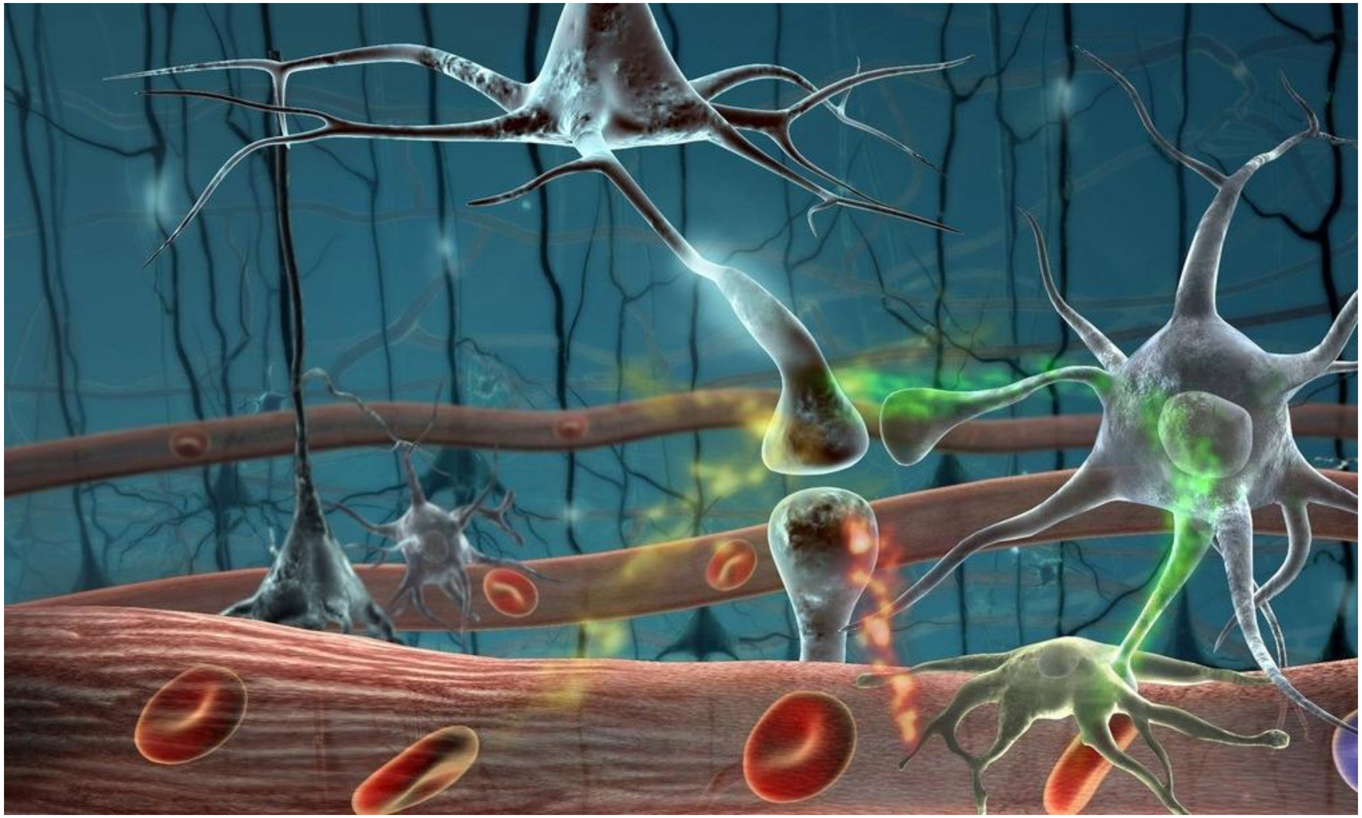


机器读心术之神经网络与深度学习 第7周

法律声明

【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

关注炼数成金企业微信

- 提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！

参考论文

Restricted Boltzmann Machines for Collaborative Filtering

Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Andriy Mnih AMNIH@CS.TORONTO.EDU
Geoffrey Hinton HINTON@CS.TORONTO.EDU
University of Toronto, 6 King's College Rd., Toronto, Ontario M5S 3G4, Canada

Abstract

Most of the existing approaches to collaborative filtering cannot handle very large data sets. In this paper we show how a class of two-layer undirected graphical models, called Restricted Boltzmann Machines (RBM's), can be used to model tabular data, such as user's ratings of movies. We present efficient learning and inference procedures for this class of models and demonstrate that RBM's can be successfully applied to the Netflix data set, containing over 100 million user/movie ratings. We also show that

Low-rank approximations based on minimizing the sum-squared distance can be found using Singular Value Decomposition (SVD). In the collaborative filtering domain, however, most of the data sets are sparse, and as shown by Srebro and Jaakkola (2003), this creates a difficult non-convex problem, so a naive solution is not going work.[1]

In this paper we describe a class of two-layer undirected graphical models that generalize Restricted Boltzmann Machines to modeling tabular or count data (Welling et al., 2005). Maximum likelihood learning is intractable in these models, but we show that learning can be performed efficiently by following an approximation to the gradient of a different objec-

当当网上的推荐


当当图书榜 | 童书 | 中小学教辅 | 教材 | 考试 | 小说 | 青春文学 | 人文社科 | 家教 | 励志 | 新书预售 | 读书社区 | 热搜排行 | 特价书市
全部商品详细分类
双12狂欢周
全部分类
搜 索
高级搜索 热搜: 曼德拉 本色 考研
图书 > 教材 > 研究生/本科/专科教材 > 理学 > 商品详情
看过本商品的还看了
张量分析（附光盘）
¥20.70
工程弹性力学与有限元法
¥16.80
理论物理学教程-弹性理论（第五版
¥33.80
实变函数论
¥17.30
数学分析（上册）——高等学校小
¥9.70
张量几何 第2版
¥40.00
张量分析[第二版] 70万种图书音像5折封顶！20万种科教类书6.9折封顶！
双12狂欢周
张量分析 第2版
TENSOR ANALYSIS
清华大学出版社
商品编号：8763182
抢 购 ¥23.50 还剩 1 天 5 小时 35 分结束
当 当 价：¥28.00
定 价：¥34.00 折扣：6.9折
顾客评分：已有486人评论，98.4%推荐
配 送 至：广东广州市海珠区 有货
下周二(12月17日)可送达，请在16小时6分钟内下单并选择“普通快递送货上门
运费说明 >>
作 者：黄克智 等
出 版 社：清华大学出版社
出版时间：2003-7-1
版 次： 页 数： 字
印刷时间： 开 本： 纸
推荐此书 点击看大图

www.stanford.edu/class/msande239/

Course Schedule

- 09/30 Overview and Introduction
- 10/07 Marketplace and Economics
- 10/14 Textual Advertising 1: Sponsored Search
- 10/21 Textual Advertising 2: Contextual Advertising
- 10/28 Display Advertising 1
- 11/04 Display Advertising 2
- 11/11 Targeting
- 11/18 Recommender Systems
- 12/02 Mobile, Video and other Emerging Formats
- 12/09 Project Presentations

Lecture Handouts

- Class information
- Lecture 1: Introduction, Supplementary notes
- Lecture 2: Marketplace design, In class presentation, Supplementary notes
- Lecture 3: Sponsored search 1, In class presentation
- Lecture 4: Sponsored search 2, In class presentation
- Lecture 5: Display advertising 1, In class presentation
- Lecture 6: Display advertising 2, In class presentation
- Lecture 7: Targeting, In class presentation
- Lecture 8: Recommender systems, In class presentation 1, In class presentation 2
- Lecture 9: Mobile, video, and other emerging formats, In class presentation 1, In class presentation 2

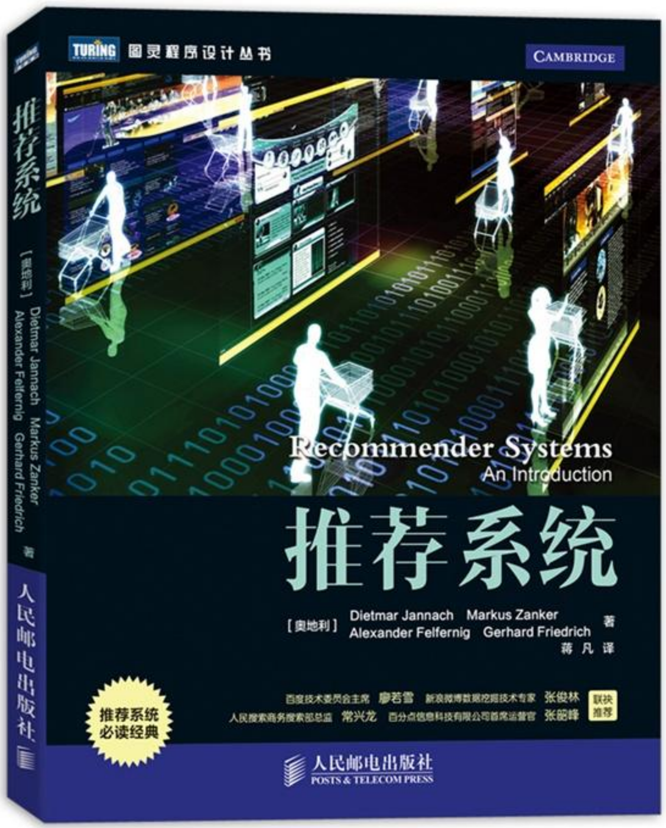
Readings & Other Links

- An interesting video on mobile advertising.
- Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords, B. Edelman, M. Ostrovsky, M. Schwarz.
- Algorithmic Game Theory, Chapter 28 (Sponsored Search Auctions). Online copy linked off Tim Roughgarden's webpage.
- Bidding for Representative Allocations for Display Advertising. Arpita Ghosh, Preston McAfee, Kishore Papineni, Sergei Vassilvitskii.

