

Разработка алгоритмов кластеризации откликов координатных детекторов для эксперимента CBM

Г.Е. Козлов¹, В.В. Иванов¹, А.А. Лебедев^{1,2}, Ю.О. Васильев³

¹Лаборатория информационных технологий, ОИЯИ, Дубна

²Франкфуртский Университет им. Гете, Франкфурт-на-Майне, Германия

³Институт Тяжелых Ионов, Дармштадт, Германия

Введение

Одним из важных проектов в физике высоких энергий на сегодняшний день является эксперимент CBM (Compressed Baryonic Matter) [1], подготовка которого ведется на строящемся в Дармштадте (Германия) ускорительном комплексе FAIR (Facility for Antiproton and Ion Research).

Для успешной реализации эксперимента необходима обработка экспериментальных данных в темпе их поступления. Получение данных осуществляется с помощью набора детекторов, часть из которых имеют ячеистую структуру. Частицы, пересекая плоскости таких детекторов, приводят к срабатыванию одной или нескольких детектирующих ячеек. Далее, исходя из этих отпечатков, программными средствами определяются наиболее вероятные координаты пролета частиц - хиты, которые используются для восстановления треков - траекторий движения частиц в установке CBM.

Число частиц, рождающихся при каждом соударении тяжелых ионов при энергиях CBM, может быть весьма велико (до нескольких тысяч), а их отпечатки могут соприкасаться и перекрываться, что создает значительные трудности на этапе восстановления хитов. Для решения этой задачи используется специальная процедура - кластеризация. Нашей целью является поиск и разработка алгоритмов кластеризации, применимых к обработке откликов координатных детекторов в эксперименте CBM. В настоящем сообщении нами рассмотрена проблема кластеризации откликов в детекторах MVD (Micro Vertex Detector) и MuCh (Muon Chamber).

Алгоритмы кластеризации

С целью определения наилучшей стратегии обработки откликов координатных детекторов и выбора подходящих алгоритмов были рассмотрены основные направления и методы кластерного анализа [2, 3]. Наиболее подходящими для решения поставленной задачи были признаны методы Уорда [4] и ближайшего соседа [5], а так же разработанный нами оригинальный алгоритм кластеризации. Использование метода Уорда для MVD детектора невозможно, в силу большого количе-

ства обрабатываемых объектов. Для детектора MuCh этот метод требует доработки и в настоящий момент не используется. В данном сообщении мы рассматриваем два алгоритма кластеризации: разработанный нами оригинальный алгоритм и алгоритм, основанный на методе ближайшего соседа.

Первый алгоритм предназначен для обработки откликов детекторов в случае, если детектирующие ячейки имеют заряд. Он способен как определять отдельные, так и разделять большинство перекрывающихся кластеров. Основной принцип работы алгоритма следующий: каждая ячейка с локальным максимумом зарядов образует отдельный кластер, и каждый кластер может включать в себя только одну ячейку с локальным максимумом зарядов. Ячейки, которые не являются локальными максимумами, присоединяются к своим соседям с наибольшим зарядом. Это дает возможность разделять перекрывающиеся кластеры с собственными максимумами зарядов без использования сложных вычислений, что особенно важно при высокой множественности данных. Работа самого алгоритма проиллюстрирована на рис. 1 и осуществляется следующим образом:

1. Все активные ячейки рассматриваются последовательно. Для каждой из них определяется сосед с наибольшим зарядом - P_{max} .
2. Если заряд соседней ячейки P_{max} больше заряда рассматриваемой P_0 , то ячейка P_0 помечается для присоединения к ячейке P_{max} и включается в ее кластер. В этот кластер также включаются все ячейки, помеченные для присоединения к P_0 .
3. Иначе, принадлежность ячейки остается неизменной.

Таким образом, на выходе мы получаем набор кластеров, подготовленных к поиску хитов. Сами хиты вычисляются как центры масс соответствующих кластеров. Стоит отметить, что при объединении ячеек используется только вертикальное и горизонтальное соседство на плоскости, диагональное соседство не учитывается. Такой подход увеличивает риск образования ложных кластеров, но, при этом, повышает возможности алгоритма к разделению перекрывающихся

