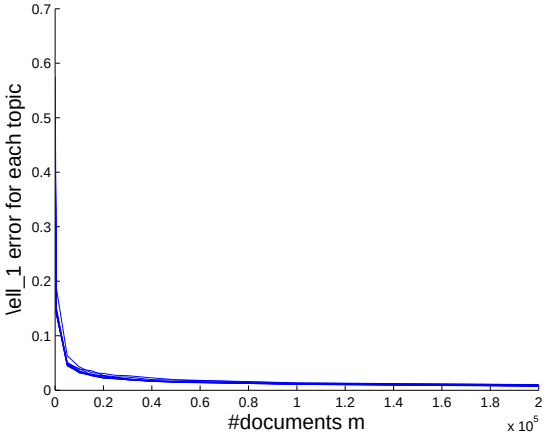
We validate the algorithm's sample complexity bound on synthetic corpus. We use a topic matrix A extracted from New York Times corpus $\in \mathbb{R}^{n\times k}$ and use i.i.d. $\theta_i \sim \mathbf{Dir}(0.1,\ldots,0.1)$ to generate topic distribution for each synthetic document. Then for each document i we sample N_i words i.i.d. from $\mathbf{Cat}(A\theta_i)$. Here we use $k = 10, n = 1000$. We set $N_i \equiv 100$, and vary m, then validate the ℓ_1 error of each topic; for simplicity, for each ground truth topic A_i, we find the nearest ℓ_1 neighbor $\hat{A}_j$ and test the value of $||A_i - \hat{A}_j||_1$.

Theorem 1 (Sample Complexity[2]). *Suppose m, n, k, γ, p is defined as above, $N_i \equiv N$, $a = \max_{i,j} \mathbb{E}(\theta_j)/\mathbb{E}(\theta_i)$ is the topic imbalance, then $\forall \epsilon > 0$*

$$m \geq O(\frac{ak^3 \log n}{N(\gamma p)^6 \epsilon}), O(\frac{(ak)^3 \log n}{N(\gamma p)^4 \epsilon^3})$$

the algorithm learns A matrix with entry wise error at most ϵ.

Since in our ground truth A is not perfectly separable, we do not expect the error dropping to 0 as $m \to \infty$. The ℓ_1 error of all 10 topics extracted wrt varying m are depicted as follows:



5 Discussion

5.1 Is Separability Assumption Reasonable?

[7] argues that separability assumption holds commonly in image segmentation, and [3] also claims it is met in topic modelling, moreover there can be several anchor words for one specific topic. But in some cases it is not quite natural: suppose we expect to extract a "overview" topic, whose support spreads widely over words, since there may be non-negligible fraction of survey articles in the corpus. Then under this assumption this is impossible since it does not have its anchor word. [1]

Another reason of proposing separablity may be from multiple solutions of the original problem. Suppose $M \in \mathbb{R}^{n\times m}$ can be factorized as products of three nonnegative matrices: $M = ARW$, $A \in \mathbb{R}^{n\times r}, R \in \mathbb{R}^{r\times r}, W \in \mathbb{R}^{r\times m}$, then obviously the rank r nonnegative factorization is not unique: either $AR \cdot W$ or $A \cdot RW$.

[1] Thank Akshay Balsubramani for pointing this out.

5.2 Comparison with SVD Approach

Note a big difference between [3] and [2] and traditional approach [5] is that it treats topic discovery a recovery problem: assuming the data are generated from a model, then recover the parameters of the model approximately, while traditional approach focus on building a model, then find the parameters that best fits the training examples, such as MAP estimate. Another recent work [1] share the principle with the former, using combination of second order moments and third order moments. For constructing their moments, it must assume the topic combination of each documents satisfies some distributional assumptions, such as Dirichlet, which is a stronger assumption, hence it is not surprising that their model may not have so good performance when topics have strong correlation. On the other hand, it does not require the existence of anchor words of each topic.

References

[1] Animashree Anandkumar, Dean P. Foster, Daniel Hsu, Sham M. Kakade, and Yi-Kai Liu. Two svds suffice: Spectral decompositions for probabilistic topic modeling and latent dirichlet allocation. *CoRR*, abs/1204.6703, 2012.

[2] Sanjeev Arora, Rong Ge, Yoni Halpern, David M. Mimno, Ankur Moitra, David Sontag, Yichen Wu, and Michael Zhu. A practical algorithm for topic modeling with provable guarantees. In *ICML*, 2013.

[3] Sanjeev Arora, Rong Ge, Ravindran Kannan, and Ankur Moitra. Computing a nonnegative matrix factorization - provably. In *STOC*, pages 145–162, 2012.

[4] Sanjeev Arora, Rong Ge, and Ankur Moitra. Learning topic models - going beyond svd. In *FOCS*, pages 1–10, 2012.

