

AETA地震预测AI算法 大赛决赛答辩

团队：4freedom



Content

01 任务分析

02 数据预处理

03 震级预测

04 震中预测

05 预测情况及总结



/01

任务分析



任务分析

决赛题目、数据与作品提交

- (1) 决赛题目 决赛开始将按周发布最新的AETA观测数据，参赛团队将发布的最新数据作为验证输入，对未来一周内川滇地区（22.00°N~34.00°N，98.00°E ~107.00°E）的 ≥ 3.5 级的地震事件进行预测。
- (2) 决赛数据 从2020年5月12日开始，按周发布最新一周的川滇地区的AETA电磁和地声均值数据用于预测。每周川滇地区实际发生的 ≥ 3.5 级的地震事件作为验证地震目录。
- (3) 决赛作品提交 在每周AETA数据更新后的当天提交对目标区域（22.00°N~34.00°N，98.00°E ~107.00°E）未来7天内是否有 ≥ 3.5 级地震的预测结果，如下表所示。

决赛预测结果提交格式						
No	Release time	Date duration	Y/N there is an earthquake	Latitude	Longitude	Magnitude
1	2020-04-12	2020-04-13~ 2020-04-19				
2	2020-04-19	2020-04-20~ 2020-04-26				
3	2020-04-26	2020-04-27~ 2020-05-03				
...	...					
i	2020-07-26	2020-07-27~ 2020-08-02				

其中，以预测结果发布的那天起，给出未来7天是否有 ≥ 3.5 级以上的地震的预测（Y/N），预测的地理范围为（22.00°N~34.00°N，98.00°E ~107.00°E），是否有震为是否有 ≥ 3.5 级以上的地震，震中给出具体的经纬度（XX.XX, XX.XX），震级为Ms。

分析任务可知，需要利用过去n天的电磁、地声数据，预测未来一周、每天川滇地区发生的地震情况，然后对比一周的预测结果，判断最大级地震是否 ≥ 3.5 ，若是则给出对应的经纬度。

由于既要给出地震震级又需要预测地震发生经纬度，我们团队将任务分为对震级的预测与对震中的预测两个问题。

/02

数据预处理



2 数据预处理

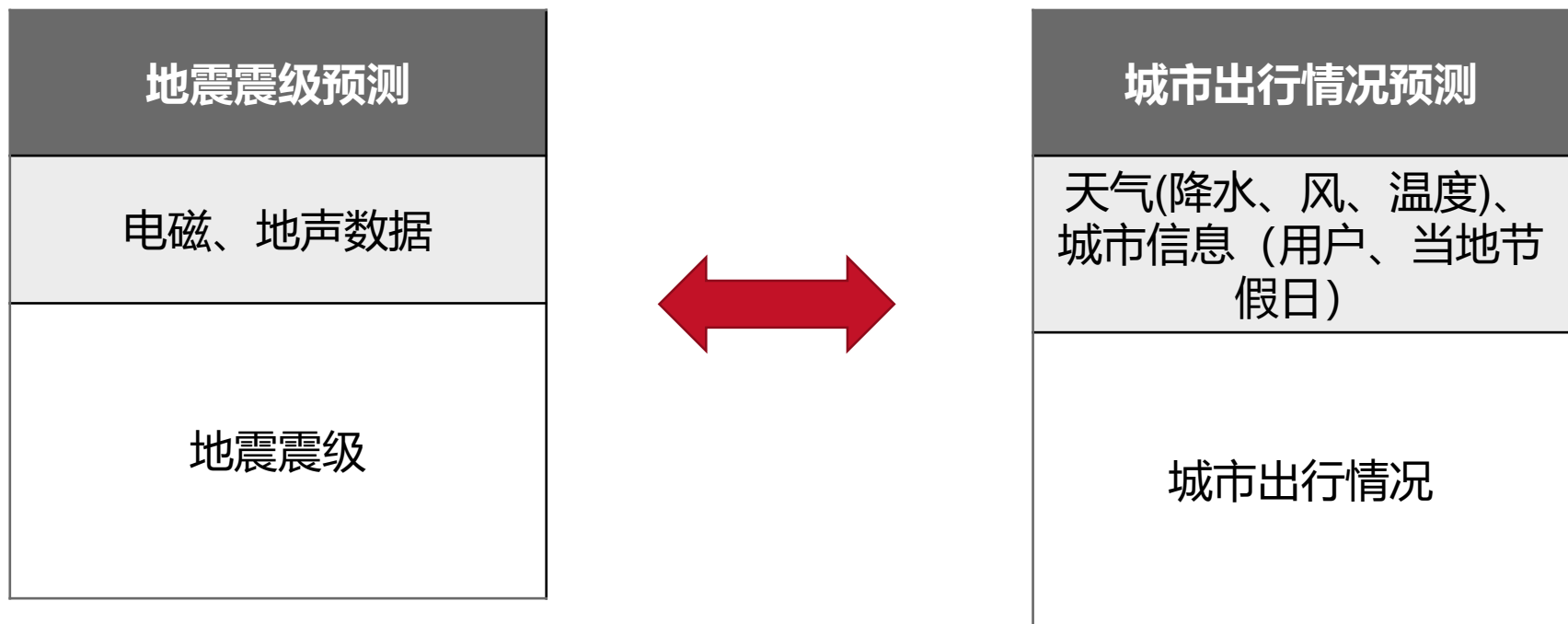
数据	处理	描述
地震目录数据	聚合处理	按天取最大值，得到川滇地区每天发生的最大级地震的震级、震中
电磁、地声数据	聚合处理	将每个站点的电磁、地声数据按天取平均值，得到每个站点每天的电磁、地声平均值
	缺失值处理	(1) 观察数据发现前325天的数据缺失值太多，为避免填充引起较大误差，因此直接删除所有站点前325天的电磁、地声数据； (2) 删除缺失值超过2/3的站点数据； (3) 利用线性插值法填补剩余缺失数据。