	1	2	3																											
Амплитуды	<table border="1"> <tr><td>5</td><td>7</td><td>4</td></tr> <tr><td>7</td><td>9</td><td>6</td></tr> <tr><td>4</td><td>6</td><td>0</td></tr> </table>	5	7	4	7	9	6	4	6	0	<table border="1"> <tr><td>5</td><td>7</td><td>4</td></tr> <tr><td>7</td><td>9</td><td>6</td></tr> <tr><td>4</td><td>6</td><td>0</td></tr> </table>	5	7	4	7	9	6	4	6	0	<table border="1"> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>48</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> </table>	0	0	0	0	48	0	0	0	0
5	7	4																												
7	9	6																												
4	6	0																												
5	7	4																												
7	9	6																												
4	6	0																												
0	0	0																												
0	48	0																												
0	0	0																												
Кластеры	<table border="1"> <tr><td>1</td><td>2</td><td>3</td></tr> <tr><td>4</td><td>5</td><td>6</td></tr> <tr><td>7</td><td>8</td><td>0</td></tr> </table>	1	2	3	4	5	6	7	8	0	<table border="1"> <tr><td>1</td><td>2</td><td>3</td></tr> <tr><td>4</td><td>5</td><td>6</td></tr> <tr><td>7</td><td>8</td><td>0</td></tr> </table>	1	2	3	4	5	6	7	8	0	<table border="1"> <tr><td>5</td><td>5</td><td>5</td></tr> <tr><td>5</td><td>5</td><td>5</td></tr> <tr><td>5</td><td>5</td><td>0</td></tr> </table>	5	5	5	5	5	5	5	5	0
1	2	3																												
4	5	6																												
7	8	0																												
1	2	3																												
4	5	6																												
7	8	0																												
5	5	5																												
5	5	5																												
5	5	0																												

Рис. 1: Принцип работы разработанного алгоритма кластеризации с использованием информации о зарядах

ся кластеров и делает его работу более быстрой за счет уменьшения количества рассматриваемых соседей.

Второй алгоритм работает без учета информации о зарядах ячеек, что является необходимым условием его использования для некоторых детекторов. Этот алгоритм основан на упрощенном методе ближайшего соседа. Данный метод предполагает последовательное присоединение объектов к кластерам в зависимости от расстояний между ними. Объект будет присоединен к некоторому кластеру, если расстояние между данным объектом и любым из объектов, уже включенных в кластер достаточно мало. Для определения близости объектов используется матрица расстояний - специальная матрица, включающая в себя расстояния между каждой парой рассматриваемых объектов.

С целью достижения максимальной скорости кластеризации в условиях отсутствия информации о зарядах ячеек, мы отказались от использования матрицы расстояний. Вместо этого, в качестве меры близости объектов - детектирующих ячеек на плоскости детектора. Таким образом, для определения кластеров может быть использована простая рекурсивная функция. Принцип работы данного подхода проиллюстрирован на рис. 2. Как и в предыдущем случае, мы учитывали только горизонтальное и вертикальное соседство. Учет диагональных соседей при таком подходе нецелесообразен, поскольку может привести к объединению соприкасающихся кластеров.

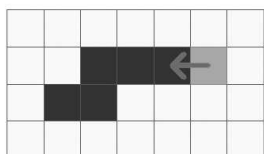


Рис. 2: Принцип работы алгоритма, основанного на методе ближайшего соседа

Кластеризация откликов в детекторе MuCh

Детектор MuCh предназначен для идентификации мюонов. Он состоит из 18 детектирующих плоскостей, объединенных в 6 станций. Между станциями располагаются адронные фильтры из железа и углерода, предназначенные для подавления фона от адронов. Таким образом, невзирая на количество попавших в детектор частиц, до последних станций долетают только мюоны. Результатом такой схемы работы детектора является неравномерное распределение кластеров в зависимости от номера станции.

При тестировании алгоритмов рассматривались центральные соударения золота по золоту при энергии 25 ГэВ на нуклон, смоделированные в программной среде эксперимента CBM - CBMROOT. В таких соударениях рождается до 1000 частиц, что приводит к высокой множественности данных, поступающих с детекторов. На мюонный детектор приходится около 6500 срабатываний ячеек, образующих порядка 3500 кластеров. При этом, до 90% от их количества сосредоточены на первых двух станциях. Кластеры на этих плоскостях располагаются достаточно близко друг к другу и могут перекрываться, что делает кластеризацию для данного детектора достаточно сложной задачей.

Рассматриваемые алгоритмы кластеризации мы сравниваем с методом, который используется в настоящее время для восстановления хитов в детекторе MuCh коллаборацией CBM. Результаты сравнения приведены на рис. 3. Как видно из приведенной гистограммы, выбор алгоритма кластеризации практически не влияет на эффективность обработки данных для последних станций детектора. При этом, на первых станциях разработанный алгоритм кластеризации с учетом зарядов ячеек показывает наилучшие результаты среди рассматриваемых. Главной проблемой данного алгоритма является образование относительно большого количества фиктивных кластеров. Их число может достигать 10-15% от общего числа найденных кластеров, что может осложнить процесс восстановления траекторий движения частиц.