[5] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[6] Weicong Ding, Mohammad H. Rohban, Prakash Ishwar, and Venkatesh Saligrama. Topic discovery through data dependent and random projections. In *ICML*, 2013.

[7] David L. Donoho and Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts? In *NIPS*, 2003.

[8] Abhishek Kumar, Vikas Sindhwani, and Prabhanjan Kambadur. Fast conical hull algorithms for near-separable non-negative matrix factorization. In *ICML*, 2013.

[9] Ben Recht, Christopher Re, Joel A. Tropp, and Victor Bittorf. Factoring nonnegative matrices with linear programs. In *NIPS*, pages 1223–1231, 2012.

A Illustrative Results

Here we present experimental results on anchor words and the top 10 entry in each topic, i.e. each column of A recovered. New York Times Corpus:

A.1 New York Times

anchor words	top 10 words
zzz_tiger_wood	zzz_tiger_wood shot player tour play round win tournament hole course
zzz_kobe_bryant	point zzz_kobe_bryant zzz_laker shot half quarter lead left scored team
cat	cat park animal county owner group restaurant problem food police
teaspoon	cup teaspoon minutes tablespoon pepper add sugar pan chicken serving
stem	cell stem human research blood scientist egg system skin patient
zzz_microsoft	zzz_microsoft company window software computer system court operating case product
zzz_israeli	zzz_israeli palestinian zzz_israel attack israelis killed zzz_west_bank soldier army military
zzz_john_mccain	zzz_john_mccain campaign zzz_george_bush republican zzz_bush political zzz_party primary zzz_republican voter
zzz_enron	zzz_enron company lay energy stock board companies partnership fund trading
zzz_met	zzz_met team player fan season million baseball night agent free
touchdown	yard game play touchdown season point defense quarter quarterback ball
zzz_governor_bush	zzz_governor_bush zzz_new_york point plan meeting lead anti record attack zzz_texas
recount	ballot election zzz_florida recount votes vote court board law voter
zzz_fed	percent economy rates zzz_fed market rate prices interest economic cut
zzz_dodger	zzz_dodger season right manager games play start left game home
priest	priest abuse church sexual allegation victim official zzz_boston children child
butter	butter food add cup flavor chopped chef minutes green makes
anthrax	anthrax mail letter building official office worker found test investigator
king	king goal game team play games shot period point minutes
gay	gay right parent women group family children member mother child
wine	wine red white french list restaurant taste expert www room
zzz_taliban	zzz_taliban zzz_afghanistan official zzz_u_s military forces government afghan war bin
beer	beer chicken skin hour body hand right wood half inside
zzz_at	company companies zzz_at percent cable customer billion business market plan
zzz_usc	team zzz_usc game player coach season guy play defense games
zzz_vicente_fox	zzz_mexico president zzz_united_states government zzz_u_s official program zzz_clinton drug zzz_vicente_fox
zzz_medicare	zzz_bush drug program plan zzz_medicare bill benefit zzz_congress billion cost
teacher	school teacher student program percent children high parent kid education
zzz_vladimir_putin	zzz_russia official zzz_russian zzz_vladimir_putin president government russian military system zzz_moscow
yankees	yankees season team million player zzz_new_york games baseball manager pitcher
chocolate	chocolate food room find look help com book feel hand
zzz_arthur_andersen	company case lawyer zzz_arthur_andersen money firm government investor statement told
zzz_rudolph_giuliani	zzz_rudolph_giuliani zzz_new_york official mayor children president public police decision court
zzz_ford	computer zzz_ford car program company problem official hour high speed
salt	salt children side left wife food husband miles eat pound
cuban	zzz_cuba father zzz_miami boy official cuban family government zzz_united_states son
zzz_eastern	com information question www american daily newspaper web zzz_eastern sport
virus	virus program system computer human window disease mail zzz_new_york problem
zzz_china	zzz_china chinese zzz_united_states zzz_american zzz_u_s military government zzz_beijing official zzz_japan
penalty	death penalty case court law lawyer zzz_texas execution cases right
ranger	ranger season game goal right play games left night point
prayer	school prayer student zzz_god high religious public com faith site
album	music song album band show record play hit rock musical
zzz_o_neal	game zzz_laker zzz_o_neal play games season point player team night
zzz_bruin	zzz_bruin play zzz_ucla point game season half left right shot
nuclear	nuclear weapon terrorist power attack plant zzz_bush group bomb scientist
zzz_red_sox	zzz_red_sox team season manager zzz_boston big baseball hit guy fan
jet	jet pick round team draft trade zzz_miami official zzz_new_york plane
zzz_abc	show zzz_abc network program zzz_nbc television zzz_cb night series executive
zzz_oscar	film movie zzz_oscar actor show award movies won fight director

