天文学中的人工智能
Artificial Intelligence in Astronomy

郭平

北京师范大学系统科学学院

图形图像与模式识别实验室

http://sss.bnu.edu.cn/~pguo

https://www.researchgate.net/profile/Ping_Guo3

Laboratory of Image Processing and Pattern Recognition
Beijing Normal University, Beijing, 100875, China

提纲

- 人工智能（AI）
- 计算智能（CI）
- 天文大数据
- 人工智能应用

什么是人工智能?

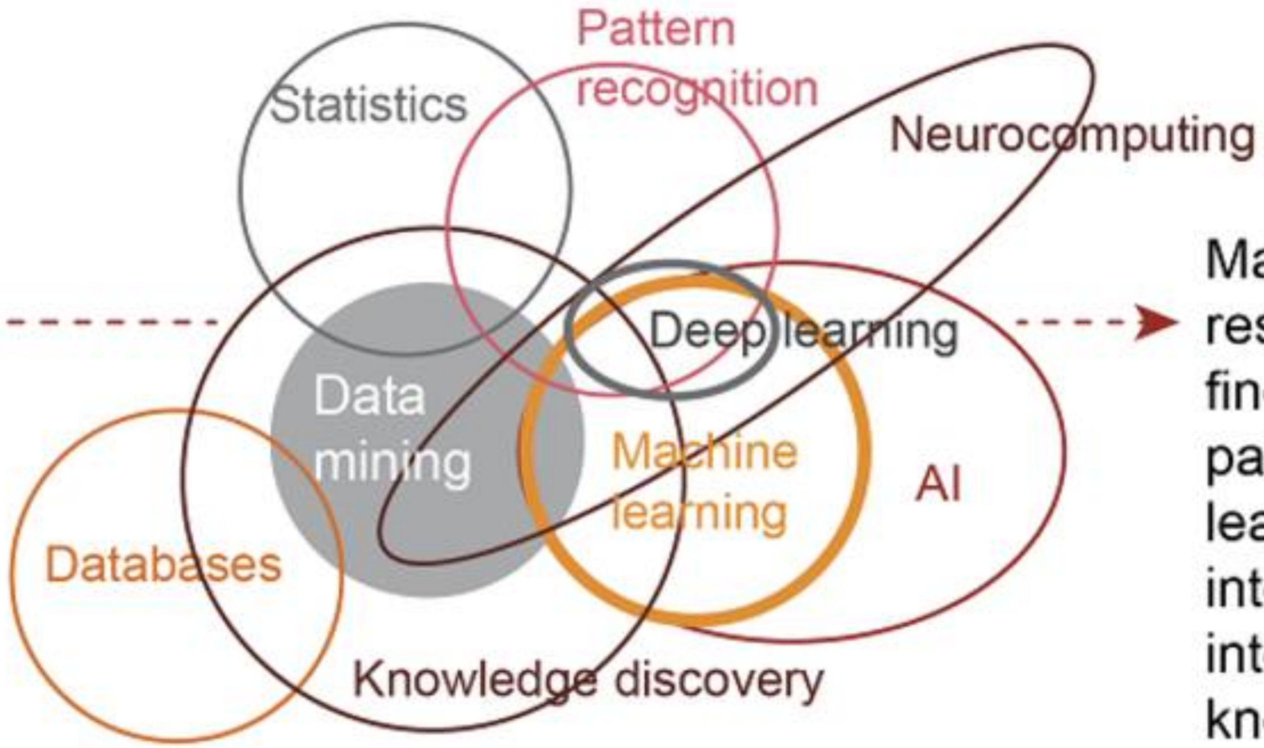
- 人工智能（Artificial Intelligence, AI）是指由人工制造出来的系统所表现出来的智能。通俗的讲就是:
  - 让机器做需要人类智能才可以做的那些事情。
  - 让机器做人类需要它做的任何事情。

- 研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的学科。

人工智能、机器学习、深度学习等之间关系

人工智能

How does machine learning relate to artificial intelligence?



Machine learning is a category of research and algorithms focused on finding patterns in data and using those patterns to make predictions. Machine learning falls within the artificial intelligence (AI) umbrella, which in turn intersects with the broader field of knowledge discovery and data mining.

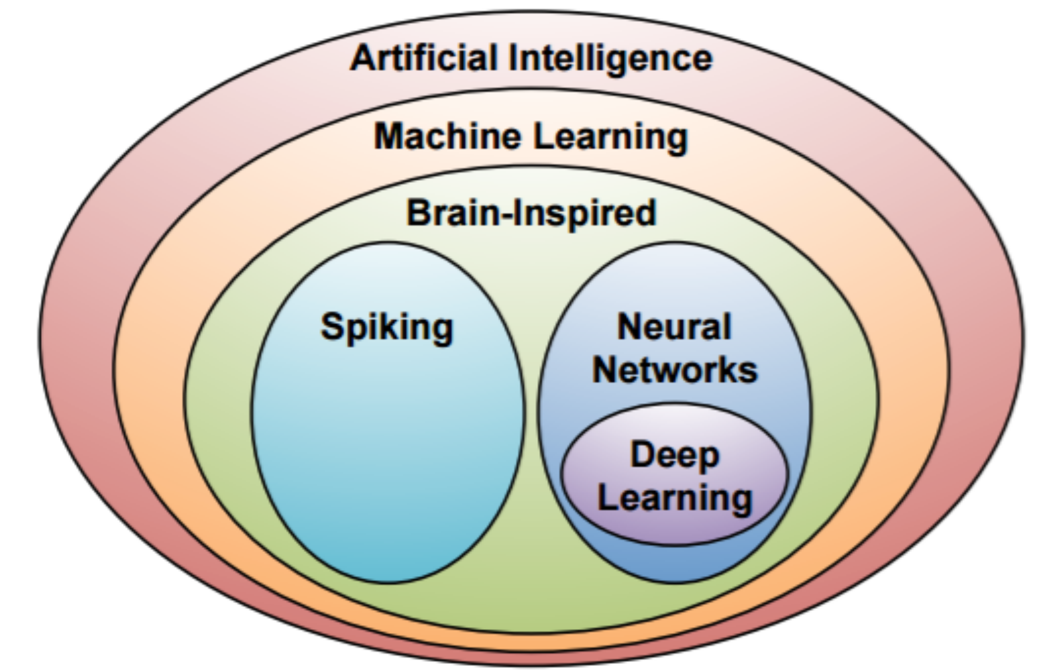
Source: SAS, 2014 and PwC, 2016

人工智能的研究内容

<u>当前主要涉及数据、模型、算法三要素</u>。

- 基础理论方面：
  - 大数据智能
  - 跨媒体感知计算
  - 混合增强智能
  - 群体智能
  - 自主协同控制与优化决策
  - 机器学习
  - 类脑智能计算
  - 量子智能计算

  ……



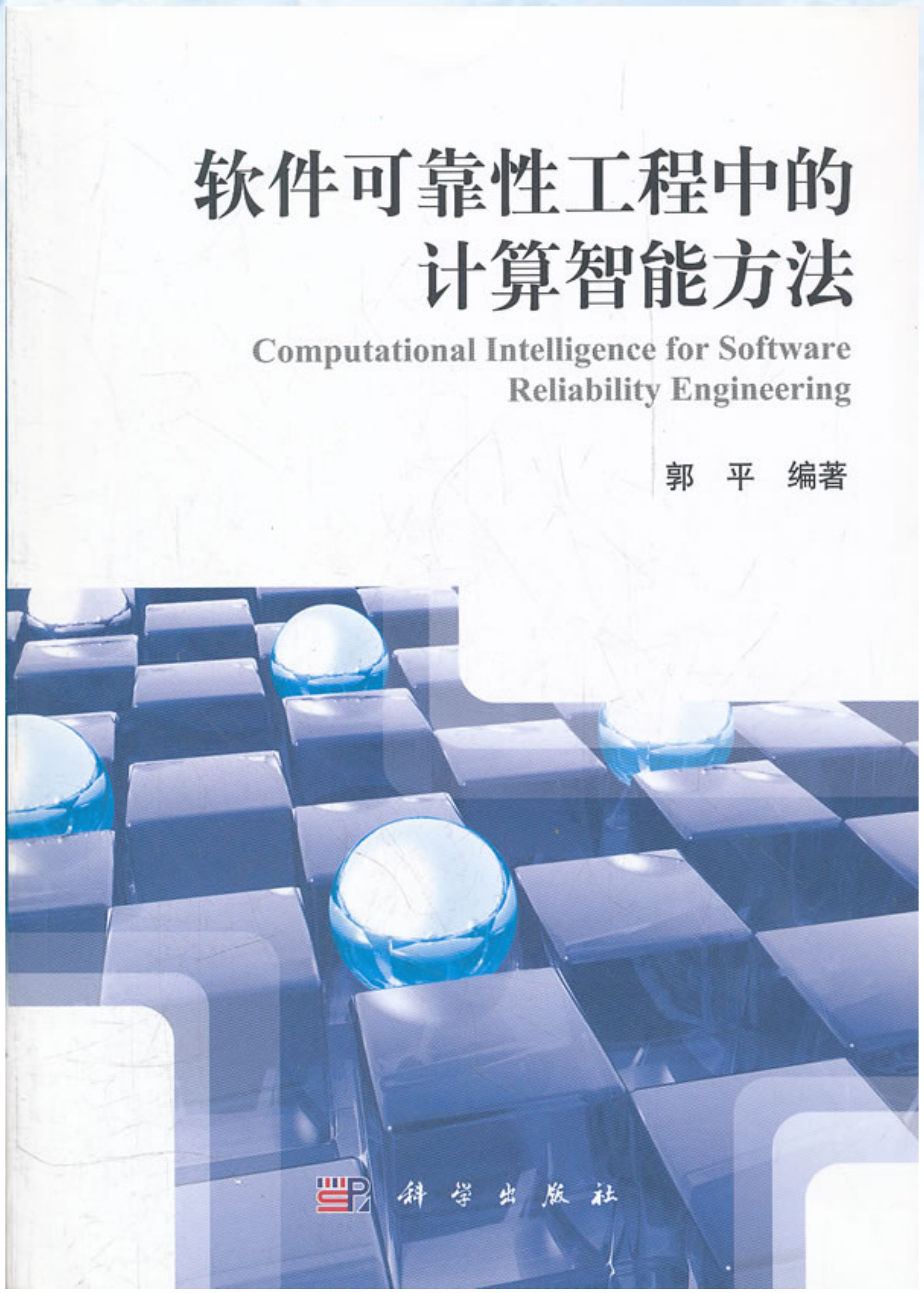
应用：

自然语言理解、数据库的智能检索、专家系统、机器定理证明、博弈、机器人学、自动程序设计、组合调度（智能优化）、感知、语音、视觉、生物特征识别（模式识别）、虚拟现实、复杂系统、大数据，等等。

计算智能（CI）概述

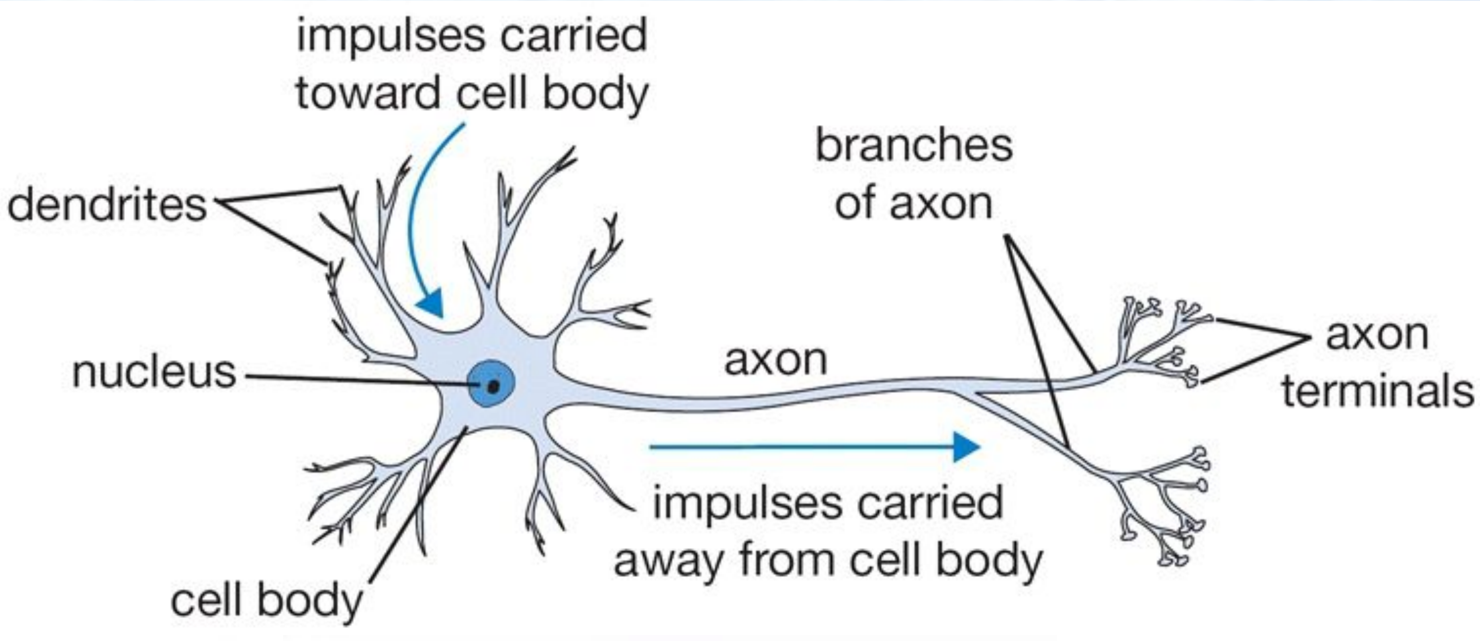
- 计算智能是人工智能发展的新阶段
- 受大自然智慧和人类智慧的启发而设计出的一类解决复杂问题方法的统称
- 计算智能一般被认为是在人工神经网络、模糊系统、演化计算三个主要分支发展相对成熟的基础上,通过相互之间的有机融合而形成的新的科学方法
- 不需要建立问题本身的精确模型,不依赖于知识表示,在观测数据上直接对输入信息进行处理
- 适于解决那些难以建立有效的形式化模型的问题
- 计算智能系统的输出通常包括预测形势或决策方案

