

数据集

名称	格式	链接	MD5
Complex Track Data	zip	百度 (提取码: cadk) , Dropbox	0a2039edccfb0b85223054f99a2d5ac5
Simple Track Data	zip	百度 (提取码: w938) , Dropbox	fbbce7270f5d5d685c2113720f0fe6a9

粒子碰撞数据集

本数据集共包含 200 多万条喷注信息，分为 **EVENT**、**JET**、**PARTICLE** 三类文件，三者为上下层级关系，一个碰撞事件（**EVENT**）会产生若干个喷注（**JET**），而一个喷注中会包含若干个粒子（**PARTICLE**）。**EVENT** 文件是对碰撞事件的描述，**JET** 文件详细说明了喷注的属性，**PARTICLE** 文件进一步描述了喷注中所含各个粒子的属性。

本次比赛分为**简单赛道**和**复杂赛道**，依次进行。简单赛道只要求选手根据**喷注属性**的数据集进行分类；复杂赛道在此基础上，又加入**喷注所含粒子的属性文件**和**碰撞事件文件**，数据的体量和维度剧增，难度也相应加大，具体如下。

复杂赛道

复杂赛道使用的数据包括喷注所在的碰撞事件、喷注的性质、喷注中所有粒子的特征，具体的训练集文件为:

complex_train_R04_event.csv、**complex_train_R04_jet.csv**、**complex_train_R04_particle.csv**。

在 **complex_train_R04_event.csv**（如下图示）中，**event_id** 字段表示质子对撞事件的编号、**number_of_jet_in_this_event** 字段表示对撞事件中所产生的喷注数量。

event_id	number_of_jet_in_this_event
1234567890abcdef	3

在 **complex_train_R04_jet.csv**（如下图示）中，**jet_id** 表示喷注的编号，**number_of_particles_in_this_jet** 表示喷注中所含的粒子数量，**jet_px**、**jet_py**、**jet_pz** 表示喷注方向，**jet_energy** 表示喷注能量，**jet_mass** 表示喷注质量，**event_id** 表示产生喷注的碰撞事件编号，**label** 表示 **jet** 所在的类别。

jet_id	number_of...	jet_px	jet_py	jet_pz	jet_energy	jet_mass	event_id
12389ab	21	84.7	2.81	1.228	85.4226	9.90909	fdsnjnrig
qwert08	19	-43.5	-22.7	-0.9435	49.5254	6.48975	plmkukjd

在 **complex_train_R04_particle.csv** 文件（如下图示）中，**particle_category** 表示粒子类别，**particle_px**、**particle_py**、**particle_pz** 表示粒子方向，**particle_energy** 表示粒子能量、**particle mass** 表示粒子质量、**jet_id** 表示粒子所在 **jet** 编号。

category	particle_px	particle_py	particle_pz	particle_energy	particle_mass	jet_id
-321	23.5467	1.11747	0.575207	23.5854	0.49368	123qwe
211	4.42681	0.0152337	0.144916	4.4314	0.13957	123qwe
130	8.22251	0.721089	0.6748569	8.29655	0.497610	123qwe

复杂赛道的测试集为：
complex_test_R04_event.csv、**complex_test_R04_jet.csv**、
complex_test_R04_particle.csv，
其中 **complex_test_R04_jet.csv** 不包括 **label**。