/03

震级预测



3.1 震级预测问题研究



问题类比，参考使用Uber2017年发表的论文《Time-series ExtremeEvent Forecasting with Neural Networks at Uber》提出的算法。

3.2 特征提取——手动提取

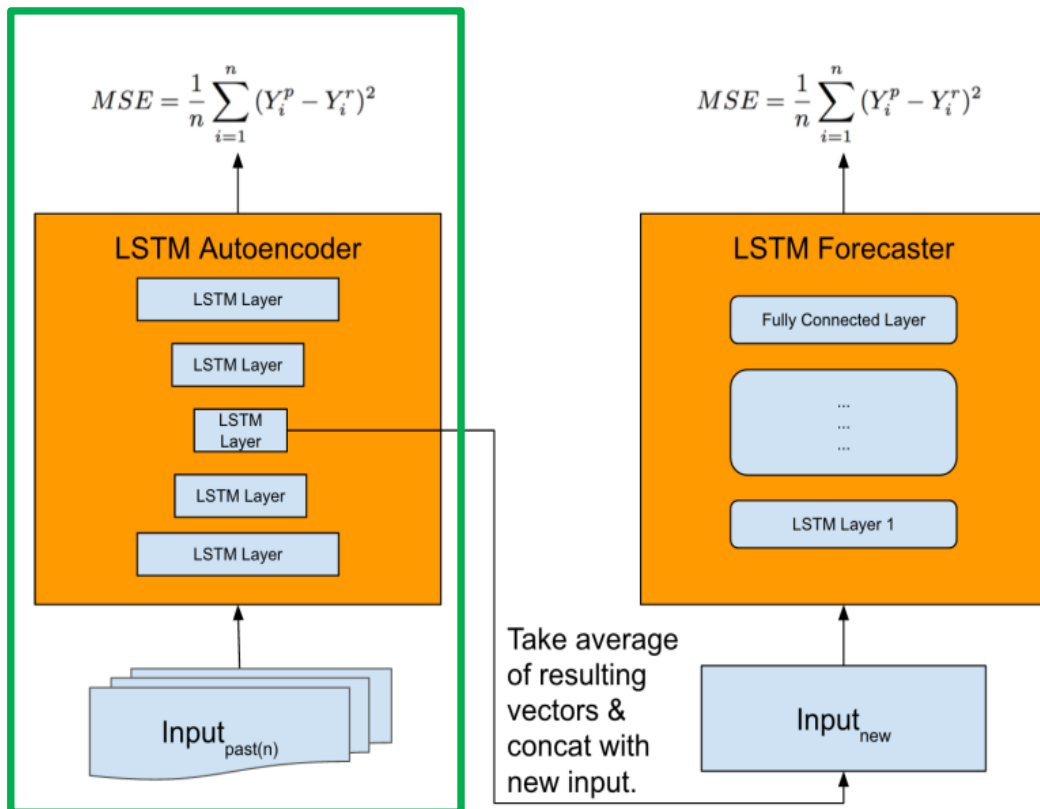
Feature	Description
Mean	Mean.
Var	Variance.
ACF1	First order of autocorrelation.
Trend	Strength of trend.
Linearity	Strength of linearity.
Curvature	Strength of curvature
Season	Strength of seasonality.
Peak	Strength of peaks.
Trough	Strength of trough.
Entropy	Spectral entropy.
Lumpiness	Changing variance in remainder.
Spikiness	Strength of spikiness
Lshift	Level shift using rolling window.
Vchange	Variance change.
Fspots	Flat spots using discretization.
Cpoints	The number of crossing points.
KLscore	Kullback-Leibler score.
Change.idx	Index of the maximum KL score.

