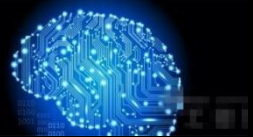




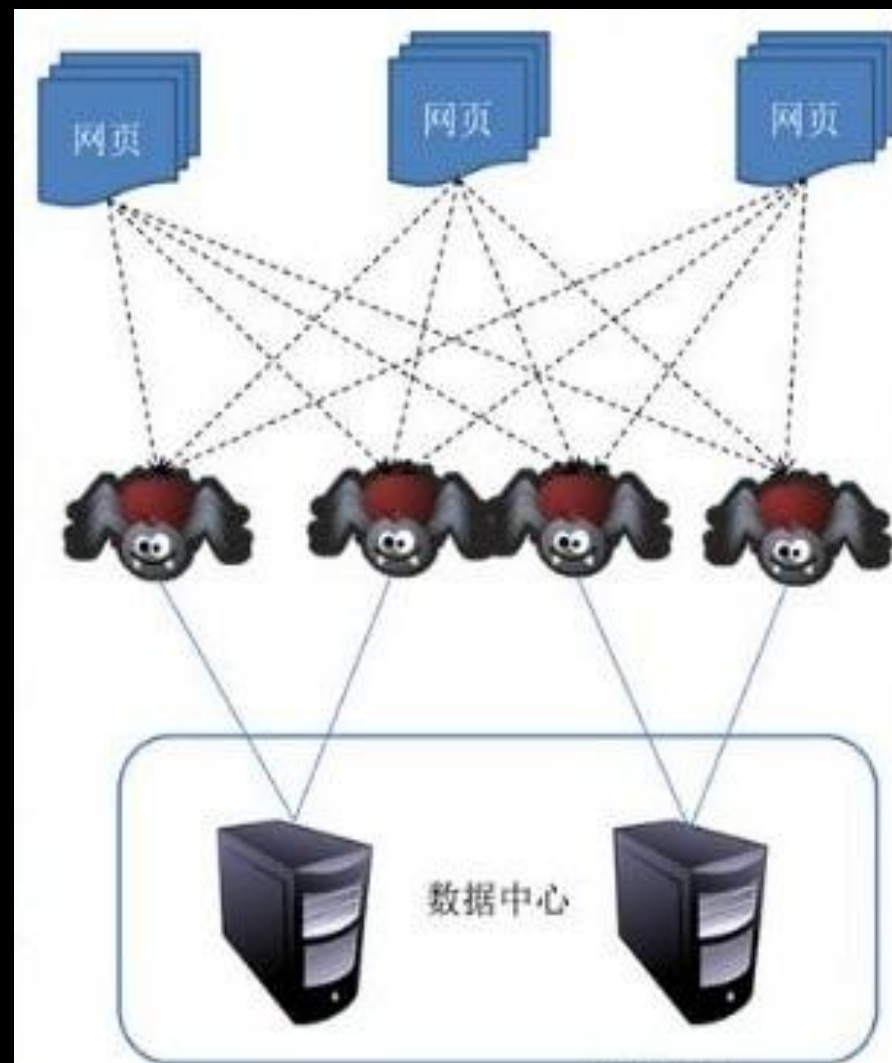
管中窥豹 ---智融链DCT团队的金融黑科技



可视化智能爬虫“章鱼帝”

可能是世界上最直观的智能爬虫定义管理软件

大幅度降低了全网数据挖掘的人工成本



仅经过简单培训的普通人员即可生成针对特定网站的智能爬虫

是世界上极少数可以低成本组织全网数据收集的有效工具



这段视频拍摄了“章鱼帝”是如何爬取“京东”商品价格数据的

<https://pan.baidu.com/s/1qYU34Ry>

这个是“章鱼帝”的应用成果实例：A股的大数据集市

<http://tushare.org/>

这个是“章鱼帝”管理软件：

[file.tushare.org:8080](file:tushare.org:8080)

**“章鱼帝”大幅度提高了数据采集效率，
是智融链上大数据集市的有效保障**



对标：kimonolabs.com

www.kimonolabs.com

是截止目前唯一一家与“章鱼帝”性能相当的可视化爬虫系统
于2016年2月被高盛旗下著名的人工智能大数据公司“Palantir”
以3000万美金的价格收购

直接访问主页是收购公告。



自然语言处理

一个自然语言处理的例子，选择最相似的前 n 门课。

1 初始课程



courses

总共 379 门课程，每行包括 3 部分内容：课程名、课程简介、课程详情。

下面所示的例子仅仅是其中的课程名：

Writing II: Rhetorical Composing.

Genetics and Society: A Course for Educators.

General Game Playing.

Genes and the Human Condition (From Behavior to Biotechnology).

A Brief History of Humankind.

New Models of Business in Society.

Analyse Numérique pour Ingénieurs.

Evolution: A Course for Educators.