软件可靠性工程中的计算智能方法

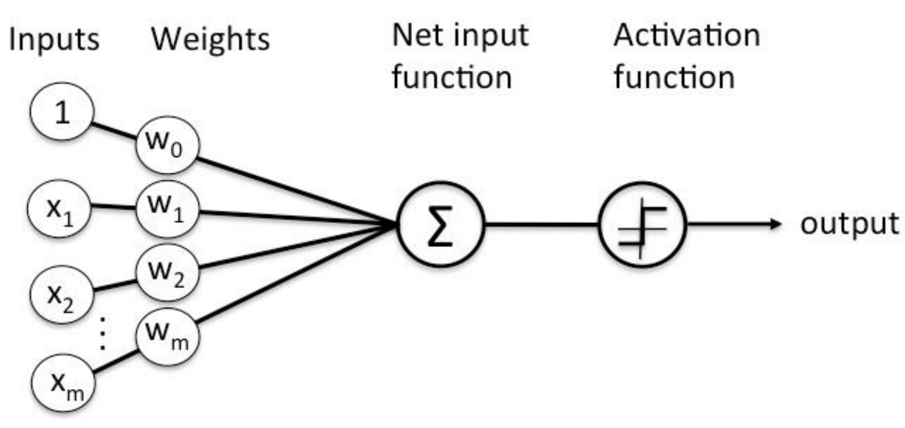


前言
第1章 软件可靠性
第2章 人工神经网络
第3章 模糊系统
第4章 演化计算
第5章 计算群体智能
第6章 人工免疫系统
第7章 统计学习方法
第8章 计算智能在软件可靠性工程中的应用
附录A 矩阵运算
附录B Gaussian积分
附录C Lagrange乘子法

生物神经元 vs. 人工神经元



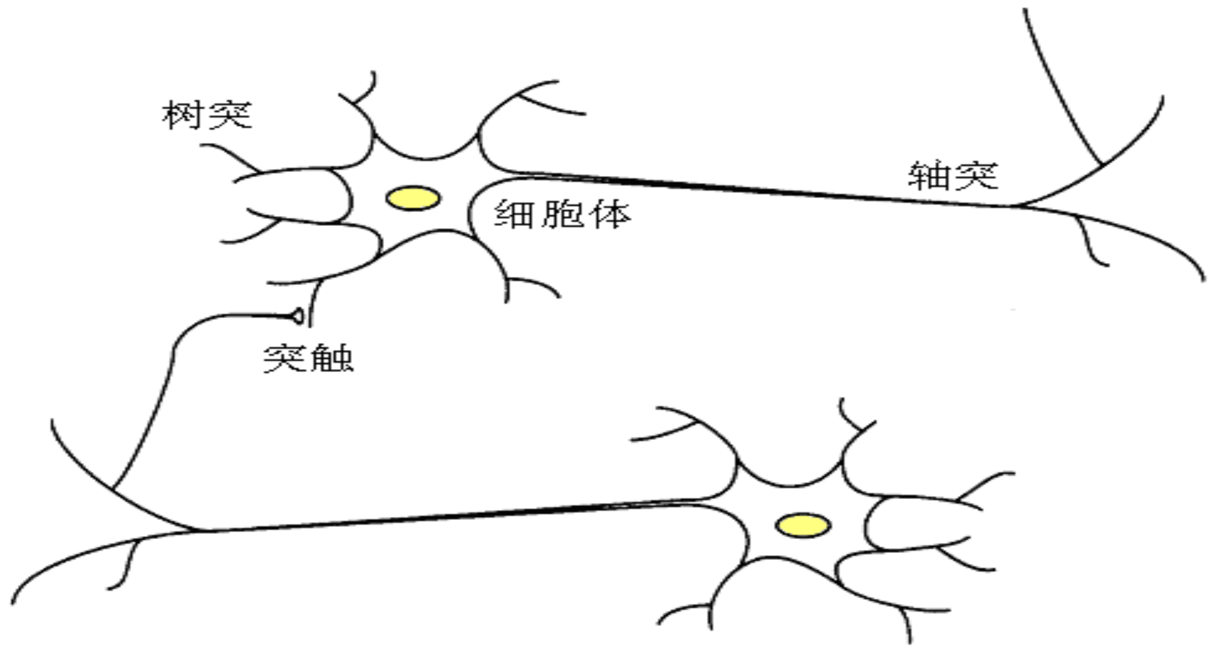
生物神经元



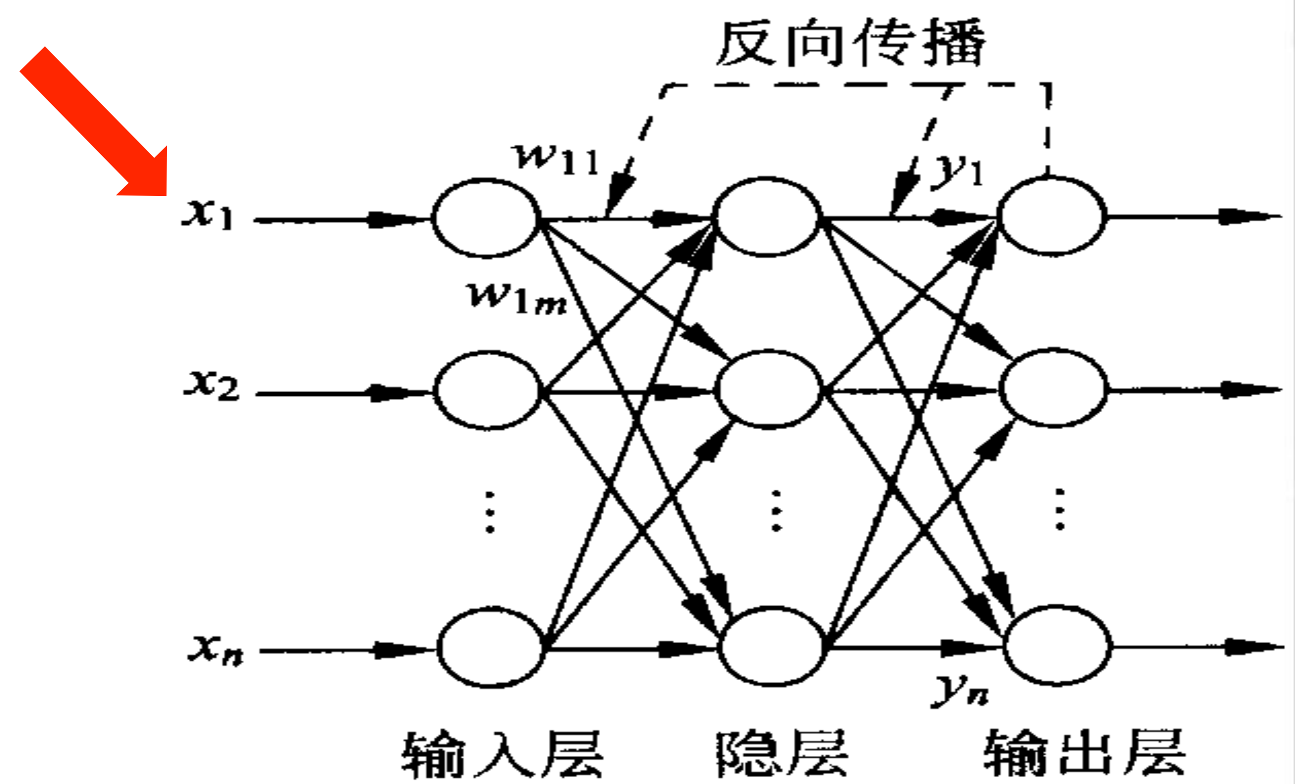
Schematic of Rosenblatt's perceptron.

人工神经元模型

生物神经网络 vs.人工神经网络



生物神经网络结构



多层前馈人工神经网络模型

图像语义分析(图像识别基础知识)

本书的主要内容包括:

- 图像表示与特征提取
- 分类判别模型与生成模型
- 图像中的目标检测与识别
- 图像语义标注
- 场景中的图像语义
- 深度学习在图像语义分析中的应用
- 图像语义分析的应用

人工智能研究学派

- 人工智能研究学派(来源于网络，不代表本人观点)
  - 符号学派（Symbolists）：是使用基于规则的符号系统做推理的人。大部分 AI 都围绕着这种方法。
  - 进化学派（Evolutionists）：是应用进化的过程，例如交叉和突变以达到一种初期的智能行为的一派。
  - 贝叶斯学派（Bayesians）：是使用概率规则及其依赖关系进行推理的一派。
  - 核保守派（Kernel Conservatives）：在深度学习之前，最成功的方法之一是 SVM。
  - Tree Huggers：是使用基于树的模型的人，例如随机森林和梯度提升决策树。
  - 联结学派（Connectionists）：这一派的研究者相信智能起源于高度互联的简单机制。
  - Swiss Posse：基本上是 LSTM，以及两种结合的 RNN 解决知觉的问题。根据 LeCun 的说法，GAN 是“最近20世纪最酷的东西”，也被声称是这帮人发明的。
  - British Alpha Goist：这帮人相信，AI = 深度学习 + 强化学习。
  - Predictive Learners：这个词是 Yann LeCun 用来形容无监督学习的，这是 AI 主要的一个未解决的领域。

人工智能研究学派

- Compressionists：认为认知和学习就是压缩（compression），这实际上也是其他学派共同的观点。
- Complexity Theorists：这一派的人采用来自物理学，基于能量的模型，复杂性理论，混沌理论和统计力学的方法。Swarm AI 可以说属于这一派。
- Fuzzy Logicians：这种方法曾经很受欢迎，但最近比较少见。最近有一个使用模糊规则在 mock dogfight 中打败一个战斗机飞行员的研究。
- Biological Inspirationalists：倾向于创造更接近于生物学中的神经元的模型。例子是 Numenta，Spike-and-Integrate，以及 IBM 的 TrueNorth 芯片。
- Connectomeist：这些人相信大脑的互连（即：Connectome）是智能的来源。有一个项目试图复制一个虚拟的蠕虫，也有一些得到雄厚资助的研究，试图以这种方式映射大脑。
- Information Integration Theorists：认为意识来源于机器的内部想象，反映了现实的因果关系。
- PAC (Probably Approximately Correct)Theorists：他们的整体思想是自适应系统可以方便地执行计算，其结果都能大致正确。简言之，在他们看来智能不需要大规模的计算。

总的来说，人工智能是一个多学科的研究领域。

天文大数据时代

哈佛教授 维克托•迈尔•舍恩伯格
"大数据时代的预言家"
《科学》《自然》等著名学术
期刊最推崇的互联网研究者

《大数据时代》一书里写到："天文学，信息爆炸的起源"

大型天文巡天项目已经积累了大量的数据，并持续产生海量的新数据多波段的天文图像、天体光谱、星表以及模拟数据

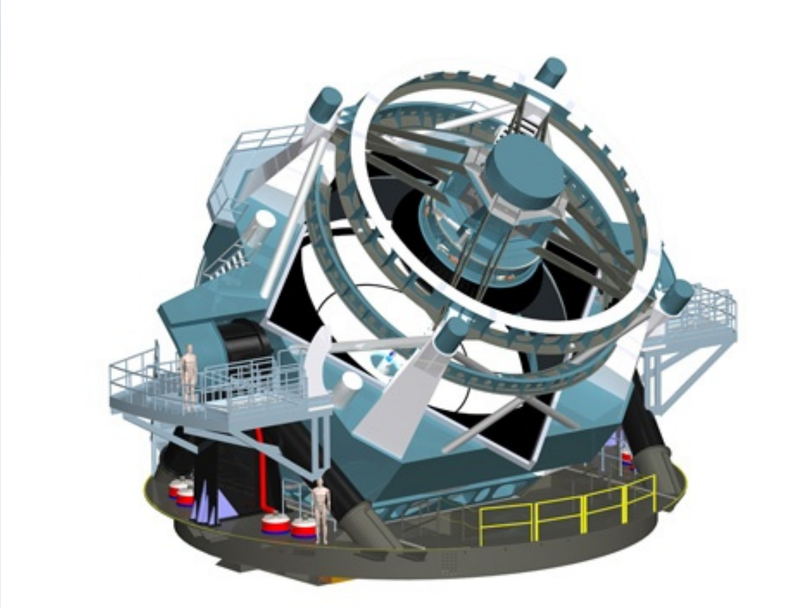
海量数据处理是天文学界迫切需要解决的现实问题，也是对传统天文研究方法提出的强有力挑战。

天文大数据来源于科学实验系统，它面向的是基础自然科学研究，不以挖掘商业价值为目的，不会危害国家安全和侵犯个人隐私，完全可以在全球范围内实现数据资源的公开和共享.

天文大数据

- **Volume**

Pan-STARRS



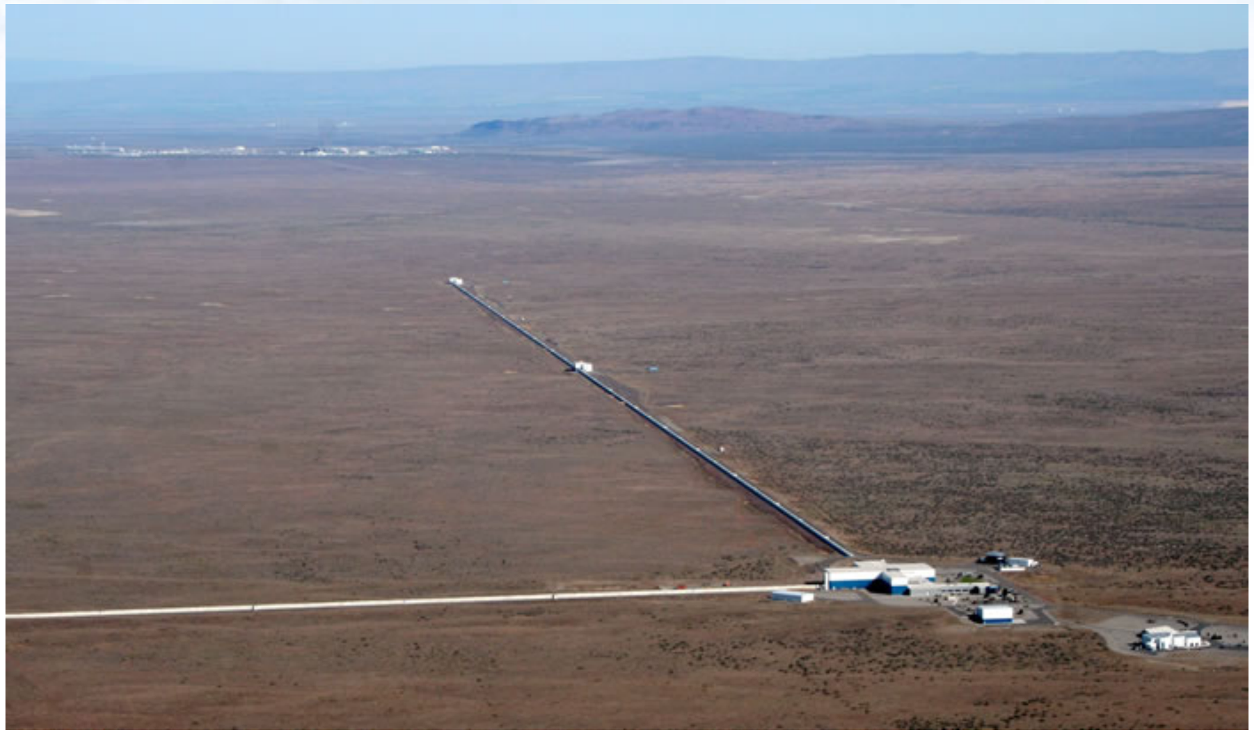
LSST

Pan-STARRS记录数以十亿的星体和星系的可见光区、红外光区图像，其摄像机每40秒曝光一次，得到14亿像素的图像，最终的数据量达40 PB

LSST被形象地描述为一部片长为10年的“宇宙电影”，其最终归档的图像文件预计将达到约70PB，星表将达到10～20PB

天文大数据

- **Volume**



LIGO

FAST

Laser Interferometer Gravitational wave Observatory
激光干涉引力波天文台

Five-hundred-meter Aperture Spherical radio Telescope
500米口径球面射电望远镜

天文大数据分析

- 天文数据的获取需要造价昂贵的专业观测设施和大量专业技术人员，获取成本极高,因此必须充分利用已有的历史数据和不断产生的新数据,发展新的方法充分挖掘其中蕴藏的科学价值

中国科学院国家天文台兴隆观测站的郭守敬望远镜（LAMOST）是世界上光谱获取率最高的望远镜，项目总投资约2.35亿元

天文大数据

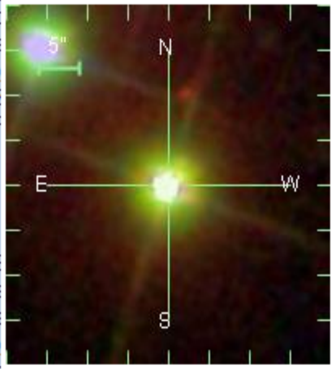
- **Variety**

Type		SDSS Object ID	
STAR		1237671262272225499	
RA, Dec		Galactic Coordiantes (*l*, *b*)	
Decimal	Sexagesimal	*l*	*b*
132.74897, 11.65616	08:50:59.75, +11:39:22.15	215.81309	31.76834

Imaging

WARNING: This object's photometry may be unreliable. See the photometric *flags* below.