Assignments

推荐系统参考书

DATAGURU炼数成金



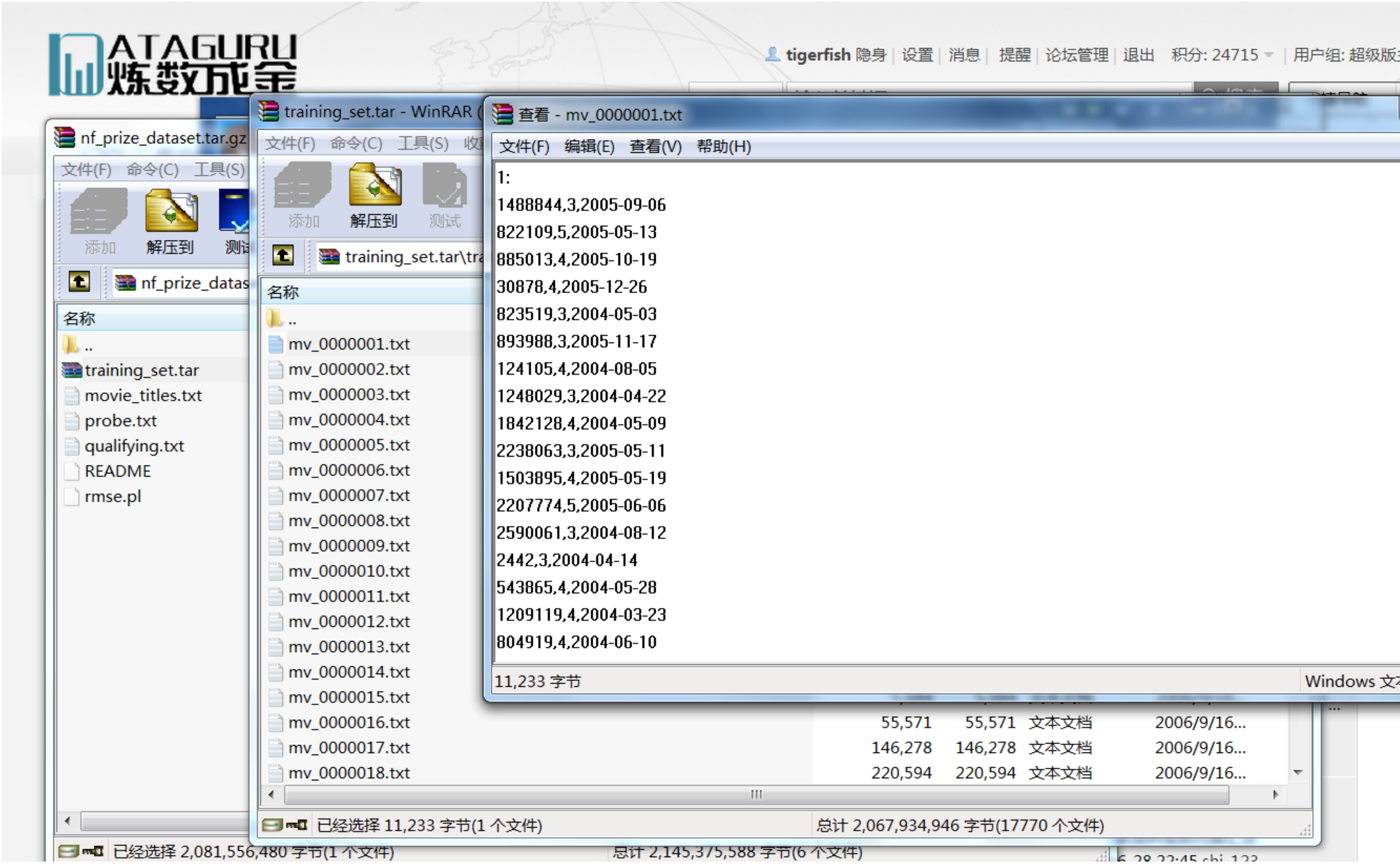
协同过滤的基本思想

- 协同过滤一般是在海量的用户中发掘出一小部分和你品位比较类似的，在协同过滤中，这些用户成为邻居，然后根据他们喜欢的其他东西组织成一个排序的目录作为推荐给你。
- 核心问题：

如何确定一个用户是不是和你有相似的品位？

如何将邻居们的喜好组织成一个排序的目录？

Netflix数据集


ATAGURU 炼数成金
tigerfish 隐身 | 设置 | 消息 | 提醒 | 论坛管理 | 退出 积分: 24715 | 用户组: 超级版主
nf_prize_dataset.tar.gz
文件(F) 命令(C) 工具(S)
添加 解压到 测试
名称
..
training_set.tar
movie_titles.txt
probe.txt
qualifying.txt
README
rmse.pl
已经选择 2,081,556,480 字节(1 个文件)
总计 2,145,375,588 字节(6 个文件)
training_set.tar - WinRAR
文件(F) 命令(C) 工具(S)
添加 解压到 测试
training_set.tar\tra
名称
..
mv_0000001.txt
mv_0000002.txt
mv_0000003.txt
mv_0000004.txt
mv_0000005.txt
mv_0000006.txt
mv_0000007.txt
mv_0000008.txt
mv_0000009.txt
mv_0000010.txt
mv_0000011.txt
mv_0000012.txt
mv_0000013.txt
mv_0000014.txt
mv_0000015.txt
mv_0000016.txt 55,571 55,571 文本文档 2006/9/16...
mv_0000017.txt 146,278 146,278 文本文档 2006/9/16...
mv_0000018.txt 220,594 220,594 文本文档 2006/9/16...
已经选择 11,233 字节(1 个文件)
总计 2,067,934,946 字节(17770 个文件)
查看 - mv_0000001.txt
文件(F) 编辑(E) 查看(V) 帮助(H)
1:
1488844,3,2005-09-06
822109,5,2005-05-13
885013,4,2005-10-19
30878,4,2005-12-26
823519,3,2004-05-03
893988,3,2005-11-17
124105,4,2004-08-05
1248029,3,2004-04-22
1842128,4,2004-05-09
2238063,3,2005-05-11
1503895,4,2005-05-19
2207774,5,2005-06-06
2590061,3,2004-08-12
2442,3,2004-04-14
543865,4,2004-05-28
1209119,4,2004-03-23
804919,4,2004-06-10
11,233 字节
Windows 文本

LSA与评分矩阵

- LSA(latent semantic analysis)潜在语义分析，也被称为LSI(latent semantic index)，是Scott Deerwester, Susan T. Dumais等人在1990年提出来的一种新的索引和检索方法。该方法和传统向量空间模型(vector space model)一样使用向量来表示词(terms)和文档(documents)，并通过向量间的关系(如夹角)来判断词及文档间的关系；而不同的 是，LSA将词和文档映射到潜在语义空间，从而去除了原始向量空间中的一些“噪音”，提高了信息检索的精确度。
- http://blog.csdn.net/wangran51/article/details/7408406
- 场景：

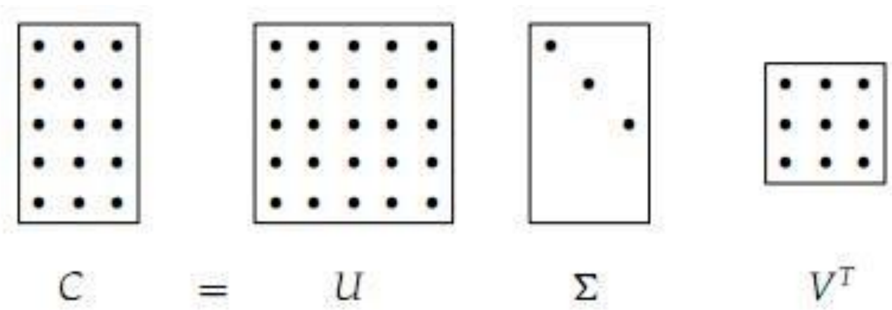
Sample Term by Document matrix

	access	*document*	*retrieval*	*information*	*theory*	*database*	*indexing*	*computer*	REL	MATCH
Doc 1	x	x	x			x	x		R	
Doc 2				x*	x			x*		M
Doc 3			x	x*				x*	R	M

