Tutorial (Ready for WCCI – FUZZ-IEEE-2018, Rio de Janeiro, Brazil, 10/July/2018)

Uncertainty Modeling in Learning from Big Data

Xizhao Wang, PhD, Professor

IEEE Fellow, CAAI Fellow

JMLC Editor-in-Chief

Big Data Institute, ShenZhen Univ.

Outline

1. A brief introduction to big data
2. General challenge of big data analytics
3. Current strategy of processing big data
4. What new challenges
   the uncertainty brings in big data learning?
5. Some of our works
6. Remarks

1. Introduction to big data (1)

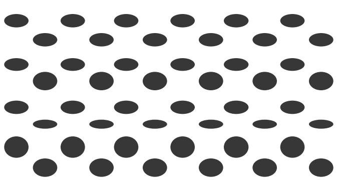
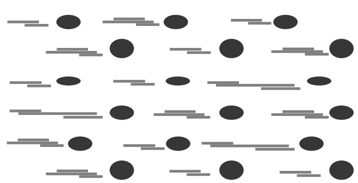
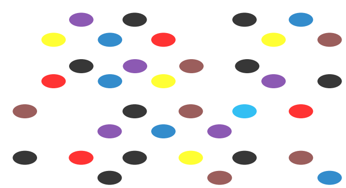
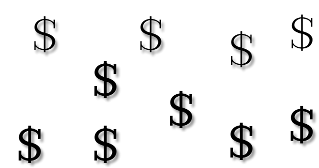
- Big data refers to datasets that are so large that conventional database management and data analysis tools are insufficient to work with them.
- Big data, which was called massive data [1], has become a bigger-than-ever problem with the quick developments of data collection and storage technologies.

[1] National Research Council, Frontiers in Massive Data Analysis, The National Academies Press, Washington, DC, 2013.

With the rapid development of data collection & storage ability, there are more and more big data. With respect to the data processing ability, it has already lead to an information explosion.

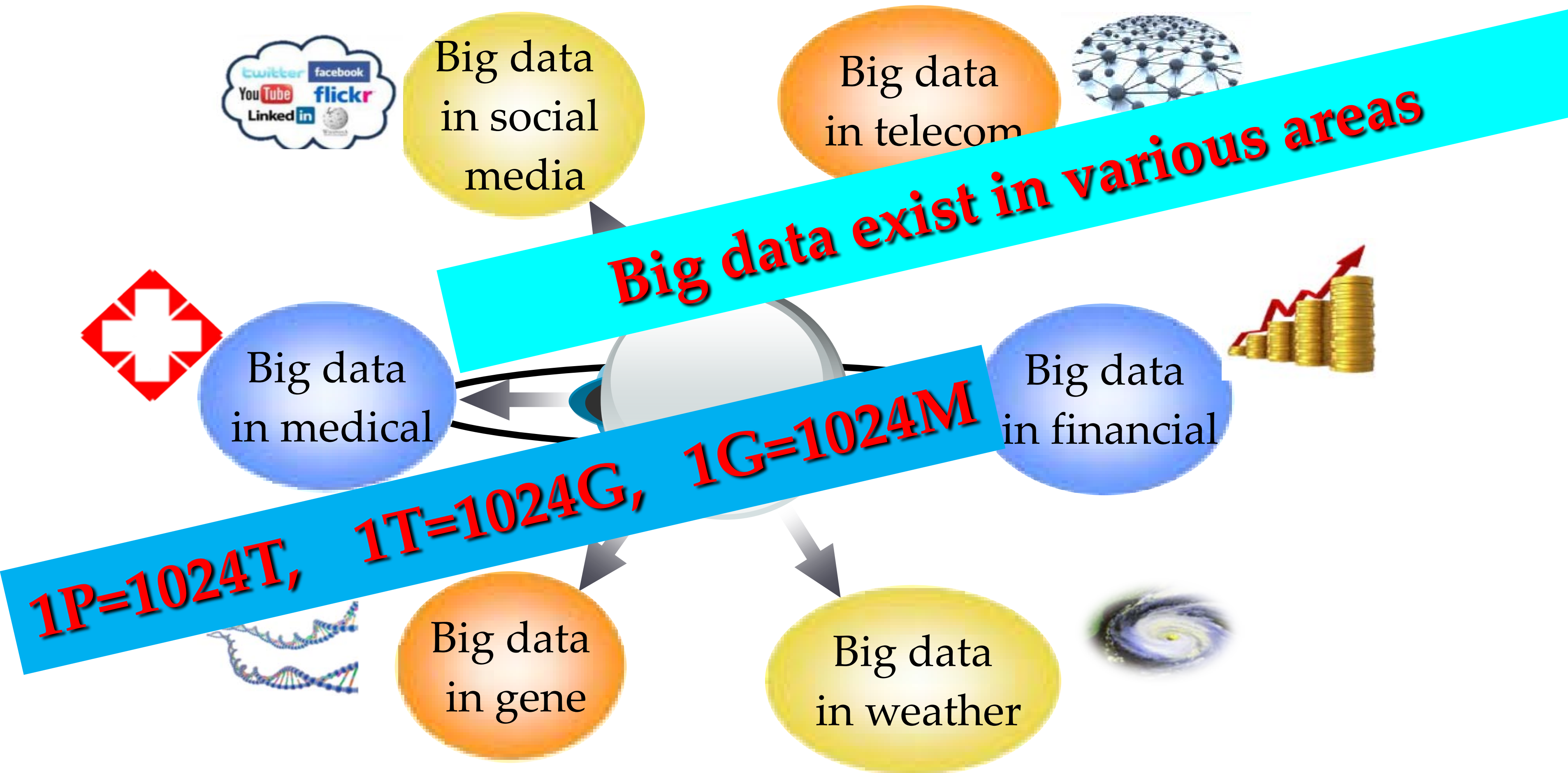
1. Introduction to big data (2)

Big data and its 5V properties

Volume	Velocity	Variety	Veracity	Value
				
Large Volume	High Velocity	Many Velocities	Difficult to Verify	High value
IDC Analyze the Future	facebook			
40 ZB by 2020, 5.2 TB per person	2.5B items/day, Over 500 TB/day		News and rumors from twitters, SMS, and blogs	Great value to countries and industries

1. Introduction to big data (3)

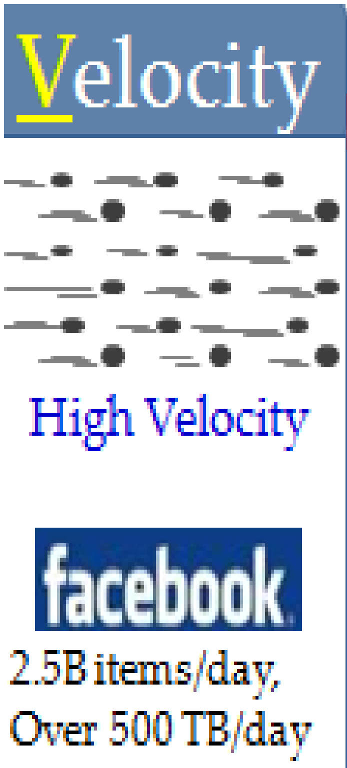
The 1st V: Volume - big and big



Big data= "Large-scale data" +"Complex types data"

1. Introduction to big data (4)

The 2nd V: Velocity (Quick change)



Tencent data storage in total: >1000PB
Incremental increase: 600TB/Day

Google data: 3 billion searches every day, 34 thousand questions per second

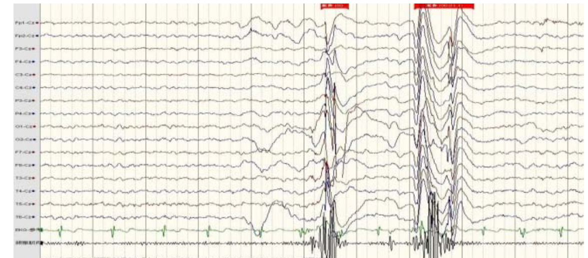
Baidu data: the number of handling searches for one day is about 5 billions, more thank Google

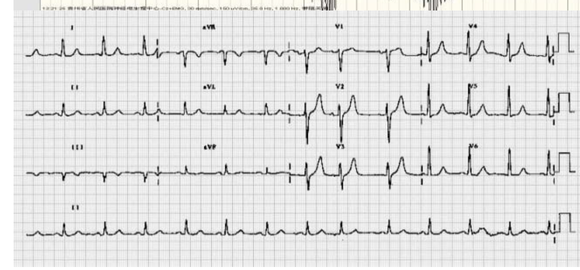
1. Introduction to big data (5)

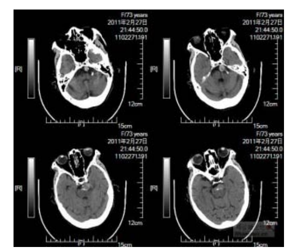
The 3rd V: Variety (Multimodality)

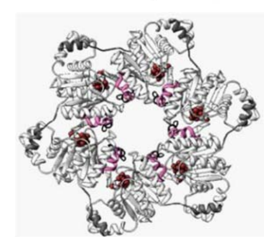
Regarding a big dataset, usually the types of feature are many

Numerical data
Symbolic data
Text
Image
Video
Time series
… … …



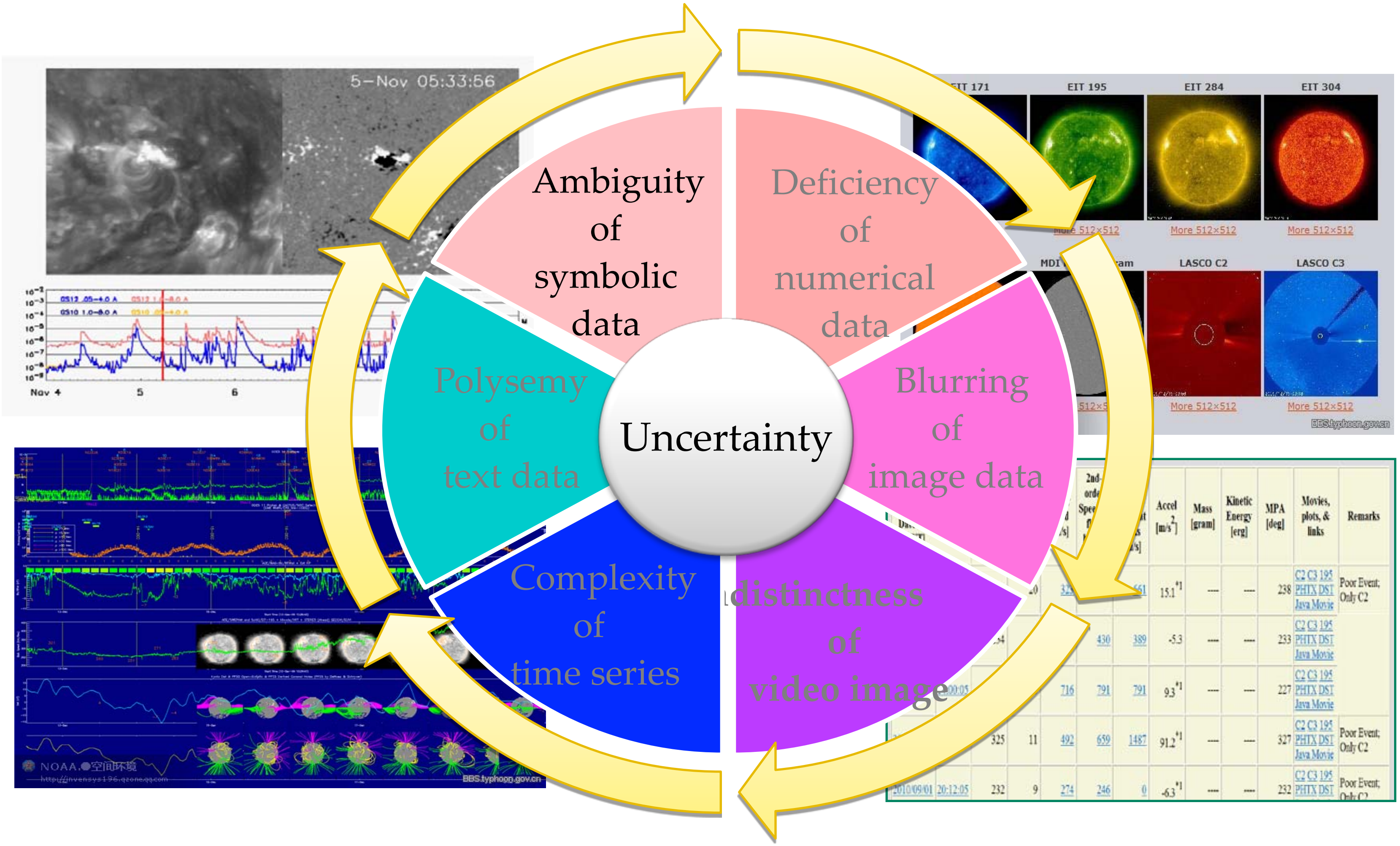






1. Introduction to big data (6)

The 4th V: Veracity (Uncertainty)



1. Introduction to big data (7)

The 5th V: Value - Big value for big data

Implicit, not explicit

Connections among the events, changing tendency, outliers, regularity.

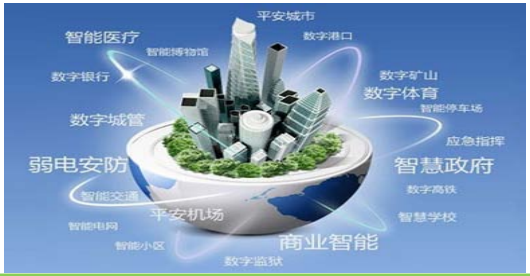
Need tools to mine.

Science research

- Astronomy
- High-energy physics
- Bioscience
- Machine design
- ……

Economy and Society

- Promoting the Internet of Things and cloud computing
- Business models of big data
- core competency of enterprises
- Influences on economic, social, and cultural
- ……



National governance

- Data assert
- National digital sovereignty
- Defense security monitoring
- Network security
- ……

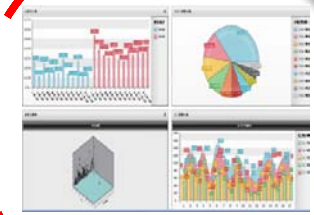
1. Introduction to big data (8)

Data analysis/mining is an important sector of big data industry chain

- model innovation
- Strategy Optimization
- grasping rules
- decision support
- ....

Data application

- data representation
- data modeling
- efficient computing
- statistical inference
- ......



Data analysis/mining

- storage pattern
- searching algorithm
- real-time processing
- ....

Data management

- data quality
- data standard
- data sharing
- data privacy
- ....

