

**浪潮-基于人工智能的地震波初至拾取项目概要设计报告**

|  |  |
| --- | --- |
| 文件编号： |  |
| 当前版本： |  |
| 编 制： |  |
| 审 核： |  |
| 批 准： |  |
| 完成日期： | 2018年8月15日 |

**浪潮（北京）电子信息产业有限公司**

版权所有 翻版必究

**目录**

[**目录** 1](#_Toc522131889)

[1 文档介绍 2](#_Toc522131890)

[1.1 文档目的 2](#_Toc522131891)

[1.2 文档范围 2](#_Toc522131892)

[1.3 读者对象 2](#_Toc522131893)

[1.4 参考文档 3](#_Toc522131894)

[1.5 术语与缩写解释 3](#_Toc522131895)

[2 地震波初至拾取平台总体设计 3](#_Toc522131896)

[2.1 地震波初至拾取平台整体设计 3](#_Toc522131897)

[2.2 运行环境配置设计 4](#_Toc522131898)

[2.2.1 软件环境配置 4](#_Toc522131899)

[2.2.2 硬件架构设计 4](#_Toc522131900)

[3 地震波初至拾取平台关键技术 5](#_Toc522131901)

[3.1 数据预处理 5](#_Toc522131902)

[3.1.1 数据提取 5](#_Toc522131903)

[3.1.2 数据拆分 6](#_Toc522131904)

[3.2 单道拾取 6](#_Toc522131905)

[3.3 多道平滑 7](#_Toc522131906)

[3.4 叠加较偏 8](#_Toc522131907)

[3.5 模型验证与性能分析 8](#_Toc522131908)

[3.5.1 模型验证 8](#_Toc522131909)

[3.5.2 性能分析 8](#_Toc522131910)

[3.6 模型部署 9](#_Toc522131911)

[4 系统优化设计 9](#_Toc522131912)

[4.1 样本集优化 9](#_Toc522131913)

[4.2 访问优化 9](#_Toc522131914)

[4.3 系统扩展优化 10](#_Toc522131915)

[5 附录 10](#_Toc522131916)

[5.1 附录1 审批记录表 10](#_Toc522131917)

# 文档介绍

## 文档目的

本文档是基于人工智能技术，对东方地球物理公司（以下称甲方）所提地震资料中提取折射波首次到达检测点时的波峰时间（初至提取）构建整体解决方案，采用神经网络算法及浪潮NF5280M5 GPU服务器构建地震波初至自动拾取平台的概要设计文档，重点论述了基于神经网络的初至波拾取模型设计和训练过程，包括系统整体框架、数据处理流程及初至拾取相关算法设计。

在下一阶段的详细设计中，程序员可参考此概要设计报告，对系统进行详细设计。在以后的软件测试以及软件维护阶段也可参考此说明书，以便于了解在概要设计过程中所完成的各模块设计结构，或在修改时找出在本阶段设计的不足或错误。

## 文档范围

1. 基于神经网络的地震波初至拾取平台总体设计；
2. 软件和硬件运行环境；
3. 初至拾取算法设计；
4. 系统优化。

## 读者对象

1. 开发、测试、维护人员；
2. 以后版本的修改人员。

## 参考文档

1. Tensorflow官方文档
2. 甲方合作选题建议及需求文档

## 术语与缩写解释

|  |  |
| --- | --- |
| **缩写、术语** | **解 释** |
| GPU | Graphics Processing Unit 图形处理器 |
| NMS | non maximum suppression非极大值抑制 |
| EM | Expectation Maximization Algorithm 最大期望算法 |
| IOU | Intersection over Union 重叠区域面积比例 |

# 地震波初至拾取平台总体设计

地震资料初至拾取技术是折射波静校正、层析成像、浅层折射波勘探及VSP等技术的关键。初至拾取目的是确定地震道上纯噪音信号和噪音与地震叠加信号之间的分界时刻，因此初至拾取正确与否在很大程度上影响后续处理的精度。在地表复杂及信噪比较低的情况下，目前仍然无法取得满意的拾取结果。同时人工的初至拾取工作繁重、校对困难，因此探寻可靠的自动拾取技术变得尤为重要。

## 地震波初至拾取平台整体设计

根据甲方所提出的需求，本项目需要实现以下功能：

1. 地震波探测区域的炮数、检波点数量、检波采样数量不同，模型读取的数据维度不同，需要构建可处理变维度数据的模型；
2. 基于历史初至数据训练模型，完成新数据初至时间预测并进行可视化；
3. 建立模型评价标准，完成自动初至拾取工具拾取结果与本方案拾取结果比较。

根据项目需求，本报告所涉及的平台整体框架采用如下图所示的四层架构：

|  |
| --- |
|  |
| 图2-1. 地震波初至拾取平台整体框架 |

其中数据层为SEG-Y格式的地震数据，为甲方提供的初始数据。运算层根据后期平台运行的实际需求确定，开发阶段采用浪潮NF5280M5 GPU服务器。在逻辑层中包含五个模块，数据预处理模块主要为训练模型做数据准备工作；单道拾取模块主要建立算法模型，完成前期初步拾取功能；多道平滑、叠加较偏、局部矫正模块对拾取结果进行小范围优化。最上层的人机交互层实现初至拾取结果可视化功能。