3 BOW 模型，建立词袋。

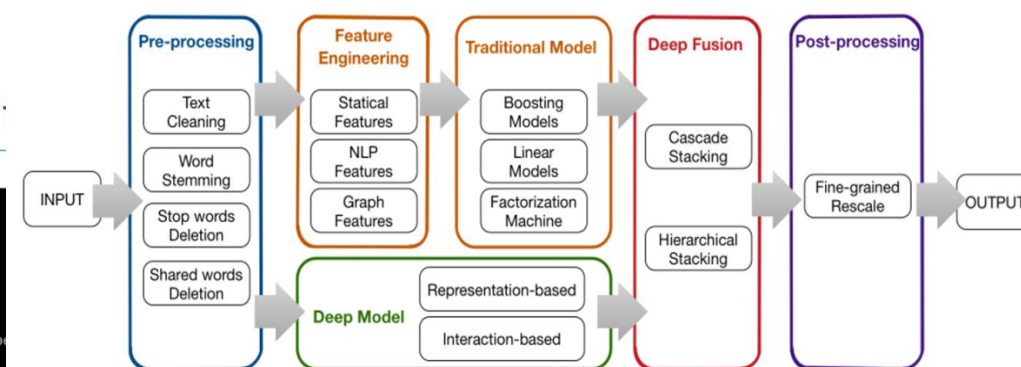
4 计算 TFIDF。

5 训练 LSI 模型。

6 寻找和 Machine Learning 课最相似的前十门。

Administrator: Node.js command prompt

```
2017-07-20 02:08:09,161 : INFO : updating model with new documents
2017-07-20 02:08:09,272 : INFO : preparing a new chunk of documents
2017-07-20 02:08:09,282 : INFO : using 100 extra samples and 2 power iterations
2017-07-20 02:08:09,283 : INFO : 1st phase: constructing (2974, 110) action matrix
2017-07-20 02:08:09,303 : INFO : orthonormalizing (2974, 110) action matrix
2017-07-20 02:08:09,418 : INFO : 2nd phase: running dense svd on (110, 379) matrix
2017-07-20 02:08:09,438 : INFO : computing the final decomposition
2017-07-20 02:08:09,438 : INFO : keeping 10 factors (discarding 72.509% of energy sp
2017-07-20 02:08:09,441 : INFO : processed documents up to #379
2017-07-20 02:08:09,443 : INFO : topic #0(3.823): -0.238*"teach" + -0.152*"nbsp" + -0.115*"mus" + -0.106*"heal" + -0.103*"comput" + -0.098*"stud" + -0.098*"network" + -0.098*"de" + 0.611*"de" + 0.380*"la" + 0.217*"e
*"que" + 0.149*"cour" + 0.144*"un" + 0.132*"les" + 0.120*"curso"
2017-07-20 02:08:09,460 : INFO : topic #1(2.703): 0.611*"de" + 0.380*"la" + 0.217*"e
```



这是一个自然语言处理NLP的算法实例
在“智融链”应用中可用来自动筛选有效的新闻，公告等自然语言数据。

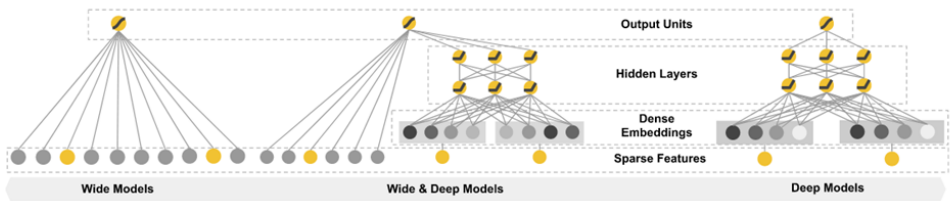


深度学习神经网络

1 例子

Wide & Deep Learning for Recommender Systems

Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishu Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Isipir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah
Google Inc.



目标是基于 5 个连续变量(age, education_num, capital_gain, capital_loss, hours_per_week)和 9 个类别变量，预测指定人的收入是否高于 50,000 dollars

2 数据集描述

<https://archive.ics.uci.edu/ml/datasets/Census+Income>

Data Set Characteristics:	Multivariate	Number of Instances:	48842	Area:	Social
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	14	Date Donated:	1996-05-01