(a) Classical time-series features that are manually derived (Hyndman et al., 2015).

如图所示，为每列原始数据计算包括 Mean、Var 等在内的 18 个特征，然后取平均值作为新的特征。

——《Time-series ExtremeEvent Forecasting with Neural Networks at Uber》

3.3 特征提取——自动提取

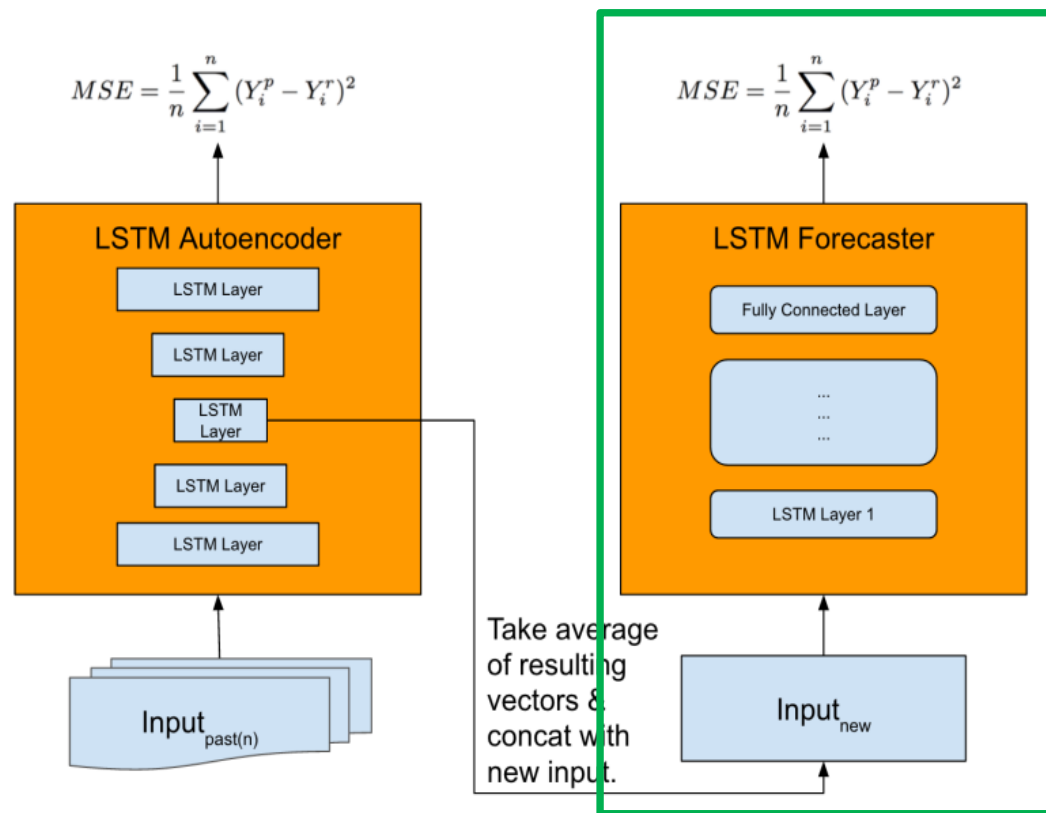


(b) An auto-encoder can provide a powerful feature extraction used for priming the Neural Network.

如图左部分所示，使用LSTM自编码器（LSTM Autoencoder）自动创建新特征。

——《Time-series ExtremeEvent Forecasting with Neural Networks at Uber》

3.4 震级预测



(b) An auto-encoder can provide a powerful feature extraction used for priming the Neural Network.