## 运行环境配置设计

### 软件环境配置

由于地震波初至拾取平台主要部署在服务器上，系统运行需要软件环境包括以下几个：

1. CentOS操作系统；
2. Python 2.7；
3. TensorFlow 1.2或以上版本

### 硬件架构设计

由于目前项目仍处于研发阶段，因此硬件平台采用浪潮NF5280M5单机多卡GPU服务器，不另做配置。

# 地震波初至拾取平台关键技术

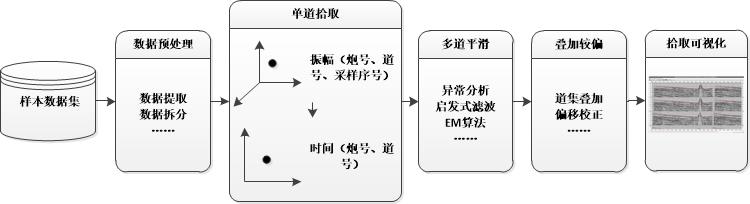
人工智能初至拾取是通过人工智能算法对样本的学习来建立一个拾取规则，以此对初至波进行拾取。该规则是否合适取决于两方面，（1）地震属性选取是否恰当；（2）人工智能算法选取与应用。

图3.1 数据流及系统流程图

图3.1中，样本数据集经过数据预处理提取地震波振幅的三维数组；对单炮单道波形，构建合理的特征集，采用神经网络算法准确拾取到初至结果；对单炮多道的初至结果，设计适用的滤波算法，剔除异常道数据，准确拟合初至波趋；寻找单道多炮数据之间的关联性，构建较正关系模型，提高初至拾取的置信度；最终将拾取结果进行可视化。

## 数据预处理

### 数据提取

地震波源数据格式为SEG-Y 数据格式，它是地震数据的最为普遍的格式之一，其格式基本结构如表3.1所示。

SEG-Y 数据中包含地震波中详尽信息，对于初至拾取场景需从中提取有效信息，本方案从SEG-Y 数据中提取道号、炮号、采样序号及对应振幅，完成数据的初步提取。

**表3.1 SEG-Y 标准磁盘文件格式**

|  |  |  |
| --- | --- | --- |
| 名称 | 长度/（字节） | 备注 |
| 文本文件头 | 3200 | EBCDIC编码（或ASCII编码） |
| 二进制文件头 | 400 | 16、32位二进制 |
| 道头 | 240 | EBCDIC编码（或ASCII编码） |
| 道数据 | 4\*采样点个数 | 32位IBM浮点数（或IEEE浮点数） |
| ……炮m | （240+4\*采样点个数）\*m | 共m炮 |

### 数据拆分

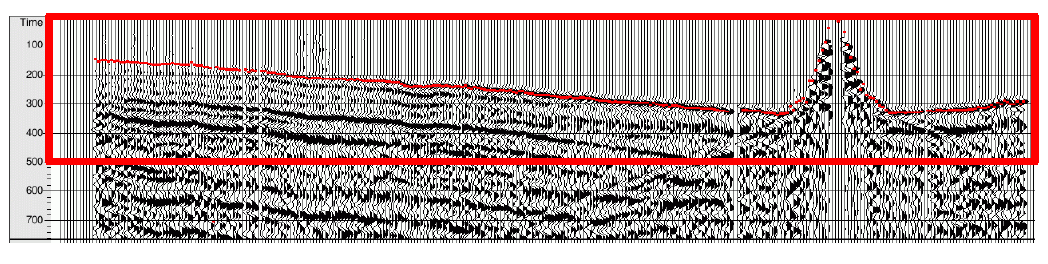
由于初至走时通常都在单道数据的前部，为了减小算法计算复杂度，可对所有数据固定采样点个数，如固定每道截取前500个采样点，如图3.2中红色框体所示。在实际应用中，需要对每一道数据进行拾取，因此初步设计的数据拆分不对道数和炮数进行修正。

图3.2 数据拆分

## 单道拾取

基于神经网络的单道初至拾取技术，需要将初至波检测问题转化为分类问题，如图3.3所示，当窗口中包含初至波时，窗口特征对应为1，否则为0。其中，每个窗口作为一个样本，窗口对应的特征作为标签。

