

# 一天构建深度学习模型进 Kaggle 赛 (MIT 教授彩标收藏版, 附六一新闻四则及历史上今天名人和事件; 欢迎赞赏、留言; 点击“阅读原文”下载 12k 字 10 页彩标 PDF) 祝您六一快乐!

秦陇纪 10 汇编

欢迎关注、收藏、转发科普作家“秦陇纪 10 数据简化 DataSimp”的 QQ 空间、微信公众、头条号、新浪博客(文章)、微博(动态)、CSDN(知识图谱)、OSC(源代码)、知乎(问答)等“数据简化 DataSimp、科学 Sciences”新媒体文章, 也欢迎加入各地各行业媒体、技术组; 转载注明出处: 秦陇纪 10 数据简化 DataSimp 公众号、头条号“数据简化 DataSimp、科学 Sciences”汇译编, 投稿邮箱 QinDragon2010@qq.com。

## 目录

一、我是如何在 1 天内构建一个深度学习模型并进击 Kaggle 比赛的深度学习 (2238 字)	1
从数学开始学习人工智能是个错误的选择	1
我为什么决定参加这个课程	3
验证结果	4
相关链接和群	6
二、纯经验知识、无实践的教学方式适合孩童教育吗?? (1316 字)	6
三、新闻四则及历史上今天名人和事件及历史上的名人和大事件 (5450 字)	7
附 i. 早报, 6 月 1 日, 星期四, 六一儿童节快乐, 各样老小孩们!	7
附 ii. 2017 年 6 月 1 日周四读报时间! 一切美好从内心开始!	7
附 iii. 2017 年 6 月 1 日 (丁酉鸡年五月初七) 周四 / 早读分享: (祝福您儿童节快乐!)	8
附 iv. 新闻早餐 第 4615 期 (文字版) 2017 年 6 月 1 日 (星期四) 农历丁酉年五月初七 己未日	8
附 v. 2017 年 6 月 1 日农历干支、节日、历史上的今天名人和事件	9
Appx. 数据简化 DataSimp 社区译文志愿者 (610 字)	9

**简介:** 22 年前, 我在斯坦福大学学习 Bernie Widrow 的神经网络课程, 这位数字滤波器之父谆谆善诱, 向我们讲解随机梯度下降的原理。学了一半的理论课程, 我还是不能完成神经网络的实现, 只学会了如何用 MatLab 构建神经网络。后来, 当我需要编写代码来训练我的网络权重时, 我才终于理解了什么是随机梯度下降。附周四新闻四则及历史上今天名人和事件; 点击“阅读原文”可下载 12k 字 10 页彩标收藏版 PDF 文档。

**作者:** 埃里克·布林约尔松(Erik Brynjolfsson), 麻省理工学院斯隆管理学院教授; 第二机器时代(*the Second Machine Age*)的作者。**来源:** 本文选自 Fast.ai 深度学习课程地址 (英文) <http://course.fast.ai/>, 第 1 部分第 2 个实践模块课程《我是如何在 1 天内构建一个深度学习模型并进击 Kaggle 比赛的深度学习》, 源码 GitHub 地址 [https://github.com/kevindeval/ai-notebooks/blob/master/invasive\\_species.ipynb](https://github.com/kevindeval/ai-notebooks/blob/master/invasive_species.ipynb), 网站 Fast.ai 是 Jeremy Howard 为结果导向型人群开设的深度学习在线课程。欢迎来到 fast.ai 的第 7 周课程“程序员深度学习实践, 第 1 部分”, 由 Jeremy Howard 杰里米·霍华德 (Kaggle 的 #1 参赛手、2 年实战经验, Enlitic 创始人) 讲授。学习如何艺术建立不需要研究生水平数学的那种富有艺术的模型。哦还有一件事...这是完全免费的! “I highly recommend this course. Jeremy is an amazing teacher” “我高度推荐这门课程。杰瑞米是个了不起的老师。”



图 1 埃里克·布林约尔松(Erik Brynjolfsson), 麻省理工学院斯隆管理学院教授

## 一、我是如何在 1 天内构建一个深度学习模型并进击 Kaggle 比赛的深度学习 (2238 字)

我读过很多关于机器学习的书, 也参加过不少这方面的课程, 但我认为 Fast.ai 是迄今为止最实用的深度学习入门课程。在本文中, 我将与大家分享我在学完前两课后的一些经验。

### 从数学开始学习人工智能是个错误的选择

22 年前, 我在斯坦福大学学习 Bernie Widrow 的神经网络课程, 这位数字滤波器之父谆谆善诱, 向我们讲解随机梯度下降的原理。

学了一半的理论课程, 我还是不能完成神经网络的实现, 只学会了如何用 MatLab 构建神经网络。后来, 当我需要编写代码来训练我的网络权重时, 我才终于理解了什么是随机梯度下降。

不过，这种窘境到现在也并未发生多大变化。大多数机器学习实战工程师都有着学术背景；当专业程序员问他们“我该如何开始学习人工智能？”，多数工程师会向他们推荐一些偏重数学的学习资料，比如《深度学习》这本书（这种情况时有发生）。

对研究人员而言，Ian Goodfellow 的这本书是一本很好的参考书，但对我们中的大多数人而言它并不是一本好的入门书籍。这本书关于梯度下降的章节使我回想起读研究生时遇到的窘境：

#### CHAPTER 4. NUMERICAL COMPUTATION

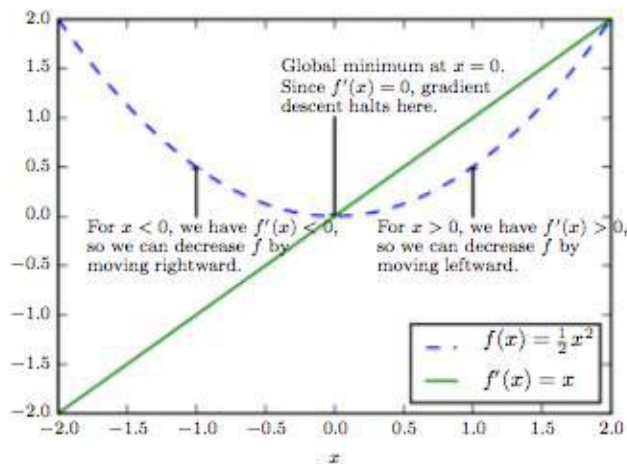


Figure 4.1: An illustration of how the gradient descent algorithm uses the derivatives of a function can be used to follow the function downhill to a minimum.

We assume the reader is already familiar with calculus, but provide a brief review of how calculus concepts relate to optimization here.

Suppose we have a function  $y = f(x)$ , where both  $x$  and  $y$  are real numbers. The **derivative** of this function is denoted as  $f'(x)$  or as  $\frac{dy}{dx}$ . The derivative  $f'(x)$  gives the slope of  $f(x)$  at the point  $x$ . In other words, it specifies how to scale a small change in the input in order to obtain the corresponding change in the output:  $f(x + \epsilon) \approx f(x) + \epsilon f'(x)$ .

The derivative is therefore useful for minimizing a function because it tells us how to change  $x$  in order to make a small improvement in  $y$ . For example, we know that  $f(x - \epsilon \text{sign}(f'(x)))$  is less than  $f(x)$  for small enough  $\epsilon$ . We can thus reduce  $f(x)$  by moving  $x$  in small steps with opposite sign of the derivative. This technique is called **gradient descent** (Cauchy, 1847). See figure 4.1 for an example of this technique.

When  $f'(x) = 0$ , the derivative provides no information about which direction to move. Points where  $f'(x) = 0$  are known as **critical points** or **stationary points**. A **local minimum** is a point where  $f(x)$  is lower than at all neighboring points, so it is no longer possible to decrease  $f(x)$  by making infinitesimal steps. A **local maximum** is a point where  $f(x)$  is higher than at all neighboring points,

上图为 Goodfellow 等人所著《深度学习》中关于梯度下降的一页。这本书有很好的研究参考价值，但是对专业程序员而言，并不是一本实用的入门书。我在圣诞节期间读完了本书三分之二的内容，但我还是不会构建神经网络。

对于研究人员或者人工智能前沿的推进而言，理解理论是至关重要的一步。但是，对于大多数人工智能商业应用而言，数学是个很大的干扰。

为什么？因为企业面临着一些更为严峻的挑战，例如选择正确的问题，组织数据，部署解决方案等。面对市场风险和执行风险，专业程序员不得不使用成熟的技术来开发应用——因此，懂得如何实现某一解决方案要比理解该方案的工作原理更为重要。

光读不练假把式，必须实践。

学习新技能的唯一方法就是：使用该技能练习构建。深度学习也一样，Jeremy Howard 的 Fast.ai 课程注重实践，是学习人工智能最快的方法。

Fast.ai 让你通过实践学习深度学习

课程的联合创始人 Jeremy Howard 和 Rachel Thomas 实现了他们的承诺——“让所有人都能运用机器学习的力量”。本文不对课程进行描述，我会推荐你阅读学习者对该课程的评价并分享我早期的经验。



### 我为什么决定参加这个课程

在参加该课程之前，我已具备良好的深度学习理论基础，并掌握了该技术各种应用的实践知识。我参加这个课程是为了获得更多实践经验。

构建简单的解决方案能帮我更加有效地筹划项目、雇用项目人员和进行工程学权衡——这是任何领导者都须具备的技能。

但是最关键的是，我想通过构建简单的深度学习解决方案来实现理论和实践的相结合。

如何建立一个算法来检测入侵物种

在一天内能达到什么样的水平？

学完前两课后，我决定参加 Kaggle 比赛。

我的目标是：

在一天结束前进入排名的前 50%。

自己动手编写每行代码，不用剪切和粘贴。

了解所有引用的库和 API。

我选择了“入侵物种监测”竞赛（Invasive Species Monitoring playground competition）。这是一个比较简单的分类问题，对初学者而言算是一个很好的起点。

还有，它听起来就很有趣。



上图为“入侵物种监测 Kaggle 竞赛”中入侵性绣球花（绣球花：一种植物的分类）。

比赛的内容是开发出识别森林和叶子图像中是否包含入侵性绣球花的算法。

早上七点，我沏了一壶咖啡后开始工作，在亚马逊 AWS 上对 GPU 进行了加速。晚上 9 点，我的眼睛开始酸痛，我的大脑跟炸了一样。

我决定将我的结果提交给 Kaggle……

我的代码

这是我在 github 上的完整笔记，里面记录了我工作计划的 7 个步骤

```

val_acc: 0.8947

running epoch 2
Epoch 1/1
2067/2067 [=====] - 82s - loss: 0.4612 - acc: 0.8665 - val_loss: 0.3063 -
val_acc: 0.8947

running epoch 3
Epoch 1/1
2067/2067 [=====] - 83s - loss: 0.3643 - acc: 0.8853 - val_loss: 0.2343 -
val_acc: 0.9079

running epoch 4
Epoch 1/1
2067/2067 [=====] - 83s - loss: 0.4234 - acc: 0.8791 - val_loss: 0.2722 -
val_acc: 0.9079