Query: "IDF in *computer*-based *information* look-up"

SVD奇异值分解

- http://blog.csdn.net/wangran51/article/details/7408414
- 项亮书P186页
- SVD与主成分分析



利用SVD求解LSA

- 分析文档集合，建立Term-Document矩阵。
- 对Term-Document矩阵进行奇异值分解。
- 对SVD分解后的矩阵进行降维，也就是奇异值分解一节所提到的低阶近似。
- 使用降维后的矩阵构建潜在语义空间，或重建Term-Document矩阵。

例子

Example of text data: Titles of Some Technical Memos

c1: *Human* machine *interface* for ABC *computer* applications
c2: A *survey* of *user* opinion of *computer system response time*
c3: The *EPS user interface* management *system*
c4: *System* and *human system* engineering testing of *EPS*
c5: Relation of *user* perceived *response time* to error measurement

m1: The generation of random, binary, ordered *trees*
m2: The intersection *graph* of paths in *trees*
m3: *Graph minors* IV: Widths of *trees* and well-quasi-ordering
m4: *Graph minors*: A *survey*

$\{X\} =$

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

例子

$$\{X\} = \{W\}\{S\}\{P\}'$$

$\{W\} =$

0.22	-0.11	0.29	-0.41	-0.11	-0.34	0.52	-0.06	-0.41
0.20	-0.07	0.14	-0.55	0.28	0.50	-0.07	-0.01	-0.11
0.24	0.04	-0.16	-0.59	-0.11	-0.25	-0.30	0.06	0.49
0.40	0.06	-0.34	0.10	0.33	0.38	0.00	0.00	0.01
0.64	-0.17	0.36	0.33	-0.16	-0.21	-0.17	0.03	0.27
0.27	0.11	-0.43	0.07	0.08	-0.17	0.28	-0.02	-0.05
0.27	0.11	-0.43	0.07	0.08	-0.17	0.28	-0.02	-0.05
0.30	-0.14	0.33	0.19	0.11	0.27	0.03	-0.02	-0.17
0.21	0.27	-0.18	-0.03	-0.54	0.08	-0.47	-0.04	-0.58
0.01	0.49	0.23	0.03	0.59	-0.39	-0.29	0.25	-0.23
0.04	0.62	0.22	0.00	-0.07	0.11	0.16	-0.68	0.23
0.03	0.45	0.14	-0.01	-0.30	0.28	0.34	0.68	0.18

$\{S\} =$

3.34								
	2.54							
		2.35						
			1.64					
				1.50				
					1.31			
						0.85		
							0.56	
								0.36

$\{P\} =$

0.20	0.61	0.46	0.54	0.28	0.00	0.01	0.02	0.08
-0.06	0.17	-0.13	-0.23	0.11	0.19	0.44	0.62	0.53
0.11	-0.50	0.21	0.57	-0.51	0.10	0.19	0.25	0.08
-0.95	-0.03	0.04	0.27	0.15	0.02	0.02	0.01	-0.03
0.05	-0.21	0.38	-0.21	0.33	0.39	0.35	0.15	-0.60
-0.08	-0.26	0.72	-0.37	0.03	-0.30	-0.21	0.00	0.36
0.18	-0.43	-0.24	0.26	0.67	-0.34	-0.15	0.25	0.04
-0.01	0.05	0.01	-0.02	-0.06	0.45	-0.76	0.45	-0.07
-0.06	0.24	0.02	-0.08	-0.26	-0.62	0.02	0.52	-0.45

$\{\hat{X}\} =$

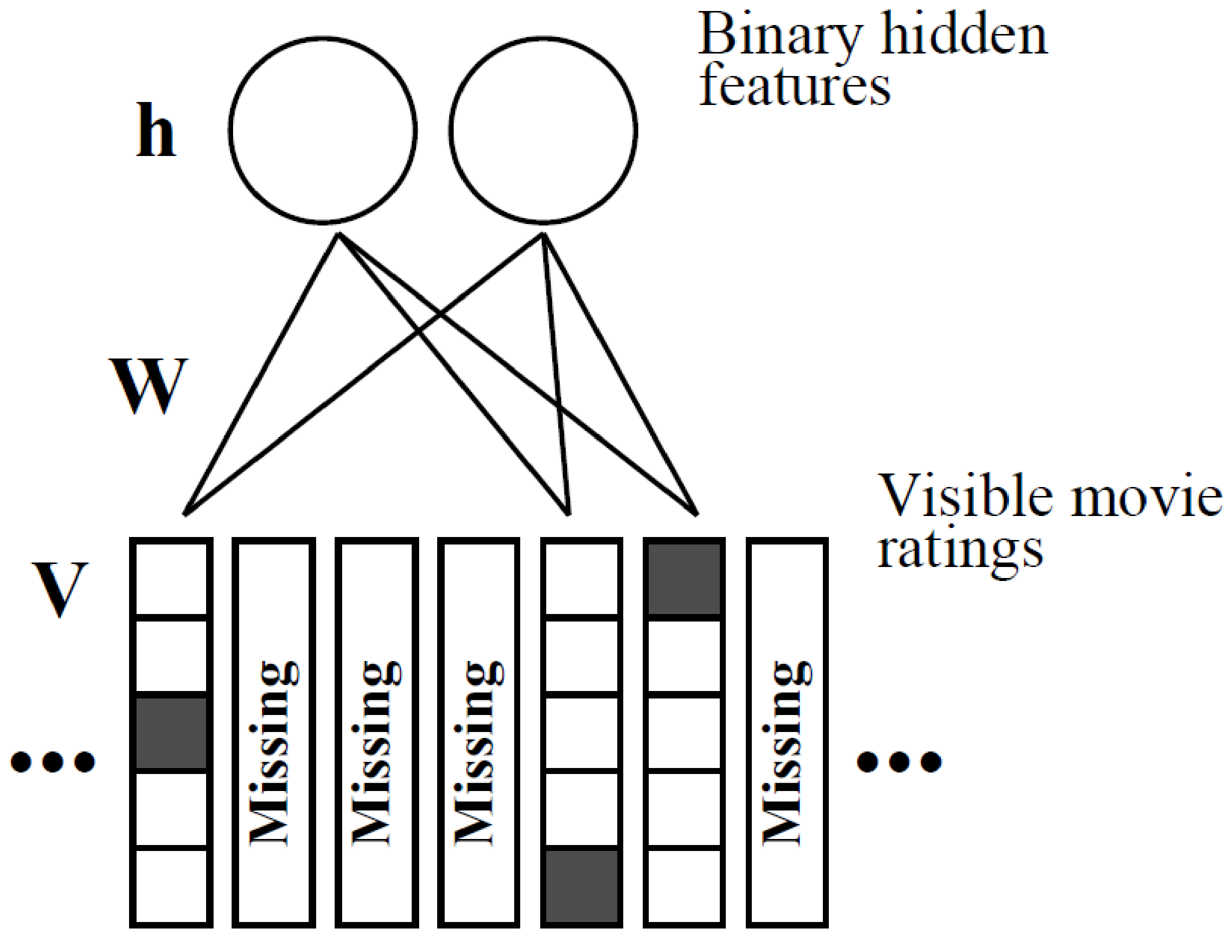
	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	0.16	0.40	0.38	0.47	0.18	-0.05	-0.12	-0.16	-0.09
interface	0.14	0.37	0.33	0.40	0.16	-0.03	-0.07	-0.10	-0.04
computer	0.15	0.51	0.36	0.41	0.24	0.02	0.06	0.09	0.12
user	0.26	0.84	0.61	0.70	0.39	0.03	0.08	0.12	0.19
system	0.45	1.23	1.05	1.27	0.56	-0.07	-0.15	-0.21	-0.05
response	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
time	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
EPS	0.22	0.55	0.51	0.63	0.24	-0.07	-0.14	-0.20	-0.11
survey	0.10	0.53	0.23	0.21	0.27	0.14	0.31	0.44	0.42
trees	-0.06	0.23	-0.14	-0.27	0.14	0.24	0.55	0.77	0.66
graph	-0.06	0.34	-0.15	-0.30	0.20	0.31	0.69	0.98	0.85
minors	-0.04	0.25	-0.10	-0.21	0.15	0.22	0.50	0.71	0.62