Flags	TOO_FEW_GOOD_DETECTIONS PSF_FLUX_INTERP DEBLENDED_AT_EDGE INTERP_CENTER SATUR_CENTER STATIONARY BINNED1 SUBTRACTED NOTCHECKED SATURATED INTERP COSMIC_RAY CHILD



Magnitudes				
u	g	r	i	z
16.40	15.78	11.64	14.96	13.91

Magnitude uncertainties				
err_u	err_g	err_r	err_i	err_z
0.01	0.02	0.00	0.01	0.01

Image MJD	mode	Other observations	parentID	nChild	extinction_r	PetroRad_r (arcmin)
53766	PRIMARY	2	1237671262272225496	0	0.08	3.42 ± 0.033

photoZ (KD-tree method)	photoZ (RF method)	Galaxy Zoo 1 morphology
		-

Cross-identifications Hide

Catalog	Proper motion (mas/yr)	PM angle (deg E)
USNO	293.07 ± 60.000	-110.8

Catalog	Peak flux (mJy)	Major axis (arcsec)	Minor axis (arcsec)
FIRST	2.65 ± 0.15	11.46	7.17

No data found for this object

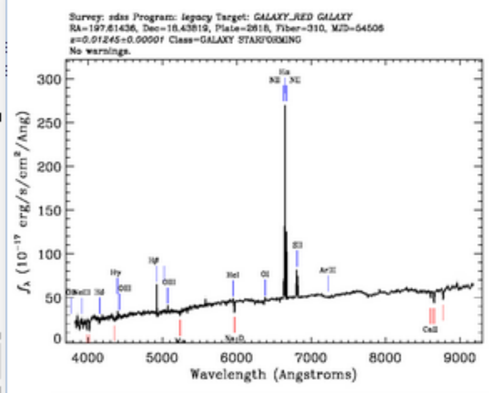
Catalog	Hubble type	21 cm magnitude	Neutral Hydrogen Index
RC3	.S?....	15.43 ± 0.300	0.98

Catalog	J	H	K_s	phQual
2MASS	13.660	12.615	12.399	EEE

Catalog	w1mag	w2mag	w3mag	w4mag	Full WISE data
WISE	10.68	10.44	6.259	4.132	Link

Optical Spectra

SpecObjID= 2947691243863304192 Interactive spectrum



Spectrograph		SDSS
class	Redshift (z)	Redshift error
GALAXY	0.012	0.00001
Redshift flags		OK
plate	mjd	fiberId
2618	54506	310

survey	programname	primary	Other spec
sdss	legacy	1	0

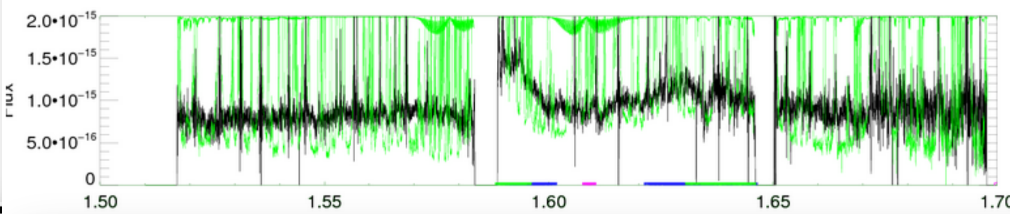
sourcetype	Velocity dispersion (km/s)	veldisp_error
GALAXY	69.11	6.176
targeting_flags		
GALAXY GALAXY_RED		

Infrared Spectra

Targeted star: 2M13102744+1826172

Instrument	APOGEE
APOGEE ID	apogee.n.s.s3.4128.2M13102744+1826172

Galactic Coordinates		RA,dec	
Longitude (L)	Latitude (B)	Decimal	Sexagesimal
330.65989	80.26965	197.61434, 18.43813	13:10:27.44, +18:26:17.27



Targeting Information

2MASS j	2MASS h	2MASS k	j_err	h_err	k_err
13.66	12.615	12.399	0.067	0.067	0.072

4.5 micron magnitude	4.5 micron magnitude error	4.5 micron magnitude source
10.541	0.02	WISE

APOGEE target flags 1	APOGEE_CHECKED APOGEE_SEGUE_OVERLAP
APOGEE target flags 2	

Stellar Parameters

Avg v_{helio} (km/s)	Scatter in v_{helio} (km/s)	Best-fit temperature (K)	Temp error
-490.187	7.26412	4509	2692.8

Surface Gravity $\log_{10}(g)$	log(g) error	Metallicity [Fe/H]	Metal error	[α/Fe]	[α/Fe] error
-0.5	0.451	-2.0	1.709	0.74	4.747

Star flags	SUSPECT_RV_COMBINATION PERSIST_JUMP_NEG PERSIST_LOW
Processing flags (ASPCAP)	SN_BAD COLORTE_BAD STAR_BAD ALPHAFE_BAD LOGG_BAD SN_WARN COLORTE_WARN STAR_WARN NFE_WARN CFE_WARN ALPHAFE_WARN

天文大数据

- **Velocity**

~2,000,000 events each night
One pair of images every 40 seconds for each spot on the sky
Real-time alert latency: 60 seconds
Nightly data generation rate: 15 TB

Science

AAAS.ORG | FEEDBACK | HELP | LIBRARIANS

All Science Journals | Enter Search Term

GUEST | ALERTS | ACCESS RIGHT

AAAS

NEWS | *SCIENCE* JOURNALS | CAREERS | MULTIMEDIA | COLLECTIONS

Subject Collections | Online Extras | *Science* Special Collections | Archived Collections | About Collections

Home > Collections > Online Extras > Special Issues 2005 > 1 July 2005: 125th Anni

About Our Sites

- About *Science*
- About *Science*NOW
- About *Science Signaling*
- About *Science Translational Medicine*
- About *Science* Careers
- About Multimedia
- About Collections

Alerts & Feeds

- E-mail Alerts
- RSS Feeds

Recent Highlights

In *Science*

- Special section: Rethinking the global supply chain
- *Science* Podcast

In a spe
publishe
Science
compani
journal's
forward -
puzzles
scientist
feature i
inquiry o
are seve
question
highlight
study of cell signaling. The Science of Aging Knowl
confronting researchers on aging. And *Science*'s Ne
their careers grappling with some of the very questi

In **Science**

INTRODUCTION AND OPEN ESSAY

What Don't We Know?
D. Kennedy and C. Norman

In Praise of Hard Questions
T. Siegfried

WHAT DON'T WE KNOW?

INTRODUCTION

What Don't We Know?

Special Section

At *Science*, we tend to get excited about new discoveries that lift the veil a little on how things work, from cells to the universe. That puts our focus firmly on what has been added to our stock of knowledge. For this anniversary issue, we decided to shift our frame of reference, to look instead at what we *don't* know: the scientific puzzles that are driving basic scientific research.

We began by asking *Science*'s Senior Editorial Board, our Board of Reviewing Editors, and our own editors and writers to suggest questions that point to critical knowledge gaps. The ground rules: Scientists should have a good shot at answering the questions over the next 25 years, or they should at least know how to go about answering them. We intended simply to choose 25 of these suggestions and turn them into a survey of the big questions facing science. But when a group of editors and writers sat down to select those big questions, we quickly realized that

WHAT DON'T WE KNOW?

Special Section

What Is the Universe Made Of

Every once in a while, cosmologists are dragged, kicking and screaming, into a universe much more unsettling than they had any reason to expect. In the 1500s and 1600s, Copernicus, Kepler, and Newton showed that Earth is just one of many planets orbiting one of many stars, destroying the comfortable Medieval notion of a closed and tiny cosmos. In the 1920s, Edwin Hubble showed that our universe is constantly expanding and evolving, a finding that eventually shattered the idea that the universe is unchanging and eternal. And in the past few decades, cosmologists have discovered that the ordinary matter that makes up stars and galaxies and people is less than 5% of everything there is. Grappling

In the dark. Dark matter holds galaxies together; supernovae measurements point to a mysterious dark energy.

form
und
sculp
They e
makes u
verse—f
But e
comparis
In the la
superno
expandi
ing dow
Is there
ing the u
All si
urement
backgro
galaxy c
cloud pr
tent, but
nary ma
together
in the un
gravity f
This
verse is
increasi
is ordina
does it r
such as t
by massi

The LSST Big Data Challenges

1. Massive data stream: ~2 Terabytes of image data per hour that must be mined in real time (for 10 years).

2. Massive 20-Petabyte database: more than 50 billion objects need to be classified, and most will be monitored for important variations in real time.

3. Massive event stream: knowledge extraction in real time for ~2,000,000 events each night.

Challenge #1 includes both the static data mining aspects of #2 and the dynamic data mining aspects of #3.

The LIGO Hanford 中控室

The LIGO GW Big Data Challenges-1

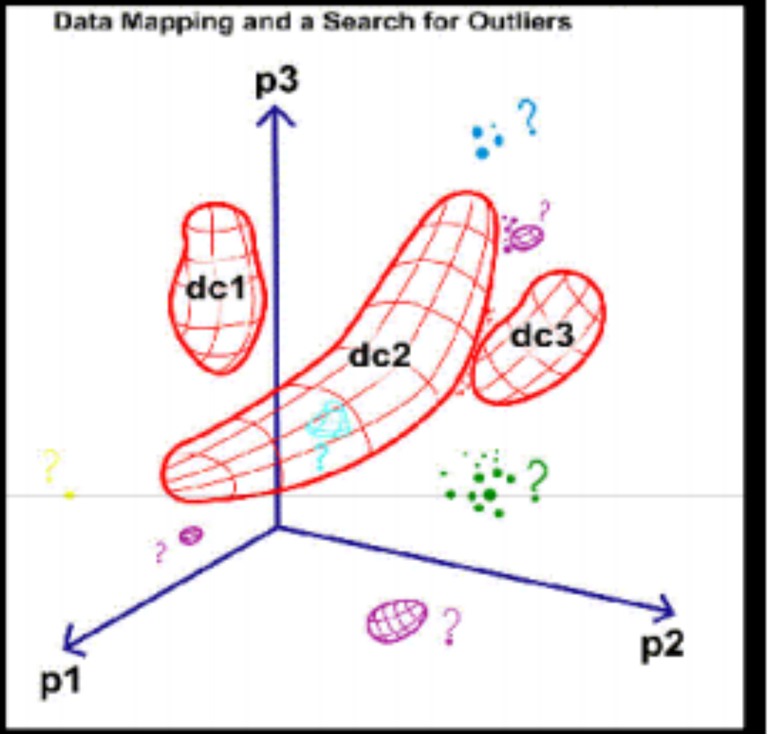
- ✧ LIGO获取的数据类型：
  - 激光干涉仪引力波探测器输出的数据
  - 各种独立的对探测器的环境和探测器设备状态进行监控的探测器和纪录仪：
    - ✓ 温度、气压、风力、大雨、冰雹、地表震动、声响、电场、磁场等环境条件进行监测
  - 对引力波探测器内部的平面镜和透镜的位置等探测器自身状态进行监测的数据。
- ✧ 在数据获取方面：
  - ✧ LIGO Hanford，DAQ的H1和H2干涉仪记录共12733个通道，其中1279个是快速通道（数字化速率在2048 Sa/秒或16384 Sa秒）。升级的LIGO的设计为记录大于300000个通道的数据采集，其中大约3000个快速通道。

The LIGO GW Big Data Challenges-2

- GW150914事件
  - 在线触发延迟是大约3 min，用了5个离线分析流水线，消耗的CPU 时间大约是5千万小时(=20,000PCs 运行100 天)。
  - 搜寻引力波信号中，采用的是匹配滤波技术。基于波形分析的技术，前提条件是已知引力波的理论模型。
- 引力波探测中，大量事件的波形是未知的，如超新星爆发及质量巨大的星体核的坍塌等，如何寻找？

Big Data Science:
Scientific KDD (Knowledge Discovery from Data)

- Characterize the known (clustering, unsupervised learning)
- Assign the new (classification, supervised learning)
- Discover the unknown (outlier detection, semi-supervised learning)