```

### Evaluate model

Evaluate some images to check our results

```

In [131]: latest_weights_filename = '77272_vgg_4.h5'

In [147]: #load weights and generate predictions of validation set for review
vgg.model.load_weights(results_path + latest_weights_filename)
valid_batches, valid_preds = vgg.test(valid_path)
valid_filenames = valid_batches.fileNames
expected_labels = valid_batches.classes

Found 228 images belonging to 2 classes.

In [252]: valid_batches.class_indices
Out[252]: {'harmless': 0, 'invasive': 1}

In [145]: #round off predictions to 0 or 1 to make labels
s = valid_preds[:5]
s

```

不用对我的代码感到惊讶——我使用的方法与课堂上学习的方法类似。任何一个具有 1 年工作经验（或以上）的程序员应该都能写出这样的代码。

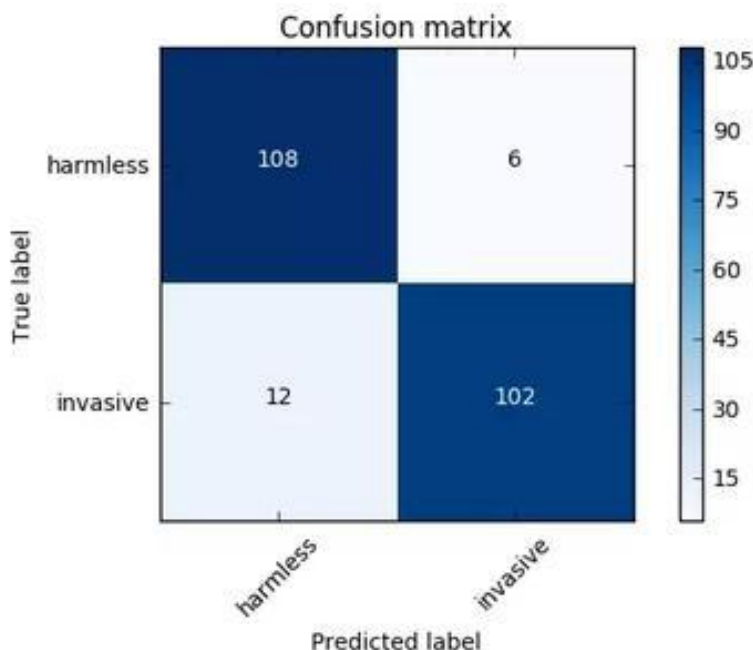
我 90% 的时间都用于学习库、组织数据和评估结果——都是一些与深度学习无关的简单流程。

### 验证结果

我使用 90% 的图像训练我的算法，并用剩余的 10% 来验证结果。

预测精度

我的预测精度达到了 91%。这是混淆矩阵：



混淆矩阵显示了我的算法检测验证集的有效性。最需要改进的地方是漏报问题——在 12 个实例中算法错误地将入侵物种识别为非入侵物种。您可以在下方查看该结果。

91% 是什么概念？就在 3 年前，最先进的计算机视觉识别技术也只能达到 80%！我在第一次尝试中就超越了它。

包含花的图像被正确识别

我检查了被算法正确识别的图像。大蓝色绣球花似乎是关键特征。

**Correctly labeled images**

```
#selected at random
correct = np.where(our_labels == expected_labels)[0]
print("found %d correct labels" % len(correct))
idx = np.random.permutation(correct)
titles = np.where(our_labels[idx[:n_view]] == 0.0, 'harmless', 'invasive')
plots_idx(idx[:n_view], titles)
```

found 210 correct labels



```
idx = np.random.permutation(correct)
titles = np.where(our_labels[idx[:n_view]] == 0.0, 'harmless', 'invasive')
plots_idx(idx[:n_view], titles)
```



某些被错误识别的图像存在明显问题

这个算法错误地识别了 18 个图像。这是算法需要改进的地方。

**Incorrectly labeled images**

```
#selected at random
incorrect = np.where(our_labels != expected_labels)[0]
print("found %d incorrect labels" % len(incorrect))
idx = np.random.permutation(incorrect)
titles = np.where(our_labels[idx[:n_view]] == 0.0, 'harmless:', 'invasive')
plots_idx(idx[:n_view], titles)
```

found 18 incorrect labels



```
idx = np.random.permutation(incorrect)
titles = np.where(our_labels[idx[:n_view]] == 0.0, 'harmless:', 'invasive')
plots_idx(idx[:n_view], titles)
```



