

Abstractive(CNN) and GRU for tendency of stock price prediction

Zuo Fangling

2019/9/20

Dataset — stock

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Q
1	代码	名称	上市地	行业板块	最新价	涨跌幅(%)	涨跌额	成交量	turnover	市值	昨日收盘	今日开盘	最高	最低	市盈率	时间
2	CGA	中国绿色农业	N	2	4.23	19.49	0.69	473816	14.3773	16855404	3.54	5.21	5.21	4	-1.23	2019/8/20
3	SINA	新浪	O	17	43.45	15.41	5.8	3686474	5.1602	3104068055	37.65	40	43.8	39.32	23.36	2019/8/20
4	WB	微博	O	17	42.31	14.1	5.23	9920093	4.4162	9504180228	37.08	39.66	42.9	38.65	15.27	2019/8/20
5	GSUM	国双科技	O	17	2.24	13.71	0.27	20809	0.0675	69045375	1.97	2	2.24	2	-0.75	2019/8/20
6	TEDU	达内科技	O	9	1.27	11.4	0.13	58693	0.1103	67576183	1.14	1.21	1.28	1.18	-0.77	2019/8/20
7	XRF	信而富	N	15	0.6489	10.1	0.0595	277726	0.4223	42671142	0.59	0.59	0.66	0.59	-0.76	2019/8/20
8	CNET	中网在线	O	17	1.4827	8.23	0.1127	163455	0.9959	24331912	1.37	1.4	1.5	1.37	-1.61	2019/8/20
9	BIDU	百度	O	17	104.22	7.78	7.52	11674431	3.3678	36128111467	96.7	100	105.3	100	11.9	2019/8/20
10	CLDC	中国贷款公司	O	15	0.6348	7.59	0.0448	257197	1.0171	16052695	0.59	0.53	0.64	0.53	-0.27	2019/8/20
11	JFIN	嘉银金科	O	15	11.41	7.54	0.8	12756	0.0238	610434992	10.61	11.14	11.58	11.14	7	2019/8/20
12	WSG	万达体育	O	18	3.85	7.54	0.27	262677	0.1876	538999987	3.58	3.62	3.88	3.61	-55	2019/8/20
13	GHG	格林酒店	N	9	11.78	7.38	0.81	40762	0.0401	1183083196	10.97	11.33	11.88	11.33	18.2	2019/8/20
14	SFUN	房天下	N	17	1.98	7.03	0.13	620824	3.4608	35339279	1.85	1.92	2.24	1.92	-15.15	2019/8/20
15	SY	新氧	O	12	15.52	7.03	1.02	173645	0.1734	1554281486	14.5	14.85	15.75	14.81	-172.44	2019/8/20
16	AMBO	安博教育	N	9	2.16	6.93	0.14	236	0.0011	46941882	2.02	2.09	2.16	2.09	9.82	2019/8/20
17	STG	尚德机构	N	9	2.49	6.87	0.16	33208	0.0194	425809922	2.33	2.39	2.66	2.39	-3.61	2019/8/20
18	SGOC	上为	O	17	0.9938	6.86	0.0638	6505	0.0081	79530481	0.93	1.01	1.05	0.99	0.43	2019/8/20
19	BNSO	恒异电子	O	0	2.2765	5.88	0.1265	501	0.0108	10727551	2.15	2.16	2.28	2.16	3.12	2019/8/20
20	YY	欢聚时代	O	17	57.81	5.88	3.21	1761478	2.1717	4688969211	54.6	56.1	58.2	55.7	5.94	2019/8/20
21	IQ	爱奇艺	O	17	18.08	5.85	1	17346730	2.3952	13094095792	17.08	17.84	18.85	17.77	-8.33	2019/8/20
22	XYF	小赢科技	N	15	2.78	5.7	0.15	35699	0.0235	415951220	2.63	2.84	2.84	2.66	3.04	2019/8/20
23	TIGR	老虎证券	O	17	4.7	5.62	0.25	596487	0.4485	625099975	4.45	4.61	4.75	4.51	-14.24	2019/8/20
24	ZNH	南方航空	N	7	30.66	5.47	1.59	14466	0.0054	8216049809	29.07	31.03	31.03	30.37	15.48	2019/8/20