Data generation

From data to industry chain of value

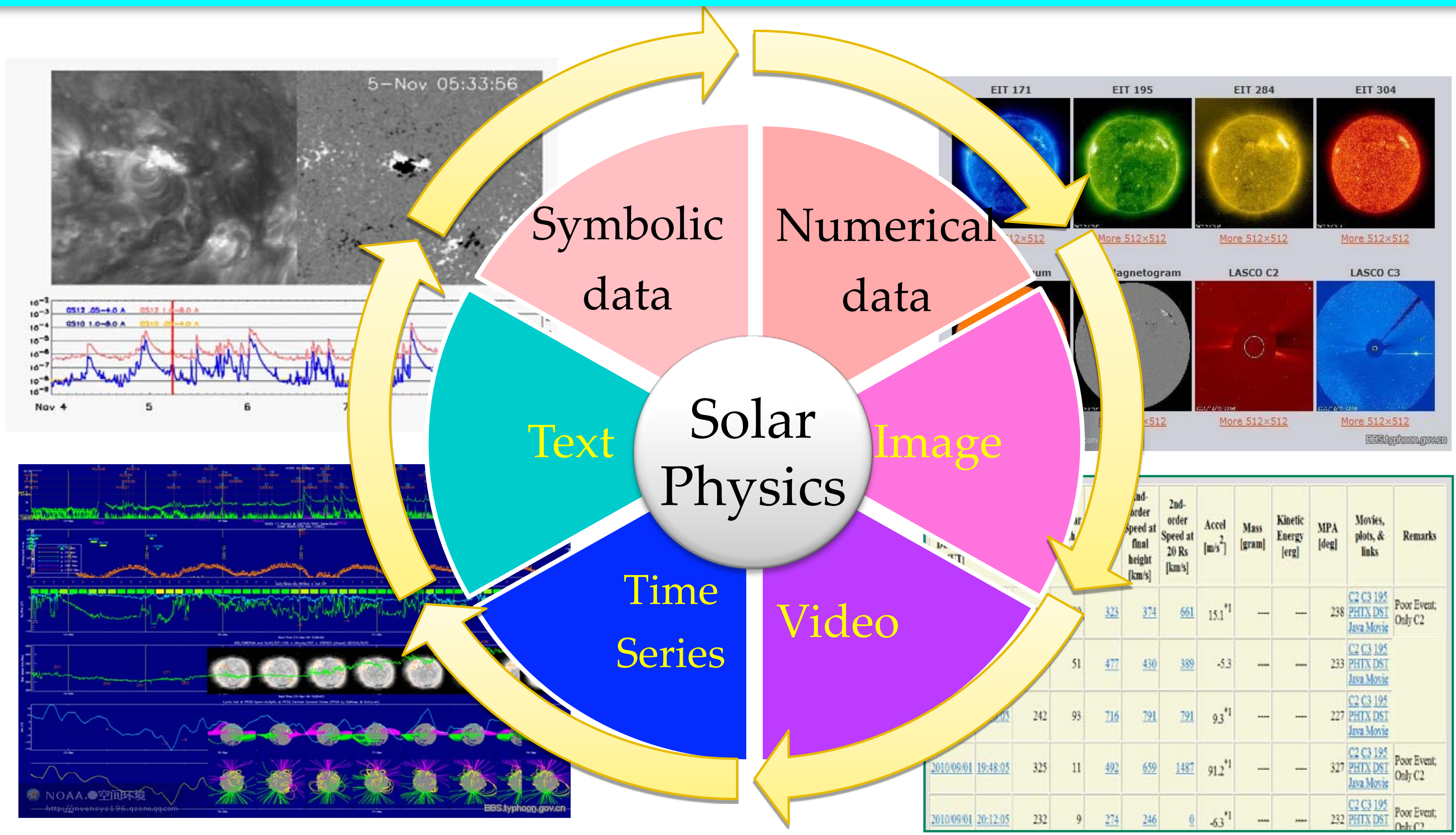
2. General challenge of big data analytics

- Big data representation
- Big dimensionality and Massive classes
- Weak relation: from a mapping to a relation
- Computation ability non-adaptive
- Poor transplant
- Hubness in high-dimensional space
- Uncertainty-strengthening of Big Data

2. Challenge of Big Data analytics (1)

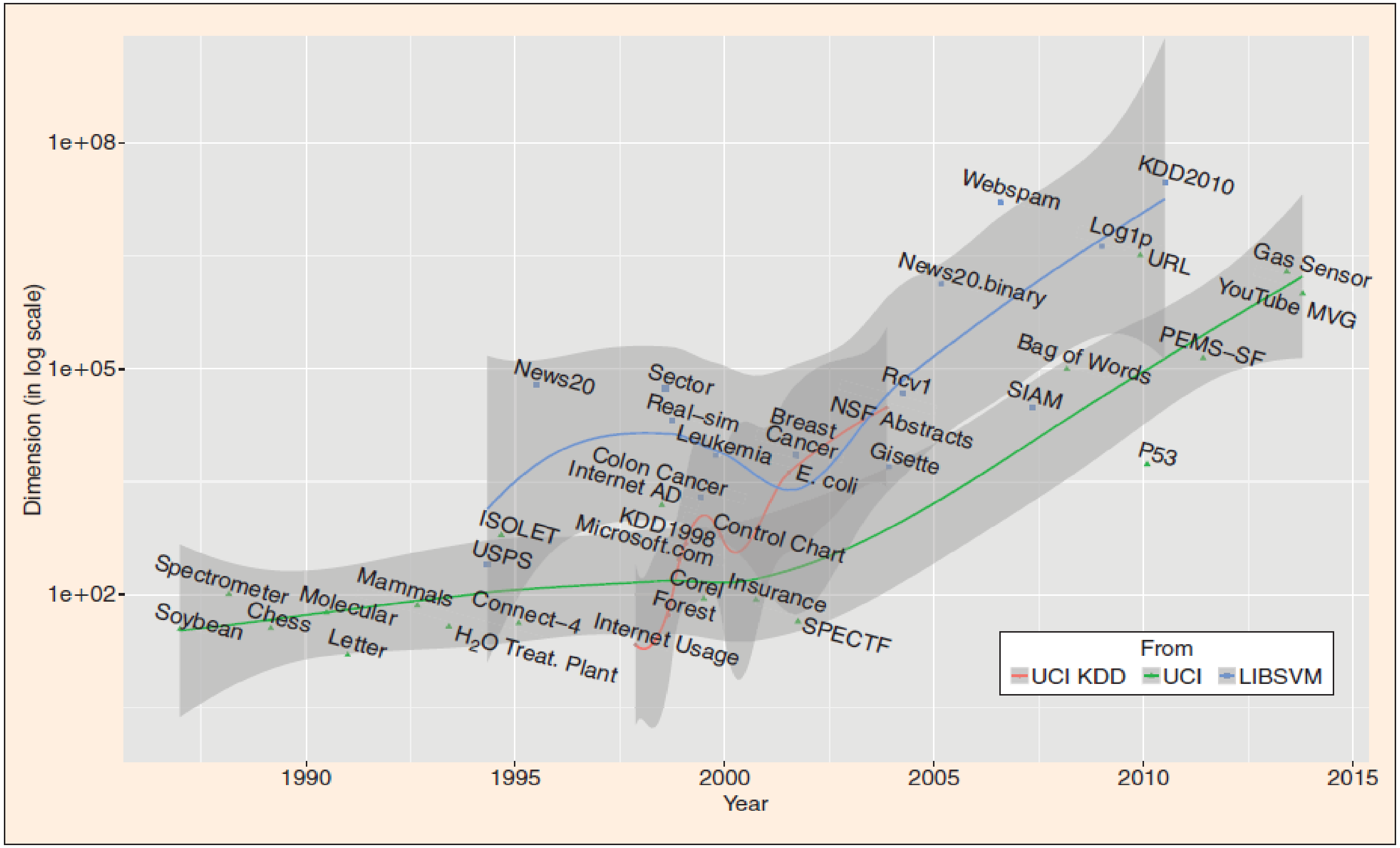
Big data representation

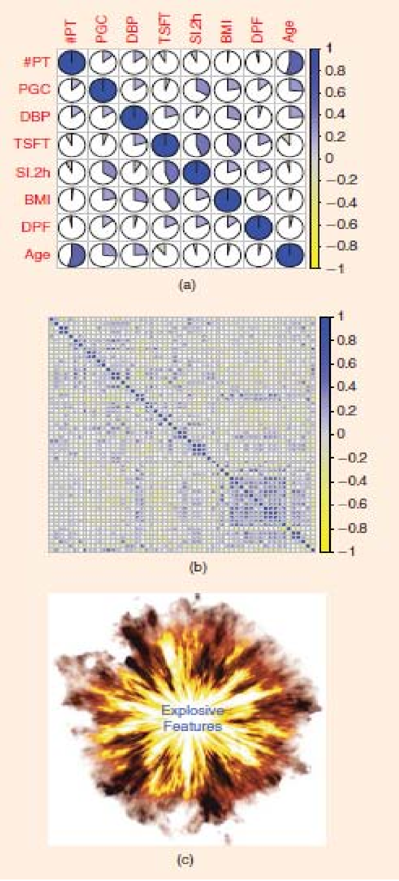
Challenge 1: Data representation



2. Challenge of Big Data analytics (2)

Challenge 2: Big Dimensionality

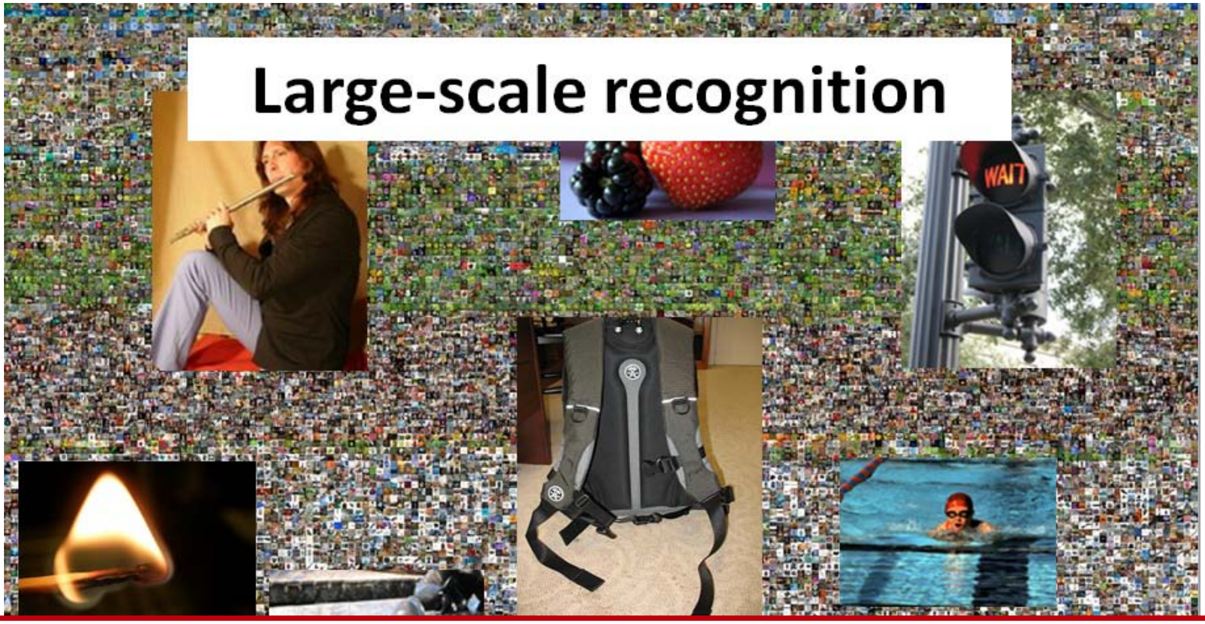




Zhai, Ong, Tsang. The emerging "Big dimensionality". IEEE CI Magazine, 2014

2. Challenge of Big Data analytics (3)
Massive Classes

Challenge 3: Massive Classes

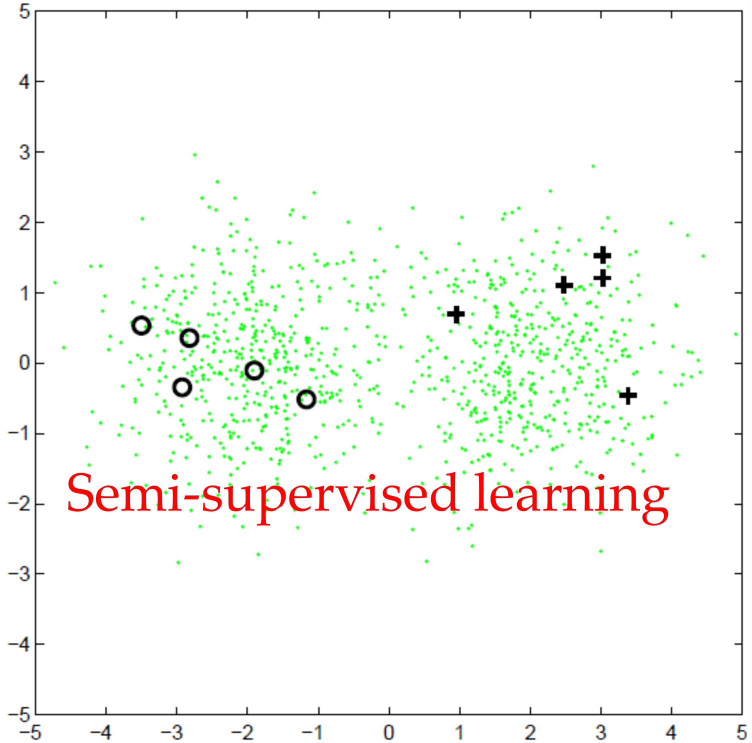


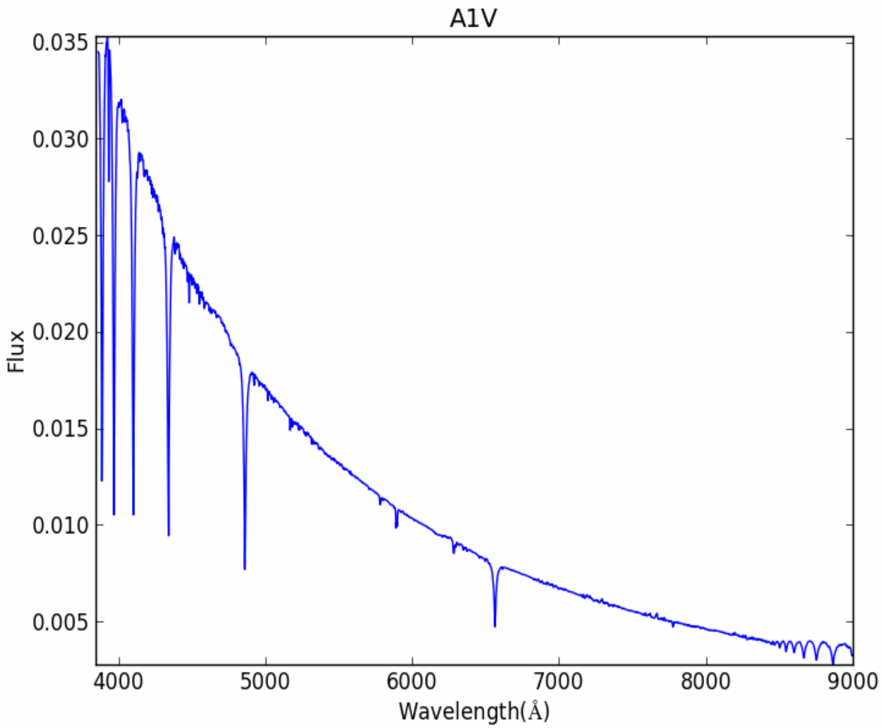
	16k ImageNet	22k ImageNet	21k Web Data	97k Web Data
Number of Classes	15,589	21,841	21,171	96,812
Number of Samples	9 million	14 million	9 million	40 million
Number of Features	1024	479	1024	1024

Gupta, Bengio, Weston. Training Highly multiclass classifiers. JMLR, 2014

2. Challenge of Big Data analytics (4)

Challenge 4: Weak relation: from a mapping to a relation





In classification tasks, the labels may be missing or labeled by error.