SVD的缺点

- 计算复杂度高，当矩阵达到1000维以上时计算已经非常缓慢，但文本分析一般都会形成非常大型的“文档-词”矩阵，从而难以实现，甚至存储都很困难

使用RBM解决协同过滤问题

- 网络结构



概率计算与能量函数

ATAGURU 炼数成金

$$p(v_i^k = 1|\mathbf{h}) = \frac{\exp\left(b_i^k + \sum_{j=1}^{F} h_j W_{ij}^k\right)}{\sum_{l=1}^{K} \exp\left(b_i^l + \sum_{j=1}^{F} h_j W_{ij}^l\right)} \quad (1)$$

$$p(h_j = 1|\mathbf{V}) = \sigma\left(b_j + \sum_{i=1}^{m}\sum_{k=1}^{K} v_i^k W_{ij}^k\right) \quad (2)$$

$$p(\mathbf{V}) = \sum_{\mathbf{h}} \frac{\exp\left(-E(\mathbf{V},\mathbf{h})\right)}{\sum_{\mathbf{V}',\mathbf{h}'} \exp\left(-E(\mathbf{V}',\mathbf{h}')\right)} \quad (3)$$

$$\begin{aligned} E(\mathbf{V},\mathbf{h}) &= -\sum_{i=1}^{m}\sum_{j=1}^{F}\sum_{k=1}^{K} W_{ij}^k h_j v_i^k + \sum_{i=1}^{m} \log Z_i \\ &\quad -\sum_{i=1}^{m}\sum_{k=1}^{K} v_i^k b_i^k - \sum_{j=1}^{F} h_j b_j \quad (4) \end{aligned}$$

模型训练

$$\Delta W_{ij}^k = \epsilon \frac{\partial \log p(\mathbf{V})}{\partial W_{ij}^k} = \\ = \epsilon \left(<v_i^k h_j>_{data} - <v_i^k h_j>_{model} \right) \qquad (5)$$

$$\Delta W_{ij}^k = \epsilon(<v_i^k h_j>_{data} - <v_i^k h_j>_T) \qquad (6)$$

模型预测

$$p(v_q^k = 1|\mathbf{V}) \propto \sum_{h_1,\ldots,h_p} \exp(-E(v_q^k, \mathbf{V}, \mathbf{h})) \qquad (7)$$

$$\propto \Gamma_q^k \prod_{j=1}^{F} \sum_{h_j \in \{0,1\}} \exp\big(\sum_{il} v_i^l h_j W_{ij}^l + v_q^k h_j W_{qj}^k + h_j b_j\big)$$

$$= \Gamma_q^k \prod_{j=1}^{F} \Big(1 + \exp\big(\sum_{il} v_i^l W_{ij}^l + v_q^k W_{qj}^k + b_j\big)\Big)$$

$$p(v_{q_1}^{k_1} = 1, v_{q_2}^{k_2} = 1, \ldots, v_{q_n}^{k_n} = 1|\mathbf{V}) \qquad (8)$$

$$\hat{p}_j = p(h_j = 1|\mathbf{V}) = \sigma(b_j + \sum_{i=1}^{m} \sum_{k=1}^{K} v_i^k W_{ij}^k) \qquad (9)$$

$$p(v_q^k = 1|\hat{\mathbf{p}}) = \frac{\exp\big(b_q^k + \sum_{j=1}^{F} \hat{p}_j W_{qj}^k\big)}{\sum_{l=1}^{K} \exp\big(b_q^l + \sum_{j=1}^{F} \hat{p}_j W_{qj}^l\big)} \qquad (10)$$

使用RBM变形

- RBM's with Gaussian hidden units
- Conditional RBM
- Conditional Factored RBM

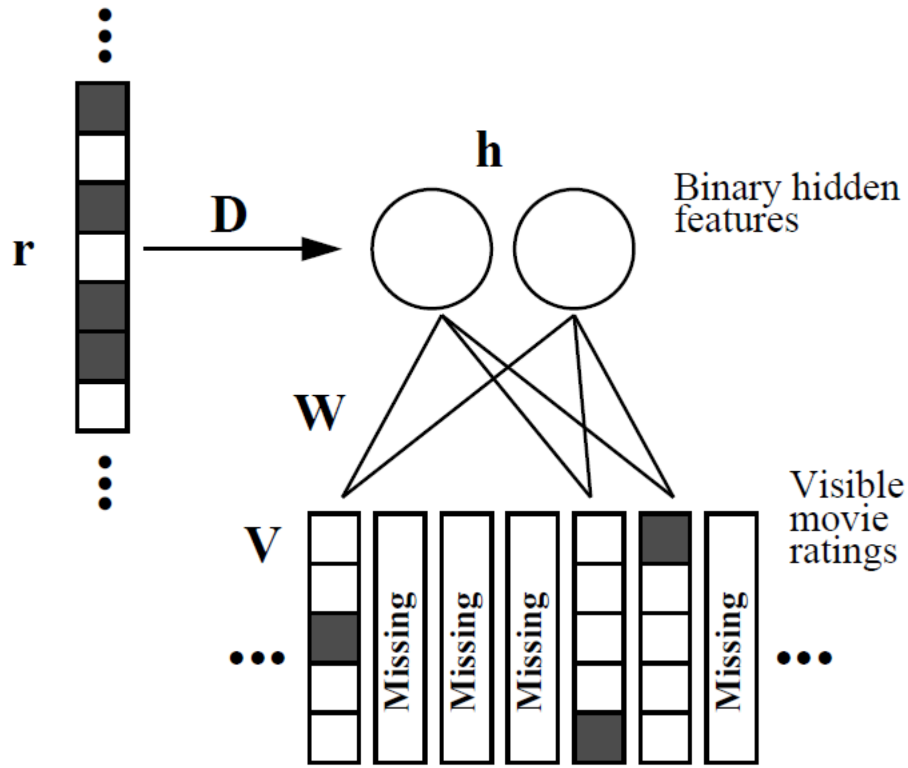


Figure 2. Conditional RBM. The binary vector **r**, indicating rated/unrated movies, affects binary states of the hidden units.

性能对比

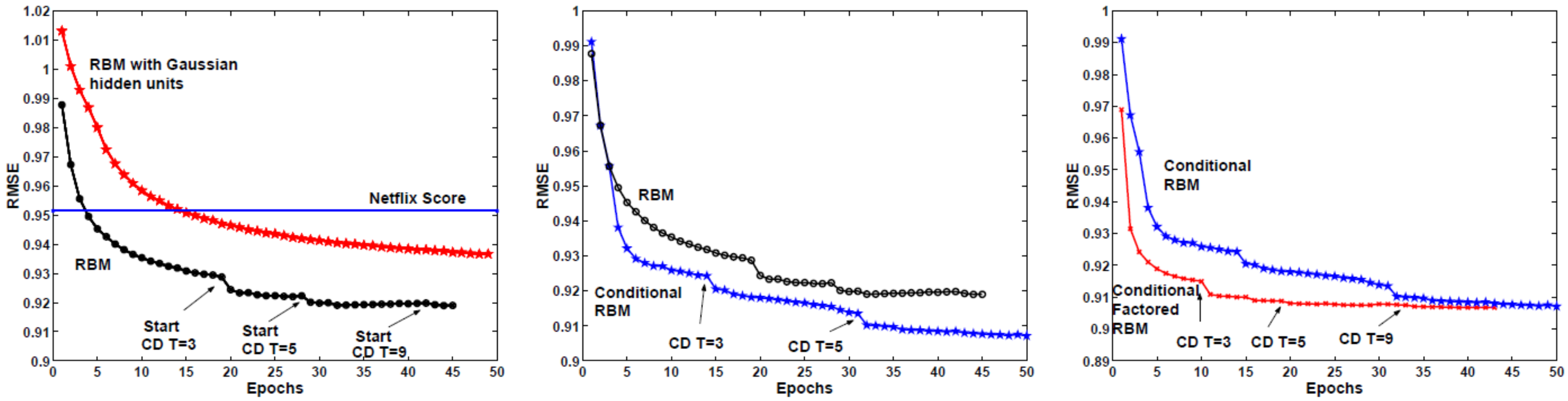


Figure 3. Performance of various models on the validation data. Left panel: RBM vs. RBM with Gaussian hidden units. Middle panel: RBM vs. conditional RBM. Right panel: conditional RBM vs. conditional factored RBM. The y-axis displays RMSE (root mean squared error), and the x-axis shows the number of epochs, or passes through the entire training dataset.