Graphic from S. G. Djorgovski

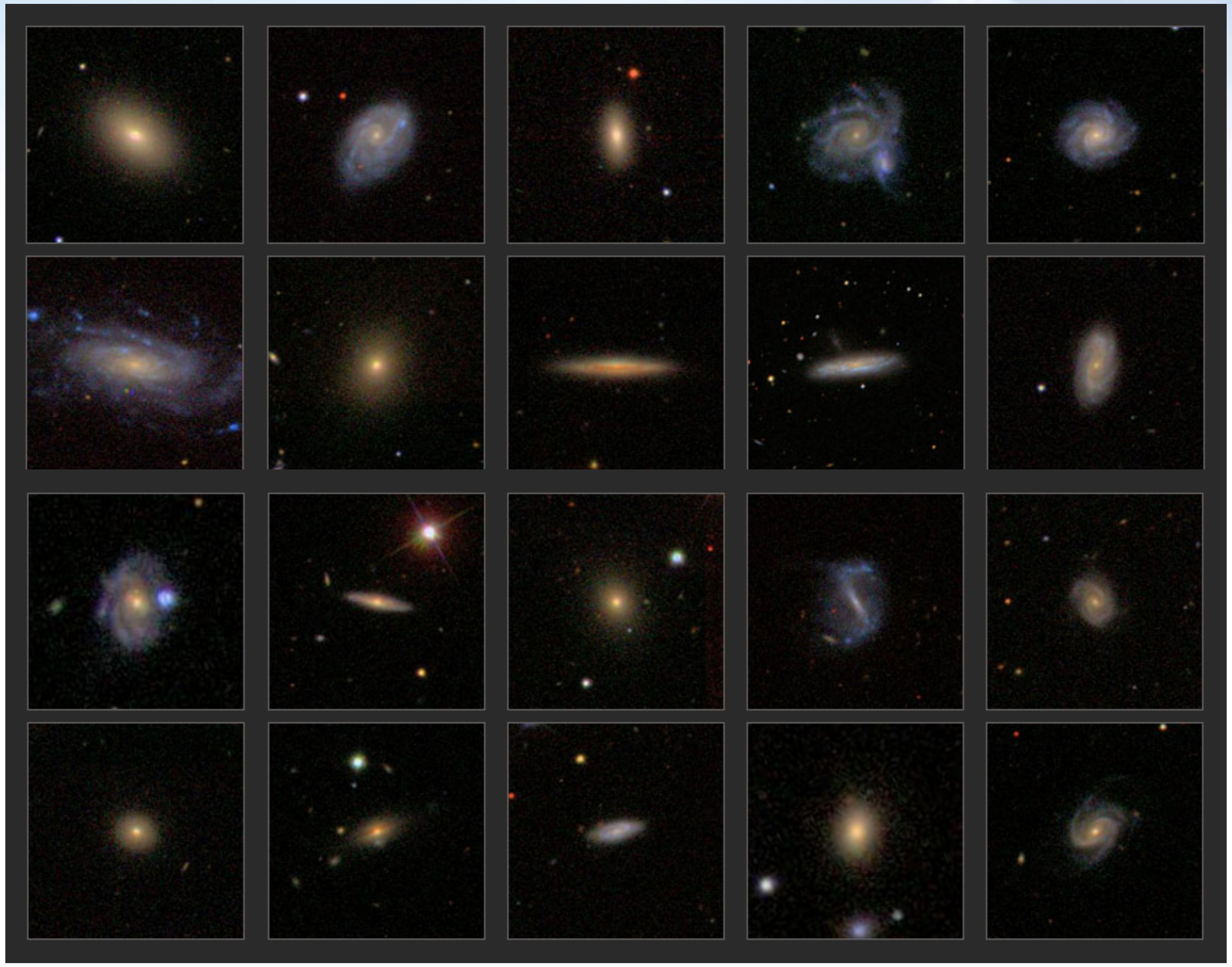
- Benefits of very large datasets:
  - best statistical analysis of "typical" events
  - automated search for "rare" events

http://www.samsi.info/sites/default/files/Borne_september2012.pdf

ANN in astronomy

- ANN has been applied with success in many astronomical tasks.
  - Morphological classification of galaxies
  - Evaluation of photometric redshifts
  - Star/galaxy classification
  - Stellar classification
  - Stellar atmospheric parameters estimation
  - Pulsar candidates identification
  - ......

Morphological classification of galaxies



Large-scale sky surveys e.g. SDSS and the forthcoming LSST, have generated or will generate vast amounts of images for millions of celestial objects.

Morphological classification of galaxies

- Galaxies exhibit a wide variety of shapes, colors and sizes, which are indicative of their age, formation, and interactions with other galaxies.
- Such studies require the accurate classification of their morphologies.
- **But, how to classify these million galaxies by their morphologies?**

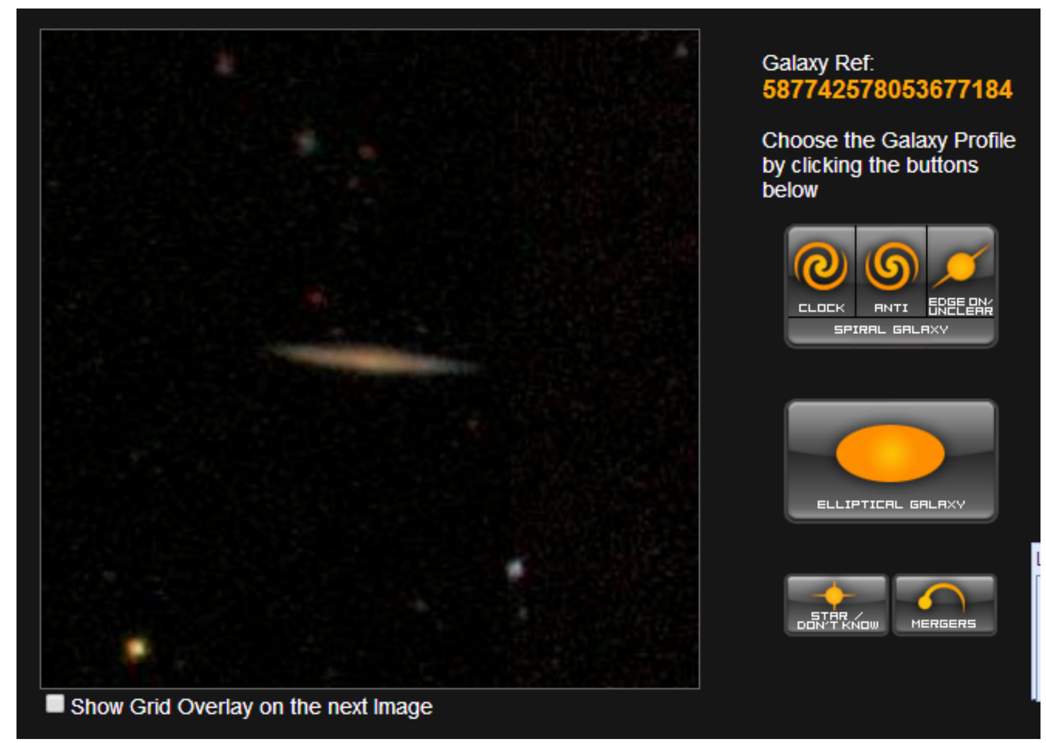
Morphological classification of galaxies

- Traditionally, it is carried out mostly via visual inspection by trained experts.

However, this is inefficient and does not scale to large numbers of images.

Morphological classification of galaxies

- The ***Galaxy Zoo project*** successfully applied a crowdsourcing strategy, inviting online users to classify images by answering a series of questions.

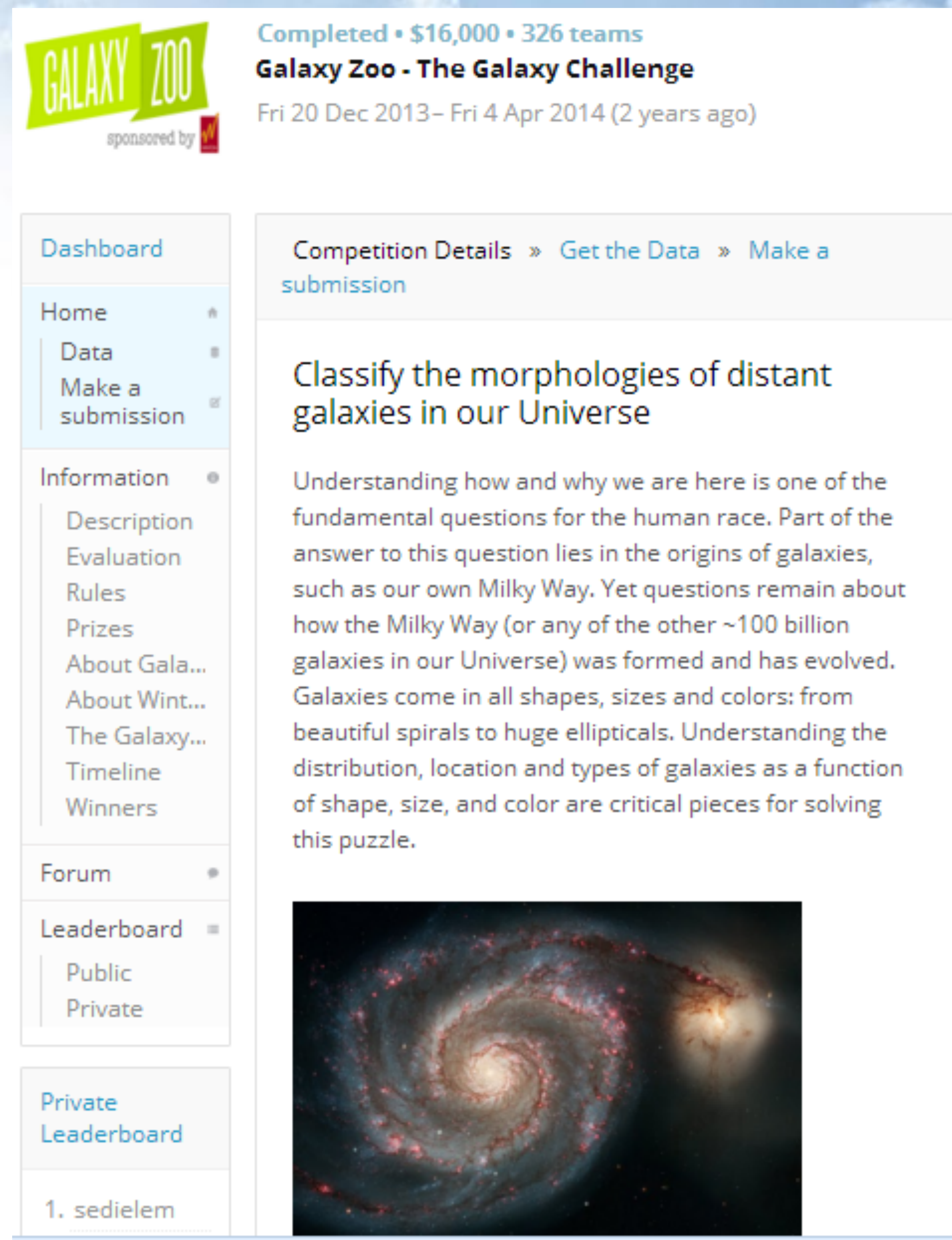


http://zoo1.galaxyzoo.org/Default.aspx

Unfortunately, even this approach does not scale well enough to keep up with the increasing availability of galaxy images.

Morphological classification of galaxies

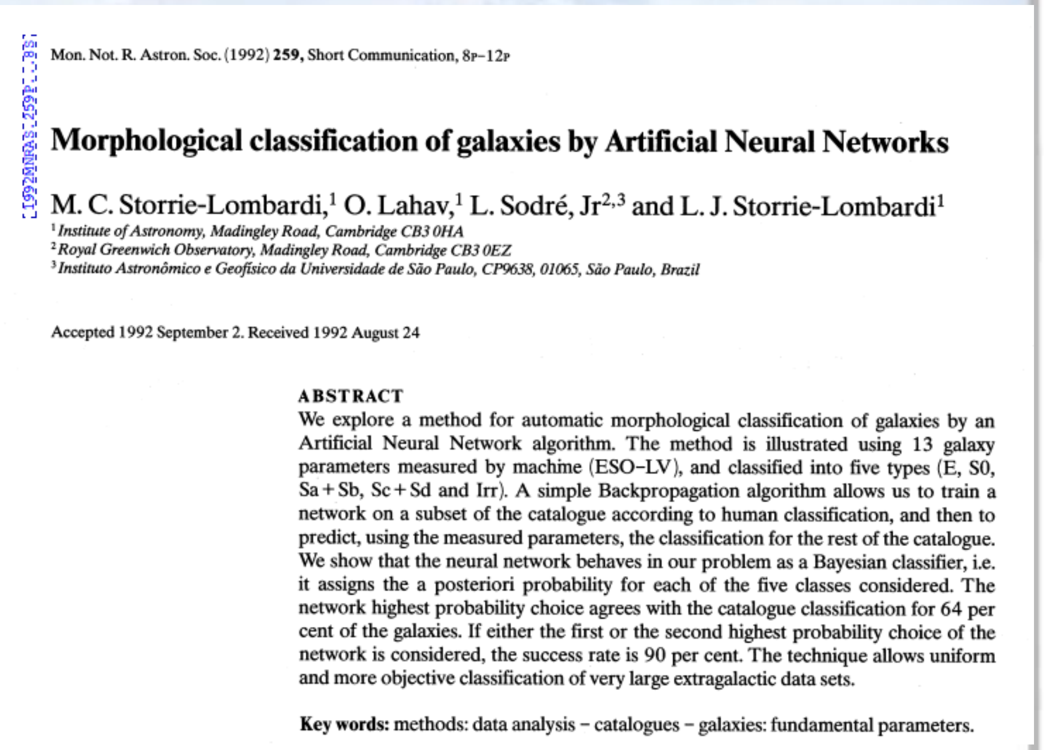
- An automated approach is becoming indispensable.
- ***Galaxy Challenge***, an international competition to build the best model for automatic galaxy morphology classification.



https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge

Morphological classification of galaxies

- ANN in morphological classification of galaxies
  - In 1992, ANN was first applied in the morphological classification of galaxies (Storrie-Lombardi et al. 1992).
  - Since then, ANN has always been used as an important technique for the morphological classification of galaxies.
  - Recently, some ***deep learning*** models have been employed in morphological classification of galaxies since we have more and more data.