漏报的图像似乎是未被正确识别的小花。我认为通过裁剪照片生成更多标记样本，可以改善算法的结果。  
胜利！我进入了前 50%！

在我的第一次尝试中，我在 52 位参赛者中排名第 21，进入前 50%。

目标达成！



Rank	Delta	Name	Score	Submissions	Time
12	+4	novice	0.98015	2	2h
13	+2	123	0.97952	14	9h
14	+2	Ricardo Quintas	0.97559	8	2d
15	+2	Sagnik Dutta	0.97530	1	5d
16	+2	zhongtao2016	0.97226	5	4d
17	+2	Ville-Ernard AI Institute	0.97140	1	2d
18	+1	Kengo Ichiki	0.96983	5	2d
19	+1	Adrien Lucas Ecoffet	0.96107	3	5d
20	+1	AZ	0.95665	1	8d
21	new	Kevin Dewalt	0.95649	1	-10s
22	+2	Jake Hawkesworth	0.94662	1	5d

**Your Best Entry**  
Your submission scored 0.95649, which is an improvement of your previous score of . Great job!

我在一天内进入了排名的前 50%。

没有最好只有更好。

Fast.ai 是否适合你？

如果你的目标是加入 Open.ai 团队或赢得 ImageNet 竞赛，那么 Fast.ai 课程并不十分适合你，因为你需要学习偏导函数。

但对于其他人，比如想要获得实用的基础知识介绍的团队领导，希望学习深度学习工程师入门知识的专业程序员，想知道如何开发 AI spec 的产品经理，或者任何想对一吨乐高进行分类的人员来说，Fast.ai 是个很好的开始。

## 相关链接和群

GitHub 地址 [https://github.com/kevindewalt/ai-notebooks/blob/master/invasive\\_species.ipynb](https://github.com/kevindewalt/ai-notebooks/blob/master/invasive_species.ipynb)

Fast.ai 深度学习课程地址（英文）<http://course.fast.ai/>

Invasive Species Monitoring 网站 <https://www.kaggle.com/c/invasive-species-monitoring>

原文地址

<https://medium.com/the-business-of-ai/fast-ai-how-i-built-a-deep-learning-application-to-detect-invasive-species-in-just-1-day-and-for-38e0ced809e9>

“数据简化 DataSimp 社区”群：223518938（需正确回答问题：数据简化的英文是什么？DataSimp）兴趣范围包括数据资源、数据分析、数据简化，数据处理分析工具，数据仓库，数据挖掘工具，报表系统等全方位知识。

## 参考文献

[1] 研 MIT. Fast.ai 深度学习课程地址（英文）[EB/OL]. <http://course.fast.ai/>.

[x] 秦陇纪. 数据科学与大数据技术专业概论；人工智能研究现状及教育应用；文献数据共词分析的神经网络训练；大数据简化之技术体系[EB/OL]. 数据简化 DataSimp（微信公众号），2017-08-23.

赞赏后，点击“阅读原文”，可下载完整 PDF。

## 二、纯经验知识、无实践的教学方式适合孩童教育吗？（1316 字）

【秦陇纪 10 言】无真实实践的纯经验知识和理论教育，不是正常的人类认知，但可以提高文化知识储备。孩童在日常生活中的语言文字交流，可以了解各种信息，但不是现实世界的实践和体验。学校简单地用纯经验知识、无实践纯文化的教学方式教育一个孩童，孩童的心智深受影响：对真实世界的认知、自身情感的体验，会形成思想层面记忆性想象为主的思维和习惯；未来将很难有自主认知能力、创造性，面对社会真实事物，得重新建立认知，比实践过来的人面对更大的挫折，甚至成年后一生都无法转变已经习惯了的“空想空谈”纯想象思维定势。

长期脱离实践和实际社会搞纯经验知识教育，对人有极大副作用。现实社会里，实质工作只有 20% 左右，一生都在空想空谈发空文者占 80% 左右，给人民造成了负担。究其原意，就是从幼儿园起，到走出学校，基本习惯了只会想象，到做事时，大家发现除了五花八门的各种想法、设想、理论，基本的实践和应用却做不来！根源在于经验知识为主的教育，让人习惯了思考和写作，实践实验操作能力被阉割了，以至于走出学校，依然习惯了空想空谈，发表文章也是空洞的说教、理论、想法、评价、访谈之类的。由于做事的权力从小就被阉割，真想做事，已经做不来了。这就是我国脱离社会实际生产生活的教育造成的最严重后果。

还不警醒？教育人脱离历史、社会、趋势，对社会不甚了解，在所谓的“教育学术界”、“文化知识范围”搞“教育教学教研”，甚至很多人把**唯书本和文学当成了教育**。那些当官的、成功的，都是早都看穿了的人，只剩下**钻学校教育牛角尖**的教书匠固执己见、在已经脱离实际很远的课堂教学上翻来覆去搞各种花样继续残害孩童。试问，一个连**家人、邻里、社区、单位基本常识**，连**自然世界、民族**都没充分接触过，手无**缚鸡之力、不能做人类技能传承者**的书生，受**应试教育**后将去做什么？

类似地方领导习惯了**土地财政**、GDP 严重依赖**房地产和基础设施**，真想发展科技、提高社会各部门效能、走向**信息化、自动化、智能工农业**，已经不懂得怎么**落地操作、无人可用**，只好延续旧思路、裹足不前了。