3 输出

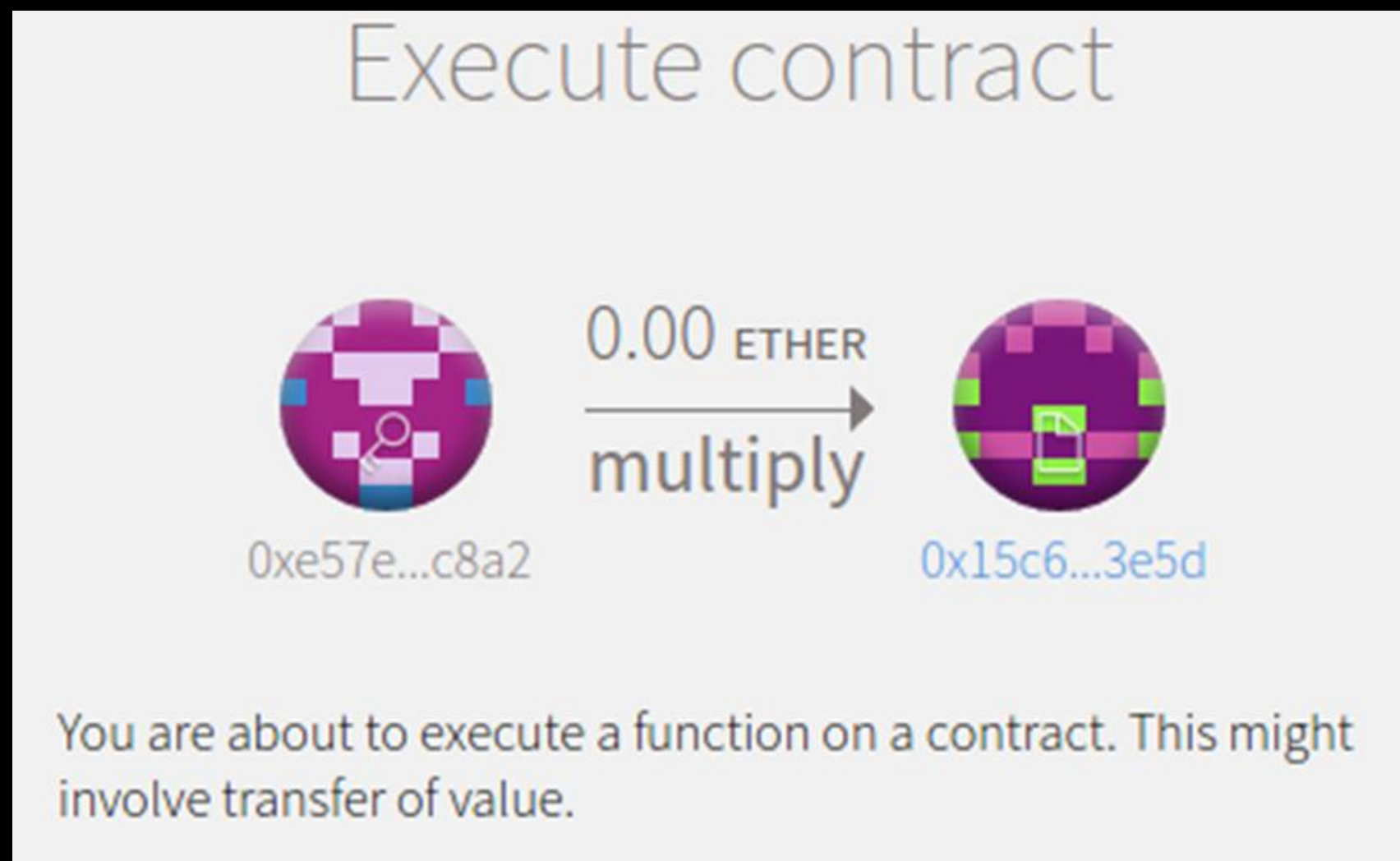
```
PU computations.
Target variables:
  variable Y/Y:0: Tensor("Y/Y:0", shape=(?, 1), dtype=float32)
=====
----- Preparing train data
  Y shape=(32561, 1), X shape=(32561, 5)
-----
----- Preparing test data
  Y shape=(16281, 1), X shape=(16281, 5)
Input data shape = (32561, 5); output data shape=(32561, 1), batch_size=32561
Test data shape = (16281, 5); output data shape=(16281, 1), validation_batch_size=16281
===== Training
Run id: wide
Log directory: /tmp/tflearn_logs/
Training samples: 32561
Validation samples: 16281
..
Training Step: 500 | total loss: 0.61789 | time: 1.134s
| wide_regression | epoch: 500 | loss: 0.61789 - binary_acc: 0.7783 | val_loss: 0.62533 - val_acc: 0.7807 -- iter: 32561/32561
..
Training Step: 1000 | total loss: 0.67440 | time: 1.104s
| wide_regression | epoch: 1000 | loss: 0.67440 - binary_acc: 0.7760 | val_loss: 0.62481 - val_acc: 0.7810 -- iter: 32561/32561
..
Training Step: 1500 | total loss: 0.64582 | time: 1.098s
| wide_regression | epoch: 1500 | loss: 0.64582 - binary_acc: 0.7803 | val_loss: 0.62627 - val_acc: 0.7807 -- iter: 32561/32561
..
Training Step: 2000 | total loss: 0.68014 | time: 0.128s
| wide_regression | epoch: 2000 | loss: 0.68014 - binary_acc: 0.7786 | val_loss: 0.62781 - val_acc: 0.7809 -- iter: 32561/32561
..
===== Evaluation
logits: (16281, ), mLn=-3.31884, max=135.066
Actual IDV
0 12435
1 3846
dtype: int64
Predicted IDV
0 14514
1 1767
dtype: int64
Confusion matrix:
actual      0      1
predictions
0      11691  2823
1       744  1023
=====
gavin@gavin-VirtualBox:~/dev/tflearn$
```

准确率为 (920+11800)/16281 = 78.1%

这是一个深度学习神经网络算法实例
在“智融链”应用中可用来预测投资品种的价格变化



区块链



详细请见“智融链”技术白皮书