Hinton里程碑式的论文

ER constituent
gurations. The
neasured SHG
onance in Fig.
ion), we find
bove the noise
signal closely
ncident power
SHG emission
nall angle with
eviations from
e SRRs (see
mall detuning
er wavelength
nces the SHG
20%. For ex-
with vertical
ve find a small
For excitation
ontal incident
out significant
is again po-

of the nonlinear optical properties of metallic

10.1126/science.1129198

Reducing the Dimensionality of Data with Neural Networks

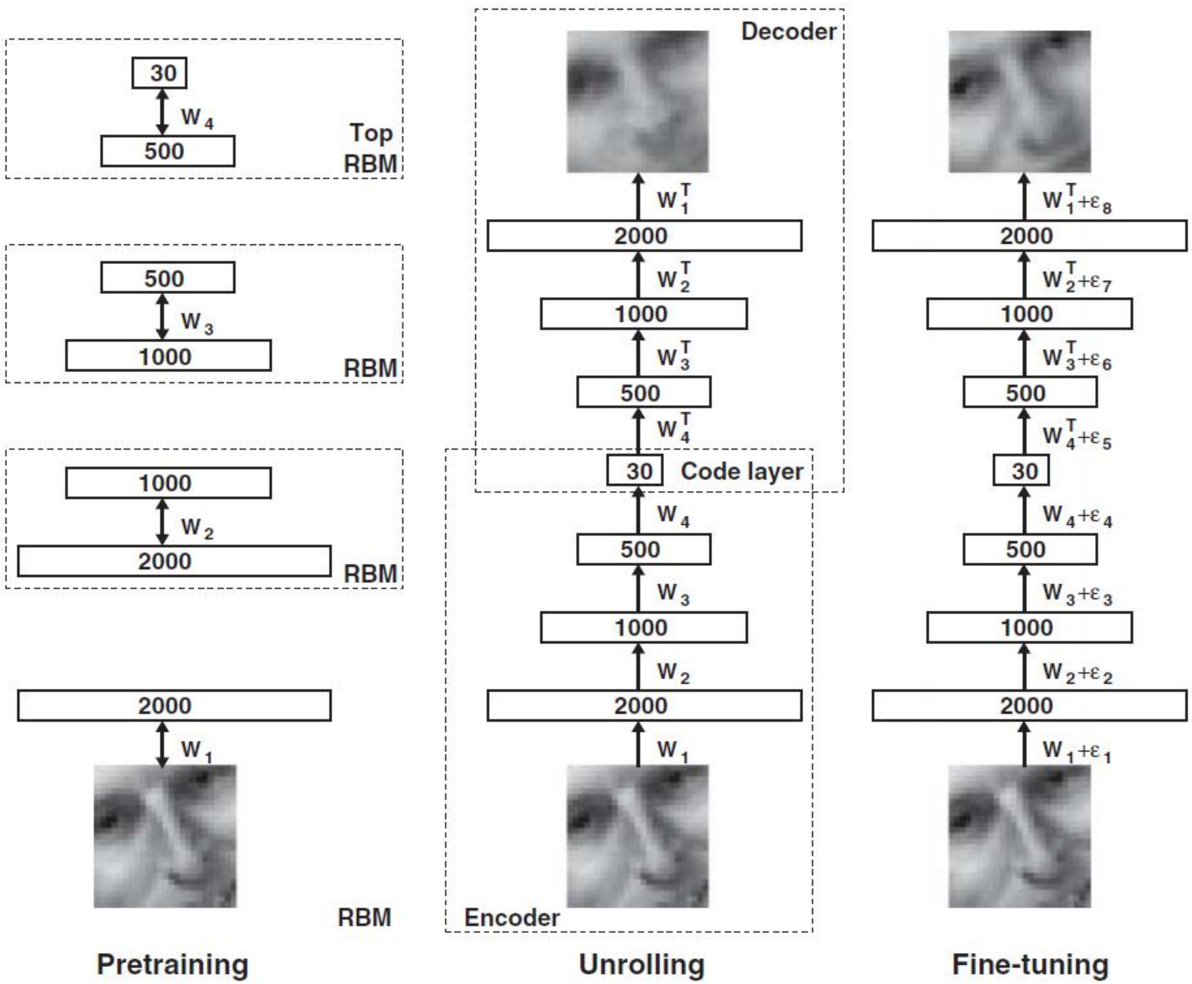
G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network

28 JULY 2006 VOL 313 **SCIENCE** www.sciencemag.org

图解“堆叠的RBM”及其作用



效果展现

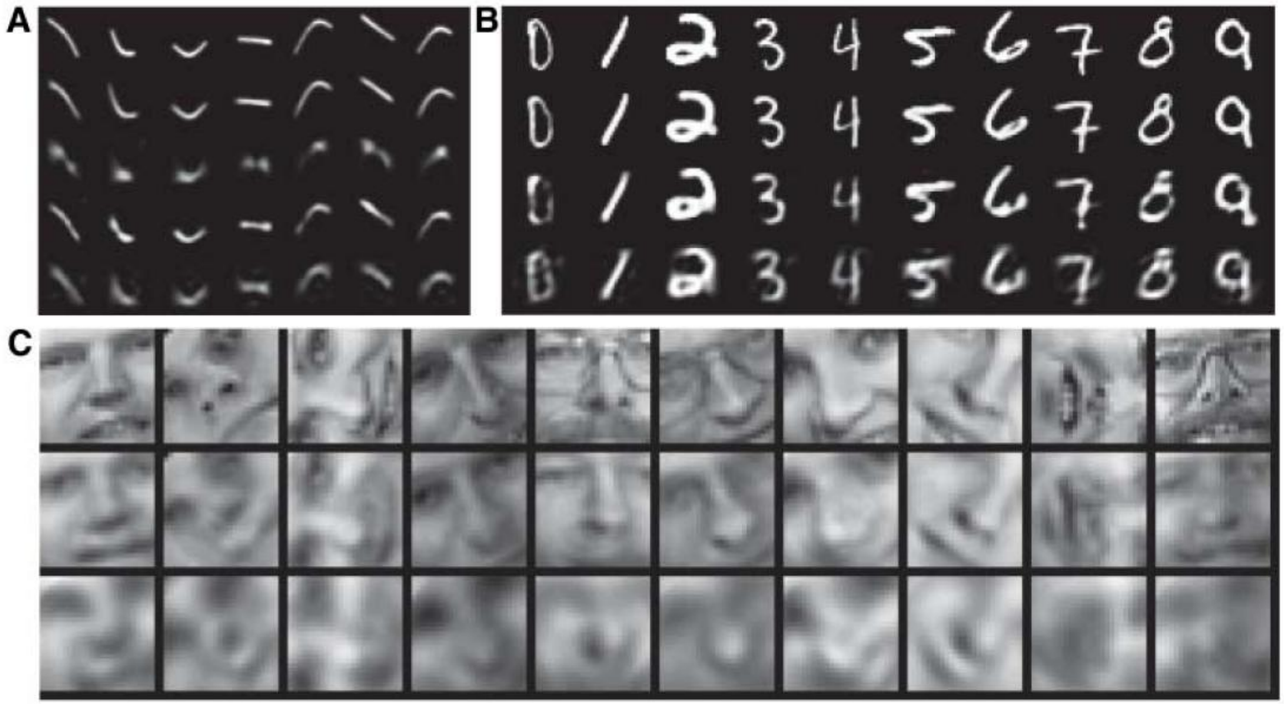
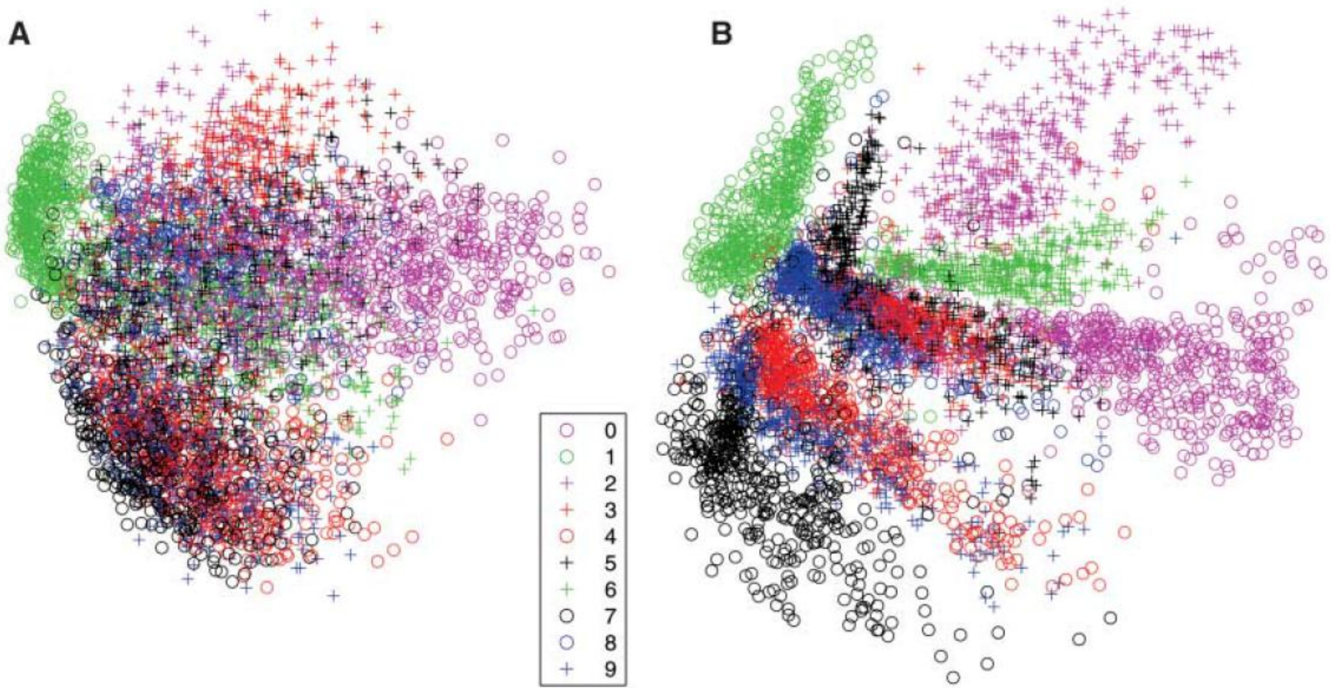