Dataset — news

	news_ty	news_title	news_time	news_content
0	能源	中国煤炭运销协会：煤炭市场下行压力加大 但也存积极因素	2019/8/16 23:43	<p>中国煤炭运销协会发布7月份煤炭市场运行综述指出，7月份，煤炭市场弱势运行。供给方面，煤炭产量与进口量均大幅增加。需求方面，全国大部分地区进入高温天气，电煤耗用水平有所升高，但受强降雨、水电高发的影响，需求不及预期。同时，受产地安全治理汽运超载、成本倒挂、区域内煤炭就地转化量增多等因素影响，市场煤炭发运积极性较低，电力企业以接收长协煤为主，电煤库存仍保持历史高位，动力煤价格小幅下降。炼焦煤市场供需基本平稳，随着中旬焦炭市场逐步好转，焦煤采购积极性回升，主流炼焦煤企业价格保持稳定。</p> <p>中国煤炭运销协会认为，后一时期，由于煤炭市场总体供应能力增强，而需求低迷，煤炭市场下行压力加大，但也存在一些积极因素：安全、环保检查力度加大；极端天气因素的不确定性；产地与港口价格倒挂等，对煤炭市场有一定的支撑。</p>

Linux+chromedriver+seleium

```
0 10 * * * python /root/news_code/banknews.py > /root/news_code/bankerror.txt 2>&1
30 10 * * * python /root/news_code/electricity.py > /root/news_code/electricityerror.txt 2>&1
00 11 * * * python /root/news_code/energy.py > /root/news_code/energyerror.txt 2>&1
30 11 * * * python /root/news_code/financenews.py > /root/news_code/financeerror.txt 2>&1
00 12 * * * python /root/news_code/healthcare.py > /root/news_code/healthcareerror.txt 2>&1
30 12 * * * python /root/news_code/house.py > /root/news_code/houseerror.txt 2>&1
00 13 * * * python /root/news_code/industry.py > /root/news_code/industryerror.txt 2>&1
30 13 * * * python /root/news_code/insurancenews.py > /root/news_code/insuranceerror.txt 2>&1
00 14 * * * python /root/news_code/it.py > /root/news_code/iterror.txt 2>&1
30 14 * * * python /root/news_code/logistics.py > /root/news_code/logisticserror.txt 2>&1
00 15 * * * python /root/news_code/mechanical.py > /root/news_code/mechanicalerror.txt 2>&1
30 15 * * * python /root/news_code/medicine.py > /root/news_code/medicineerror.txt 2>&1
00 16 * * * python /root/news_code/metal.py > /root/news_code/metalerror.txt 2>&1
30 16 * * * python /root/news_code/neconsumption.py > /root/news_code/neconsumptionerror.txt 2>&1
00 17 * * * python /root/news_code/opconsumption.py > /root/news_code/opconsumptionerror.txt 2>&1
30 17 * * * python /root/news_code/raw.py > /root/news_code/rawerror.txt 2>&1
00 18 * * * python /root/news_code/telecom.py > /root/news_code/telecomerror.txt 2>&1
30 18 * * * python /root/news_code/utilities.py > /root/news_code/utilitieserror.txt 2>&1

45 9,12,15,18,20,21 * * * python /root/stockcode/sina_concept.py > /root/stockcode/console.txt 2>&1
01 23 * * * root shutdown -h now
01 9 * * * root reboot
```