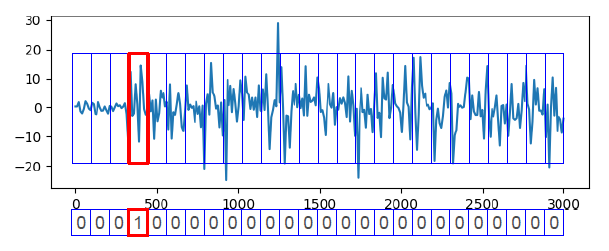


图3.3 单尺度窗口情形

为了提高初至拾取检测的准确度，可采用多尺度窗口对单道地震记录进行拾取，再经过NMS算法缩小候选窗口。

NMS算法实现过程：

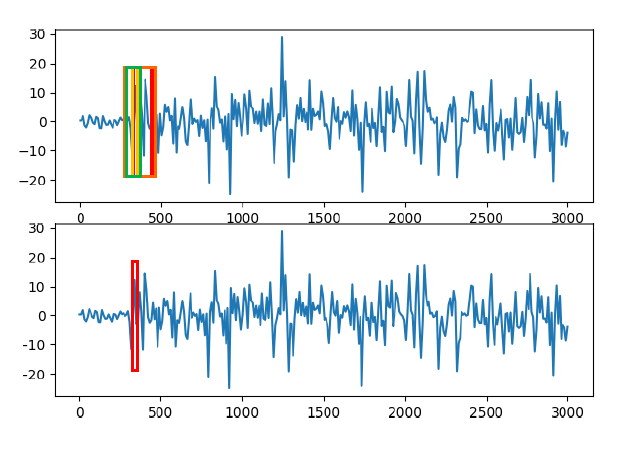
1. 置信度排序并选出置信度最高的窗口；
2. 计算其他窗口与最高置信度窗口的IOU，若IOU大于阈值，则去除；
3. 未处理的窗口中重新选置信度最大的窗口，重复第（2）（3）步，直到没有未处理过的窗口为止。

图3.4 多尺度窗口情形

## 多道平滑

单道拾取结果不可避免存在不切合实际的数据（异常值）或者噪声，为了提高数据的质量，有必要对拾取结果进行平滑处理，使拾取结果更接近真实数据分布。

通过相邻道的初至时间趋势和初至时刻能量特征，剔除并平滑异常初至走时。

多道平滑涉及传统滤波方法，样条插值法，一分类法（one-classSVM），密度聚类（DBSCAN），决策树等。

## 叠加较偏

由于地下地壳差异，部分检波点无法接收到特定起爆点传来的地震波，因此需要对不同炮点的多炮数据进行叠加，以增强检波信号，提高初至拾取准确度。

偏移校正：因为每次放炮点的位置不同，地震波传播到各检波点的时间也有所差异，但是可以根据放炮点位置偏移，估算地震波的偏移量。此外，通过计算不同炮的波形特征，通过特征匹配算法也可以得到对应的偏移校正量。

道集叠加：经过偏移校正后的道集数据可以相互叠加以增强检波信号，如采用均值、加权平均法、EM算法等。

## 模型验证与性能分析

### 模型验证

在完成上述模型构建与训练后，需要对其进行如下验证：

1. 人工修改任务量验证

比较初至拾取工具拾取结果与本方案拾取结果，确定人工修改工作任务量，验证本方案拾取结果可行性。

1. 模型前馈实时性验证

在服务器端模拟采用不同前端设备部署上述模型时，前馈计算的耗时，以寻找满足设计实时性需求的所有可行方案。

### 性能分析

1. 参数敏感性

参数敏感性分析旨在寻找影响初至拾取精度的重要参数集及其影响规律。当需求中的初至拾取精度发生变化时，通过配置敏感参数集快速适应新的拾取需求。

1. 最大误差场景

该指标用以测试模型及所搭配的系统可以正常运行和处理的极限环境。包括场景复杂度、光照强度、温湿度极限等。

1. 最佳适用场景

该指标用以测试系统最佳工况，并根据实际需求，调试模型参数以使系统最佳工况覆盖大多数应用场景。

1. 系统可扩展空间

测试系统配置瓶颈并统计端口数量，给出合理的系统最大扩容、功能扩展空间估计，根据甲方未来业务增长的实际需求评估系统合理的硬件配置及软件接口预留。

## 模型部署

本项目根据客户需求，将模型部署在客户所提供服务器上，在后台构建内部云计算架构，为业务部门提供计算服务。

# 系统优化设计

在上述原型系统的基础上，为了提升系统计算速率和性能，可以在项目工期内有针对地对系统相应模块进行升级和优化。

## 样本集优化

由于炮点位置不同，地震波在不同炮情况下传播到各检波点的时间也有所差异，可以根据炮点位置偏移，在单道拾取前对样本集进行偏移校正，使炮点位置差异对拾取结果影响最小化。

## 访问优化

为了提升I/O访问效率和GPU与CPU、GPU与GPU之间的通信速率，可以采用搭载了NVLink的高性能计算节点。

## 系统扩展优化

为了合理分析系统性能瓶颈和可扩展范围，本方案建议采用Teye系统对整体解决方案进行分析，从系统层面给出扩展优化诊断结果。在此基础上与甲方进一步沟通，结合未来甲方业务发展的实际需求，提供系统扩展的优化方案。

# 附录

## 附录1 审批记录表

|  |  |  |  |
| --- | --- | --- | --- |
| 角色 | 签名 | 日期 | 备注 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |