

用AI从新型数据中挖掘Alpha 吴海山 合一创投首席数据科学家



百度大数据实验室 Baidu Big Data Lab



市场信息通过报纸、电 报、电话

PC和互联网:彭博、路透、 万得等金融资讯提供平台

移动互联网和物联网时 代的新型分析平台







- 宏观:不能快速反应市场变化、频 率低,如CPI、PMI每个月更新一次
- 微观: 实地尽调费时费力难以客观; 公开的数据已经很难有alpha

新的金融市场分析平台能够:

lo]

- 实时收集并清洗传感器产生的海量 新型数据
- 通过人工智能进行分析挖掘,实时 监测市场,识别投资信号



自度大数据实验室 aidu Big Data Lab

THE SPECTRUM OF DIFFUSION



FULLY DIFFUSED

FULLY COMMODITIZED

These are some of the table stakes fo anyone undertaking market analysis.

LOOKING FOR THIS?	TRY HERE:
Stock Prices (US)	End-of-Day Stock Prices quandl.com/data/EOD (from QuoteMedia)
Stock History (Europe)	London Stock Exchange Prices quandl.com/data/XLON (from Exchange Data International)
Fundamentals (US)	Core US Fundamentals quandl.com/data/SF1 (from Sharadar)
Fundamentals (Europe)	Global Fundamentals quandl.com/data/RB1 (from Robur)
Futures (US)	Continuous Futures quandl.com/data/SCF (from Stevens Analytics)
Futures (Europe)	Eurex Futures

e table stakes for	DIFFUSING NO	W				
narket analysis.	Not yet looking at	Not yet looking at these types of data?		NASCENT		
TRY HERE: Time to start. End-of-Day Stock Prices quandl.com/data/EOD (from QuoteMedia) LOOKING FOR THIS? TRY HERE: London Stock Exchange Prices Sentiment Data AlphaOne Sentiment quandl.com/data/AOS (from Accern)	Time to start. LOOKING FOR THIS? TRY HERE:		Get a jump on the competition by seek- ing out sources in these industries.			
	INDUSTRY	PROMISING PROVIDERS				
quandl.com/data/XLON (from Exchange Data International)	Advertiser Spending	Total US Ad Spend quandl.com/data/BL1	maritime)	www.windward.eu		
Core US Fundamentals quandl.com/data/SF1 (from Sharadar) Analysis	(from Borell) Ursa Space www.ursaspace.com	Drone Imagery	3D Robotics www.3dr.com			
Global Fundamentals quandl.com/data/RB1 (from Robur)	Economic Data	ata CLS quandl.com/data/CLSH	Wearable Tech	www.samsara.com Sensoria		
Continuous Futures quandl.com/data/SCF (from Stevens Analytics)	Transportation	North American Commodi- ties Transport guandl.com/data/RR1	Food Prices in Develop- ing Countries	www.sensoira.com Premise www.premise.com		
Eurex Futures quandl.com/data/BCEUX from Barchart)	(from Transmatch)	Ag lech	Vellus Labs www.telluslabs.com			
	MODERATELY COMMODI	TIZED				
			UNTAPPED	>		

Eventually, all untapped data becomes fully commoditized and a necessity.



容

Alternative Data – The New Trend in Financial Data

Alternative Data Sources







One day in Beijing





Photos are creating a real CDA THE DATA AND THE PROVIDE THE PRO



- 卫星数据分析的商业应用
 - OrbitalInsight



US Retail Traffic

Quantifying sales trends of major U.S. retail chains in real time through the number of cars in parking lots.



China Economic Index

Providing independent data about China's economy through vehicle counting, building shadows and construction rates.



Oil Inventory

Analyzing shadows on floating lids of oil tanks around the world can capture how much oil is truly available.



Agriculture

Tracking forests and fields from the sky can anticipate crop shortages, predict deforestation, and indicate poverty.

Source: Orbitalinsight

- US Micro Retail Traffic Quantifying retail traffic patterns in parking lots of over 90 major retailers as a proxy for micro-level traffic
- US Retail Traffic Index (US Macro Retail Traffic) Measuring nationwide retail traffic at over 100,000 parking lots.







The height of an oil tanker by measuring the shadow at a certain hour of day and angle of satellite. Higher buildings might also suggest higher levels of urbanization, which <u>may</u> (or <u>may not</u>) be accompanied with economic growth Source: Orbitalinsight

The Orbital Insight poverty map includes data generated from the car-counting algorithm, the shadow detection classifier (as a proxy for building height), the development classifier (as a proxy for % coverage of human construction), and agricultural productivity estimates.