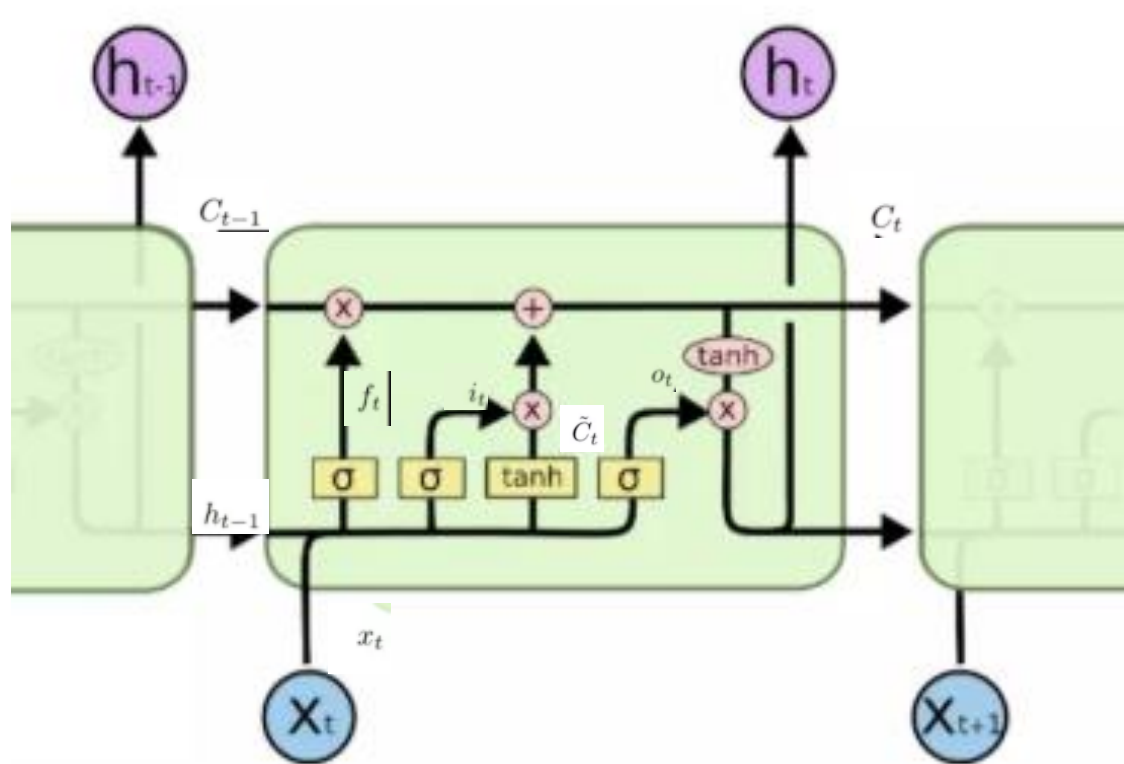
~

Abstractive(CNN)

- Capture a long-distance dependency
- Parallel computing, the modeling of RNN depends on the historical information of the sequence.
- Higher Calculation efficiency, $O(n)$ && $O(n/k)$, k is convolution kernel-size

from <https://www.jianshu.com/p/0f2396f0d98f>

LSTM



LSTM中的重复模块包含的四层交互神经网络层

$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

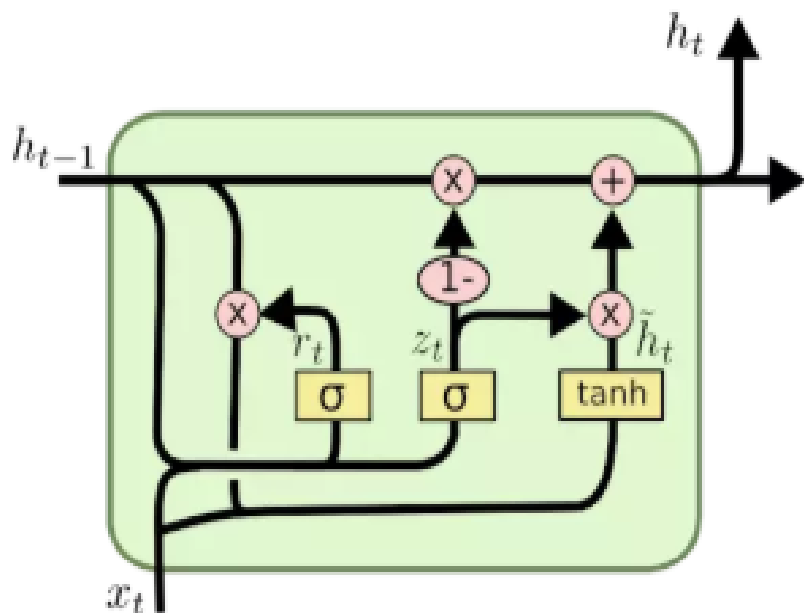
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

GRU

- forget gate and the input gate are combined



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

GRU

<https://www.jianshu.com/p/9dc9f41f0b29>

Research plan

- Introducing time series (the previous price and news)
- Try emotional score —— API,

https://ai.baidu.com/tech/nlp/sentiment_classify

Future plan

- 《Convolutional Sequence to Sequence Learning》 , download address : <https://arxiv.org/pdf/1705.03122.pdf>
- Labeled training dataset
- Python code