Fig. 2. (**A**) Top to bottom: Random samples of curves from the test data set; reconstructions produced by the six-dimensional deep autoencoder; reconstructions by "logistic PCA" (*8*) using six components; reconstructions by logistic PCA and standard PCA using 18 components. The average squared error per image for the last four rows is 1.44, 7.64, 2.45, 5.90. (**B**) Top to bottom: A random test image from each class; reconstructions by the 30-dimensional autoencoder; reconstructions by 30-dimensional logistic PCA and standard PCA. The average squared errors for the last three rows are 3.00, 8.01, and 13.87. (**C**) Top to bottom: Random samples from the test data set; reconstructions by the 30-dimensional autoencoder; reconstructions by 30-dimensional PCA. The average squared errors are 126 and 135.

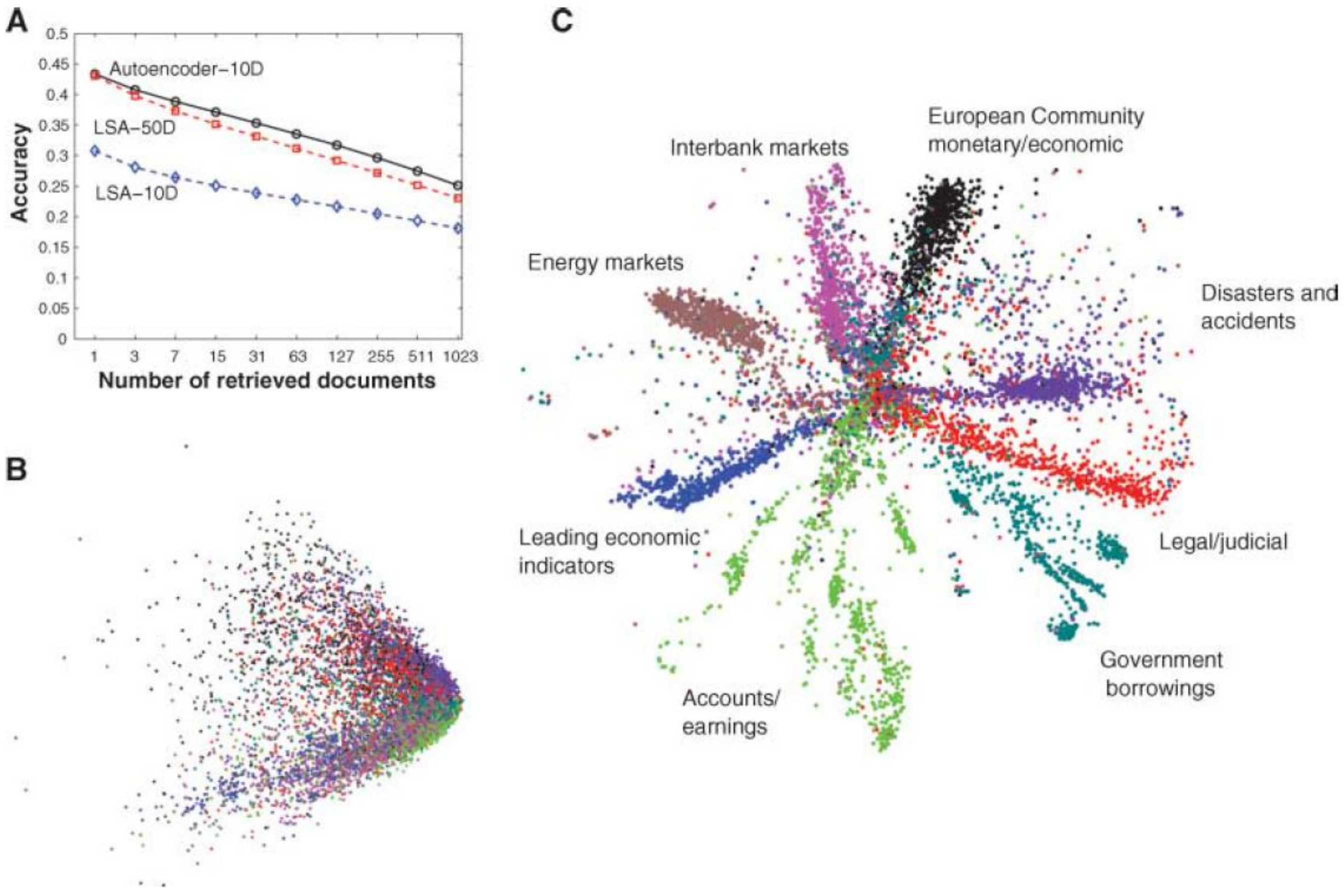
降维效果

Fig. 3. (**A**) The two-dimensional codes for 500 digits of each class produced by taking the first two principal components of all 60,000 training images. (**B**) The two-dimensional codes found by a 784-1000-500-250-2 autoencoder. For an alternative visualization, see (*8*).



降维效果

Fig. 4. (A) The fraction of retrieved documents in the same class as the query when a query document from the test set is used to retrieve other test set documents, averaged over all 402,207 possible queries. (B) The codes produced by two-dimensional LSA. (C) The codes produced by a 2000-500-250-125-2 autoencoder.



参考论文

Semantic Hashing

Ruslan Salakhutdinov
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 3G4
rsalakhu@cs.toronto.edu

Geoffrey Hinton
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 3G4
hinton@cs.toronto.edu

ABSTRACT

We show how to learn a deep graphical model of the word-count vectors obtained from a large set of documents. The values of the latent variables in the deepest layer are easy to infer and give a much better representation of each document than Latent Semantic Analysis. When the deepest layer is forced to use a small number of binary variables (e.g. 32), the graphical model performs "semantic hashing": Documents are mapped to memory addresses in such a way that semantically similar documents are located at nearby addresses. Documents similar to a query document can then be found by simply accessing all the addresses that differ by only a few bits from the address of the query document. This way of extending the efficiency of hash-coding to approximate matching is much faster than locality sensitive hashing, which is the fastest current method. By using semantic hashing to filter the documents given to TF-IDF, we achieve higher accuracy than applying TF-IDF to the entire document set.

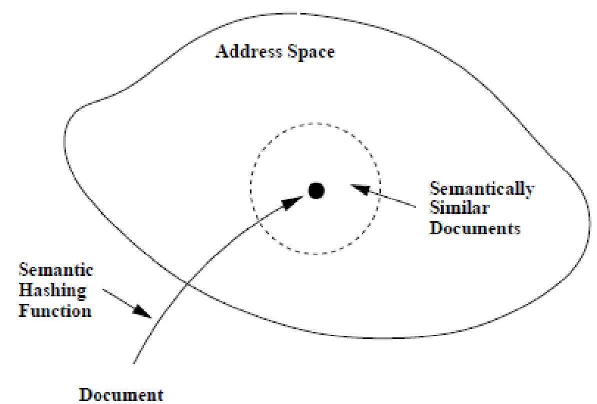


Figure 1: A schematic representation of semantic hashing.

cessfully applied in the domain of information retrieval. A simple and widely-used method is Latent Semantic Analysis (LSA) [5],

词向量与文本检索

- TF-IDF词频计算
- 高频截断
- 将文档转化为词向量
- 计算词向量的相似度，得到文档的相似度，筛选出相似度在门槛值以上的目标文档作为检索结果
- 弱点：