Source: Orbitalinsight







Ba

基于全卷积深度网络的多模融合AOI识别算法

基于深度学习的卫星图像数据,自动识别经济区域,如工厂、写字楼等(AOI)







卫星图像的AOI检测

• 基于FCN进一步做后处理(Deeplab算法)

- 使用atrous convolution (hole algorithm)
- Atrous Spatial Pyramid Pooling (ASPP)
- 使用CFR做图像后处理





卫星图像的AOI检测

• 使用路网数据来做FCN模型的预训练(pretraining),进一步提高模型效果







Our preliminary result





IoU: 0.80137

IoU: 0.92203

red line is predicted area yellow line is ground truth area





Track the growth of of employees and consumers





Identify work place and home place by mining mobility trajectories





百度大数据实验室 Baidu Big Data Lab



Track the growth of employees: four cases



百度大数据实验室 Baidu Big Data Lab



Employment Index of 2,000 Industrial Parks





百度大数据实验室 laidu Big Data Lab



Consumer Index of 4,000 large shopping centers







From location search to foot traffic





- 问题建模:给定用户当前的定位位置,确定用户到访的POI
 - 将时空bayes模型和LSTM模型相融合,极大提高了 POI到访的准确度



基于时空Bayes+LSTM的用尸POI到 访预测



基于时空Bayes+LSTM的用尸POI到 访预测

• 实验效果对比

	K=1	K=3	K=5
Nearest	18%	24%	35.7%
Learning to Rank (Microsoft[1])	23.4%	45.6%	53.1%
Learning to Rank (Facebook[2])	29.0%	51.3%	58.9%
Bayes+LSTM	42.8%	65.6%	75.4%

[1] Lian, Defu, and Xing Xie. "Learning location naming from user check-in histories.", Microsoft

2013, Facebook

[2] Shaw, Blake, et al. "Learning to rank for spatiotemporal search. ",



How can we track the consumer growth with search query data on Baidu





Apples Revenue (Greater China) Forecasting

We chose about 30 flagship Apple Stores in mainland China and tracked the change of location queries related to these stores. (Our model predicted a 34.1% decline in Greater China)







China Box office forecasting using location query data related to cinemas







China Box-office Fraud Detection











Consumer Index in catering industry(covering 3million restaurants)





Consumer Index in tourism industry(cover ng 300k touris attractions)







China Economic Index

China Economic Index has published on Bloomberg terminal

BDCHETP 100.7 As Of 07/31/16						
Baidu Economic Index - Employment - Ind.	. Baidu Inc.	90/Gol to Export				
BDCHEIP Index	Economic Indices Description		ottinge -	l.l.e.		
Series Baidu Economic Index - Employment - 1	Industrial Darks	91) Sedicit 92) S	Settings •	WOr	la Economic S	latistic
Θ (SA Index)		94) Standard Views 95) Custom	Dango lun - 2016 lu	2016 Course Daidu Inc		
(SA, Index)		Way Indicators	Range Jun 2016 - Jul	L 2016 Source Baldu Inc	Index CA	
Country China Source Baidu Inc.		*Ney Indicators	11) Bigdata Consumption Index	10) Bigdata Employment Inde	ation Thuex SA	
		* Drices	10 Diguata consumption index	Ticker	^ Jul 2016	lun 20
1) Description 2) Release Details 3) Related Indicator	rs 4) Related News	■Labor Market	2014=100	Ticker	500 2010	5411 20
Latest 100.7 St	art Date Jan 2014	Retail & Wholesale Sector	- Consumption			
Period Jul 2016 Fr	requency Monthly	■Industrial Sector	Shopping Centres	✓ BDCHCSC In	100.2	92
Base Year 2014=100	ext Update Oct 5, 2016	Whole Economy Activity	Sales of Automobiles	✓ BDCHAUS In	103.3	104
		Services Sector	Catering Services	🗠 BDCHCA Index	88.7	85
5) General Notes 6) Detailed Notes	7) Historical Chart	■Surveys/Cyclical Indicators	LTourism	⊯ BDCHCTM In	130.2	119
Paidu Economic Index uses and positioning data and		■Housing and Real Estate				
balluu Economic Index uses geo-positioning uata allu	Survey n.a.	■Personal/Household Sector				
map-search data to measure economic activities in		■Intl Trade & BoP	*			
China. It covers 2,000 industrial parks, 4,000		■Government Finance & Debt				
shopping centres, and millions of Points of Inter-	est /	Monetary Sector			2) Lege	end
(POI) nationwide. It starts from January 2014.		Financial Indicators	3) Maximize Chart	4) Clear 5) Edit Chart V 6) Da	ate Range 🔻 /) worl	KDENCN
Please refer to http://bdl.baidu.com/mobimetrics	5	*Demographics	Sales of Automobiles 103.30			-11
for more details and methodology.	or		130			-10
for more details and methodotogyr	⁶⁵			\sim \wedge \wedge \vee	$\forall \wedge \land$	/ 10
	- 80					/ 95
						-90
	Mar Jun Sep Dec Mar Jun Sep Dec Mar Jun 2014 2015 2016		90-			-85
	8) See in context (ECST)		80 - V			-80
	9) Transform Data (FCWB)		Dec Mar Jun	Sep Dec Mar Jun Sep	Dec Mar Jun	·







Mobimetrics: mapping ghost cities in real estate boom of China

Track the growth trends of residents



Detect home/work location via trajectory mining

Detect residential area via POI mining





Detected vacant area



大数据实验室 Big Data Lab





Vacant area in Rushan, Weihai





Ba

3度大数据实验室 aidu Big Data Lab



Vacant area in Erdos





国内外知名媒体报道