Mon. Not. R. Astron. Soc. (1992) **259,** Short Communication, 8P–12P

Morphological classification of galaxies by Artificial Neural Networks

M. C. Storrie-Lombardi,[1] O. Lahav,[1] L. Sodré, Jr[2,3] and L. J. Storrie-Lombardi[1]
[1] *Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*
[2] *Royal Greenwich Observatory, Madingley Road, Cambridge CB3 0EZ*
[3] *Instituto Astronômico e Geofísico da Universidade de São Paulo, CP9638, 01065, São Paulo, Brazil*

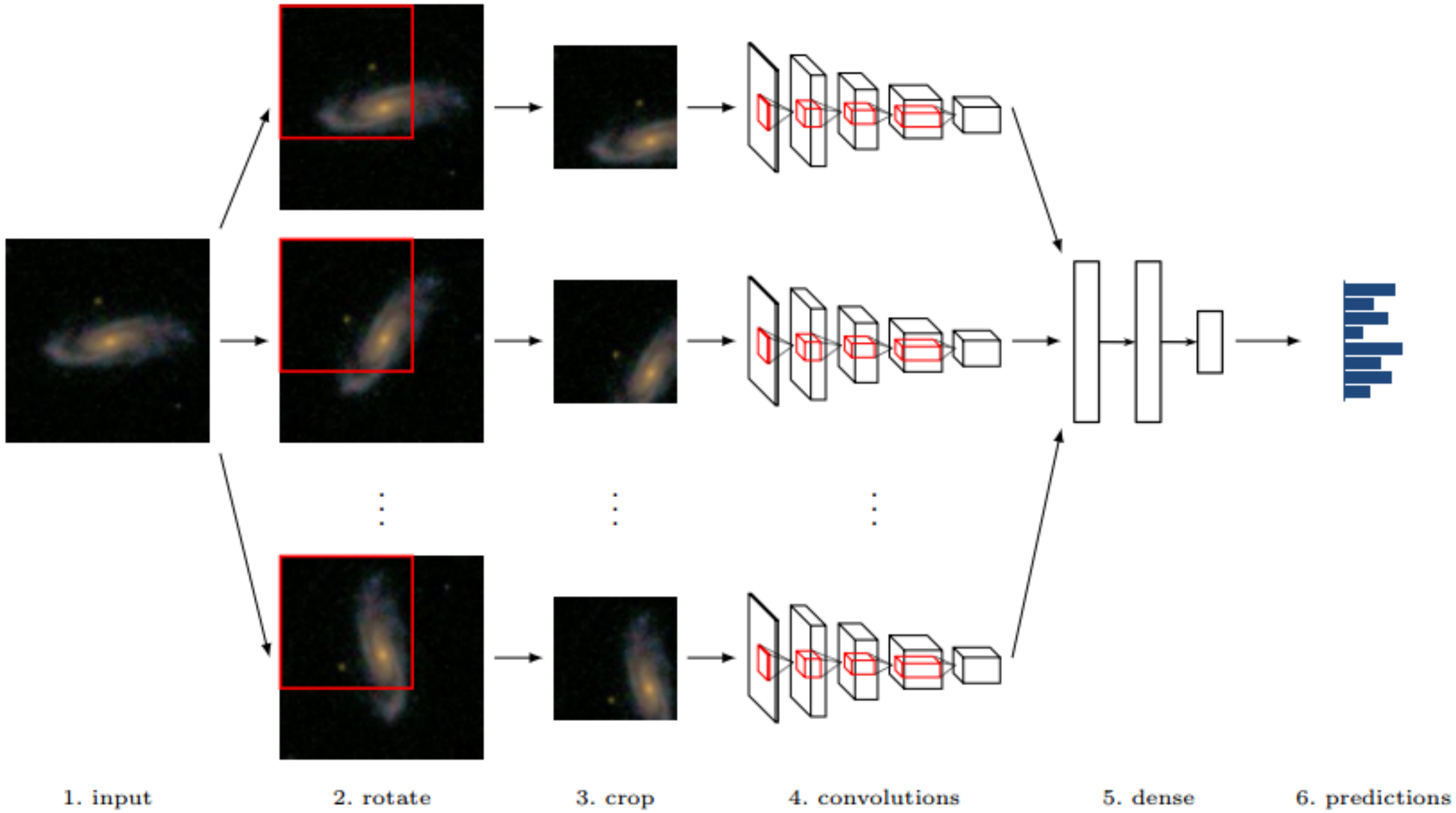
Accepted 1992 September 2. Received 1992 August 24

ABSTRACT
We explore a method for automatic morphological classification of galaxies by an Artificial Neural Network algorithm. The method is illustrated using 13 galaxy parameters measured by machine (ESO–LV), and classified into five types (E, S0, Sa + Sb, Sc + Sd and Irr). A simple Backpropagation algorithm allows us to train a network on a subset of the catalogue according to human classification, and then to predict, using the measured parameters, the classification for the rest of the catalogue. We show that the neural network behaves in our problem as a Bayesian classifier, i.e. it assigns the a posteriori probability for each of the five classes considered. The network highest probability choice agrees with the catalogue classification for 64 per cent of the galaxies. If either the first or the second highest probability choice of the network is considered, the success rate is 90 per cent. The technique allows uniform and more objective classification of very large extragalactic data sets.

Key words: methods: data analysis – catalogues – galaxies: fundamental parameters.

Morphological classification of galaxies

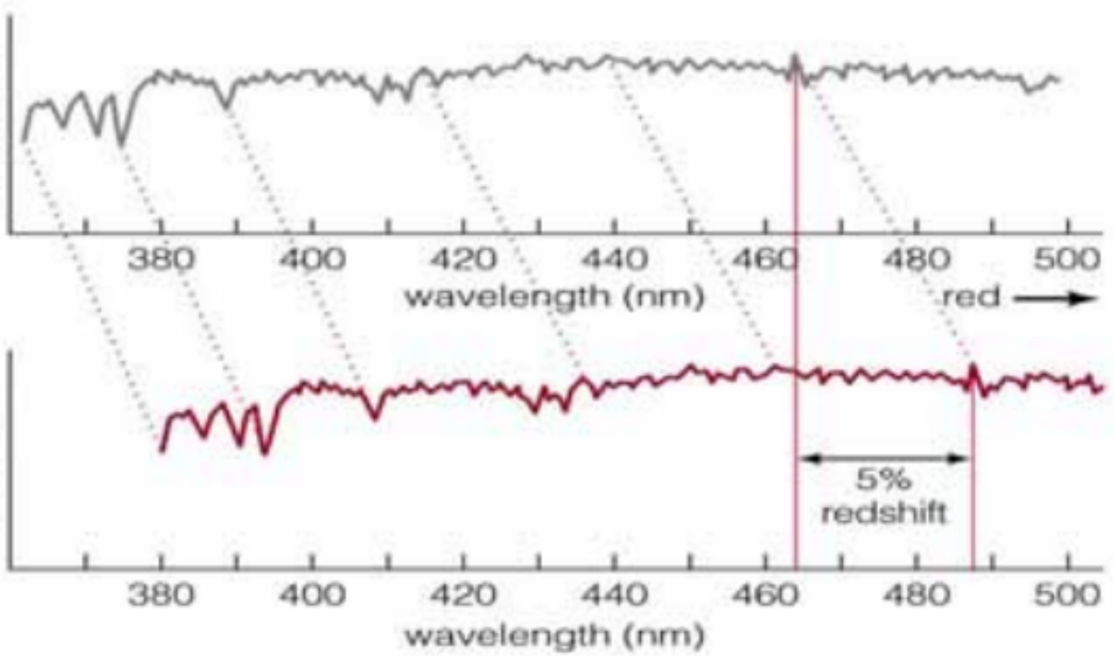
- Convolutional neural network (CNN) has been used for galaxy morphology classification in the context of the Galaxy Challenge (Dieleman et al. 2015).



Their model achieves state-of-the-art performance and wins the first place in the Galaxy Challenge.

Photometric redshifts evaluation

- The photometric redshift technique is to use the medium-band or broad-band photometry of galaxies to approximate its redshift.
- Although it is less accurate than those obtained with spectra, it is low cost.



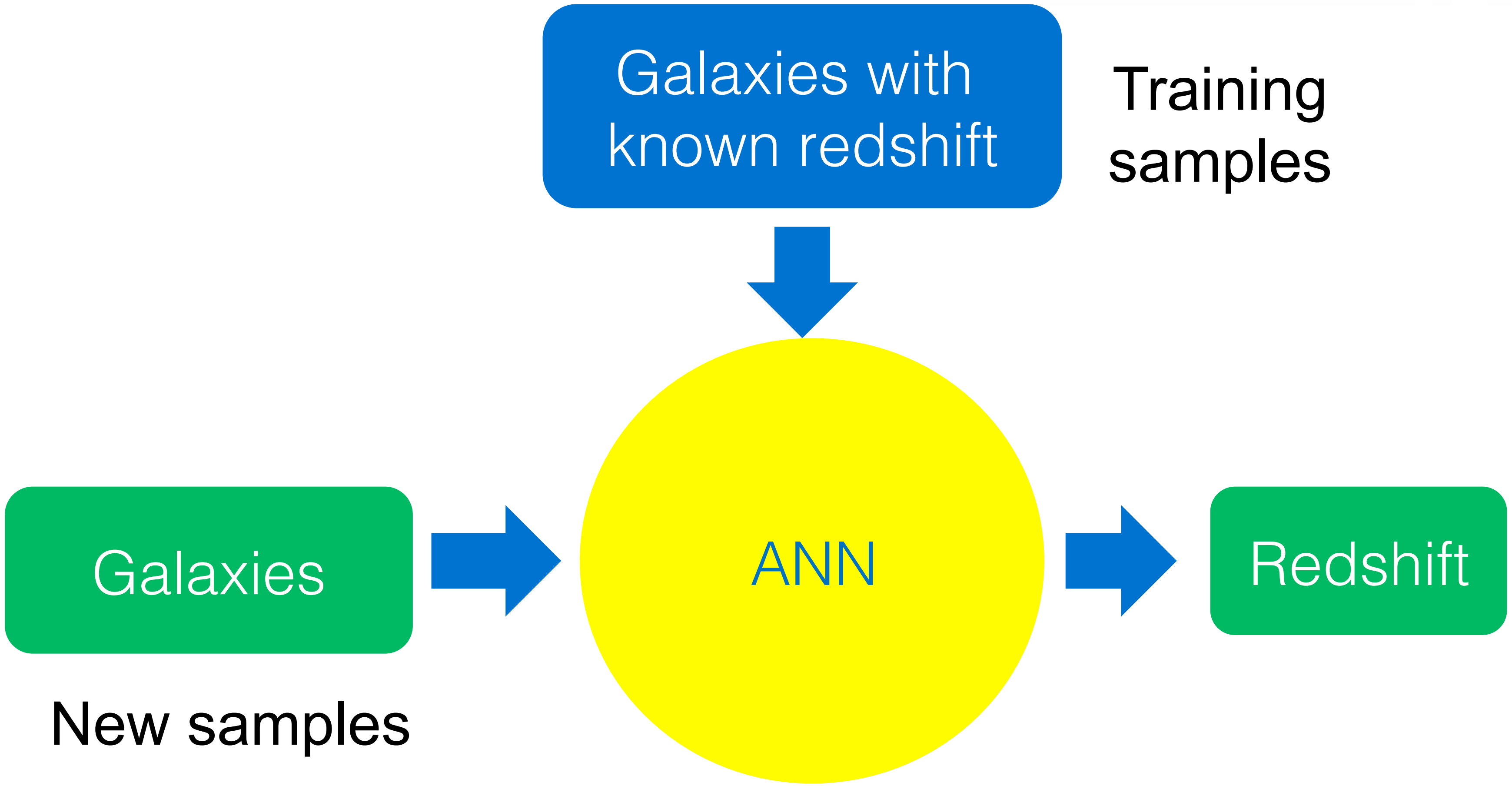
Redshift is a crucial cosmological parameter denoting the distance of galaxies according to the Hubble Law.

Photometric redshifts evaluation

- Traditionally, the approach to evaluate the photometric redshift is the template-fitting technique.
  - This method is unintelligent and highly depending on the quality and representativeness of the template.
- Another method is through the use of machine learning techniques.
- It can be viewed as a regression task in the view of machine leaning.

Photometric redshifts evaluation

- ANN takes fluxes in expert specific optical and near infrared filters as input and output the estimated redshift.