anchor words	top 10 words
zzz_government	zzz_government official right law police security officer court case evidence
zzz_gore	zzz_bush zzz_gore president campaign zzz_white_house million zzz_republican republican democratic percent
firefighter	firefighter building fire zzz_new_york family official police officer home found
zzz_aol	web com zzz_aol mail online zzz_internet site program internet www
index	percent stock market companies fund point index investor average high
zzz_navy	zzz_navy military room computer system information ship civilian water program
zzz_india	zzz_india zzz_pakistan group attack official government zzz_united_states terrorist million military
zzz_mccain	campaign zzz_bush zzz_mccain zzz_republican republican zzz_senate bill vote political voter
zzz_palestinian	official zzz_palestinian security plan officer forces police zzz_cia proposal settlement
medal	zzz_olympic medal women run won win gold games zzz_u_s final
fish	fish water river word small put scientist home boat www
zzz_iraq	zzz_iraq zzz_bush zzz_u_s war zzz_united_states zzz_american oil military weapon attack
zzz_black	percent zzz_black black white film group american student part look
privacy	information government companies privacy law company mail web computer zzz_internet
zzz_ncaa	team season game player point games coach tournament play zzz_ncaa
racing	race car driver team won racing track season win start
bird	bird light building zzz_new_york million book water look night winter
starring	film movie minutes hour play starring director character movies actor
cancer	women patient cancer drug test doctor percent treatment study disease
airline	flight passenger airline percent airport security ticket government business airlines
tax	tax percent cut million taxes income money plan pay spending
zzz_al_gore	zzz_al_gore campaign zzz_george_bush president democratic presidential voter zzz_clinton zzz_texas vice
zzz_giant	team player season zzz_giant coach play league games zzz_nfl guy
union	union worker government member labor percent official job leader president
rocket	team rocket game point season high play shot plan million
auction	site million auction company web house zzz_internet com money stock
italian	italian restaurant zzz_new_york food show american zzz_italy part century owner
soccer	player team play game soccer goal zzz_u_s games women fan
con	company companies deal customer power agreement player president statement business
cent	company million percent quarter cent analyst share billion sales price
zzz_dick_cheney	president zzz_dick_cheney zzz_george_bush administration zzz_bush zzz_white_house military republican zzz_clinton
gun	gun law officer police children court right case control violence
flag	flag white red black blue vote member right bill american
zzz_yasser_arafat	palestinian zzz_yasser_arafat zzz_israel peace leader israeli zzz_bush violence government arab
protein	protein human memory brain problem companies scientist food research drug
jew	jew jewish zzz_israel religious group zzz_american american show war history
zzz_fbi	official zzz_fbi attack terrorist law investigation case agent information federal
birthday	birthday song school family company home mother mail holiday wife
chip	chip computer art web com design need mail software technology
dog	dog show history house told student book friend family home
zzz_iran	government zzz_iran political country war zzz_afghanistan right nation women american
patriot	million patriot season player game coach team play quarterback zzz_nfl
publisher	book author publisher writer word right magazine sales published newspaper
zzz_bill_clinton	zzz_bill_clinton president zzz_george_bush zzz_white_house political office term presidential zzz_american republican
zzz_aid	drug million zzz_aid percent government official study patient zzz_united_states money
farmer	farmer bill farm program government percent food country water billion
bankruptcy	million company companies bankruptcy executive business bill debt billion plan
shirt	shirt women look girl school young show home wear right
inning	run inning hit game home ball homer left field lead
sauce	oil sauce add minutes cup tablespoon red cooking cut green