那些**思想意识玩家、政府闲官员、伪学者、行业骗子、纯资本家、广告商、脱离社会实际、外行领导内行**的人，无所谓谁是实干的、实干的是谁，干不干都非此类**耍嘴皮子、玩文字游戏、满脑子理论知识、搞关系为生**的人负责，他们只搞社群组织或坐收渔利。但喊口号、讲理论也是必不可少的。**做事要氛围、凝聚力、目标**，喊喊口号只是**做事**中最简单、最先的一步，达成了认识、大家都知道做了，这个时候需要真干事的人。如果**把事情做成**可分为十个**阶段或十等份**，那么，**动动嘴皮、喊喊口号**只是前面的一份，**尝试**能占五六份，找对方法**做成、优化**能占两三份，最后一份是大家**认可、使用**才能算做成了。

我们要大步往前走，要做**具体的生产实验研究**，不停留在**低级肤浅的思想意识或文字口号**上！对于人才的培养，绝不可像过去条件所限采取**过分简单化搞教育**；更不可把大学及研究生层次的教学，搞成**教书本和考试**。对于科研，不能拿发文章的数量来衡量，学术文章脱离了**社会和产业**，彻底玩弄概念和理论，还有什么作用；真的有价值的东西能是发文章的吗？要实干，要做一个环节、一个阶段的事情，需要更有效的管理体制。

### 三、新闻四则及历史上今天名人和事件及历史上的名人和大事件(5450 字)

附 i. 早报, 6 月 1 日, 星期四, 六一儿童节快乐, 各样老小孩们!

- 1、天津港内一纸制品公司起火 系中外合资企业 过火面积超过一千平米;
- 2、美军两艘**航母联合演习**向朝鲜施压, 外交部回应: 望各方克制;
- 3、最高检: 去年检方批捕涉**校园欺凌和暴力**犯罪案件 1180 人;
- 4、军报: **军队和武警部队**全面停止有偿服务工作计划于 2018 年完成;
- 5、北京超 **10%吸烟学生**每日吸 10 支以上 7 成学生购烟未被拒;
- 6、中央财政拨付 **4.85 亿补助资金** 支持东北华北抗旱等;
- 7、国赔新标准: 侵犯公民人身自由权**每日赔偿 258.89 元**;
- 8、**人民币汇率走出寒冬**: 高估压力已近完全释放, 趋稳将成新常态;
- 9、《网络安全法》今日起施行 **贩卖 50 条个人信息可入罪**;
- 10、**北京二手房价大面积下跌** 部分区域跌幅超 20%;
- 11、公安部部署启动机动车号牌管理改革, 推广应用**全国统一选号系统**和号牌生产管理系统; ;
- 12、联合国秘书长: 若美国退出全球领导角色会有其他国家取而代之;

【微语】在你渐渐迷失在你的人生道路上的时候, 记得这句话: 千万不要因为走的太久, 而忘记了我们为什么出发。

附 ii. 2017 年 6 月 1 日周四读报时间! 一切美好从内心开始!

- 1、[海外网] 韩国**新世界集团**旗下**易买得超市**将全面退出中国市场。这是自易买得超市传出将关闭在华业务以来, 集团高层首次出面证实。据悉, 易买得超市 1997 年首次进入中国市场, 至今已有 20 年的在华经营历史。
- 2、[人民网] 节后上班第一天的“**打虎**”节奏实在太过瘾。接连 6 只受审大老虎一审宣判, 其中宁波市原市长卢子跃、国家统计局原局长王保安、洛阳市委原书记陈雪枫受贿均超过 1 亿元, 三人被判无期。另有四川省原副省长李成云获刑 16 年, 武钢董事长邓崎琳获刑 15 年, 中国电信原董事长常小兵获刑 6 年。
- 3、[南湖晚报] 众所周知, 在浙江嘉兴使用固定电话, 区号是“0573”, 但是**平湖和嘉善**却有几家企业的固定电话开头却是“**021**”的**上海区号**。作为嘉兴电信的试点项目之一, 预计今年六七月份将推广到嘉兴全市, 届时嘉兴的企业如果想使用上海“021”号码, 可以到电信的中心厅进行登记办理。
- 4、[华尔街日报] 福布斯杂志公布了**全球上市公司 2000 强排行榜(2017 版)**。中国石化以 2557 亿美元的年营收, 蝉联了全球最大上市石油公司冠军; 荷兰皇家壳牌则超越中国石油, 晋升到了全球第二的位置。
- 5、[国家旅游局] 阿根廷目标: 2019 年吸引**中国游客 40 万人次**。中国与阿根廷双方就互发 10 年多次有效旅游、商务签证达成一致。阿根廷旅游属于市场上比较小众的产品, 主要针对的是中国高端客群。此次阿根廷发放 10 年多次签, 意味着中国高端出境游市场正在逐渐孕育成熟, 潜力很大。
- 6、[中新网] 近日, **菜鸟网络**宣布设在法国的海外仓正式开通运营。未来, 菜鸟法国海外仓将和西班牙海外仓一起面向全欧服务, 大幅提升欧洲消费者在线购买中国商品的 72 小时达比例。
- 7、[新华网] 端午节期间, 美国知名电商企业亚马逊跨境电商直邮货物同时从芝加哥、洛杉矶、纽约三大港口起飞, 首次空运至郑州新郑国际机场, 该批商品共涉及近千种商品。这是美国亚马逊在我国中西部地区开通的第一个直邮口岸。
- 8、[36 氪] 市场调研机构 Stratechery 送出一份调查数据显示, 微信的流行在一定程度上让国人购买 iPhone 的兴趣降低。Stratechery 创始人本·汤普森强调, 在中国国内手机市场, 操作系统并没有这么重要, 想法应用才是最重要的是, 比如微信已经在国内变得无处不在。国内用户忠诚度降低是苹果最头疼的, 但这样的情况在国外却基本没有发生。
- 9、[新浪体育] 阿森纳官方宣布续约温格, 两年合约留教授至 2019。本赛季阿森纳在联赛中的成绩并不优异, 最终他们排名联赛第五, 无缘下赛季的欧冠联赛,。不过温格率队拿到了足总杯冠军, 温格获得了高层的力挺与支持, 最终他得到了续约合同。
- 10、心灵点滴: “一条鱼向前方游去, 众鱼纷纷阻挡: ‘危险, 千万别游向那里!’”为什么危险呢? 那鱼问。众鱼心悸道: ‘那里有很多诱饵, 我们不少同类有去无回。’鱼说, ‘诱饵搁在那里, 它不会伤害你, 只有当你禁不起诱惑, 去吞食诱饵时, 那才危险。’在这个世界上, 危险的不是诱饵, 是一颗禁不起诱惑的心。”