如图右部分所示，将手动方式提取的特征与自动方式提取的特征连接，然后使用简单LSTM的神经网络进行地震震级预测。

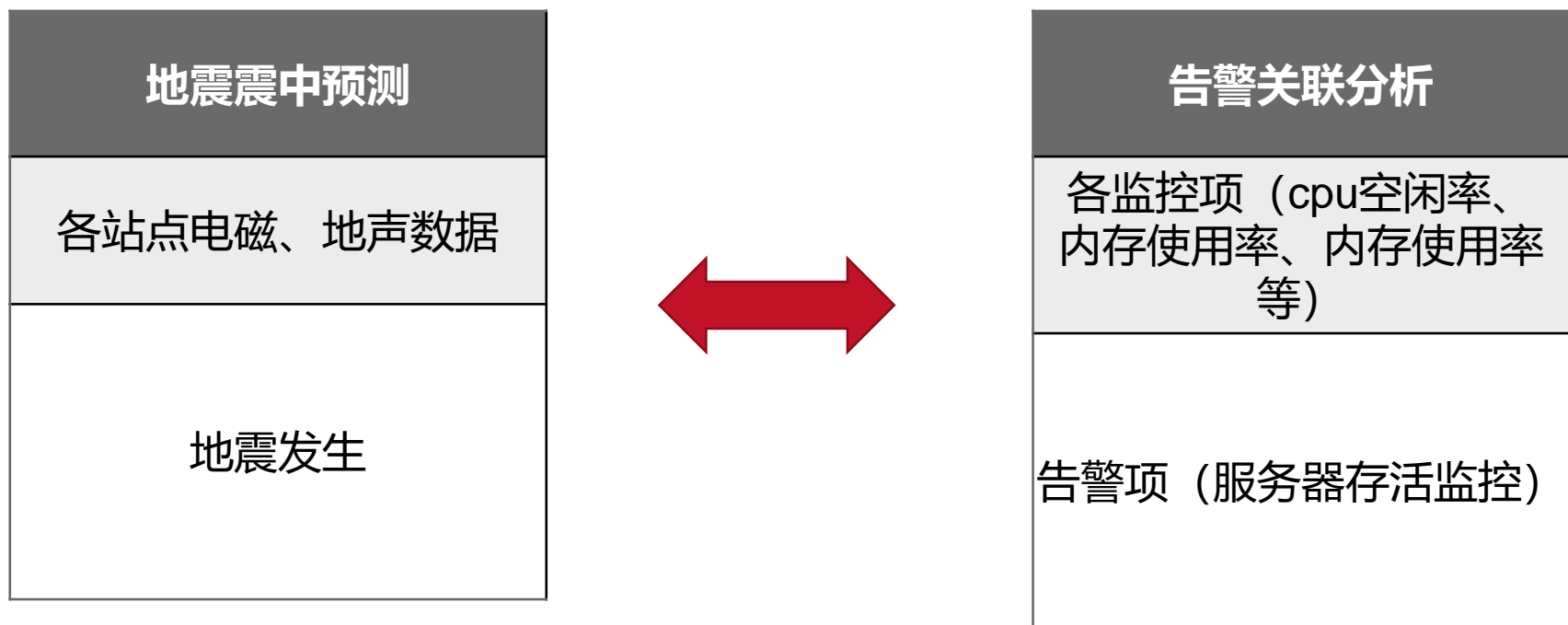
——《Time-series ExtremeEvent Forecasting with Neural Networks at Uber》

/04

震中预测



4 震中预测



问题类比，参考使用微软2014年在SIGKDD会议上发表的论文《Correlating Events with Time Series for Incident Diagnosis》中提出的报警关联分析方法。通过告警关联分析得到哪些站点的数据与地震事件最相关，取最相关的几个站点经纬度平均值作为震中。（基于假设：离震中越近的检测站点数据越与地震事件相关）