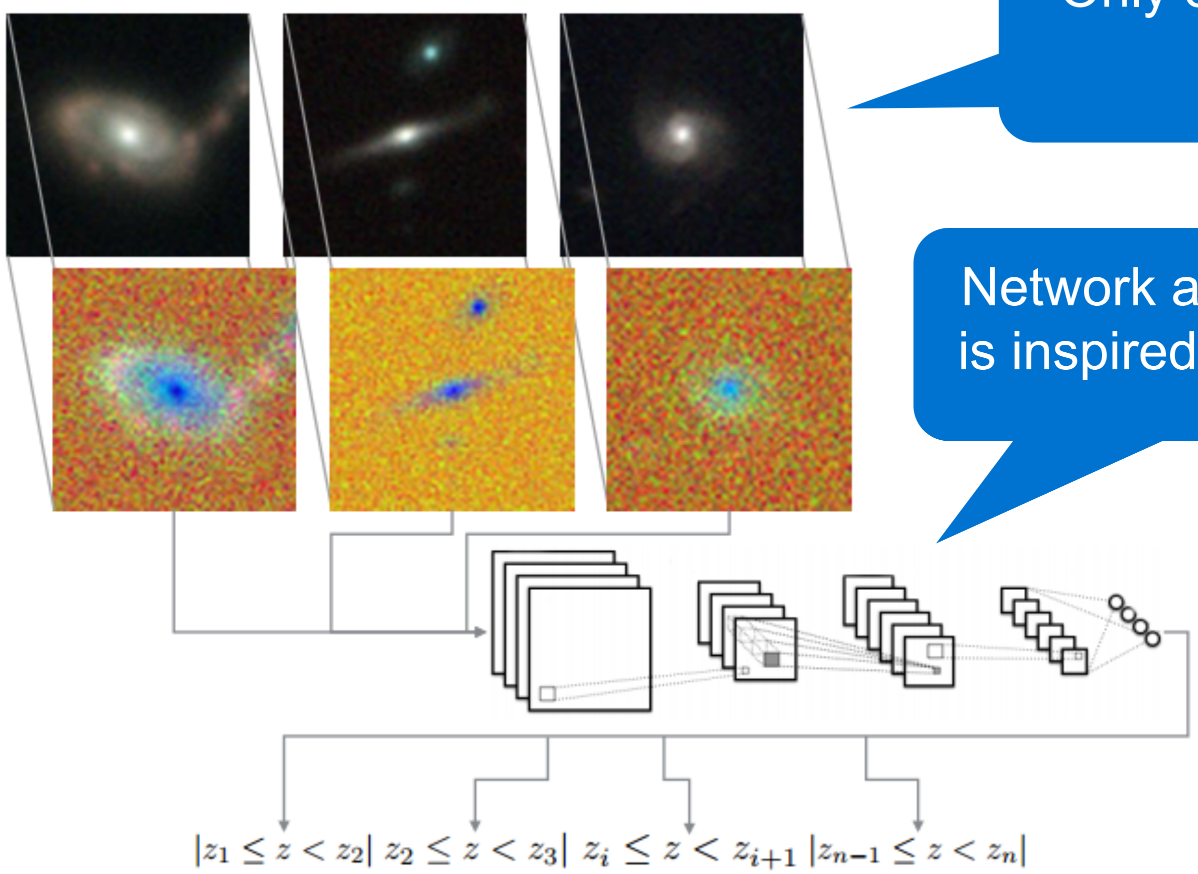


Photometric redshifts evaluation

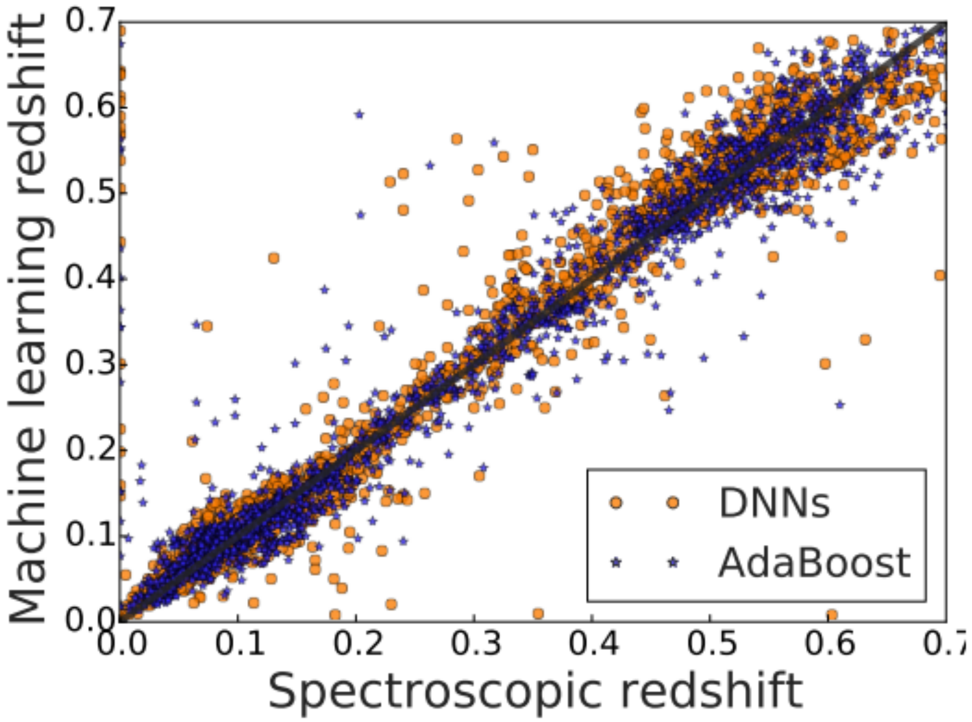
- CNN has been used for photometric redshifts evaluation (Hoyle 2016).

Only 33,167 training samples

Network architecture is inspired by AlexNet



This is an end-to-end method and removes the user from feature selection

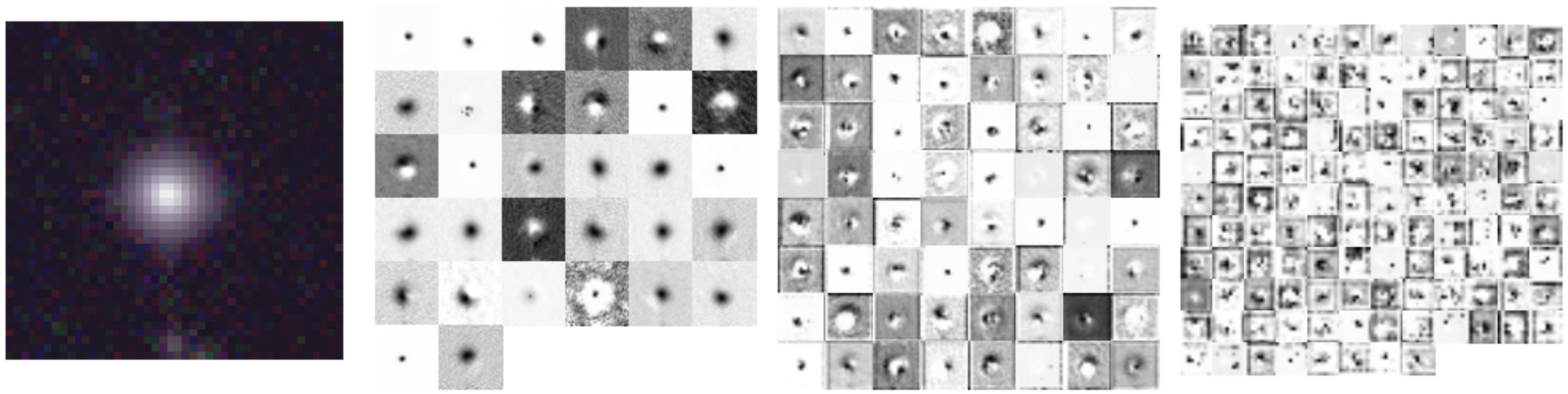


Star/galaxy classification

- The problem of source classification is fundamental to astronomy.
- The basic task behind star/galaxy classification is to classify objects into stars and galaxies automatically using optical and near infrared images.
- ANN was first employed in star/galaxy classification task by Odewahn et al. in 1992.
  - They train networks using 14 experts-designed features as the input and obtain a promising classification performance.
- Subsequently, ANN has always been used in this kind of task.

Star/galaxy classification

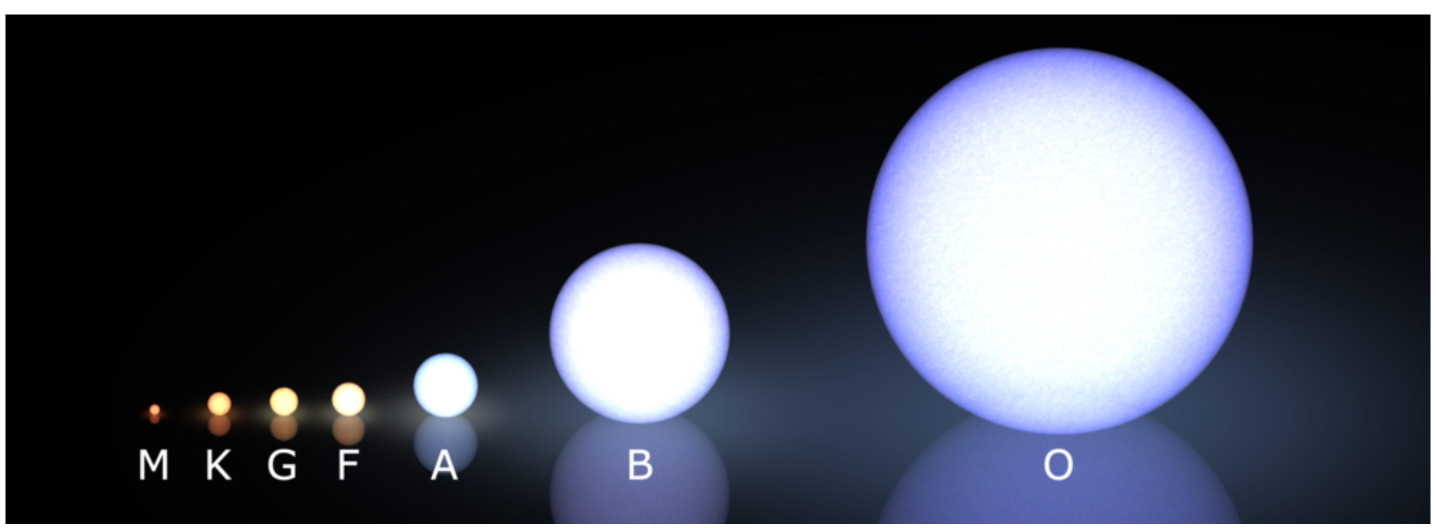
- It's necessary to pay attention to the utilization of unlabeled data.
  - Self-organizing map (SOM) (Miller and Coe 1996, Cortiglioni et al 2001).
- New techniques or idea of computer vision and pattern recognition community could be applied naturally in this visual binary classification task.
  - Recently, deep convolutional neural networks has been employed in the task of star-galaxy classification (Kim et al. 2016).

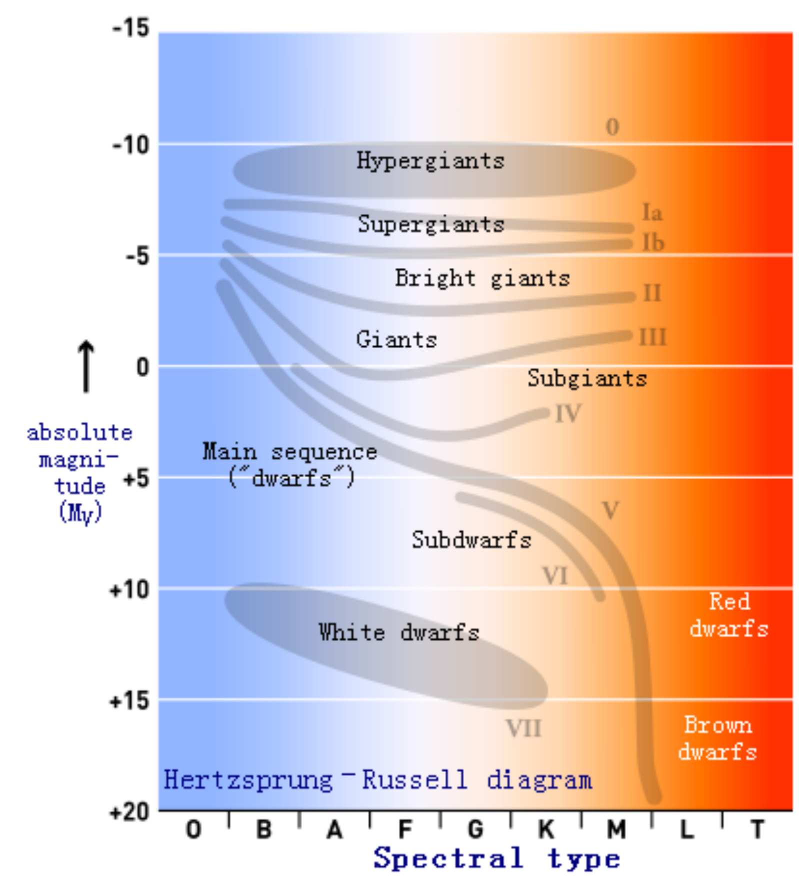


(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

Stellar classification

- In astronomy, stellar classification is to classify stars based on their spectral characteristics.
- Most stars are currently classified under the Morgan–Keenan (MK) system.
- In the MK system, a luminosity class is added to the spectral class.

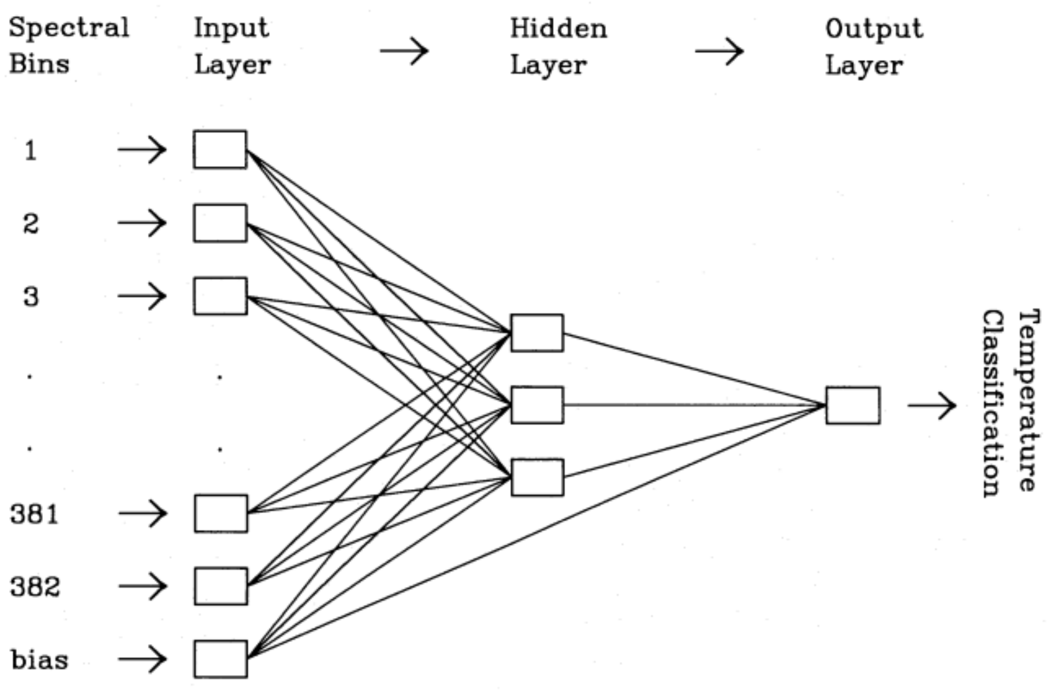




Stellar classification

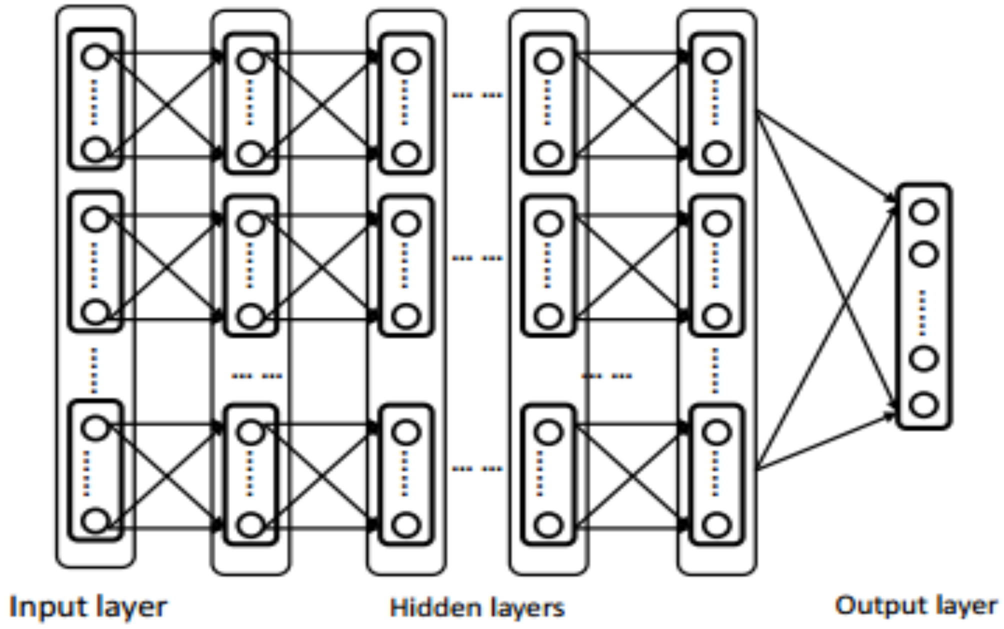
- For the small data set, MK classification can be conducted by expert by comparing the stars to be classified with the standard ones.
- For large scale data set, this way is infeasible and automated approaches are required.
- ANN has been used as common tool to perform the classification task since it was first used by in 1994 (Von et al. 1994)

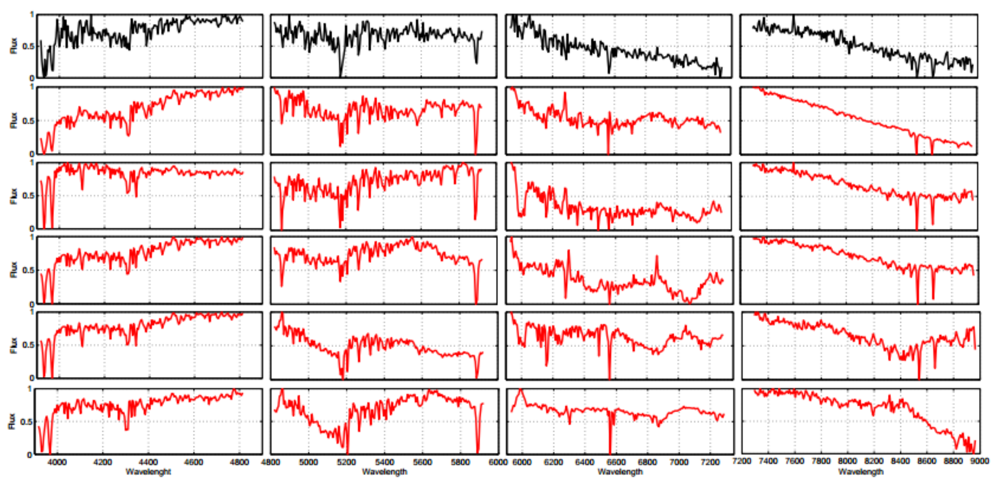
The complete spectra, the line and the continuum are used as the input to the NN, respectively.