1 高词汇量时产生大计算量

2 不同词汇的频率与相似性有关这个假设很有疑问

3 同义词，近义词无法识别

语义Hash的思路

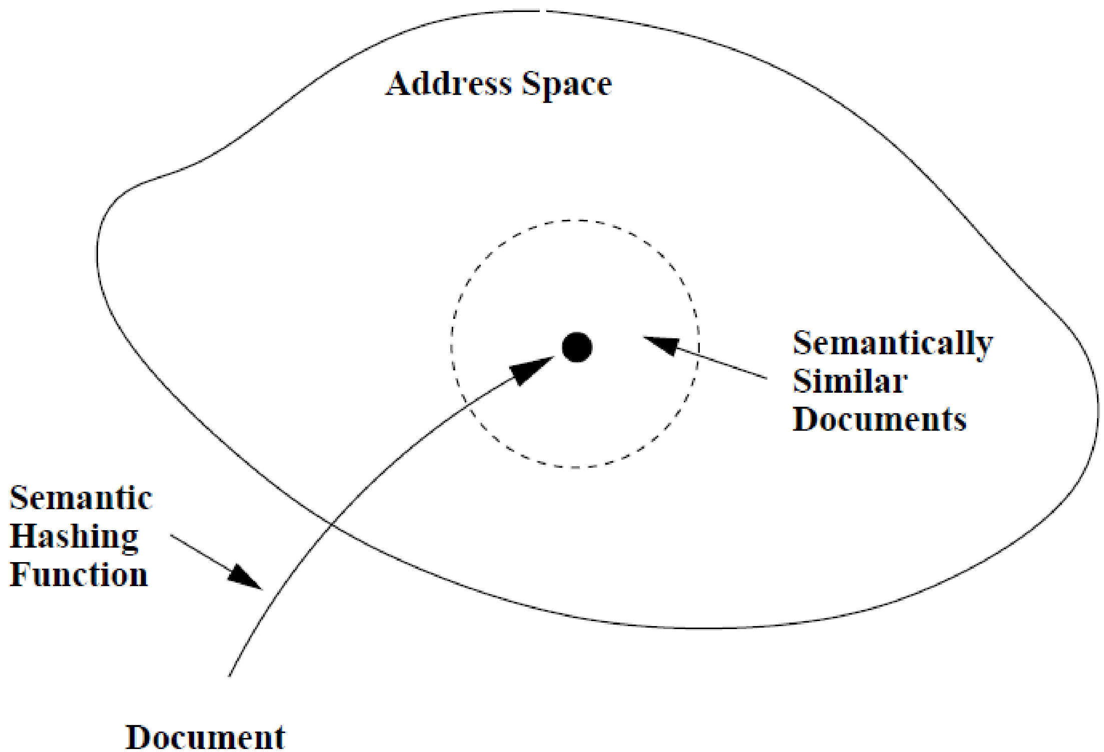


Figure 1: A schematic representation of semantic hashing.

THE CONSTRAINED POISSON MODEL

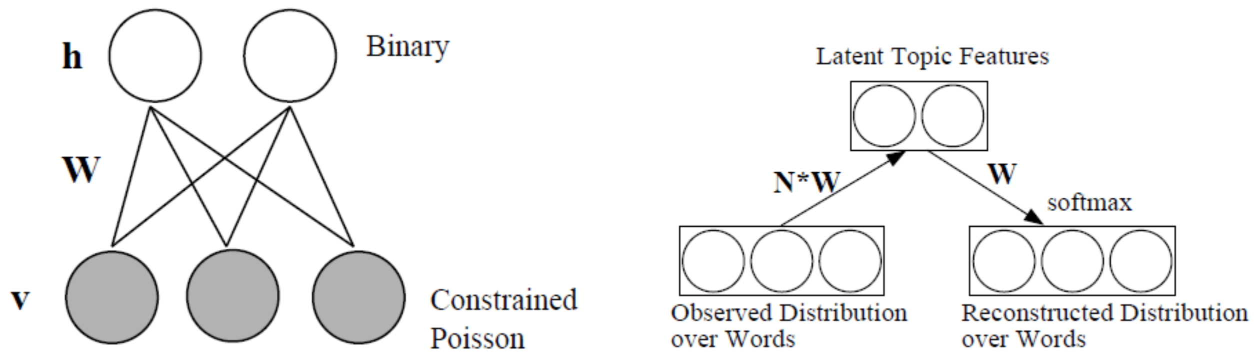


Figure 3: The left panel shows the Markov random field of the constrained Poisson model. The top layer represents a vector, h, of stochastic, binary, latent, topic features and and the bottom layer represents a Poisson visible vector v. The right panel shows a different interpretation of the constrained Poisson model in which the visible activities have all been divided by the number of words in the document so that they represent a probability distribution. The factor of N that multiplies the upgoing weights is a result of having N *i.i.d.* observations from the observed distribution.

概率计算与能量函数

$$p(v_i = n|\mathbf{h}) = \mathrm{Ps}\left(n, \frac{\exp\left(\lambda_i + \sum_j h_j w_{ij}\right)}{\sum_k \exp\left(\lambda_k + \sum_j h_j w_{kj}\right)} N\right) \quad (1)$$

$$p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum w_{ij} v_i) \quad (2)$$

$$p(\mathbf{v}) = \sum_{\mathbf{h}} \frac{\exp\left(-E(\mathbf{v}, \mathbf{h})\right)}{\sum_{\mathbf{u},\mathbf{g}} \exp\left(-E(\mathbf{u}, \mathbf{g})\right)} \quad (3)$$

$$\begin{aligned} E(\mathbf{v}, \mathbf{h}) = & -\sum_i \lambda_i v_i + \sum_i \log\left(v_i!\right) \\ & -\sum_j b_j h_j - \sum_{i,j} v_i h_j w_{ij} \end{aligned} \quad (4)$$

训练

$$\Delta w_{ij} = \epsilon \frac{\partial \log p(\mathbf{v})}{\partial w_{ij}} = \epsilon(<v_i h_j>_{data} - <v_i h_j>_{model})$$

$$\Delta w_{ij} = \epsilon(<v_i h_j>_{data} - <v_i h_j>_{recon}) \qquad (5)$$

Recursive Pretraining 与 Fine-Tuning

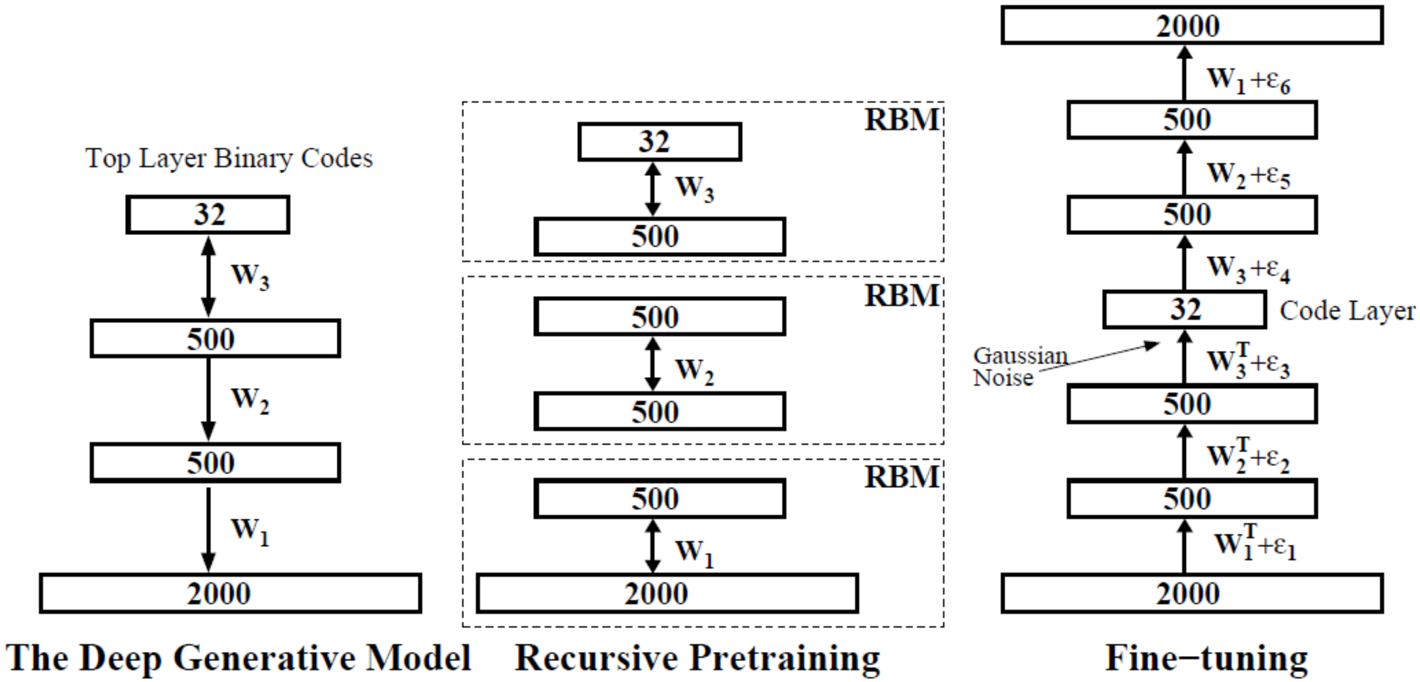


Figure 2: Left panel: The deep generative model. Middle panel: Pretraining consists of learning a stack of RBM's in which the feature activations of one RBM are treated as data by the next RBM. Right panel: After pretraining, the RBM's are "unrolled" to create a multi-layer autoencoder that is fine-tuned by backpropagation.