/05

预测及得分情况

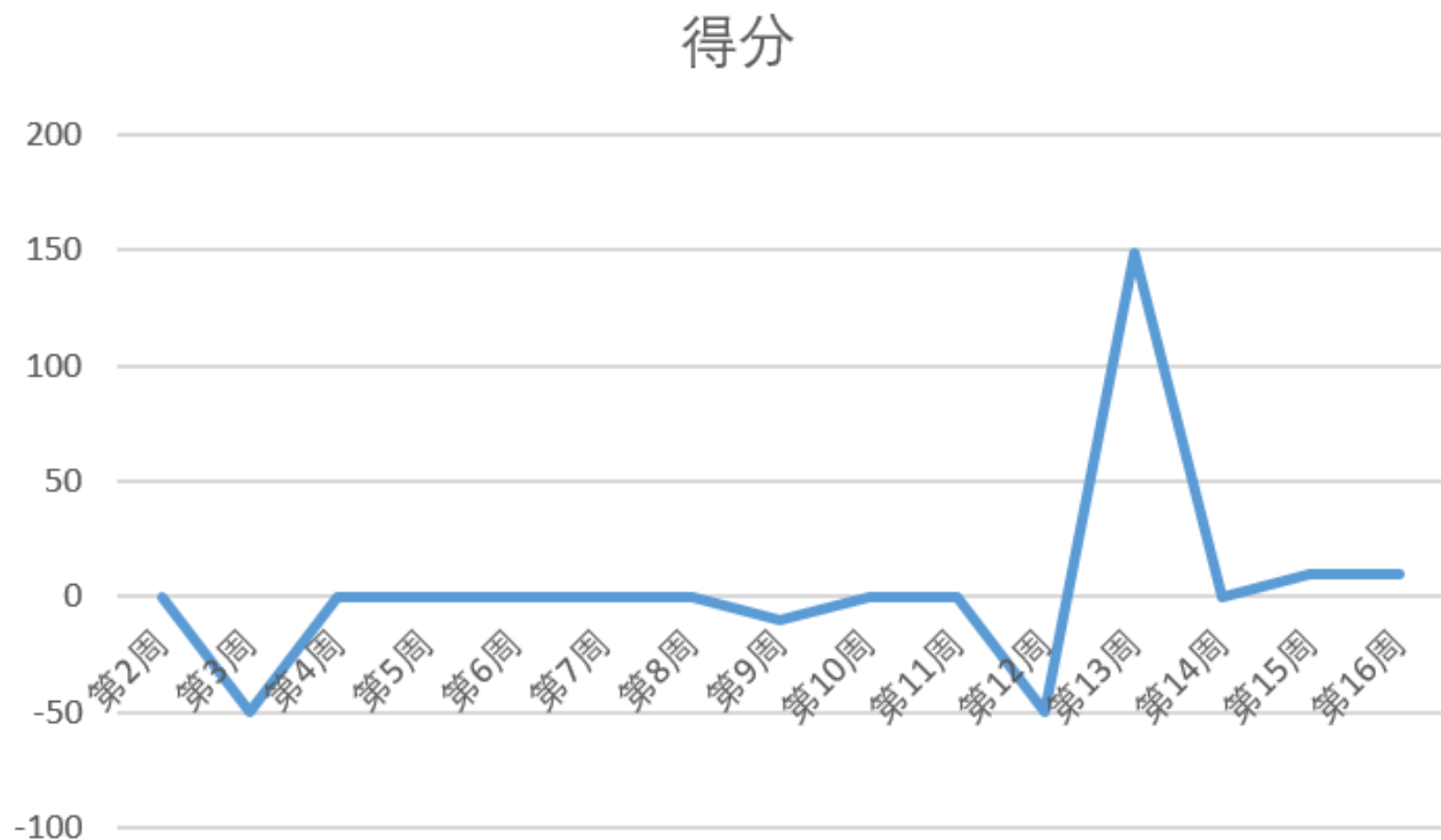


5.1 预测情况

date	第n周	得分	实际震级	预测震级	震级误差	震中误差
05.25-05.31	第2周	0	3.8	3.5	0.3	701.19
06.01-06.07	第3周	-50	4.3	0		
06.08-06.14	第4周	0	4	3.9	0.1	251.89
06.15-06.21	第5周	0	4.7	3	1.7	667.76
06.22-06.28	第6周	0	3.8	3.9	0.1	669.92
06.29-07.05	第7周	0	4.5	3.1	1.4	382.23
07.06-07.12	第8周	0	4.4	3.1	1.3	312.11
07.13-07.19	第9周	-10	0	4.1		
07.20-07.26	第10周	0	4.1	3.8	0.3	1008.61
07.27-08.02	第11周	0	3.7	3.5	0.2	653.13
08.03-08.09	第12周	-50	3.9	0		
08.10-08.16	第13周	149.222	3.8	3.9	0.1	36.41
08.17-08.23	第14周	0	0	4.1		
08.24-08.30	第15周	10	0	0	0	
08.31-09.05	第16周	10	0	0	0	
平均误差					0.6	520km

9次有震预测
4次震中误差<400km;
平均震级误差： 0.6
平均震中误差： 520.39km
2次无震准确预测

5.2 得分情况



最好成绩

第13周分数: 149.222:

震级误差: 0.1

震中误差: 36.42km



Thanks

团队：4freedom