Stellar classification

- In this kind of works, the input to networks can be various.
  - Spectra (with and without the continuum)
  - Spectra + PCA, LLE…
  - Spectral indices
  - Single band images
- Recently, DNN is used to learn features from the spectra automatically. (Wang et al. 2016)





Stellar atmospheric parameters estimation

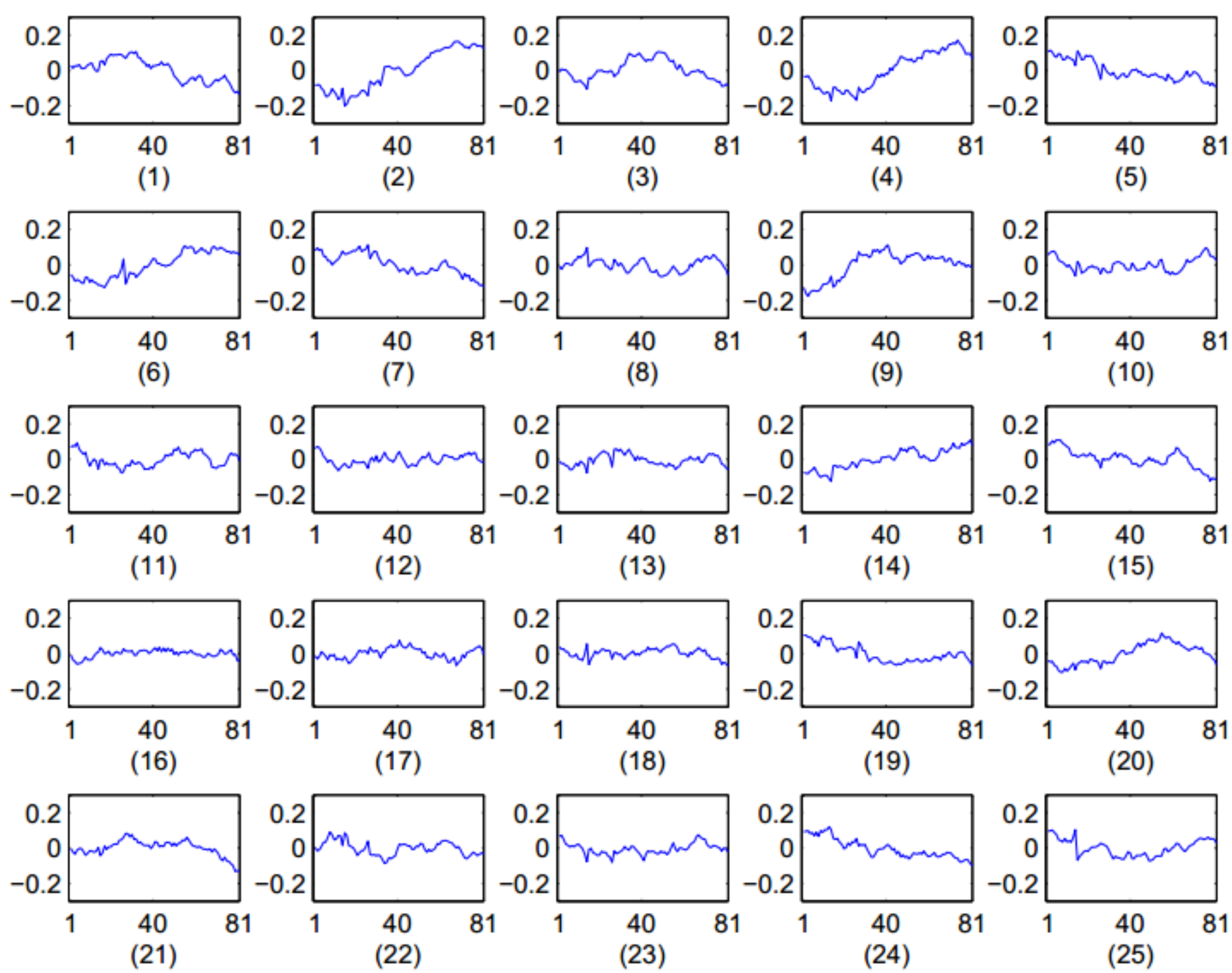
- The stellar parameters estimation is to obtain the atmospheric parameters from the stellar spectra.
- These parameters, such as Teff (the effective temperature), log g (the surface gravity) and [Fe/H] (the metallicity) reflect the intrinsic physical properties of stars, such as masses, ages, and elemental abundances.
- From the view of point of machine learning, this task can be regarded as a regression, in which the representative spectra along with known parameters are used to train a model. This model is then used to predict the unknown stellar atmospheric parameters.

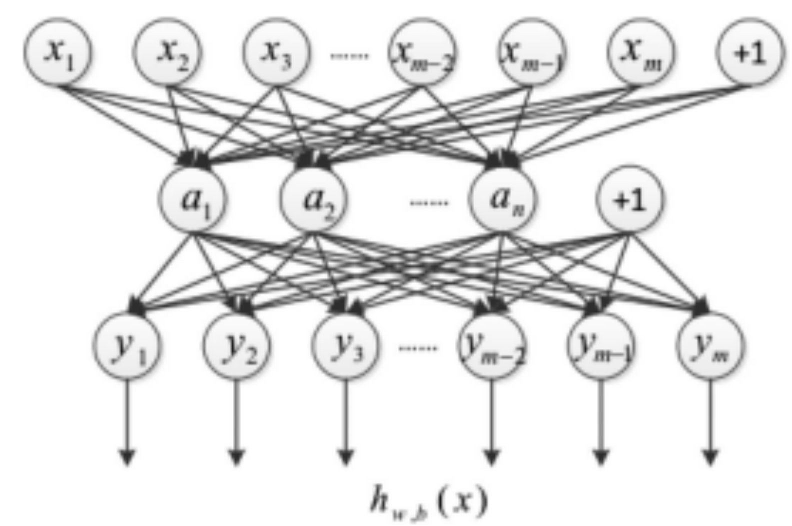
Stellar atmospheric parameters estimation

- ANN is a commonly used model for stellar parameters estimation.
  - Gulati et al. train an ANN with synthetic spectra to assign the effective temperatures for G-K dwarfs.
  - Bailer-Jones investigates the performance of ANN in the stellar parameters estimation with low SNR spectra.
  - Manteiga et al. take coefficients of fast Fourier transforms (FFTs) and discrete wavelet transform (DWT) as input features to an ANN which is trained to estimated the parameters Teff, log g, [Fe/H] and [α/Fe].

Stellar atmospheric parameters estimation

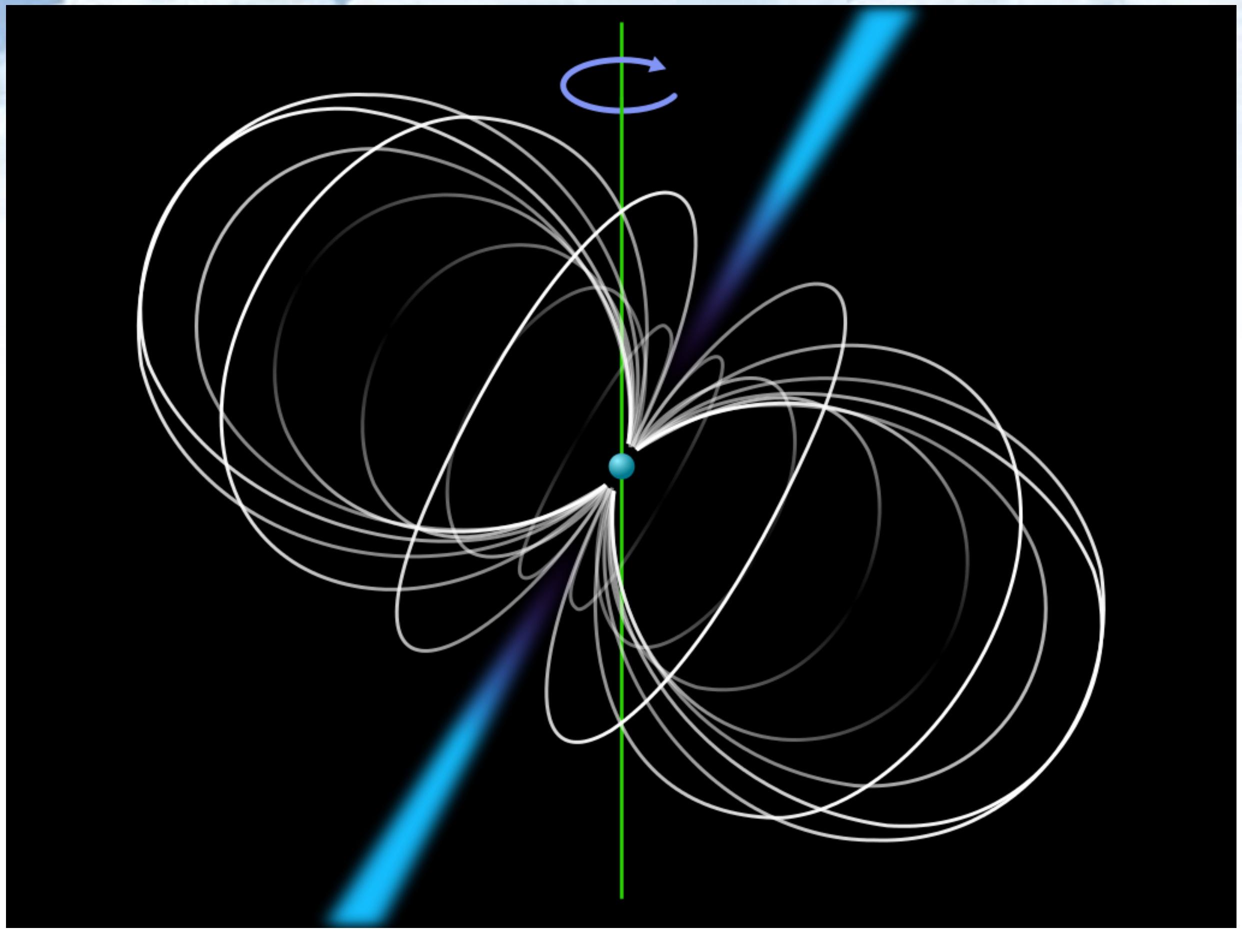
- Yang & Li use the auto-encoder to learn local representations from stellar spectra. Then they use a BP network to estimate stellar parameters.





Visualization of learned BSEs (filters) . Each BSE is learned in an autoencoder and then used to extract features in spectra.

Pulsar Search



P. Guo, F. Duan, P. Wang, Y. Yao. and X. Xin, Pulsar Candidate Identification with Artificial Intelligence Techniques. *arXiv :1711.10339*, 2017. (DCGAN+SVM)

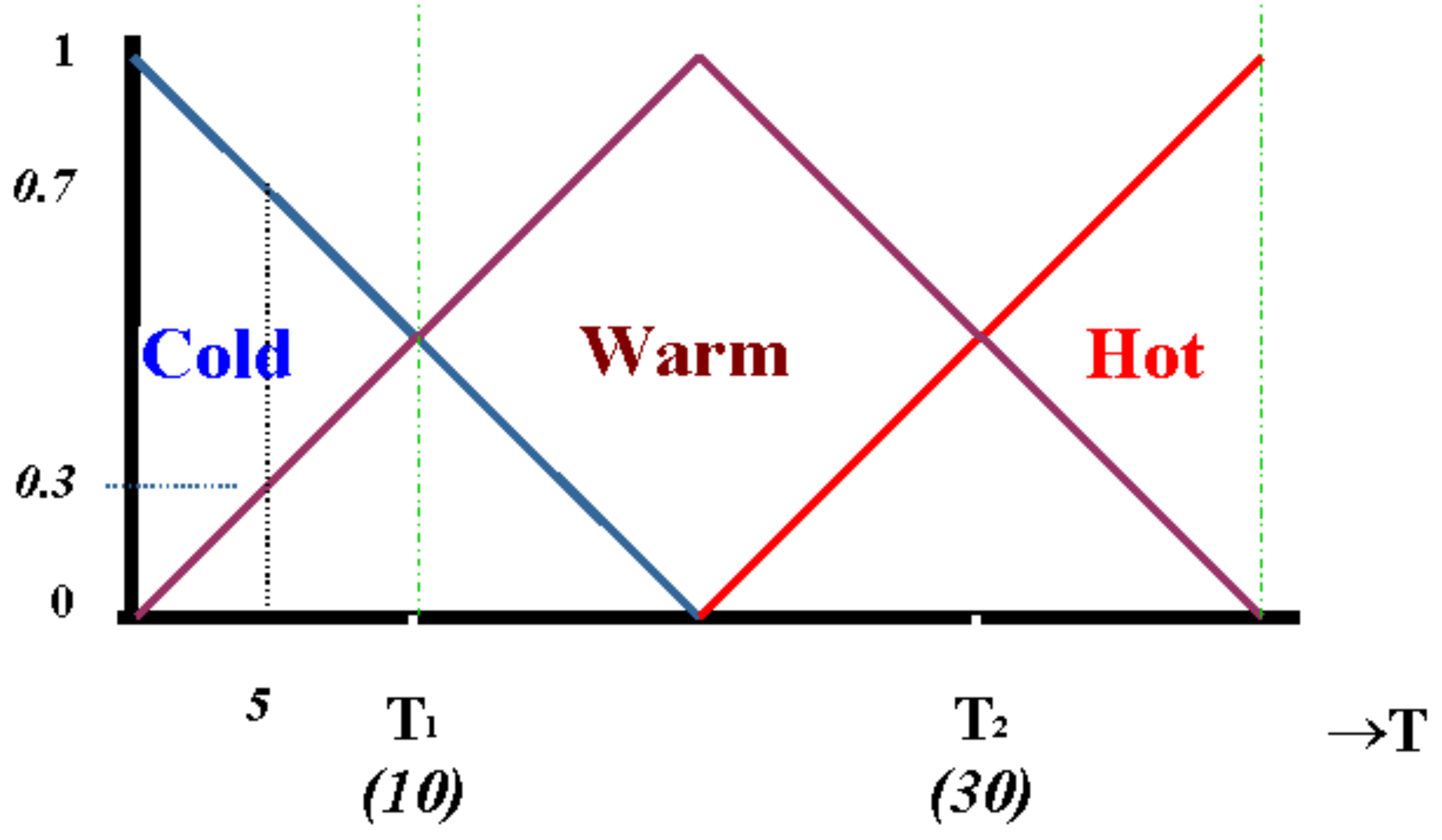
- A pulsar is a highly magnetized rotating neutron star or white dwarf that emits a beam of electromagnetic radiation.
- The first pulsar was observed on November 28, 1967, by Jocelyn Bell Burnell.
- Applications
  - Maps
  - Precise clocks
  - Probes of the interstellar medium
  - Probes of space-time
  - Gravitational waves detectors