A.2 NIPS abstracts

anchor words	top 10 words
iiii	iiii cell iii border effect bar visual field responses receptive
skill	learning action task skill function reinforcement policy robot agent loss
song	learning neuron song network input motor auditory system synaptic temporal
hint	hint learning function examples error input performance network information method
clause	unit network function clause proof set hidden constraint bound word
rat	cell direction head rat place firing field model environment rate
routing	network routing learning function path neuron node neural algorithm pattern
actor	algorithm learning function action critic policy actor method parameter approximation
interneuron	neuron model interneuron input response layer connection pattern synaptic network
dominance	orientation ocular dominance map pattern cortical visual cortex development center
fuzzy	cell network neural system function input rules fuzzy output rule
obd	network weight error training unit function learning set neural method
sheet	network unit function learning structure prediction input problem hidden weight
wind	model wind field data parameter point front direction likelihood mean
wta	input output current unit wta net voltage pattern information circuit
parietal	function neuron visual field position object eye layer representation cortex
mistake	algorithm learning function bound weight concept distribution examples number mistake
pyramid	network image neural pyramid level learning images function result algorithm
muscle	model movement control trajectory arm motor hand dynamic muscle inverse
signature	network neural input training set feature signature output layer features
instruction	instruction block system number features schedule performance memory execution program
release	neuron synaptic spike synapses pattern parameter synapse release dynamic probability
som	data algorithm vector map space set point cluster input number
harmony	node nodes grammar tree field rule rules link parent graph
bipolar	cell signal noise current filter processing optimal voltage bipolar visual
option	policy action option optimal set reward step sutton markov reinforcement
star	system pattern neural star data recognition sensor noise memory image
vor	eye model head gain input system vor velocity learning visual
charge	weight chip circuit analog neuron neural charge system transistor transfer
adaboost	algorithm function error margin classifier adaboost learning cost training boosting
document	document word query information term retrieval matrix text clustering distribution
eeg	data eeg component signal ica noise analysis subject artifact independent
bootstrap	model network data prediction distribution error training neural bootstrap set
insect	insect behavior neuron neural food energy animal leg system sensor
disparity	layer disparity unit model output input local data image pixel
lesion	network unit lesion neural performance result size number effect pattern
facial	image images face system facial recognition representation action expression performance
player	learning network algorithm player function weight result game carlo monte
stack	network neural learning string stack neuron training symbol unit grammar
speaker	speaker network recognition speech training neural classifier performance word tdnn
knn	classifier training set data algorithm pattern method classification problem performance
owl	network system unit model sound learning signal data motor owl
student	learning error training student generalization unit weight noise teacher vector
edelman	object view model recognition features feature image unit representation images
gamma	network model memory neural gamma layer input output problem number
oscillator	oscillator system pattern neural phase fig delay connection neuron visual
snn	system neural hmm net training layer speech snn segment error
tangent	distance tangent vector transformation set space point algorithm method class
packet	packet data problem system application channel rate power result error
item	item data model list context term classification representation memory feature
composite	task action algorithm learning composite optimal set problem function solution

anchor words	top 10 words
spin	point model cluster correlation function data spin distribution method number
buffer	unit input weight learning pattern rule hidden output layer rules
refractory	spike firing neuron model rate cell spikes train signal period
cues	target cues cue location subject memory stimulus sound vector spectral
odor	cell input model odor activity olfactory information oscillation feature receptor
ridge	model parameter algorithm data adaptive estimate linear ridge regression kernel
hit	training pattern network unit input set output hidden number hit
axon	circuit pulse axon width input delay threshold voltage output velocity
stock	learning parameter training data system set task performance experiment return
style	model data parameter style mixture likelihood content observation prior factor
lgn	cell model visual input cortical cortex field neuron orientation lgn
cochlea	model circuit frequency filter analog cochlea parameter system chip sound
unlabeled	data set model training examples neural labeled class unlabeled classification
syllable	word model syllable control frequency phoneme human representation activation effect
splines	function model algorithm data set number error basis approximation network
ranking	algorithm problem learning set error ranking selection loss number result
motion	motion direction visual system signal stage moving velocity field filter
yang	model data algorithm learning information representation input method number component
plant	network control neural model system controller learning error forward output
pomdp	algorithm policy action learning problem states function model optimal probability
receiver	signal noise information net channel error neural system receiver algorithm
children	model set children rules variables test examples perceptron prediction bayesian
texture	image texture region images feature local orientation features model contrast
svm	vector set function kernel problem training support svm solution algorithm
regularizer	network function input weight data neural training unit output term
rotated	network output input digit rotation pattern neural training image classification
message	network neural classifier set system weight problem solution message number
maass	function network input neural neuron threshold weight output net bound
road	network unit input hidden training output layer neural weight road
committee	network input set error training committee weight performance learning output
leaf	tree learning function node trees decision probability data nodes algorithm
asynchronous	system dynamic algorithm function neural network asynchronous method result point
codebook	vector algorithm pattern class output classification performance distribution problem error
saliency	map saliency image visual noise task feature features element attention
lip	point system linear space image images speech local function information
perturbation	learning network weight function gradient algorithm neural error problem perturbation
conductances	neuron information channel noise rate synaptic cell firing voltage membrane
hasselmo	learning input pattern function network synaptic region representation unit layer
wavelet	gaussian filter component distribution image images data wavelet coefficient density
binding	binding object direction representation activation temporal mechanism evidence visual role
character	character word input recognition system field training output set net
mjolsness	model data point parameter algorithm object graph problem level method
impulse	function neuron neural signal threshold impulse circuit algorithm network input
consonant	network unit training hidden speech set neural output layer input
phone	system model context word hmm recognition speech training data parameter
attractor	network attractor input unit system learning neural dynamic parameter pattern
service	call control problem channel set service decision link cost performance
hyperparameter	model data gaussian distribution parameter mean method bayesian set prior
demonstration	learning model task function control reinforcement system robot learn dynamic