2. Challenge of Big Data analytics (5)

Challenge 5: Computational ability not scalable

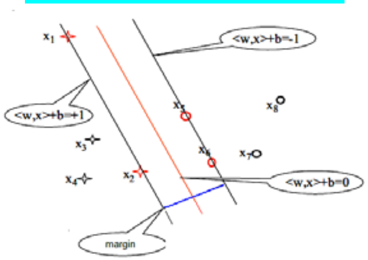
Computational Ability

NOT match to

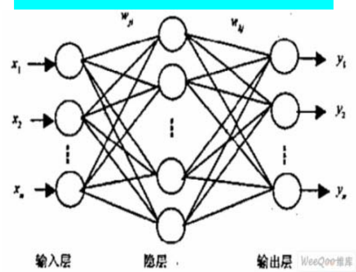
Problem Scale (Complexity)

Traditional machine learning methods

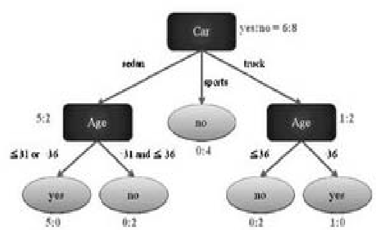
SVM



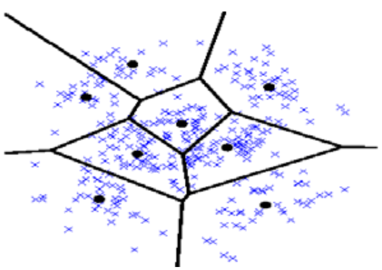
NN



DT



C-Mean



The computational cost of many algorithms≥O(*N*2)

Only efficient to small data!!!

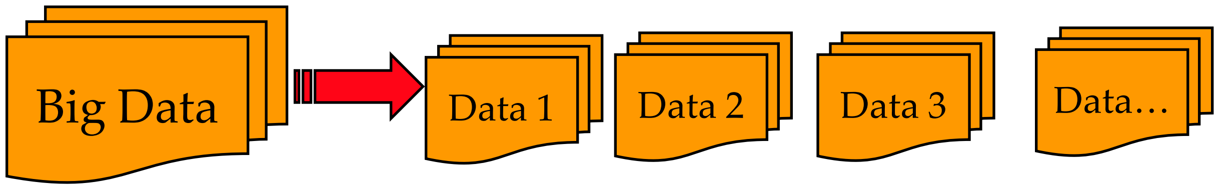
How to apply these methods to Big Data???

3. Current handling strategy for big data computing

- Divide-and-Conquer
- Parallelization
- Incremental
- Sampling
- Granular Computing
- Feature selection for big data
- Hierarchical classes

3. Current strategy for big data computing (1)

Divide-and-Conquer



It is general strategy
Becoming big to small
Processing in every small-block
Separate results are then fused together.

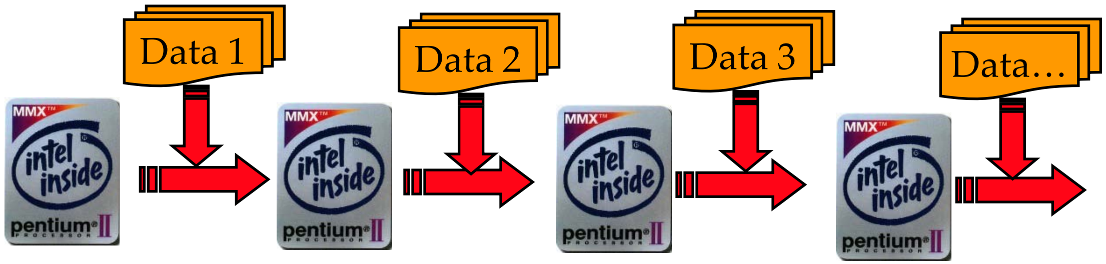
Michael Jordan in his keynote speech at the 14th Computing in the 21st Century Conference highlighted: Divide-and-Conquer and Statistical Inference is the fundamental strategy of processing Big Data

3. Current strategy for big data computing (2)
Parallelization

(1) Parallelization: another fundamental strategy
(2) Parallelization: Not decrease workload
(3) Parallelization: reduce working hours
(4) Parallelization: Not suitable for all problems

3. Current strategy for big data computing (3)

Increment



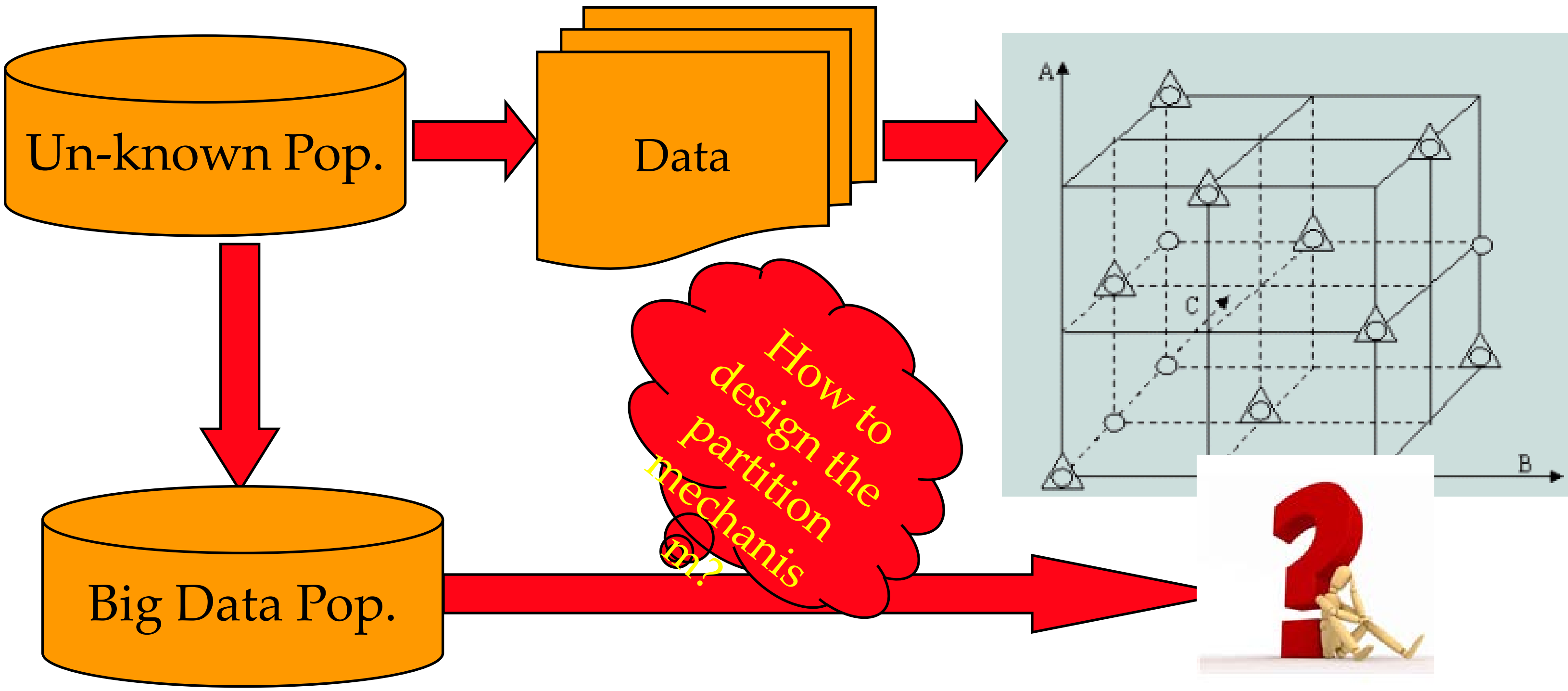
A general strategy for big data processing
Batch-data or streaming-data
A step-by-step learning process
Training only on the a new-coming data-block
A data-block used for training once only

Requirement: Good memory of the algorithm

3. Current strategy for big data computing (4)

Sampling

- A basic strategy: from big to small
- An old technique in Prob. & Statistics
- Relation between a sample and the population
- New advances with big data era is coming



3. Current strategy for big data computing (5)

Very low complexity learning algorithm

A very low complexity learning algorithm

Suitable for both classification and regression

Can efficiently handle Big Data

Fast training speed
(This strategy does focus on this advantage)

Simple structure

Support the online sequential learning tasks

Strong generalization ability

3. Current strategy for big data computing (6)

Granular Computing

Granular Computing: another specific strategy to process big data

It is still to becoming big to small, by selecting an appropriate granularity

Contents lists available at ScienceDirect

Information Sciences

/ins

CrossMark

Data-intensive applications, challenges, techniques and technologies: A survey on Big Data

C.L. Philip Chen *, Chun-Yang Zhang

Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China

6. Underlying technologies and future researches

The advanced techniques and technologies for developing Big Data science is with the purpose of advancing and inventing the more sophisticated and scientific methods of managing, analyzing, visualizing, and exploiting informative knowledge from large, diverse, distributed and heterogeneous data sets. The ultimate aims are to promote the development and innovation of Big Data sciences, finally to benefit economic and social evolutions in a level that is impossible before. Big Data

6.1. Granular computing → Granular computing

When we talk about Big Data, the first property of it is its size. As *granular computing* (GrC) [142] is a general computation theory for effectively using granules such as classes, clusters, subsets, groups and intervals to build an efficient computational model for complex applications with huge amounts of data, information and knowledge, therefore it is very natural to employ granular computing techniques to explore Big Data. Intuitively, granular computing can reduce the data size into different level of granularity. Under certain circumstances, some Big Data problems can be readily solved in such way.

3. Current strategy for big data computing (7)

Feature selection for Big Dimensionality

How to scale to ultrahigh dimensional feature selection task on big data

- **only a small subset of features or kernels are involved in the sub-problem optimization**
- **avoids the storing of all base kernels or the full explicit feature mappings.**
- **a modified accelerated proximal gradient method to accelerate**
- **several cache techniques are proposed to further enhance the efficiency.**

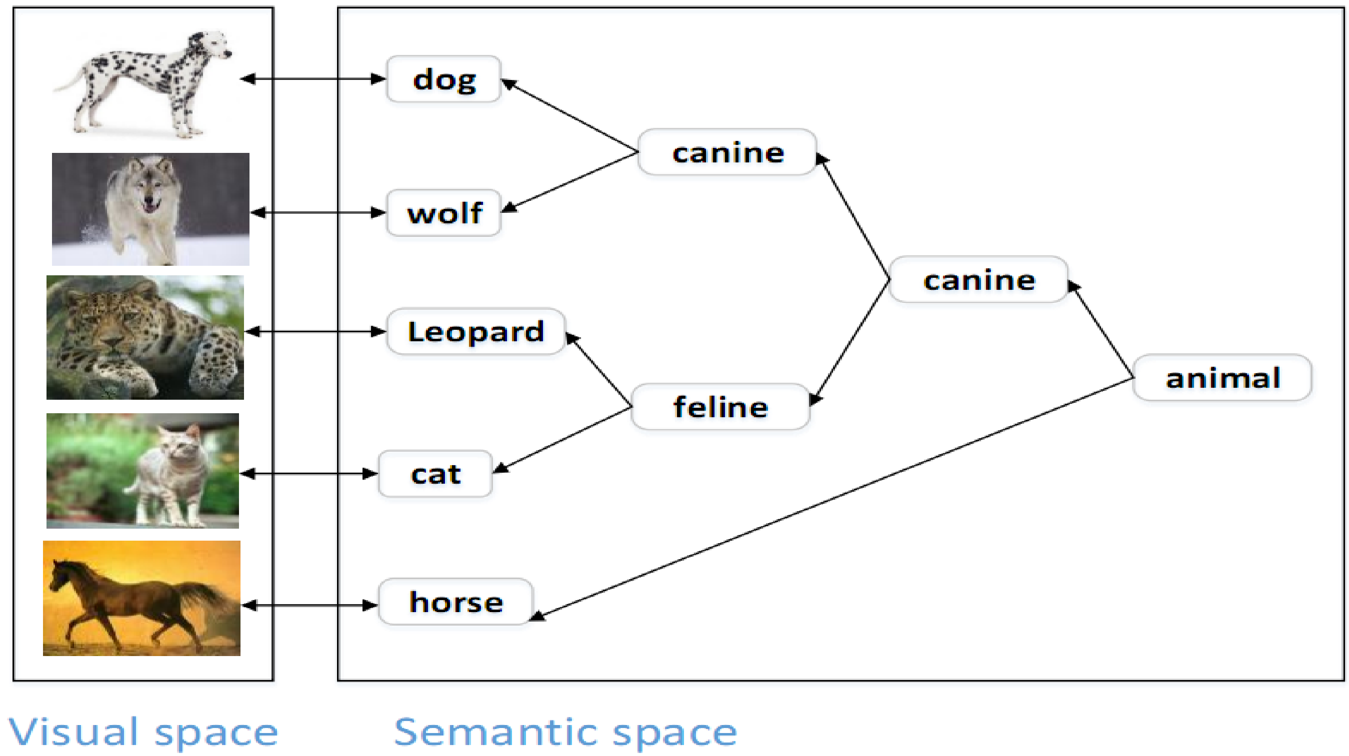
million training examples ($O(10^7)$)

100 trillion features ($O(10^{14})$)

Tan M, Tsang I W, Wang L. Towards ultrahigh dimensional feature selection for big data [J]. Journal of Machine Learning Research, 2014, 15(2):1371-1429.

3. Current strategy for big data computing (8)
Constructing hierarchical classes for Massive class problem

- In real word, objects are often organized into a class hierarchy - a tree.
- In the semantic space, classes are not isolated. Some classes are similar or grouped.
- Boosting the classification efficiency.



4. What new challenges the uncertainty brings in big data learning

Any impact of uncertainty on big data learning?

Including uncertainty's
representation, measure, and processing

What role it plays?
How does it influent the big data reduction?
Why does it work?