Others

- In addition to the applications mentioned above, ANN has also been applied in many other astronomical data analysis tasks.
  - Solar activity detection (Borda et al. 2002)
  - Source detection in astronomical images (Masias et al. 2012)
  - The classification of gamma-ray bursts (Balastegui et al. 2001)
  - Removing cosmic-ray in CCD images (Waniak et al. 2006)
  - The weak lensing measurement (Nurbaeva et al. 2015)
  - The identification and study of Gravitational Wave (Carrillo et al. 2015, Kim et al. 2015)

Fuzzy Logic

- FL describes the linguistic expressions in daily life, such as tall, high, far, etc. It considers everything as a matter of degree, and uses membership degree to quantitatively describe the uncertainty.



FL in astronomy

- Fuzzy set theory was first applied in analysis and prediction of solar activity, using the technique of fuzzy clustering (Hu & Jiang 1985).
- FL has been applied in astronomy tasks including
  - Classification tasks
  - Analysis and prediction of solar activity
  - Other task, e.g., telescope control system, celestial navigation

Classification tasks

- Classification in astronomy are not such obvious and usually carry with noises, hence fuzzy set theory is supposed to be a useful tool in this area, such as
  - Galaxies morphological classification.
    - Fuzzy set was first applied in the morphological classification of the faint galaxies. (Spiekermann 1992)
  - Pulsating stars classification
    - Dumitrescu et al. used fuzzy hierarchical clustering to get the groups and possible structure of pulsating stars and then classify unknown samples by establishing the membership of an in a given class.
  - Star/Galaxy classification (Mähönen et al. 2000, Cortiglioni et al 2001)
- The advantage of fuzzy classifier is that the object can be seen how strongly it belongs to this class through the output membership values

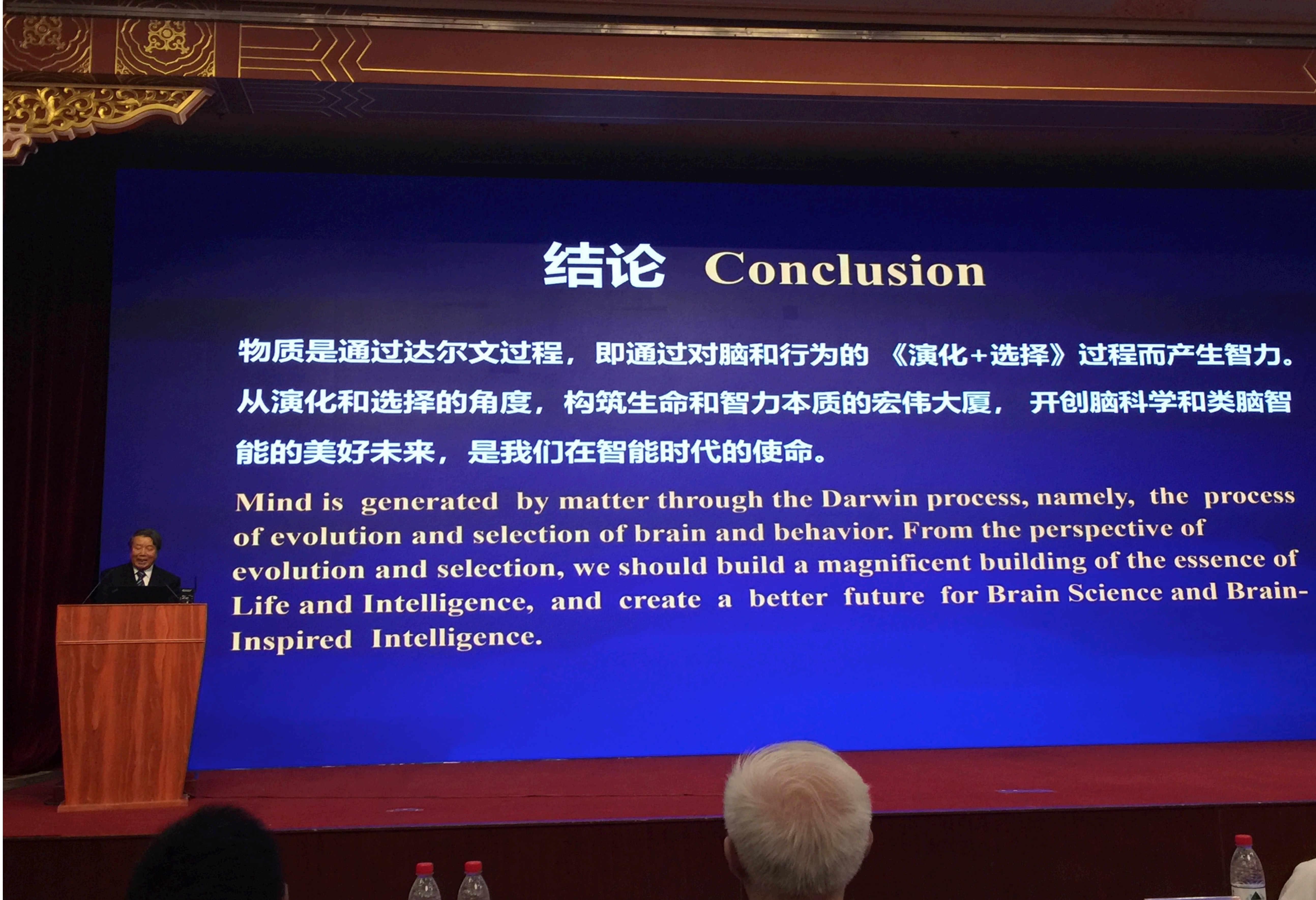
Analysis and prediction of solar activity

- The main task is to forecast the activity of the active regions as well as solar flares and coronal mass ejection (CME).
  - The activity prediction
    - Zhang used fuzzy system to predict maximum sunspot number and epoch.
    - Liu et al. predicted solar flares through fuzzy clustering of the active regions
    - Zhou et al. applied fuzzy comprehensive evaluation to forecast activity.
  - The solar image processing
    - Revathy et al. proposed a fuzzy technique to get the active regions from solar images.
    - Barra et al. segmented solar images into coronal holes, quiet sun and active regions via a fuzzy clustering algorithm.
    - Druckmüller put forward a noise adaptive fuzzy method to do image processing.

Others

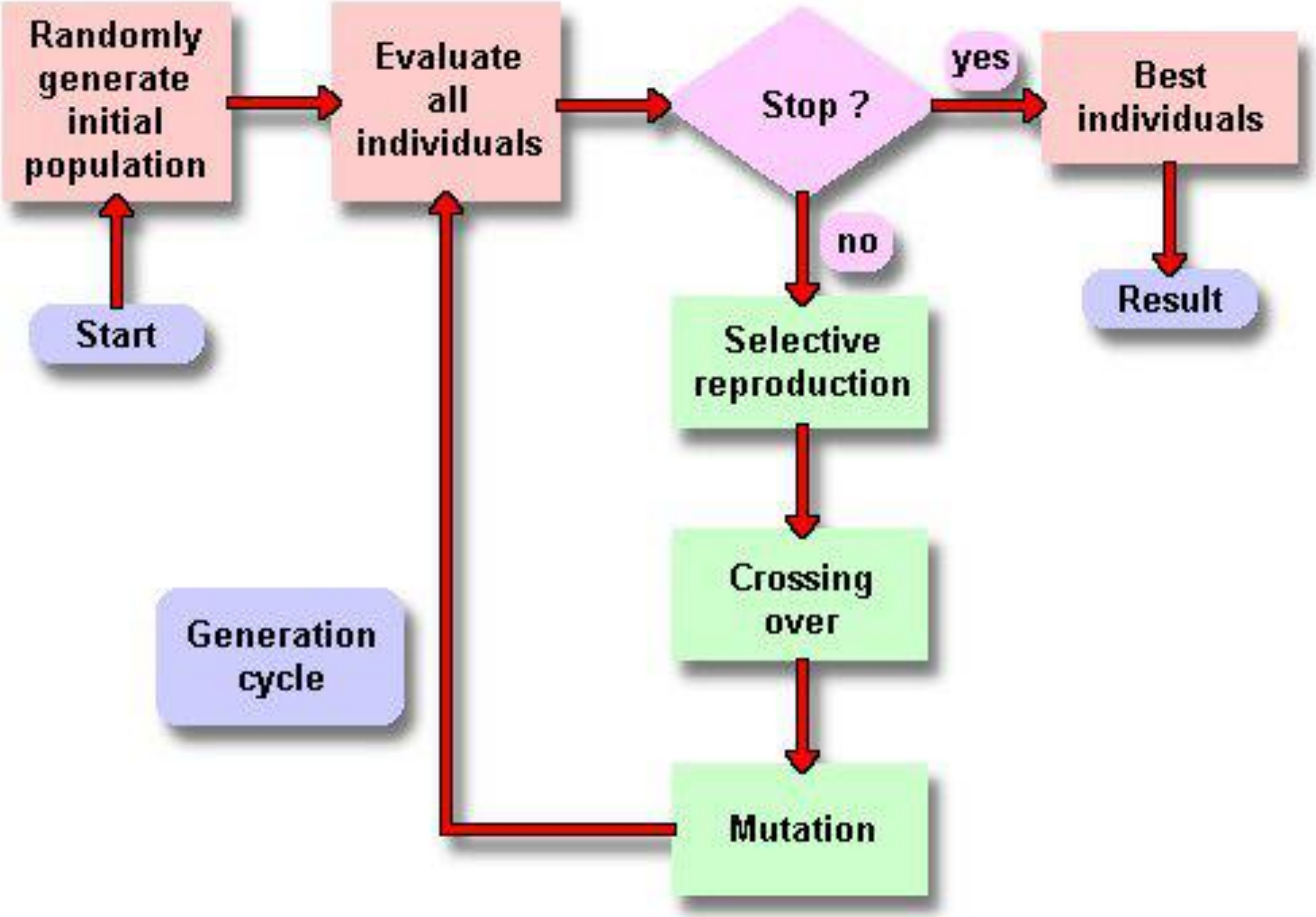
- Telescope control system (Attia 2009)
- Celestial navigation (Zhang et al. 2012)
- Removing cosmic-ray in images (Shamir 2005)
- Objects detection in astronomical images (Shamir & Nemiroff 2005).
- Deriving photometric redshifts (Speagle & Eisenstein 2015)

郭爱克院士在中国第二届系统科学大会演讲稿， 2018.5.12



Evolutionary computation

- Evolutionary Algorithms (EA) include a group of heuristics based on the mechanism for the natural selection proposed by biology Darwinism and have been proven to be effective to the complex optimization problems.



EA in astronomy

- From an astrophysics point of view, the applications of evolutionary computation in astronomy can be grouped into four fields:
  - Study of distant bodies for optical and radio band
  - Solar dynamics, helioseismology and stellar evolution
  - Study of galaxies and stellar population.
  - Monitoring and study of high energy events and associated theory

Study of distant bodies

- Search for planets with potential for life (Mayor & Queloz 1995).
- Identifying the planetary companion of stars (Lazio 1997, Lazio & Cordes 1993, Rozenkiewicz & Gozdziewski 2010).
- Curve fitting for galaxies rotation (Charbonneau & Knapp 1995).
- Characterization of extra-solar planets (Chwatal et al. 2008).
- Determining the orbits of multiple asteroids (Vachier et al. 2012)

Study of solar dynamic and helioseismology

- The first works where evolutionary computing techniques applied in study of stellar dynamic was proposed by Metcalfe in 1999.
  - Genetic algorithm is used to simulate binary systems by adjusting the observed light curves with those generated by theoretical models.
  - This task is computationally expensive. The author implemented a distributed GA on a grid of twenty-five workstations.

Study of galaxies and stellar population

- Study the evolutionary history of the galaxy NGC4449 (Theis 1999)
  - Mathematically, the problem can be approached by an optimization model of fitting an N-body simulation for a given dataset.
- Study the influence of interaction between and the galaxy NGC5195 and the Messier M51 (Wahde et al. 2001).
- EA + ANN to classify of distant galaxies (Cantu & Kamath 2002).
- Genetic algorithm + simulated annealing to estimate the parameters of stellar population synthesis (Han 2015).

Study of supernovae and others high-energy events

- Modeling and explaining real observations of gamma-ray burst (Zwart et al. 2001).
- Detecting gravitational waves.
  - Lightman et al. proposed the implementation of EA to maximize the chances of detecting gravitational waves and minimize the risk of false alarms in LIGO in 2006.
  - The first paper using genetic algorithms to the analysis of data from LISA was introduced by Crowder & Cornish in 2006.
    - Evolutionary approach is used to isolate potential sources of gravitational waves from the possible tens of thousands of overlapping signals in the LISA data stream.

人工智能不仅仅是深度学习

✧天文学中的计算智能方法

- International Journal of Computational Intelligence Systems，special -issues： “Computational Intelligence in Astronomy”
- Publication date：March 2018；Vol. 11. Edited by Prof. Ping Guo, Yuping Wang, Hailin Liu and Yiu-ming Cheung.

https://www.atlantis-press.com/journals/ijcis/special-issues

Computational Intelligence in Astronomy: A Survey

Ke Wang, Ping Guo, Fusheng Yu, Lingzi Duan, Yuping Wang, Hui Du

Thanks