数据集

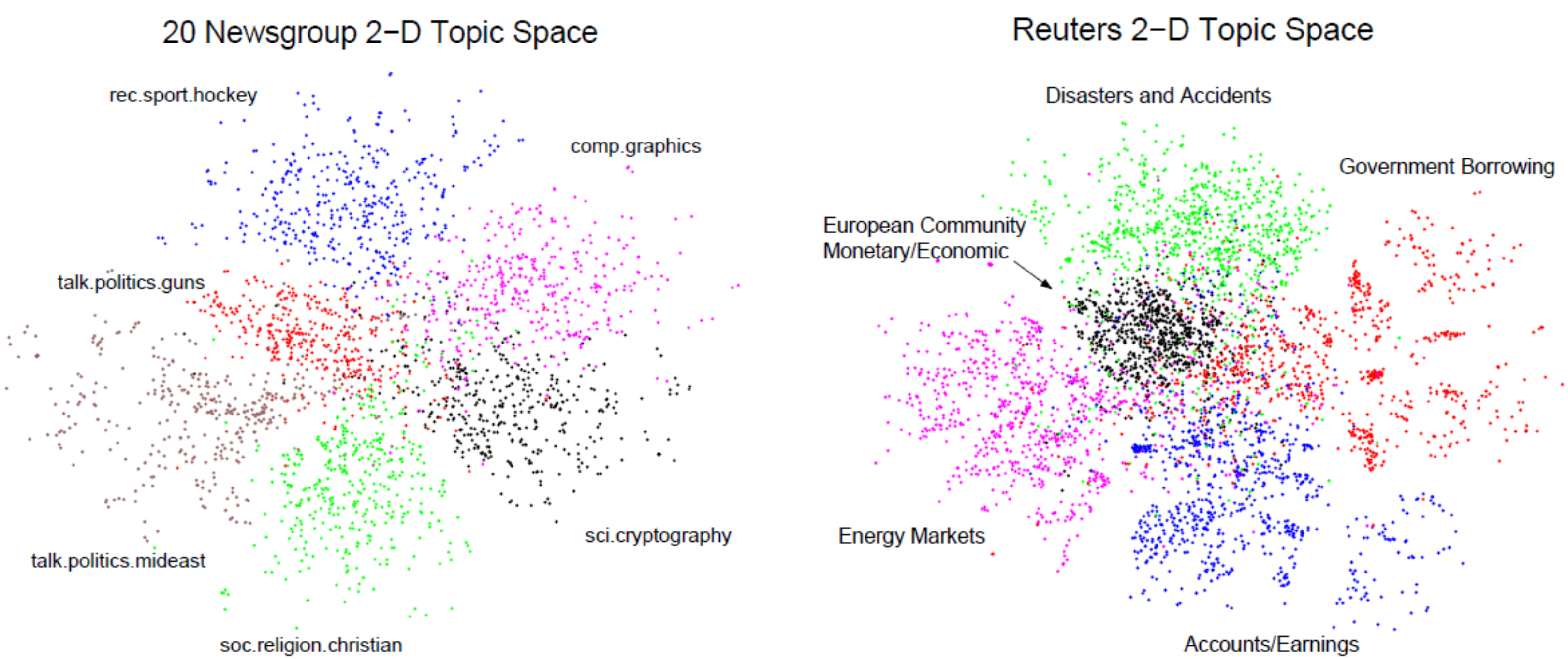


Figure 5: A 2-dimensional embedding of the 128-bit codes using stochastic neighbor embedding for the 20 Newsgroups data (left panel) and the Reuters RCV2 corpus (right panel). See in color for better visualization.

结果对比

20 Newsgroups

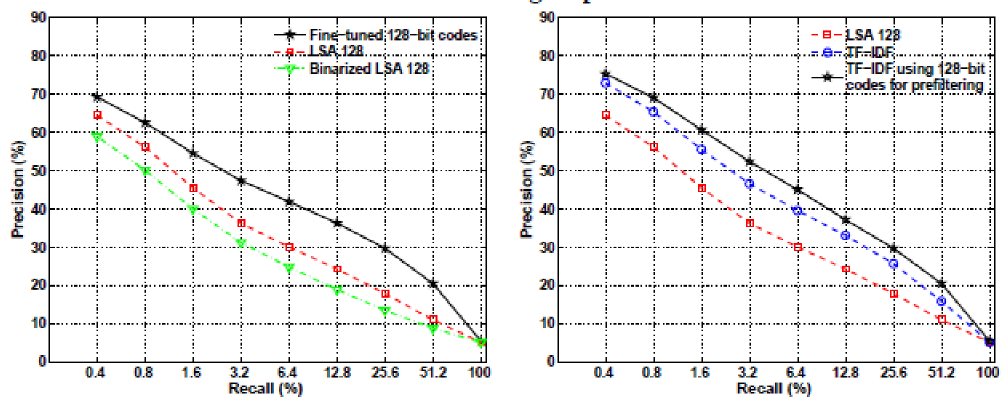


Figure 6: Precision-Recall curves for the 20 Newsgroups dataset, when a query document from the test set is used to retrieve other test set documents, averaged over all 7,531 possible queries.

Reuters RCV2

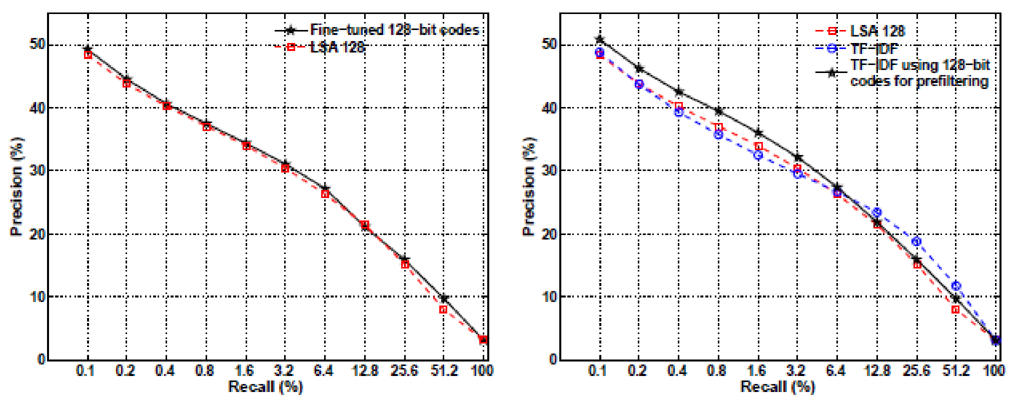


Figure 7: Precision-Recall curves for the Reuters RCV2 dataset, when a query document from the test set is used to retrieve other test set documents, averaged over all 402,207 possible queries.

生成式模型 vs 判别式模型？

Classification using Discriminative Restricted Boltzmann Machines

Hugo Larochelle LAROCHEH@IRO.UMONTREAL.CA
Yoshua Bengio BENGIOY@IRO.UMONTREAL.CA
Dept. IRO, Université de Montréal C.P. 6128, Montreal, Qc, H3C 3J7, Canada

Abstract

Recently, many applications for Restricted Boltzmann Machines (RBMs) have been developed for a large variety of learning problems. However, RBMs are usually used as feature extractors for another learning algorithm or to provide a good initialization for deep feed-forward neural network classifiers, and are not considered as a standalone solution to classification problems. In this paper, we argue that RBMs provide a self-contained framework for deriving competitive non-linear classifiers. We present an evaluation of different learning algorithms for RBMs which aim at introducing a discriminative component to RBM training and improve their performance as classifiers. This approach is simple in that RBMs are used directly to build a classifier, rather than as a stepping stone. Finally, we demonstrate how

image data (Gehler et al., 2006) or as a good initial training phase for deep neural network classifiers (Hinton, 2007). However, in both cases, the RBMs are merely the first step of another learning algorithm, either providing a preprocessing of the data or an initialization for the parameters of a neural network. When trained in an unsupervised fashion, RBMs provide no guarantees that the features implemented by their hidden layer will ultimately be useful for the supervised task that needs to be solved. More practically, model selection can also become problematic, as we need to explore jointly the space of hyper-parameters of both the RBM (size of the hidden layer, learning rate, number of training iterations) and the supervised learning algorithm that is fed the learned features. In particular, having two separate learning phases (feature extraction, followed by classifier training) can be problematic in an online learning setting.

In this paper, we argue that RBMs can be used successfully as stand-alone non-linear classifiers along-

炼数成金逆向收费式网络课程

- Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情，请看我们的培训网站 http://edu.dataguru.cn

Thanks

FAQ时间