4. Uncertainty-based learning for big data (1)

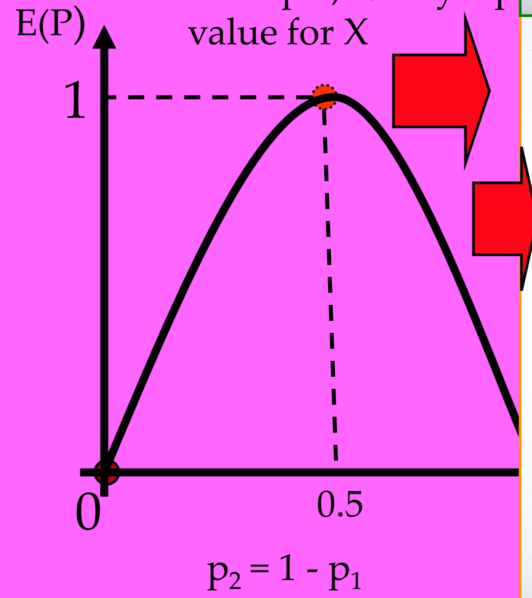
Definition of Uncertainty

Uncertain

Shannon entropy

Cl

Example, totally 2 p
value for X



In 1
Amb
subs

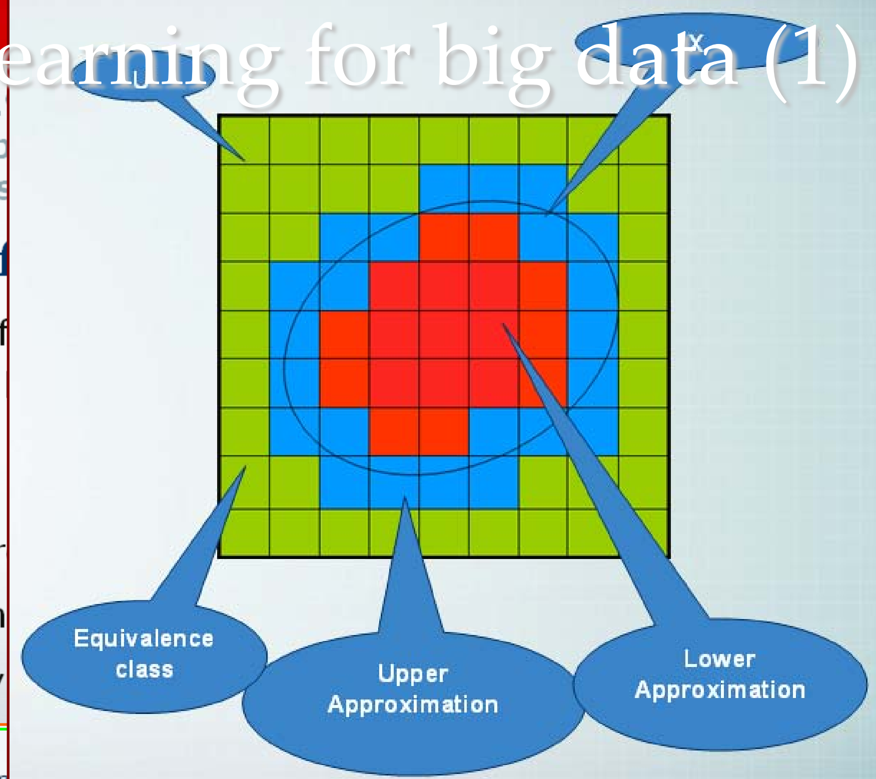
- Def

For f
(Am

wher

mem
any i,

Spec



- $(\underline{R}X, \overline{R}X)$ is called **Rough Set** of X,

where $\underline{R}X = \{x \in U \,|\, [x]_R \subseteq X\}$, and

$$\overline{R}X = \{x \in U \,|\, [x]_R \cap X \neq \varnothing\}$$

Rough-degree:

$$\beta_R(X) = 1 - \frac{|\underline{R}X|}{|\overline{R}X|}$$

4. What new challenges the uncertainty brings in big data learning

Summary of Uncertainty Definition

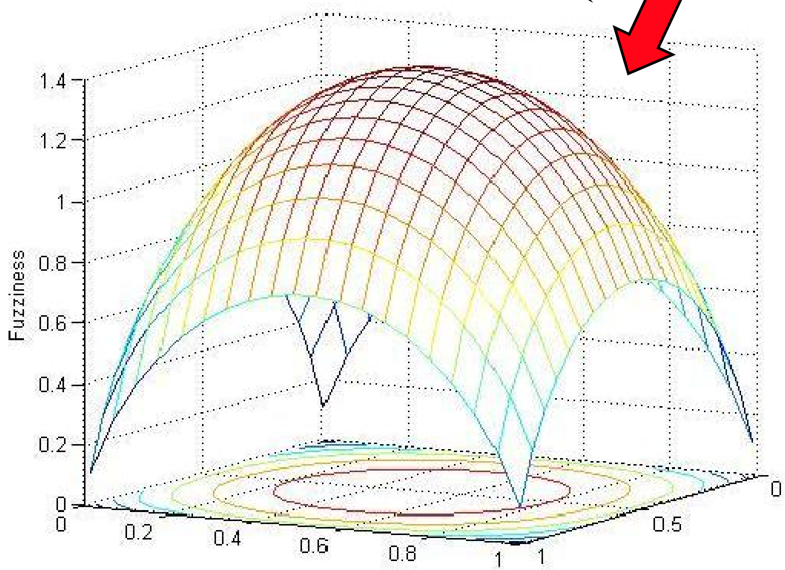
Uncertainty	of an object	with description
Shannon entropy	Probability distribution	Uncertainty caused by randomness
Classification entropy	Crisp set	Impurity of the class distribution in a set
Fuzziness	Fuzzy set	Uncertainty of a linguistic term
Non-specificity	Fuzzy set	Non-specificity when choosing one from many available choices.
Rough-degree	Rough set	Upper / lower approximation

4. What new challenges the uncertainty brings in big data learning

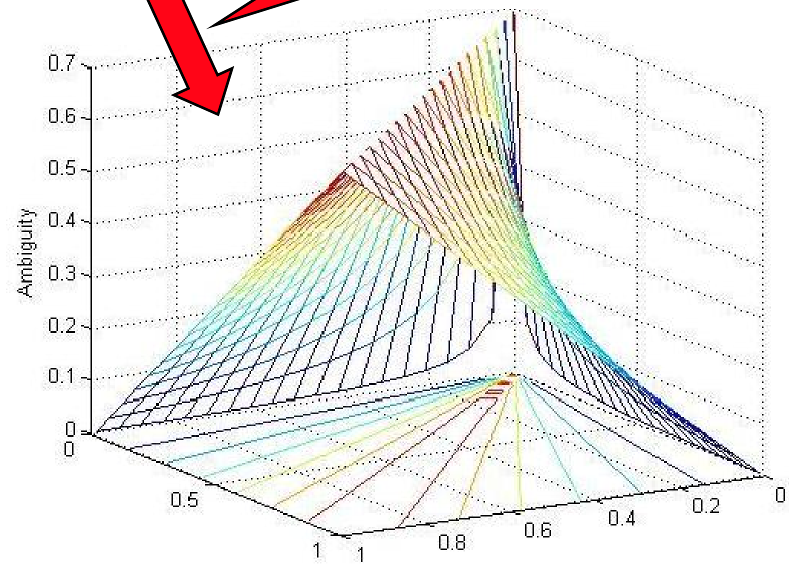
Fuzziness

Fuzzy set $A=\{\mu_1,\mu_2\}$

Ambiguity



(a) Fuzziness



(b) Ambiguity

Fig. a. Fuzziness with the formula

$$E_v(A) = -\frac{1}{2}\sum_{i=1}^{2}\left[\mu_i \log_2 \mu_i + (1-\mu_i)\log_2(1-\mu_i)\right]$$

Fig. b. Ambiguity with the formula

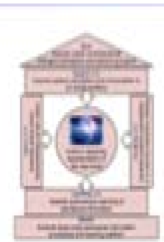
$$E_a(A)=\mu_1/\mu_2$$

Relation between 2 uncertainties: fuzziness & ambiguity

4. What new challenges the uncertainty brings in big data learning

Uncertainty – an important challenge for big data machine learning

Future basic theories of AI are uncertain information processing and new machine learning methods.



Science and technology innovation 2030 major projects - artificial intelligence 2.

7 tasks of artificial intelligence are put forward and #6 of them is "eliminating data bias, otherwise it is better not to use them".

White House report
« Preparing for the Future of Artificial Intelligence »

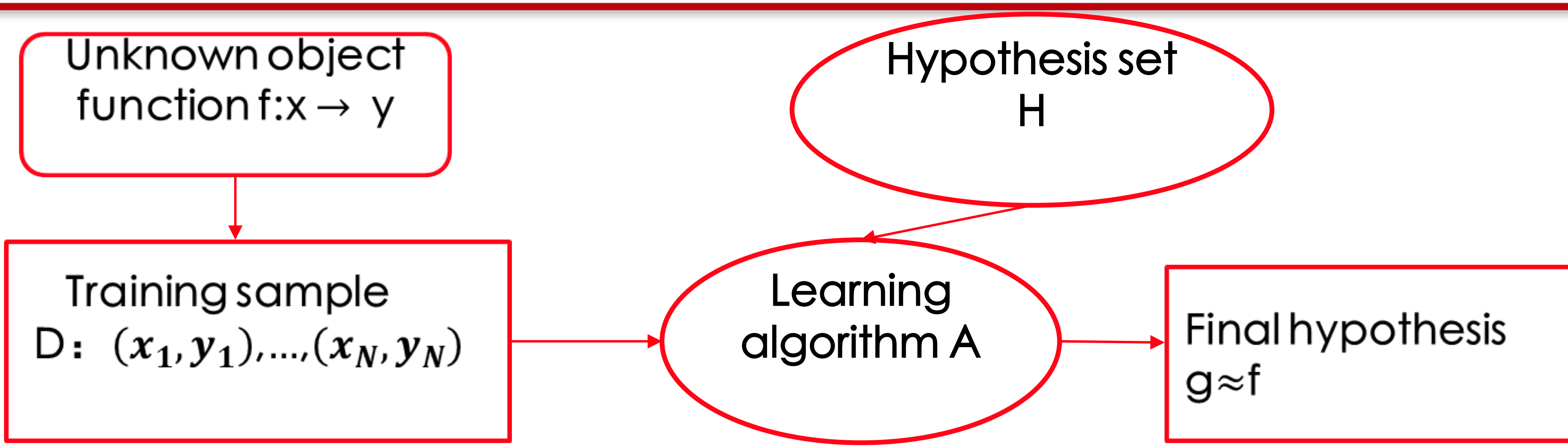
At the present stage, intelligent system needs to be adaptive in the open environment and robust to noise.

Thomas G. Dietterich
AI expert, President AAA116,
founding president of International Machine Learning Institute

How to model uncertainty under big data environment, to explore machine learning theory induced by uncertainty, and to develop fast and robust machine learning algorithms to cope with these challenges is the central task of machine learning in the era of big data.

4. What new challenges the uncertainty brings in big data learning

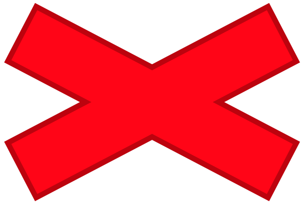
What challenges does uncertainty bring to big data machine learning?



Basic hypothesis of traditional learning

- Samples i.i.d
- Features basically complete
- Class distribution is balanced and supervised information is accurate
- Noise ~ Gauss distribution

Big data machine learning



4. What new challenges the uncertainty brings in big data learning

What challenges does uncertainty bring to big data machine learning?

In big data environment, uncertainty makes original hypothesis of machine learning destroyed and traditional learning algorithm unavailable.

Data uncertainty

- Samples are not i.i.d () → NOT i.i.d, Data relation cannot be ignored in modeling.
- Noise characteristics (misdiagnosis rate 30%-70%) → Gauss distribution assumption is destroyed in the data and error modeling process.
- Incomplete data (missing rate) → Data completeness assumption is destroyed
- Decision long tailed distribution (80% samples, 20% labels) → Large scale categories make data long tailed and cannot be ignored in modeling

Model uncertainty

- Model underdetermined (P= 3 billion base pairs) → Hypothesis $p/n \to 0$ and existence and uniqueness of solution are destroyed.
- Solution space is huge (Zeiler&Fergus, 140M) → f space is complex, nonlinear, discontinuous, and huge in parameter space
- Uncertainty of model output () → Different model outputs have same decision

Uncertainty of large-scale problem solving strategy

- Sampling (Facenet, 200M, human face) → A sampling strategy guided by learning is urgently needed
- Divide and conquer (whether assembled) → Divide-and-Conquer learning: sub-model combinations are the original model?
- Approximation (circulant matrix VS kernel matrix) → Approximation strategy brings new uncertainties

4. What new challenges the uncertainty brings in big data learning

Research status

Data uncertainty

The uncertainty of data mainly includes the uncertainty of feature space and the uncertainty of decision space

- Data noise (TNNL2014 AAAI2016)
  Single Gaussian→Gaussian Mixture
- Data missing (ICML2013 NIPS2015)
  Data deletion and Data imputation
- Data inconsistency (TKDE 2012)
  Inconsistency of rough set partition
- Not i.i.d. (TPAMI2016 TYCB2017)
  "High spectrum image denoising medical images"
- The category of long tail (JMLR TIP)

Mostly aimed at single uncertainty research

Problems: lack of data uncertainty measurement index system and its impact mechanism on learning

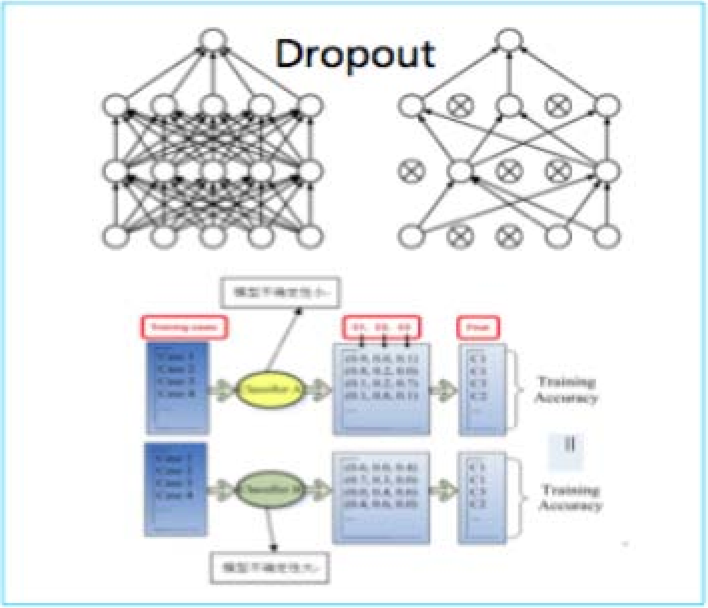
4. What new challenges the uncertainty brings in big data learning

Research status

Model uncertainty

It is mainly reflected in model underdetermined, huge solution space and uncertainty of model output

- Model to be underdetermined (Nature 2014)
  - Regularization
  - Augmented data
  - Data dimension reduction
- Huge solution space (JMLR 2014)
  - Pooling
  - Droupout
  - Sparse
  - Random weighting
- Uncertainty of model output (TFS2015)
  - Ensemble Learning
  - Multiple classifier system



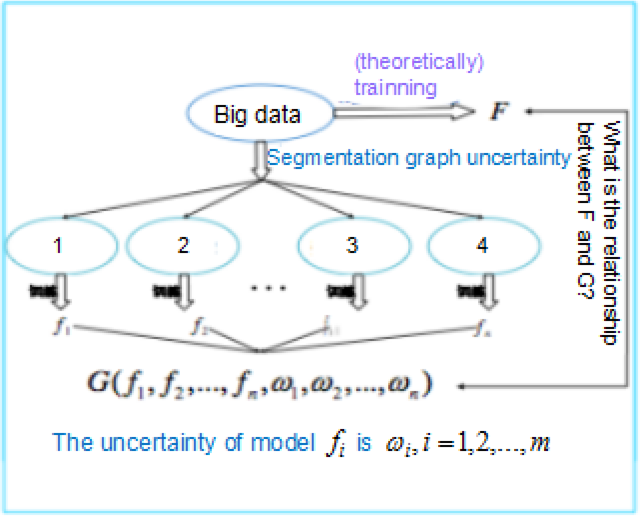
Problems: we have not yet systematically established the relationship between the uncertainty of model and model performance

4. What new challenges the uncertainty brings in big data learning

❖ The uncertainty of large-scale computation

Data itself has uncertainty meanwhile there exists uncertainty of the model in the process of stochastic optimization and divide-and-conquer learning

- Nonconvex smooth object function (Boyd)
  ADMM
  Stochastic optimization
- Approximate calculation (NIPS 2011)
  Stochastic gradient descent
  Stochastic coordinate optimization
- Sampling (VLDB Journal 2014)
  Gibbs Sampling
  Sparse Sampling
- Divide and conquer algorithm (ICML2014)
  Design for constructive algorithms



Problems: Sub-models fusion is the population model? If not, how to get an estimation and an error bound?

5. Some of our works related to learning from big data with uncertainty

5.1 Our 1st work is the study on relationships between outputted-uncertainty and classifier-generalization. More details can be found from the following 2 papers and 1 book. The major idea is presented in the following pages.