Кластеризация откликов в детекторе MVD

Детектор MVD планируется оснастить четырьмя плоскостями. Для активных элементов (ячеек) рассматривается два варианта реализации: с учетом и без учета зарядов частиц. Мы исследуем оба случая, используя соответствующие алгоритмы кластеризации. При отсутствии зарядов мы также учитываем возможность применения поро-

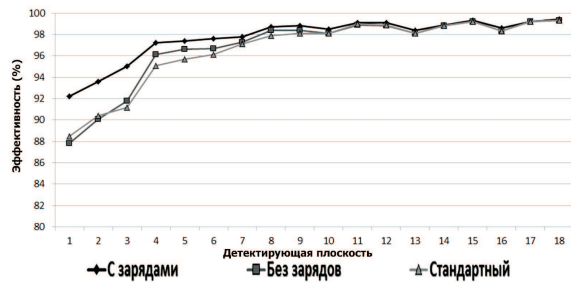


Рис. 3: Эффективность кластеризации откликов детектора MuCh

га активации ячеек. При тестировании на Au+Au соударениях все рассматриваемые алгоритмы показали высокие результаты по точности и эффективности кластеризации и определения хитов [6].

Помимо частиц, рождающихся при соударении ионов золота по золоту, при прохождении высокоинтенсивного пучка через мишень образуется большое количество дельта-электронов. Их число зависит от частоты взаимодействия и может значительно превышать количество частиц от соударений Au+Au. Поэтому мы исследовали эффективность алгоритмов кластеризации при различных значениях частоты взаимодействия (рис. 4). Частота взаимодействия варьировалась от 10 до 120 кГц с шагом в 10 кГц, а также были рассмотрены критические значения: 300 и 500 кГц.

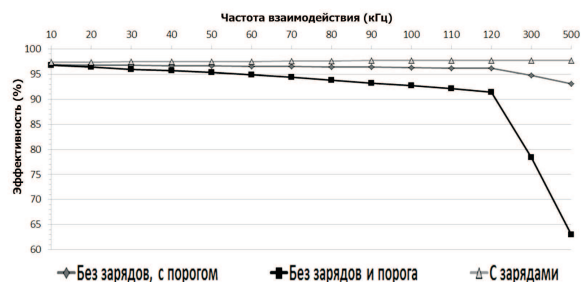


Рис. 4: Эффективность кластеризации откликов детектора MVD

Наилучшие результаты показал алгоритм, использующий заряды ячеек. Но при повышении числа активных ячеек (10^5 и более) скорость работы данного алгоритма сильно падает и обработка одного события может занимать до нескольких десятков секунд. Алгоритм беззарядовой кластеризации, используемый совместно с оптимальным порогом активации ячеек, продемонстрировал очень близкие результаты, лишь незначительно уступая алгоритму кластеризации с учетом зарядов. Таким образом, информация о зарядах ячеек не является обязательной для успешной кластеризации откликов детектора MVD. Но, при отсутствии порога активации ячеек, эффективность данного алгоритма серьезно падает.

Заключение

Мы рассматриваем проблему кластеризации откликов координатных детекторов, используемых в эксперименте CBM. С целью решения этой задачи нами были разработаны специализированные алгоритмы кластеризации. Тестирование алгоритмов на данных MuCh и MVD детекторов показало их применимость для решения поставленной задачи. Алгоритм кластеризации с использованием зарядов ячеек в обоих случаях показал наилучшую эффективность, но он имеет проблемы с образованием фиктивных кластеров и низкой скоростью работы при больших объемах данных (от 10^5 объектов и выше), что делает его использование для MVD детектора при высокой частоте взаимодействия невозможным.

При обработке откликов MVD детектора наилучшие результаты продемонстрировал алгоритм, основанный на методе ближайшего соседа с использованием порога активации ячеек. Незначительно уступая в эффективности, он не имеет проблем с фиктивными кластерами и его скорость работы значительно выше, чем у второго разработанного нами алгоритма. Таким образом, каждый из рассмотренных алгоритмов обладает своими преимуществами и недостатками, но в целом способен производить кластеризацию откликов координатных детекторов. Для детектора MuCh наилучшие результаты достигаются с использованием зарядов ячеек, тогда как для детектора MVD возможна кластеризация без учета зарядов.

Список литературы

- [1] Compressed Baryonic Matter in Laboratory Experiments, The CBM Physics Book, 2011 (<http://www-alt.gsi.de/documents/DOC-2009-Sep-120-1.pdf>)
- [2] Козлов Г.Е. Исследование алгоритмов кластеризации откликов детекторов с ячеистой структурой. Сообщение ОИЯИ, P10-2012-123, Дубна, 2012. 18 с.
- [3] G.Kozlov Study of clustering algorithms for detectors with a pad structure // CBM Progress Report 2011, GSI, Darmstadt, 2012. p. 69
- [4] J.H.Ward Hierarchical grouping to optimize an objective function, J. Amer. Statist. Assoc., Vol. 58, 1963, p. 236-244
- [5] R.Sibson SLINK: An optimally efficient algorithm for the single-link cluster method, The Computer Journal, Vol. 16, 1973, p. 30-34
- [6] G.Kozlov, V.Ivanov, A.Lebedev, I.Vassiliev: Study of clustering algorithms for the MVD detector // CBM Progress Report 2012, GSI, Darmstadt, 2012. p.91