美好的一天从内心开始!



**附 iii. 2017 年 6 月 1 日（丁酉鸡年五月初七）周四 / 早读分享：（祝福您儿童节快乐！）**

1、【李克强总理离京访问德、比并出席中欧领导人会晤】昨天总理离京，正式访问德国并举行中德总理年度会晤、赴布鲁塞尔举行第 19 次中国-欧盟领导人会晤并对比利时进行正式访问。李总理此访将与德、欧、比领导人共同发出推动构建开放型世界经济、支持贸易投资自由化便利化、维护多边主义、维护世界和地区和平稳定的积极信号。

2、【安邦深夜再发公开信：已起诉财新特约作者郭婷冰】5 月 3 日，安邦集团在其官方网站刊登致胡舒立女士的公开信，31 日夜再次刊登致胡的公开信，表示，5 月 26 日在加拿大法院对财新传媒特约作者提起诉讼。

3、【惠誉警告百度卷入影子信贷 面临违约风险】继 5 月 18 日穆迪将百度 A3 的发行人评级及高级无抵押债券评级列入下调复评名单后，惠誉也将百度列入负面观察名单，理由均为担心高速增长金融服务业务将增加公司的声誉风险及或有负债风险。（百度财务数据显示，至 3 月底，其全资所有、负责理财及小额贷款业务的金融服务事业群组 "FSG" 总资产达 250 亿元人民币，相当于百度总资产的 12%。而这一资产规模去年底为 120 亿元，表明在三个月内翻了一倍多。）

4、【今年内或将正式公布并实施养老保险综合改革方案】这份酝酿了四年多的方案，将对关系到每位参保人切身利益的养老保险统筹层次、完善个人账户以及延迟退休等重要改革作出制度安排。

5、【微软的创始人比尔·盖茨认为影响世界的三个领域】第一个是人工智能领域、第二个是能源领域、第三个是生物科学领域。

6、【中美科学家发现：以纳米材料激发免疫系统抗癌】纳米免疫颗粒的生产方法、大批量生产工艺、剂量测试等研究正在中山医院进行。科学家希望纳米医学能够为癌症患者带来福音。

7、【日本将下调债务占 GDP 比重的目标】因为安倍政府似乎难以在 2020 财年实现基本预算盈余的目的。

8、【美联储本月加息预期增大】据 GME "美联储观察"：美联储 6 月加息概率为 91.2%，9 月加息概率为 93.4%。

9、【今年 IPO 申请被否达 27 家 多家企业涉嫌商业贿赂折戟首发】5 月 31 日，证监会审核了 7 家公司的首发申请，浙江时代电影院线股份有限公司被否。从今年被否企业的情况看，持续盈利、规范运行、信息披露方面受到发审委询问的次数最多，另外，还有多家企业因涉嫌商业贿赂而折戟首发。

10、【仙言潮声】童心就是一块糖果带来的甜蜜，就是一件玩具带来的惊喜，就是一次旅行带来的兴奋……她简单、纯粹，但时时刻刻都充满着期待和快乐！

美好的一天从拥有一颗童心开始！

**附 iv. 新闻早餐 第 4615 期（文字版）2017 年 6 月 1 日（星期四）农历丁酉年五月初七 己未日**

每天三分钟 知晓天下事

**A、【国内】**

1) 全国“扫黄打非”部门突出整治校园周边文化市场和净化未成年人上网环境两项重点，打出护苗组合拳迎“六一”；中国儿童道路交通安全蓝皮书发布，关注儿童座椅安全；

2) 刘云山：深入学习习近平讲话贯穿的马克思主义观点；

3) [法制与反腐] 国家统计局原局长王保安受贿超 1.5 亿元，被判无期徒刑；湖南湘潭教育学院两名正处级干部被开除党籍并降为科员；多地推出升级版“禁酒令”约束公务员更严更细；