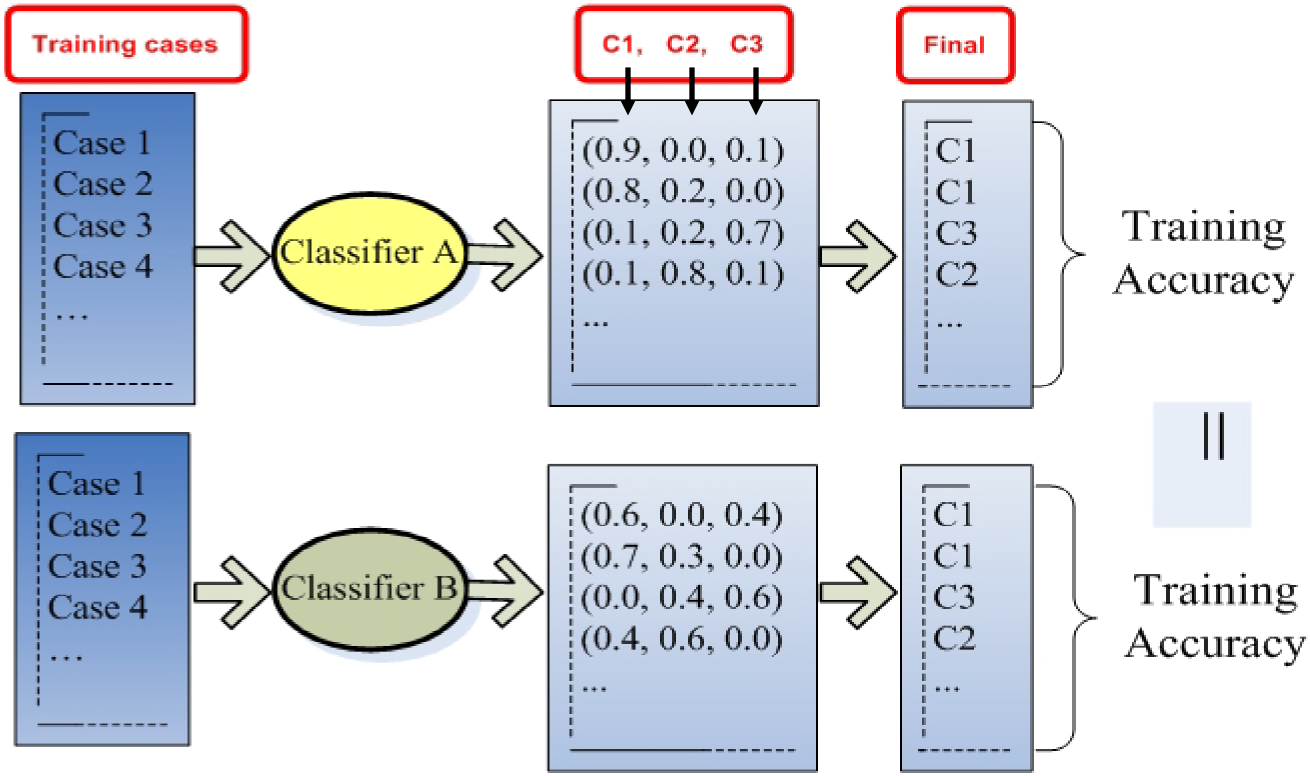
Xi-zhao Wang, Hong-Jie Xing, Yan Li, et al, A Study on Relationship between Generalization Abilities and Fuzziness of Base Classifiers in Ensemble Learning, IEEE Transactions on Fuzzy Systems, 2015, 23(5): 1638-1654

Xizhao Wang, Ran Wang(*), Chen Xu, Discovering the Relationship Between Generalization and Uncertainty By Incorporating Complexity of Classification, IEEE Transactions on Cybernetics, DOI: 10.1109/TCYB.2017.2653223

Xizhao Wang and Junhai Zhai, Learning with Uncertainty, CRC Press Talor & Francis Group, 2016

5. Some of our works related to learning from big data with uncertainty

Basic idea



Same training accuracies, but different uncertainties
Do you think the uncertainty as a key factor?

5. Some of our works related to learning from big data with uncertainty

Experimental verification

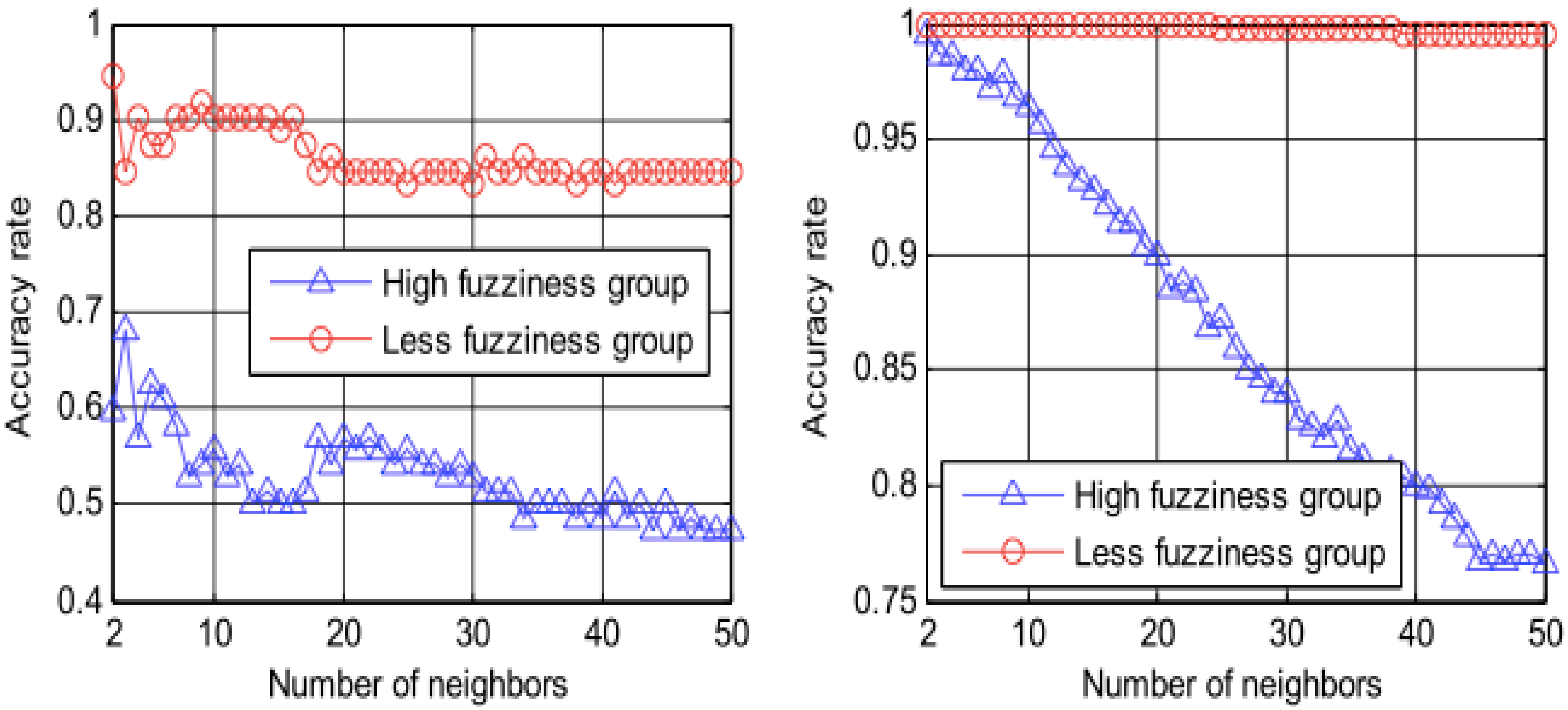


Fig. 7. Difference of testing accuracy rates between the high-fuzziness group and low-fuzziness group.

5. Some of our works related to learning from big data with uncertainty

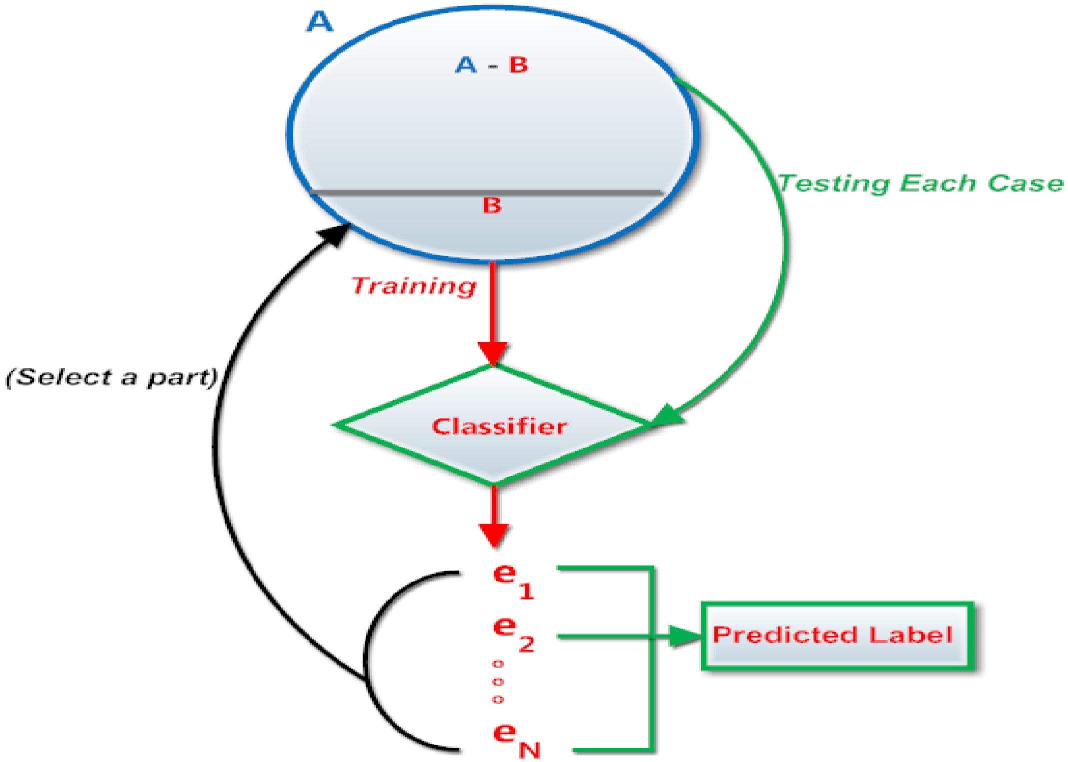
5.2 Our 2nd work is the study on semi-supervised learning based on uncertainty measure for big data. More details can be found from the following 3 papers. The major idea is presented in the following pages.

Xi-zhao Wang, Ling-Cai Dong, Jian-Hui Yan, Maximum ambiguity based sample selection in fuzzy decision tree induction, IEEE Transactions on Knowledge and Data Engineering, 2012, 24(8): 1491-1505

Xi-zhao Wang, Rana Aamir and Ai-Min Fu, Fuzziness based sample categorization for classifier performance improvement, Journal of Intelligent & Fuzzy Systems 29 (2015) 1185–1196

Ran Wang, Xi-zhao Wang (*), Sam Kwong, and Chen Xu, Incorporating Diversity and Informativeness in Multiple-Instance Active Learning, Accepted in May 2017; IEEEE Transactions on Fuzzy Systems

5. Some of our works related to learning from big data with uncertainty



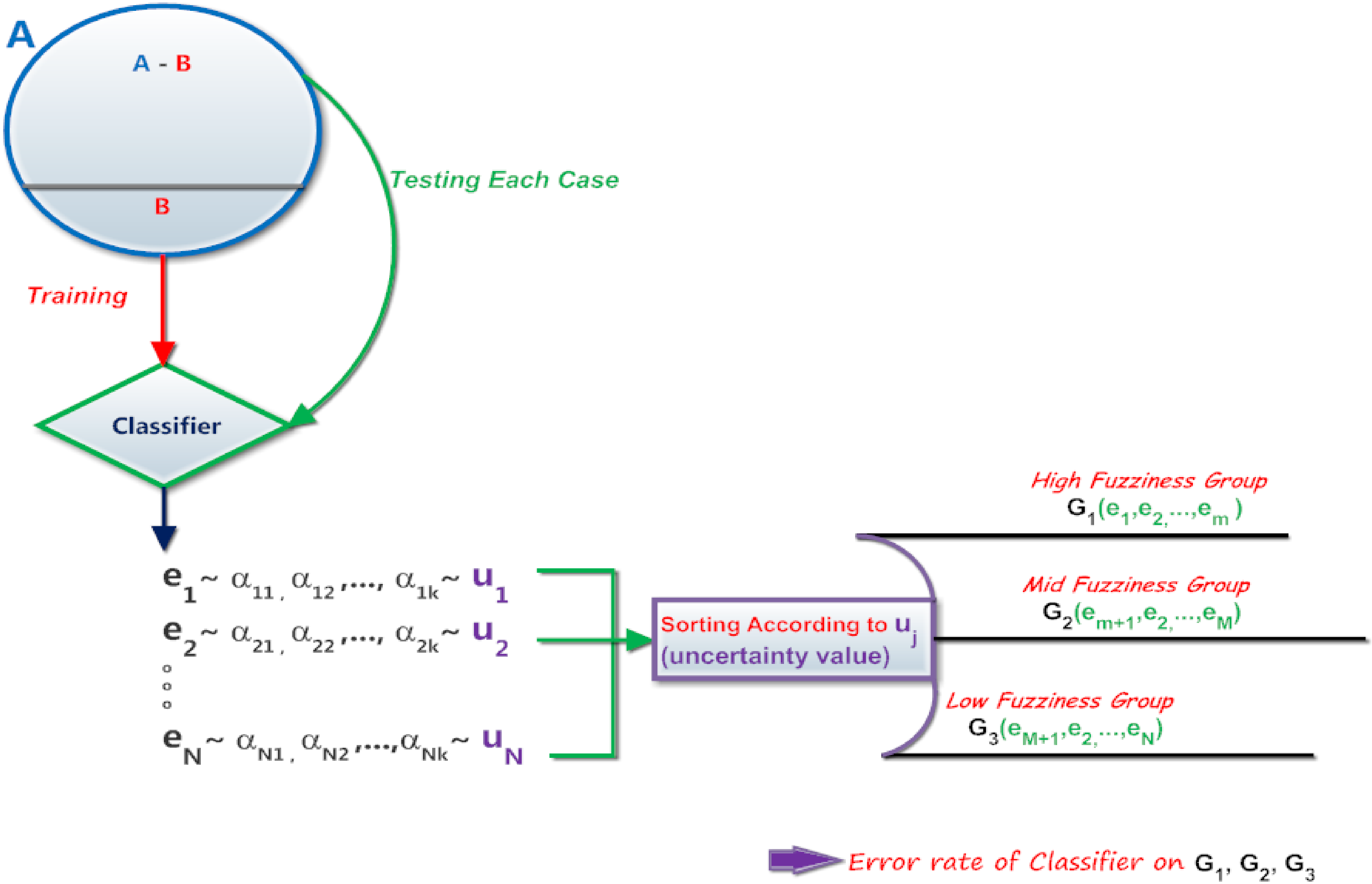
- Impossible to label each case
- Unnecessary to label most cases
- A strategy of Divide-and-Conquer
- An approach to semi-supervised learning
- High fuzziness samples important
- Explained as extreme entropy principle

After One Loop : $B \uparrow B_1$, $A - B \downarrow A - B_1$

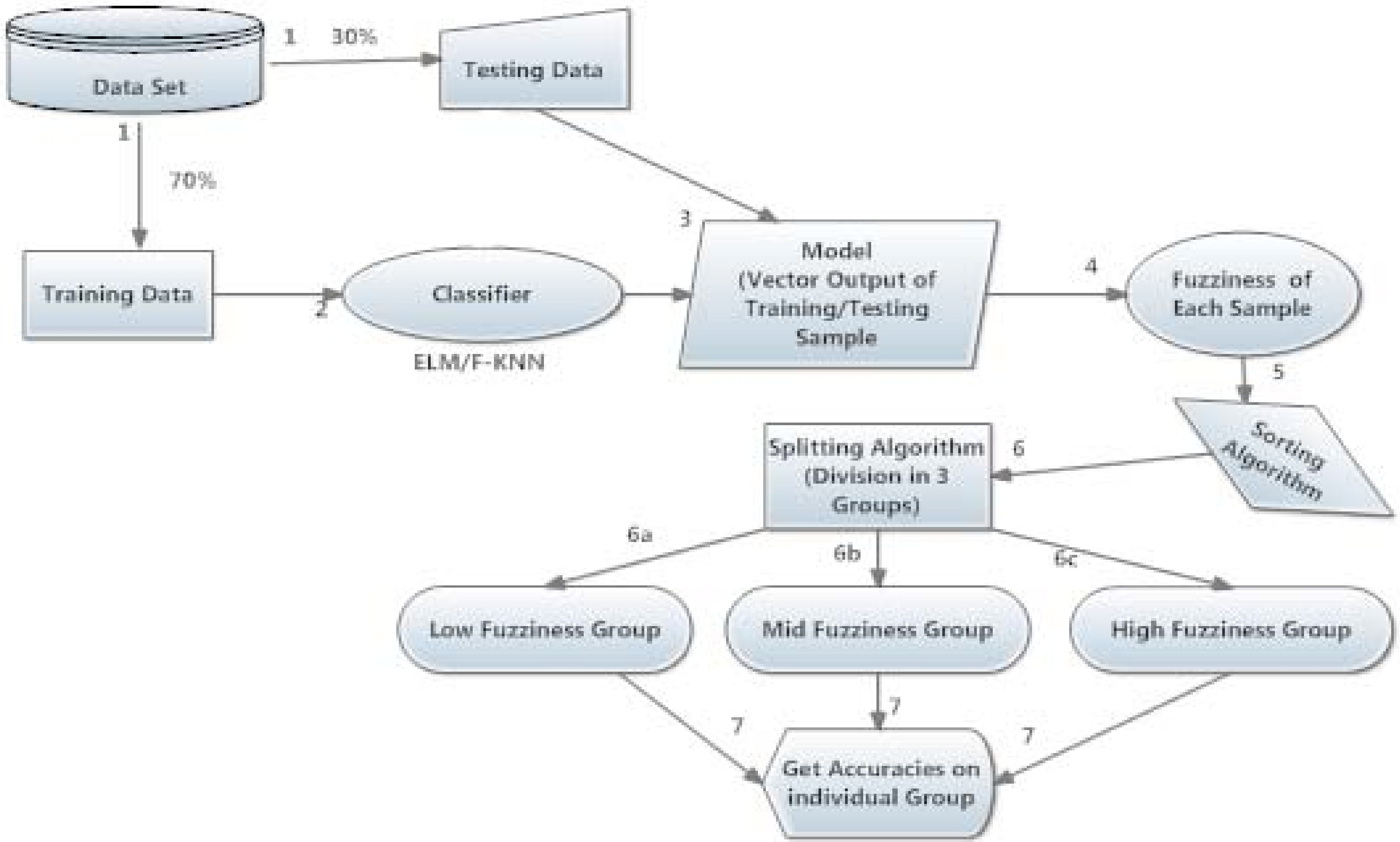
Expect : Classifier from B_1 is better than Classifier from B

1. **Any requirement for the trained classifier?**
2. **What selection strategy?**

5. Some of our works related to learning from big data with uncertainty



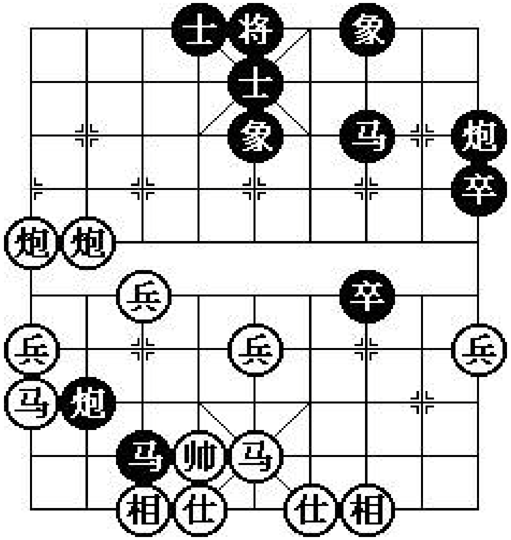
5. Some of our works related to learning from big data with uncertainty



5. Some of our works related to learning from big data with uncertainty

We collect a data set for Chinese Chess-Game Scene Classification – CCGSC GGSC: the file-size is 1.86G, including more than 10^7 records of playing chess-game, and more than 10^9 scenes of chess-game

Three types of scenes:
A-Good
B-Good
No-result



That is a semi-supervised learning, with unstructured data.
Large amount of scenes need to label – labeling needs senior experts (Chess Masters) to do for complicated scenes, and so, quite expensive cost.

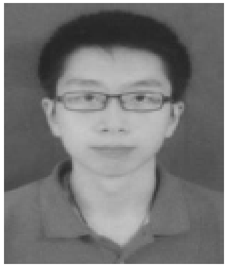
The experimental results show a very high prediction

5. Some of our works related to learning from big data with uncertainty

5.3 Our 3rd work is on deep learning with uncertainty. More details can be found from:

Xizhao Wang, Tianlun Zhang, Ran Wang(*), Non-Iterative Deep Learning: Incorporating Restricted Boltzmann Machine into Multilayer Random Weight Neural Networks, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Volume: 47 Issue: 8, DOI: 10.1109/TSMC.2017.2701419

Abstract—A general deep learning (DL) mechanism for a multiple hidden layer feed-forward neural network contains two parts, i.e., 1) an unsupervised greedy layer-wise training and 2) a supervised fine-tuning which is usually an iterative process. Although this mechanism has been demonstrated in many fields to be able to significantly improve the generalization of neural network, there is no clear evidence to show which one of the two parts plays the essential role for the generalization improvement, resulting in an argument within the DL community. Focusing on this argument, this paper proposes a new DL approach to train multilayer feed-forward neural networks. This approach uses restricted Boltzmann machine (RBM) as the layer-wise training and uses the generalized inverse of a matrix as the supervised fine-tuning. Different from the general deep training mechanism like back-propagation (BP), the proposed approach does not need to iteratively tune the weights, and therefore, has many advantages such as quick training, better generalization, and high understandability, etc. Experimentally, the proposed approach demonstrates an excellent performance in comparison with BP-based DL and the traditional training method for multilayer random weight neural networks. To a great extent, this paper demonstrates that the supervised part plays a more important role than the unsupervised part in DL, which provides some new viewpoints for exploring the essence of DL.



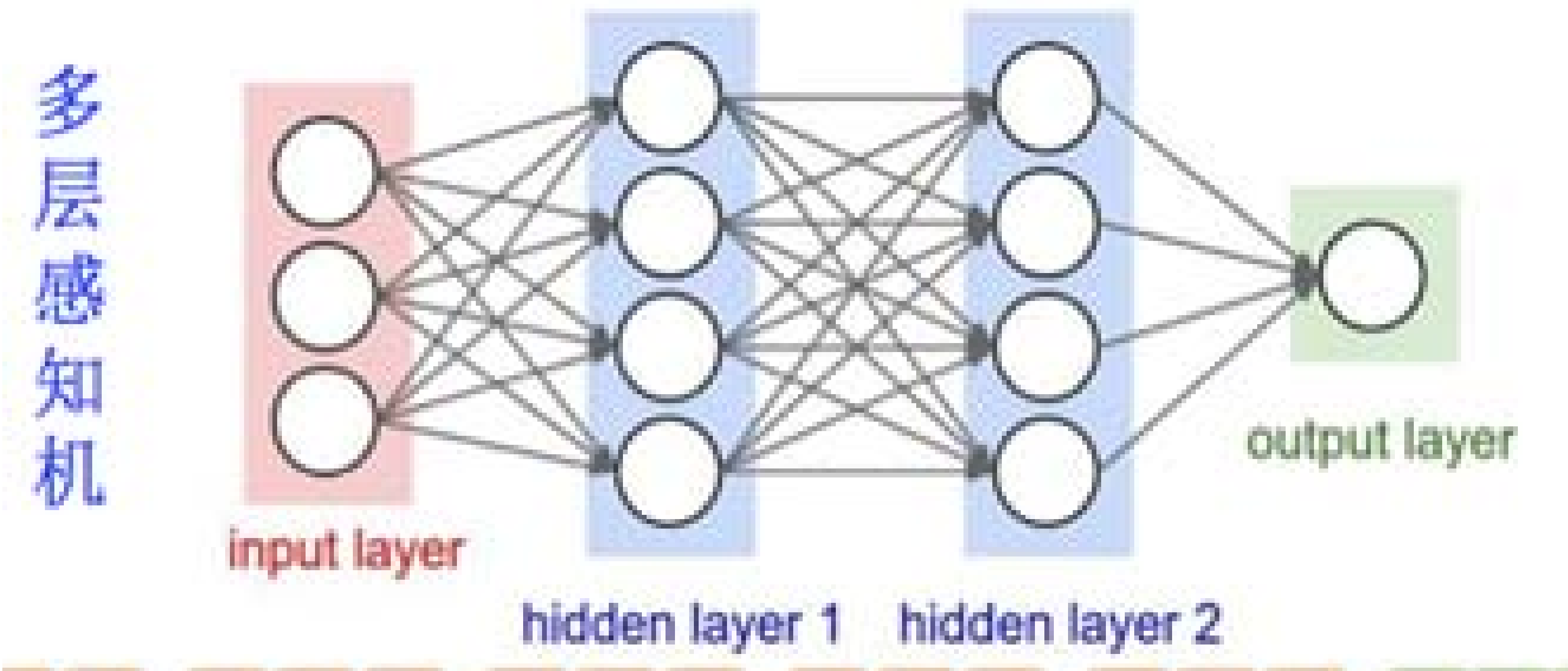
Tianlun Zhang received the bachelor's degree in information management and information system from the Industrial and Commercial College, Hebei University, Hebei, China, in 2014, and the M.Sc. degree in software engineering from the College of Mathematics and Information Science, Hebei University, Hebei, in 2016. He is currently pursuing the Ph.D. degree with the Information Science and Technology College, Dalian Maritime University, Dalian, China.

His current research interests include machine learning and pattern recognition.

Ran Wang (S'09–M'14) received the B.Eng. degree in computer science from the College of Information Science and Technology, Beijing Forestry University, Beijing, China, in 2009, and the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, in 2014.

From 2014 to 2016, she was a Post-Doctoral Researcher with the Department of Computer Science, City University of Hong Kong. She is currently an Assistant Professor with the College of Mathematics and Statistics, Shenzhen University, Shenzhen, China. Her current research interests include pattern recognition, machine learning, and fuzzy sets and fuzzy logic and their related applications.

5. Some of our works related to learning from big data with uncertainty



1. RBM+BP

The most fundamental training mechanism given by Hilton et al

2. RBM + Random Weight

Recently popular training mechanism

5. Some of our works related to learning from big data with uncertainty

Basically, the essences of RBM + RWA (Random Weight Assignment) are two-fold.

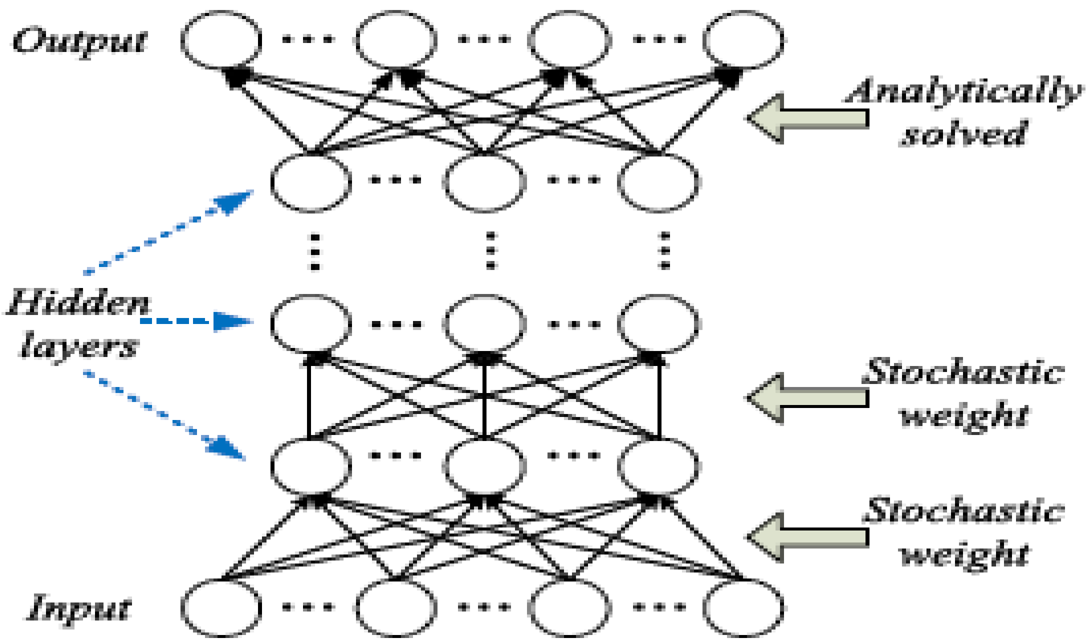
The first is the RBM weight initialization, which was first introduced into DL by Hinton.

The second is the RWA mechanism , which proves that most connecting weights are not necessarily to be iteratively tuned for a fully connected feed-forward neural network.

Viewing the essence of DL proposed by Hinton, *we can find that the* key parts are also twofold, i.e., 1) the RBM-based weight initialization and 2) the BP-based weight tuning. One interesting question is: which part is more important?

Our major idea:
phases 1) and 2) can be separately handled.
And phase 2) is not necessary by BP.