4) 中共中央批准：陈刚同志任河北省委员、常委；贵阳书记陈刚履新河北，曾任职北京朝阳；陕西发布干部任职公示，徐启方拟任宝鸡市委书记；

5) 中国记协发布《中国新闻事业发展报告（2016 年）》，新兴媒体法规监管不断完善；

6) 工信部首次将民营铁塔纳入共建共享协调机构；爱立信基于物联网技术积极布局智能家居；

7) 山西临汾交警查车闹出人命，警方：交通事故；

8) [军事] 北京卫戍区“老虎团”一连指导员直接越级提升成营教导员；

9) [港澳] 广深港高铁香港段工程已完成 90%；中银香港与微信支付（香港）共促港移动支付；澳门 2017 年第 1 季度本地生产总值增长 10.3%；香港各界“筑梦之旅”贵州专列团在港启动；

10) [台湾] 国民党党魁交接陷僵局，洪秀柱停开昨日中常会盼以大局为重；阿联酋航空要台籍空姐不得使用“青天白日旗”，并以五星红旗替代，台媒炸锅；台北国际电脑展开幕，人工智能和物联网成亮点。

**B、【国际】**

1) 李克强离京正式访问德国、比利时；李克强抵达柏林对德国进行正式访问；习近平主席特使王勇出席肯尼亚蒙内铁路通车仪；

2) 联合国秘书长：若美退出全球领导角色，他国会取代；联合国拒为日慰安妇立场站台，要其重新审教科书；

3) 美军首次洲际弹道导弹拦截测试成功，曾称明确针对朝鲜；

4) 特朗普再狠批德国军费没付够，德国政党怒了；默克尔启动欧洲独立，美国治下世界秩序渐崩塌；特朗普告诉各国领导人：有事打我手机；

5) 特朗普缩减预算波及 600 万“难民”，救济金恐不再发放；

6) 前日本驻韩大使出书，称文在寅是“最差总统”，尚未发售便引起广泛关注；

7) 阿富汗首都使馆遭重大爆炸袭击，伤亡情况不详；伊拉克西部自杀式袭击致 13 人亡，一名记者殉职；

8) 绰号“胖叔”的泰国猴子因游客投食太多体重飙升，被送去减肥营运动。

**C、【财经证券】**

1) 央行行长助理：MPA（宏观审慎评估体系）并非金融市场波动源头；

2) 学者：延迟退休等养老顶层设计方案确定，未对外公布；

3) 京沪楼市成交量大幅下跌，房价将更明显回落；楼市调控初见成效，一线城市租金首降；

4) 中国城市公用事业附加费取消，电价每度下降 1 分 1，可节约全社会用电成本 350 亿；

5) 证监会主席刘士余喊话后 33 家公司仍“一毛不拔”，已披露分红预案公司中 75% 进行现金分红；

6) 昨日收盘：沪指 3117.18/+0.23% 深成指 9864.84/+0.06% 创业板 1763.73/+0.23% 恒指 25660.65/-0.16%。

**D、【文教体娱】**

1) 第 27 届全国书博会在廊坊开幕（5 月 31 日至 6 月 2 日）：66 家单位携 5000 种好书亮相；

2) 褚君浩院士：做得好科普，科研才可能有大格局；安徽新规：驴友违规遇险自付救援费；

3) 世界百大著名运动员：C 罗居首压詹皇+梅西，中国宁泽涛、马龙和林丹三将入围；

4) 李隼暂代孔令辉行女乒主教练职；刘国梁钦点李晓霞恩师接替孔令辉，王楠张怡宁都服他；刘国梁称迎战世乒赛信心不变；亚



冠-上港客场 3-2 苏宁总分 5-3 晋级。

#### E、【生活服务】

1) 北京 ETC 发行量超 300 万套，处于国内领先水平（注：ETC 是不停车电子收费系统）；北京 654 个事项可网上申报审批，有的事项已实现“全球通办”；北京 9 家社区儿童中心提供早教；

2) 上海二手房交易新规，6 月起网签增加身份核验环节；

3) 广东 6 月一批新规即将落地：高温补贴今起发放，指定病种报医保不设起付标准；深圳最低工资调至 2130 元/月，东莞上调污水处理费；

4) 山东淄博淡水鱼商户成立自律组织，鱼“身份”可追溯；

5) 西安实行**商品房销售记分管理**，超 5 分将暂停网签；6 月中旬起西安市经适房满 5 年补 15%差价可买商品房（西部网）。

#### F、【健康养生】

1) 身体大部分的不健康，来自于“管不住嘴，迈不开腿”，即吃得太多，运动太少，解决方案当然就是：管住嘴，迈开腿，记住：晚上少吃点，对身体是绝对有好处的。

2) 湿非一病、百病兼之；湿气一除，百病全无。头发爱出油、面部油亮、浑身无力、常感到疲倦、大便稀黏腥臭，粘稠（不易冲掉），说明身体有了湿气；目前，艾灸、穴位贴敷除湿效果较好（翠梭健康）。

（编辑：西安知非 自新华、中新、腾讯、凤凰、东方财富网）

### 附 v. 2017 年 6 月 1 日农历干支、节日、历史上的今天名人和事件

**导语：**2017 年 6 月 1 日是儿童节，它不仅是我们中国儿童的节日他还是全世界儿童的节日。这个节日是为了保障儿童的权益而设立的。那么 2017 年 6 月 1 日是的节日、星座、出生名人呢？订阅、关注秦陇纪 10 公众号哦。

#### 2017 年 6 月 1 日节日

6 月 1 日是公历一年中的第 152 天(闰年第 153 天)，离全年结束还有 213 天。

2017 年 6 月 1 日节日：**国际儿童节**、香港无冷气日、肥肥日、墨西哥**植树节**、**全球父母节**、缅甸植树节、京都薪能、堆沙节、蝴蝶会、奥地利吹奏乐节、洪水节、肯尼亚自治节、**国际牛奶日**、里斯本国际节、西萨摩亚独立日、多哥植树节、马来西亚佳佳节。