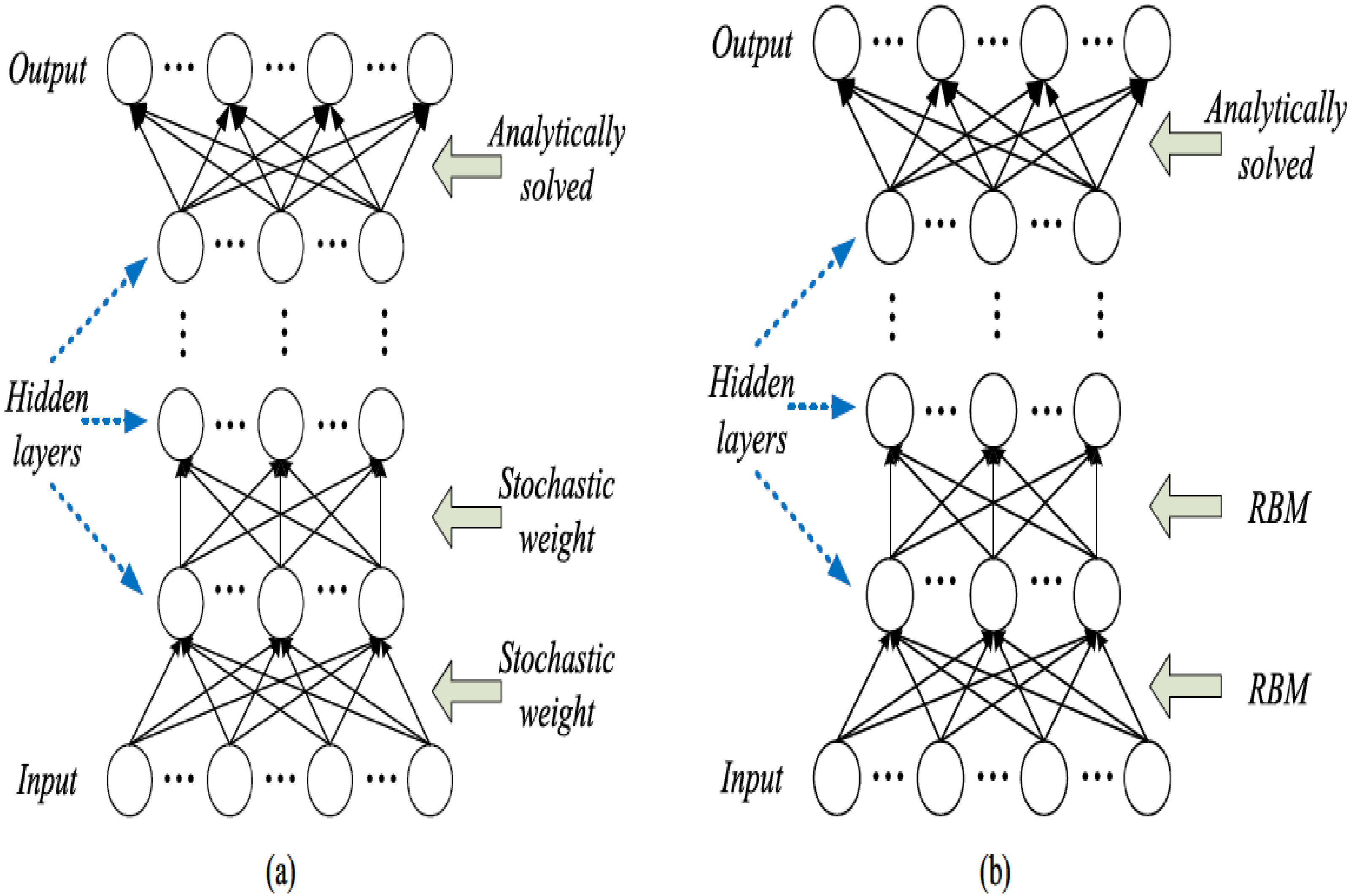
5. Some of our works related to learning from big data with uncertainty



(a)

- The central idea is to improve the RWA1 training algorithm (i.e., Algorithm 3) by replacing the random assignment of weights with RBM-based weight initialization for all hidden layer nodes. The determination of weights for output layer in this scheme is identical to that in RWA1.

5. Some of our works related to learning from big data with uncertainty



5. Some of our works related to learning from big data with uncertainty

Algorithm 4: RBM-GI [Fig. 3(b)]

Input:

Training set $\mathbb{X} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N} \in \mathcal{R}^L \times \{1, \ldots, C\}$;

Tag matrix $\mathbf{T}_{N\times C} = [\mathbf{y}_1^{\mathrm{T}}, \mathbf{y}_2^{\mathrm{T}}, \ldots, \mathbf{y}_N^{\mathrm{T}}]^{\mathrm{T}}$;

The number of hidden layers n;

The n-dimensional vector $[m_1, m_2, \ldots, m_n]$ where $m_j, j = 1, 2, \ldots, n$ is the number of nodes in the j-th hidden layer.

Output:

Output function.

Get the input feature matrix $\mathbf{X}$ by (16), let $\mathbf{H}_0 = \mathbf{X}$;

Take $\mathbf{H}_0$ as the input, call Algorithm 1 to learn the parameters between the input layer and the first hidden layer to get the weights and bias $(\mathbf{W}_1, \mathbf{b}_1)$;

for $i = 1$ *to* n **do**

Get the output matrix of the i-h hidden layer

$$\mathbf{H}_i = \text{sigmoid }(\mathbf{W}_i\mathbf{H}_{i-1} + \mathbf{b}_i);$$

Take $\mathbf{H}_i$ as the input, call Algorithm 1 to learn the parameters between the ith hidden layer and the $(i+1)$th hidden layer to get $(\mathbf{W}_{i+1}, \mathbf{b}_{i+1}, \mathbf{b}_i')$, where $\mathbf{b}_i'$ is called the reconstruction bias of $\mathbf{b}_i$;

end

Calculate the weight matrix β of the output layer

$$\beta = \left(\mathbf{H}_n^{\mathrm{T}}\mathbf{H}_n\right)^{-1}\mathbf{H}_n^{\mathrm{T}}\mathbf{T};$$

The output function is given by

$$\begin{aligned}
\mathbf{H}_1 &= \text{sigmoid}(\mathbf{W}_1\mathbf{H}_0 + \mathbf{b}_1) \\
\mathbf{H}_2 &= \text{sigmoid}(\mathbf{W}_2\mathbf{H}_1 + \mathbf{b}_2) \\
&\ \vdots \\
\text{Output} &= \beta\mathbf{H}_n.
\end{aligned}$$

5. Some of our works related to learning from big data with uncertainty

TABLE II
COMPARATIVE RESULTS ON THE SELECTED UCI BENCHMARK DATASETS

ID	RWA1 Testing Accuracy (%)	RWA1 Training Accuracy (%)	RWA1 Training Time (s)	DBN Testing Accuracy (%)	DBN Training Accuracy (%)	DBN Training Time (s)	RBM-GI Testing Accuracy (%)	RBM-GI Training Accuracy (%)	RBM-GI Training Time (s)	Hidden Structure
1	95.99	98.67	0.6552	95.72	99.40	469.68	**96.83**	98.48	7.64	100-100-200
2	88.19	94.13	0.5304	**97.15**	98.20	5,484.90	95.94	98.19	30.87	100-100-100
3	83.40	88.80	1.1388	**86.40**	87.12	220.90	84.00	89.16	146.28	200-300-300
4	86.20	88.16	1.3572	**88.60**	87.47	270.69	88.20	88.78	146.72	200-300-300
5	91.98	96.86	2.1684	**94.67**	98.37	1,980.60	91.38	96.05	176.05	256-256-300
6	74.10	75.13	0.3900	**85.30**	84.60	11,055.00	84.75	86.79	4.57	100-100-100
7	73.30	74.80	1.2324	62.05	64.10	3,406.80	**74.45**	76.04	159.65	100-100-100
8	78.23	77.96	1.1388	81.29	85.90	751.40	**83.86**	83.32	153.49	100-100-100
9	83.14	86.47	0.8268	**87.10**	91.43	4,906.00	84.91	87.63	97.84	100-100-200
10	95.23	95.63	0.9984	**96.03**	96.53	15,481.00	91.55	92.67	125.69	50-50-50
11	57.20	54.89	0.6708	60.40	57.20	4,074.90	**64.33**	64.14	79.03	50-50-100
12	**99.91**	99.97	0.3120	99.10	100.00	170.78	99.81	99.61	47.08	50-50-50
13	96.64	96.57	3.6816	**98.01**	97.93	2,877.60	97.25	97.14	203.72	10-10-30
14	81.33	91.05	0.1248	**100.00**	100.00	29.29	85.82	92.17	1.20	100-100-100
15	**100.00**	99.89	0.8736	**100.00**	100.00	155.64	**100.00**	99.97	148.93	150-150-150
16	75.67	76.58	5.1168	**83.85**	84.27	14,462.00	76.06	76.44	17.94	150-150-150

Note: For each dataset, the highest testing accuracy is in bold face.

5. Some of our works related to learning from big data with uncertainty

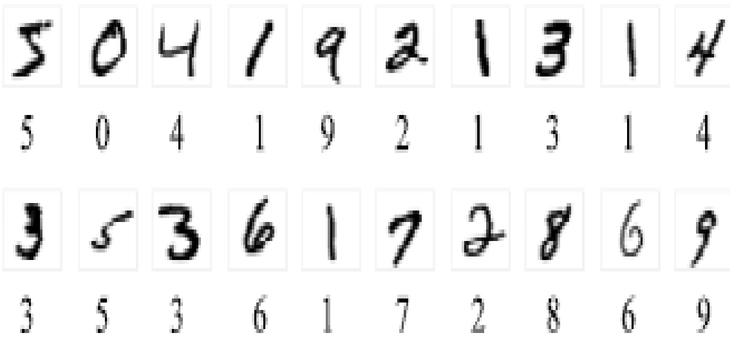


Fig. 6. Sample images in MNIST dataset.

Fig. 7. Sample images in ORL dataset.

TABLE III
PERFORMANCE COMPARISON ON MNIST DATASET

Method	Training Accuracy (%)	Testing Accuracy (%)	Training Time (s)
RWA1	99.28	97.76	1339.47
DBN	100.00	98.75	47230.00
RBM-GI	99.30	97.58	1373.99

TABLE IV
PERFORMANCE COMPARISON ON ORL DATASET

Method	Training Accuracy (%)	Testing Accuracy (%)	Training Time (s)
RWA1	100.00	97.50	12.00
DBN	94.50	85.00	7628.00
RBM-GI	100.00	97.50	80.49

5. Some of our works related to learning from big data with uncertainty

5.4 Our 4th work is on the model tree. More details can be found from the following papers.

Ran Wang, Sam Kwon, Xizhao Wang, and Qing-Shan Jiang, Segment Based Decision Tree Induction with Continuous Valued Attributes, IEEE Transactions on Cybernetics, 2015, 45(7): 1262-1275

Xi-zhao Wang, Yu-Lin He, Dabby D. Wang, Non-Naive Bayesian Classifiers for Classification Problems with Continuous Attributes; IEEE Transactions on Cybernetics, 2014, 44(1): 21-39

Xi-zhao Wang, Ran Wang, Hui-Min Feng, Huachao Wang, A new approach to classifier fusion based on upper integral, IEEE Transactions on Cybernetics, 2014, 44(5): 620-635

Ran Wang, et al, Learning ELM-tree from big data based on uncertainty reduction, Fuzzy Sets and Systems, Volume 258, pages 79-100 (2015)

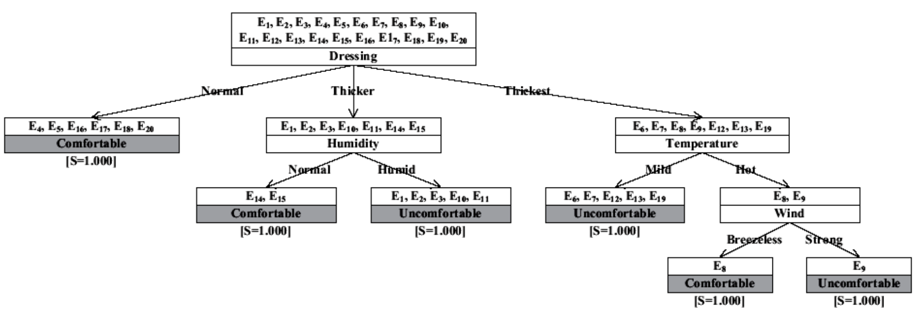
5. Some of our works related to learning from big data with uncertainty

Big Data Classification with Mixed Attributes

$(D_1, D_2, \cdots, D_m, C_1, C_2, \cdots, C_n)$

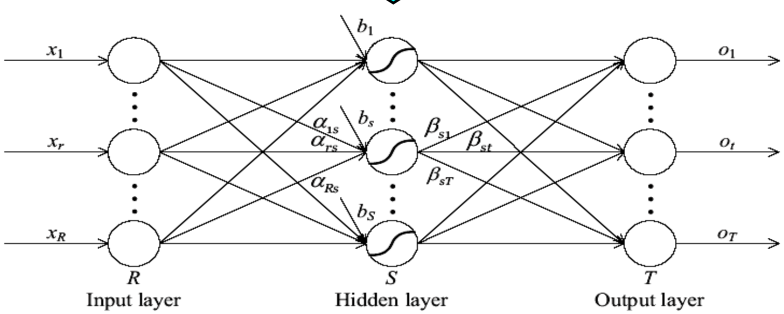
Constructing fuzzy decision tree based on the uncertainty reduction upon the set of discrete features

$(D_1, D_2, \cdots, D_m)$



Training RWN on the set of continuous features

$(C_1, C_2, \cdots, C_n)$



solve β_{ij} to make $H\beta = Y$

RWN Tree

5. Some of our works related to learning from big data with uncertainty

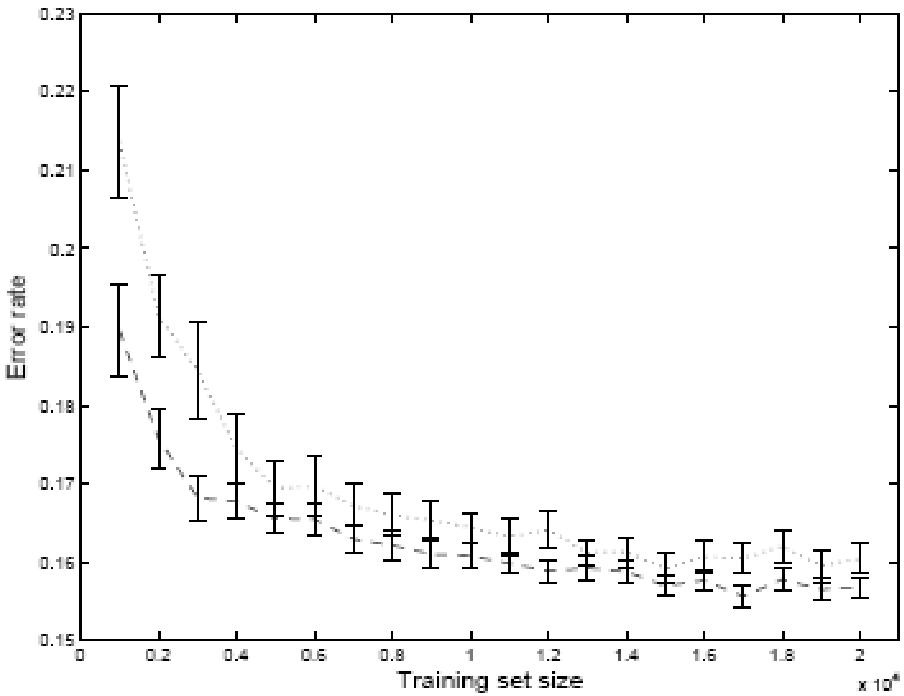
Execution time (seconds) of parallel ambiguity-based RWNTree on 4 big datasets.