#### 2017 年 6 月 1 日农历

公历：2017 年 6 月 1 日 星期四

农历：二〇一七年 五月小 初七日

回历：1438 年 9 月 6 日

干支：丁酉年 乙巳月 己未日

八字：丁酉 乙巳 己未 甲子

五行：山下火 佛灯火 天上火 海中金

生肖：属鸡

星座：双子座

星宿：翼宿(翼火蛇)

值神：明堂(黄道日)

冲煞：羊日冲(癸丑)牛 煞西

#### 2017 年 6 月 1 日星座

双子座(5 月 21 日-6 月 21 日)

6 月 1 日性格：6 月 1 日出生的人通常若不是公众人物，就是走在时代尖端的人。不管属于哪一类，他们都是在社会竞赛中备受瞩目的焦点，或是见证这一切的人。他们对历史的兴趣，通常仅止于那些对“现在”有影响的话题。事实上，这类型的人可以把所有的精力都花在注意流行趋势、穿着打扮、谈论时髦话题，以及研究各种形式的艺术上。

#### 6 月 1 日出生名人

1926 年——**玛丽莲·梦露**出生，好莱坞影星(1962 年逝世)

1937 年——**摩根·弗里曼**，好莱坞影星

1950 年——**夏琳·邓肯**，美国歌手。

1953 年——**习近平**，中共中央政治局常委、中央书记处书记、国家副主席、中央党校校长。

1970 年——台湾籍多栖明星**刘若英**出生、

1970 年——美国足球运动员**阿莱克西·拉拉斯**出生

#### 6 月 1 日历史大事件

1831 年——**詹姆斯·罗斯**发现北磁极

1941 年——纳粹德国占领希腊克里特岛

2001 年——**尼泊尔**皇室血案，王子迪彭德拉将尼泊尔比兰德拉国王、王后艾什瓦尔雅等多名王室成员杀害后自杀身亡。

2003 年——**三峡工程二期**竣工，开始蓄水

2007 年——中国在西昌卫星发射中心用“**长征三号甲**”**运载火箭**成功发射“鑫诺三号”通信卫星，这次发射是长征系列运载火箭第 100 次发射。

2008 年——中国大陆所有超市、卖场开始实行**有偿提供塑料袋**

**儿童节**是所有儿童的生日，希望各位家长能够抽出时间、陪伴自己的小孩，度过一个快乐的儿童节。

## Appx. 数据简化 DataSimp 社区译文志愿者（610 字）

“**数据简化 DataSimp**”社区翻译组、媒体组缺少志愿者，当下需要：①设计黑白静态和三彩色动态社区 LOGO 图标；②翻译**美欧 IT 大数据、人工智能、编程开发技术**文章的至少投一篇高质量首译**美欧数据科学技术论文**，正式成为**数据简化 DataSimp 社区贡献者**。非诚勿扰，季度无贡献者自动退出。加入**数据简化 DataSimp 社区**，请在公众号后台留言，或加 QinlongGEcai（请备注：姓名-单位-职务-手机号）微信。社区筹备详情，请阅读本公众号文章《科研江湖中的一眼清泉之数据简化 DataSimp 社区及学会》。

**Data Simplification and Sciences** Wechat and Toutiao Public Account, QinDragon2010@qq.com, 2017.6.1Thu, Xi'an, Shaanxi, China:

LIFE

请关注“数据简化 DataSimp、科学 Sciences”微信公众号！15 时 59 分 1 秒 2017 年 6 月 1 日星期四

Life begins at the end of your comfort zone. -- Neale Donald Walsch

### THE DAY

The strength of purpose and the clarity of your vision, along with the tenacity to pursue it, is your underlying driver of success. -- Ragy Tomas

长按下面二维码“识别图中二维码”关注[公众号：数据简化 DataSimp](#)（搜名称也行）。



长按下面二维码“识别图中二维码”关注[公众号：科学 Sciences](#)（搜名称也行）。



主编：[秦陇纪](#)（b80），科普家、IT 老师，数据简化 DataSimp 新媒体（筹技术、社区、期刊）创始人，研究方向：教育技术、大数据、人工智能、知识工程，数据挖掘、信息抽取、数据简化、知识库构建。欢迎高质量大数据、人工智能、知识工程、自然语言处理领域研究者给“数据简化 DataSimp”社区投稿；或贡献数据采集、处理、分析、简化、应用各环节开源代码。

（西安[秦陇纪 10 数据简化 DataSimp](#)汇编，欢迎有志于数据简化之传媒、技术的伙伴加入全球“数据简化 DataSimp”社区！转载出处：

[秦陇纪 10 数据简化 DataSimp](#)公号“[数据简化 DataSimp](#)、[科学 Sciences](#)”汇译编，[投稿邮箱 QinDragon2010@qq.com](#)）

欢迎留言，赞赏；点击“[阅读原文](#)”下载 [PDF](#)