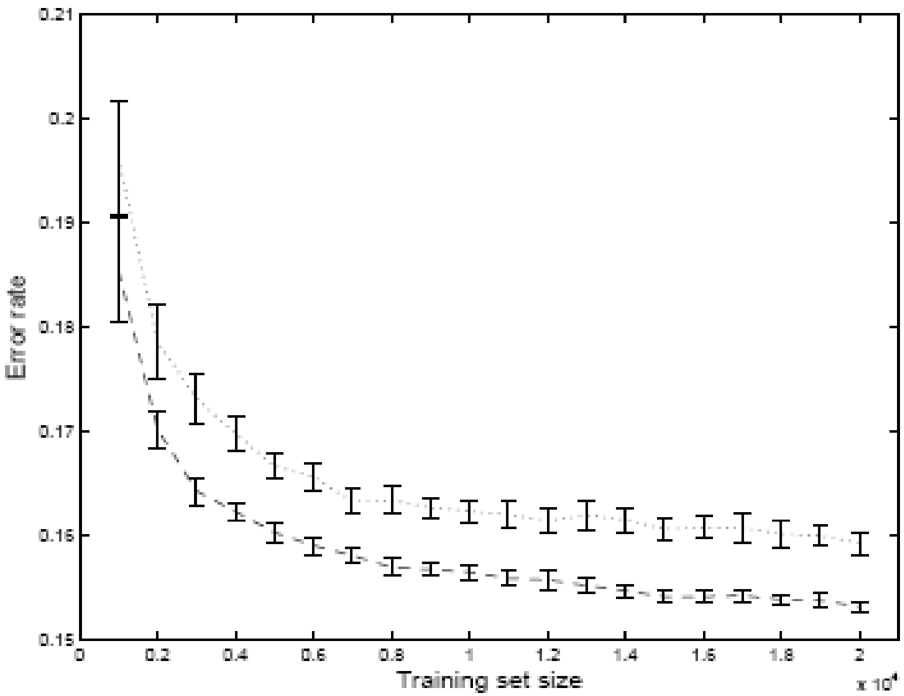
Dataset	2 computers	4 computers	6 computers	8 computers
Image Segment 【19 attributes, 7 classes, 2310000 instances】				
2-times	3143.89	1977.50	945.01	488.23
4-times	8340.78	5159.94	2439.82	1148.51
6-times	11 922.16	7261.19	3646.30	1752.65
8-times	14 890.93	9034.98	4393.87	2738.84
Magic Telescope 【10 attributes, 2 classes, 19020000 instances】				
2-times	66 615.66	38 604.97	10 776.83	3513.07
4-times	170 802.54	94 024.51	28 241.36	8315.77
6-times	235 026.29	131 217.35	40 562.60	12 446.06
8-times	293 555.18	164 999.23	48 942.07	24 052.94
Page Blocks 【10 attributes, 2 classes, 5473000 instances】				
2-times	5756.21	2592.83	860.94	268.98
4-times	14 494.77	6516.00	2110.56	661.30
6-times	21 217.74	9475.43	3149.15	991.34
8-times	25 652.75	11 972.18	3829.53	1947.15
Wine Quality-White 【11 attributes, 6 classes, 4898000 instances】				
2-times	6239.44	4195.61	2738.31	1969.97
4-times	16 040.40	10 713.09	6973.18	4703.07
6-times	23 659.38	16 009.42	10 335.04	7061.85
8-times	28 012.47	18 839.69	12 496.50	9875.35

Parallel environment: 1 host computer and 8 servant computers where each computer is with a Pentium4 Xeon 3.06G Hz CPU, 512MB RAM, and Red Hat Linux 9.0 operation system.

5. Some of our works related to learning from big data with uncertainty

Data Set Characteristics:	Multivariate	**Number of Instances:**	4000000	**Area:**	Computer
Attribute Characteristics:	Categorical, Integer	**Number of Attributes:**	42	**Date Donated**	1999-01-01
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	33087





6. Some remarks

Remark 1: Big data problem

That is a problem if only if the data-size or the feature number is very large.

1. Rich-tail phenomenon

- ✓ **Minority Class – (very small number of cases)**
- ✓ **Not in regularly sized data**
- ✓ **In big data, summation of all minority will not be minority**

2. Hubness in high-dimensional space

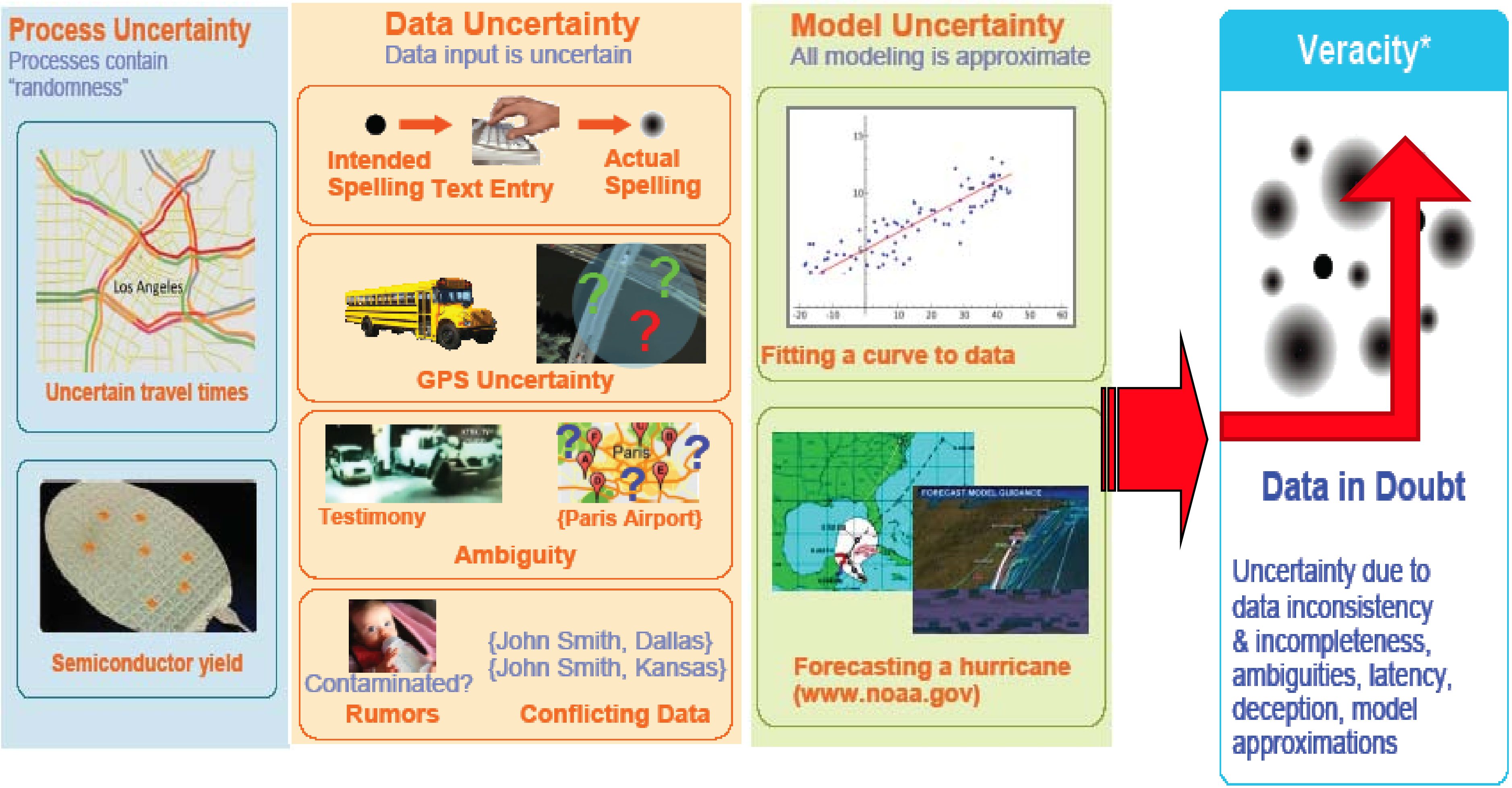
Radovanovic, et al. Hubs in space: Popular nearest neighbors in high-dimensional data. JMLR 2010

3. Uncertainty reduction

Large number of samples make same statistical features such as means and variances stationary according to large number theorem of probability.

6. Some remarks

Remark 2: Uncertainty-boosting of big data



6. Some remarks

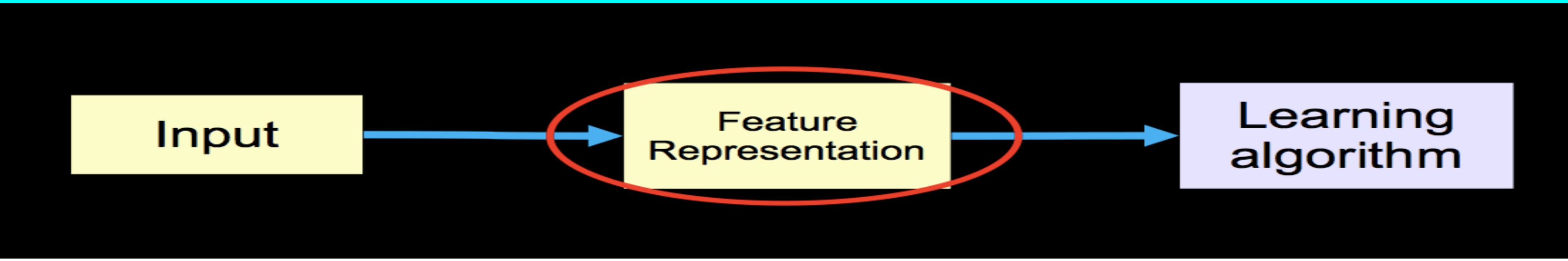
Remark 3: Look back at big data learning

Previous viewpoint: when the data size is becoming really big, i.e., the samples are indeed sufficient, we generally think that model is unnecessarily complicated and a simpler model may work very well.

Current viewpoint: After several years' practice, we confirm the previous thinking is not correct. Even if big data, a complicated model is still required.

6. Some Remarks

Remark 4: Regarding big data, feature extraction and learning algorithm: which one is more important?



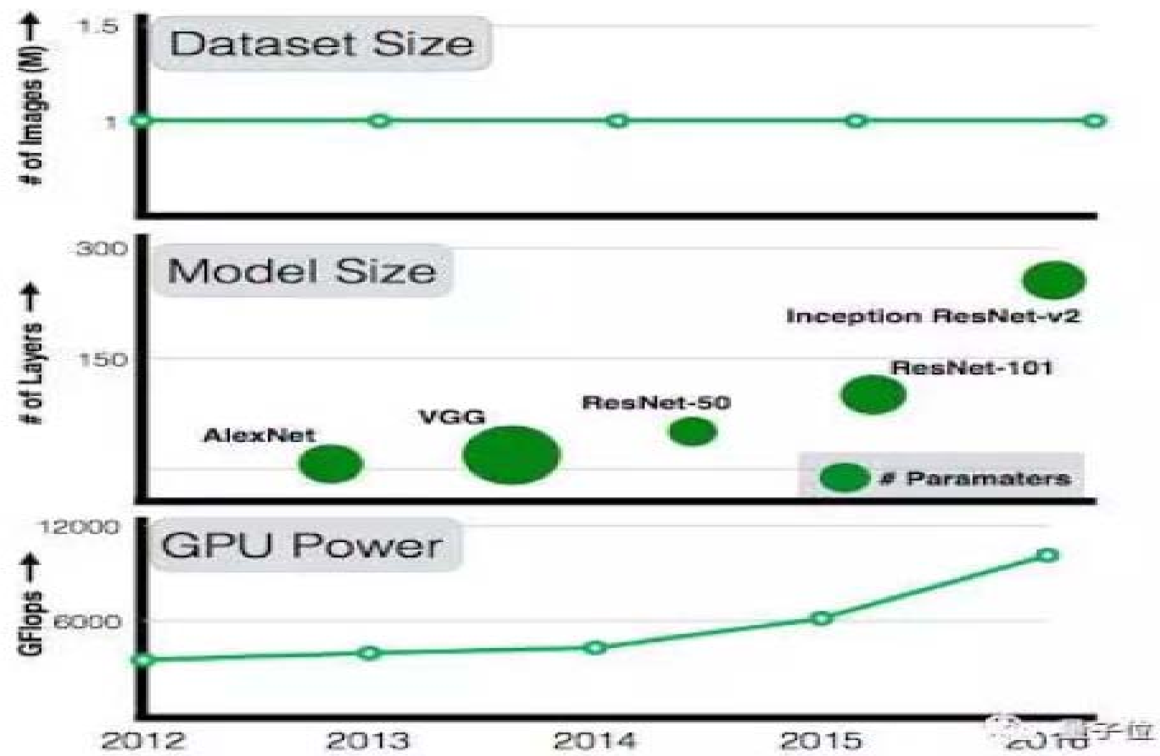
If the algorithm changeable, do you think the gradient descent based BP is the best training algorithm?

And furthermore, once the features are extracted and the structure of neural network is fixed, the iteration in the training process is necessary?

6. Some Remarks

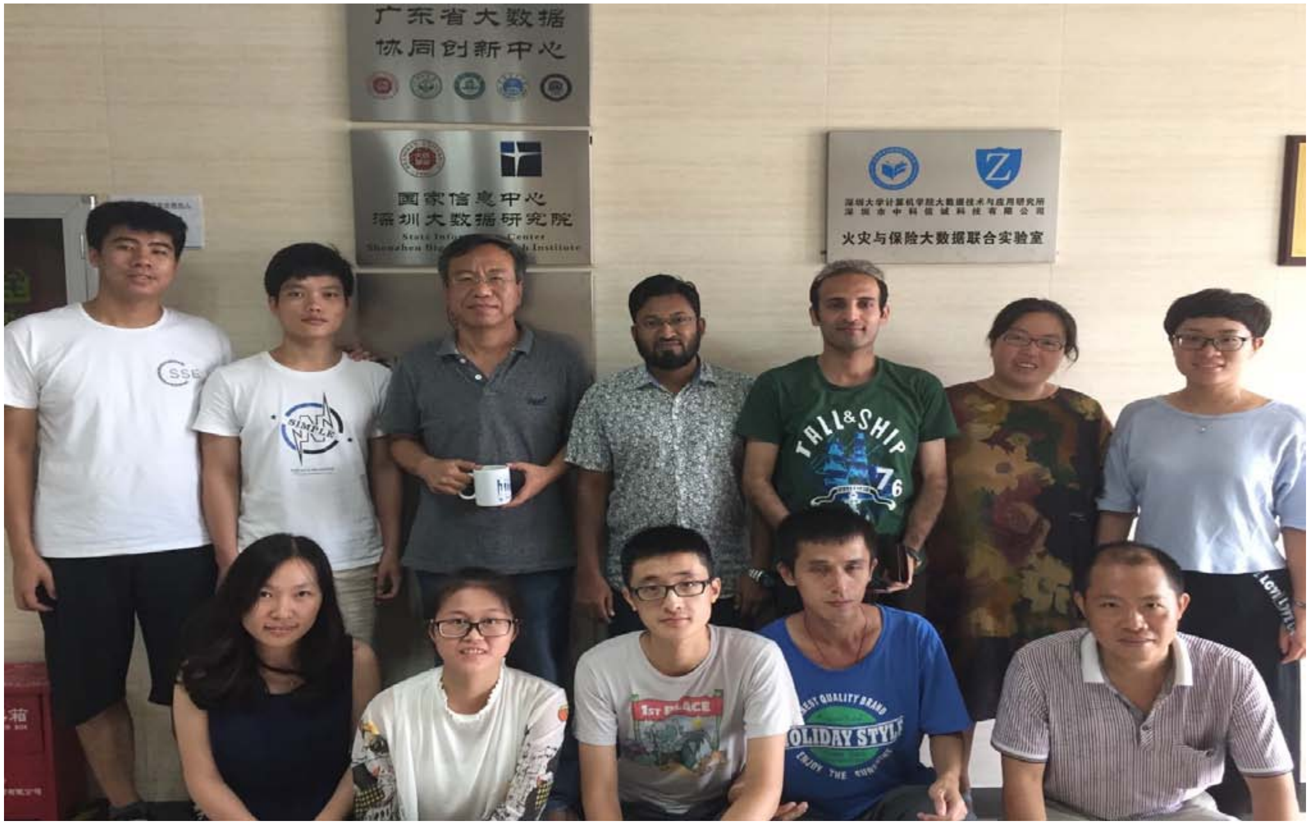
Remark 5: Any role does the Big Data play in DL?

Google deep learning experiment over 300 million pictures – conclusion: the more pictures, the better performance



△ 过去5年间，GPU计算力和模型复杂度都在持续增长，但训练数据集的规模没有任何变化

My thanks go to my colleagues, students and team members



Thank you four your attention!
Any Questions