

РАЗРАБОТКА ПРИКЛАДНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА ГИСТОГРАММ В СРЕДЕ *ROOT*

Филиал «Протвино» университета «Дубна»
Кафедра информационных технологий

Исследования, выполняемые в науке и технике, включают в себя обработку полученных данных с их детальным анализом. Обработка этих данных заключается в представлении данных в графическом виде (более наглядном для понимания), оценке параметров функциональных зависимостей и их погрешностей, статистической проверке гипотез об адекватности используемых моделей.

Исходя из этого мы видим, что любое исследование имеет большое количество переменных, которые нужно контролировать, обрабатывать и анализировать. А при работе с большими массивами данных необходимо иметь возможность быстро сравнивать и анализировать любые данные, например, распределения, полученные на подвыборках данных за разные периоды времени или при разных условиях. Это требует удобного и функционального графического интерфейса и инструментария для их быстрого анализа.

Основная задача данной работы — создать программный продукт, который сможет автоматизировать процесс оценки большого набора данных, будет удобен в использовании и даст возможность объективно анализировать данные.

Для достижения поставленной цели надо решить следующие задачи:

- выполнить формализацию исходных данных в контексте используемой среды
- выполнить преобразование исходных данных в соответствии с устанавливаемыми критериями
- построить графические изображения сравниваемых данных и результат их анализа
- разработать инструментарий для анализа данных
- создать наглядный, удобный и понятный пользовательский интерфейс и инструментарий
- построить графические изображения данных и результат их анализа, а также организовать вывод подписей данных

Для создания пакета была выбрана среда разработки *ROOT*[1], созданная в Европейском центре ядерных исследований, так как она имеет всю необходимую функциональность и разрабатывалась с учетом обработки огромного количества данных.

Программное обеспечение для работы с большими массивами данных должно давать возможность быстрого сравнения и анализа распределений, полученных на подвыборках данных за разные периоды времени или при разных условиях. Интерфейс программы (рис. 1) должен быть удобным и содержать следующие инструменты:

- инструмент для поиска и отображения структуры входных данных;
- задание типа операции над гистограммами — деления, вычитания гистограмм, объединения бин, выполнение нормировки на количество записей и событий, нахождение различий как в наборах гистограмм, так и в отдельно выбранных гистограммах;
- инструмент для вывода множества гистограмм, возможность ввода названий гистограмм и редактирования их координат.

Входные данные находятся в файлах с расширением *root* и имена этих файлов появляются в окошке графического интерфейса (рис. 2).

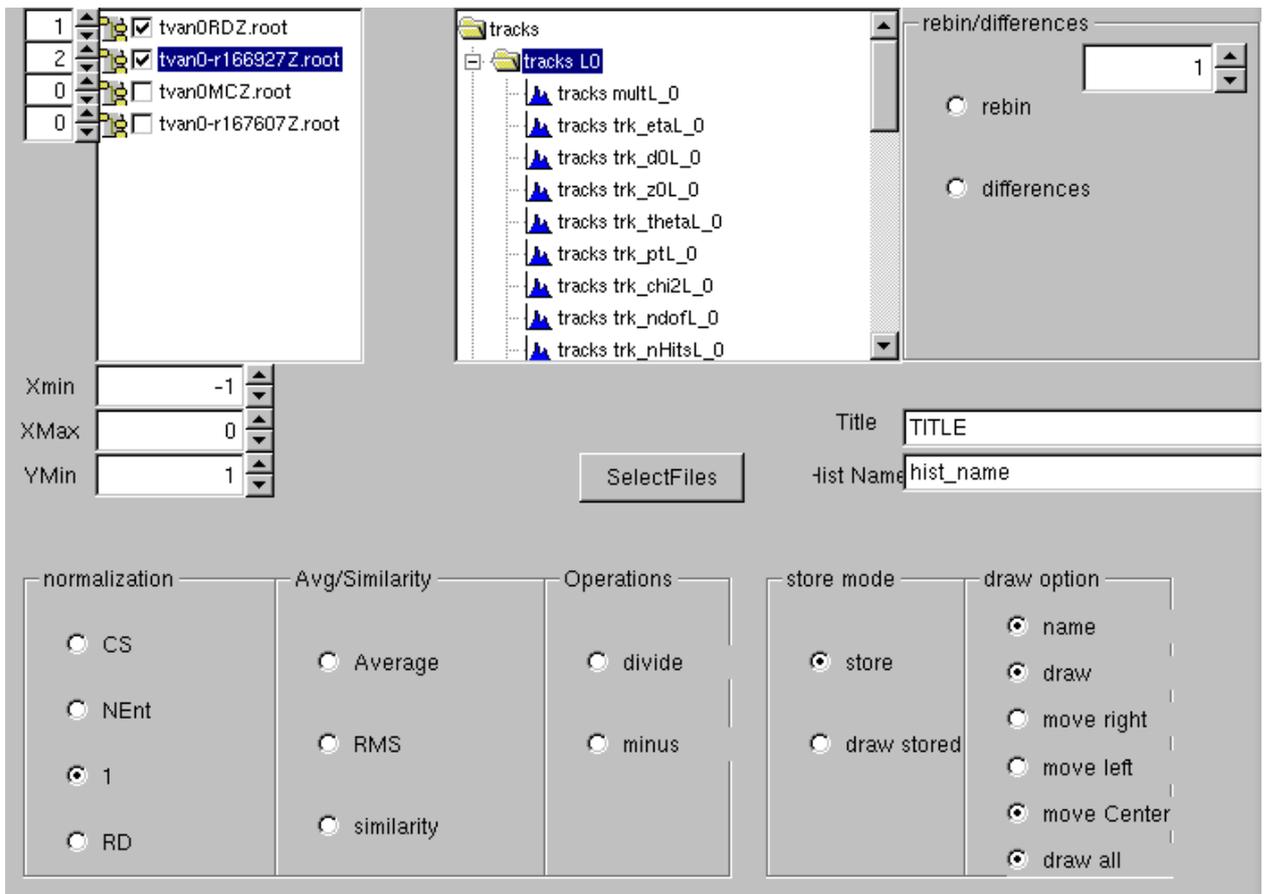


Рис. 1. Графический интерфейс программы для сравнения гистограмм

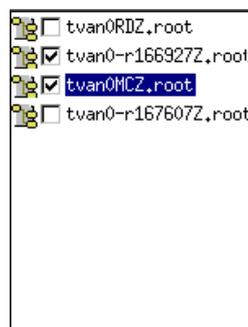


Рис. 2. Окно графического интерфейса для выбора нужных файлов

При нажатии на имя нужного файла он загружается в память программы, отмечается в окошке как выбранный для обработки. Гистограммы из этого файла будут рисоваться определённым цветом в соответствии с порядком выбора.

В другом окошке появляется структура, отражающая содержание входного файла — директории и имена гистограмм в них. Далее, с помощью простого нажатия имени гистограммы, можно выполнять различные операции. На рис. 3 приведена панель управления этими операциями. Левая колонка кнопок нужна для определения нормировок, с которыми будут рисоваться гистограммы. Например, если нажать кнопку 1, то гистограммы будут нарисованы так, как они определены в исходных файлах. Если же нажать кнопку *Nent*, то предварительно каждая гистограмма будет пересчитана таким образом, что полная площадь под ней будет равна 1. В такой моде удобно сравнивать форму гистограмм. Выбрав моду нормировки и нажав на имя выбранной, мы в новом окне увидим нарисованными несколько гистограмм (по числу выбранных файлов). Цвет каждой соответствует порядку выбора входных файлов. Пример такого изображения приведён на рис. 4.

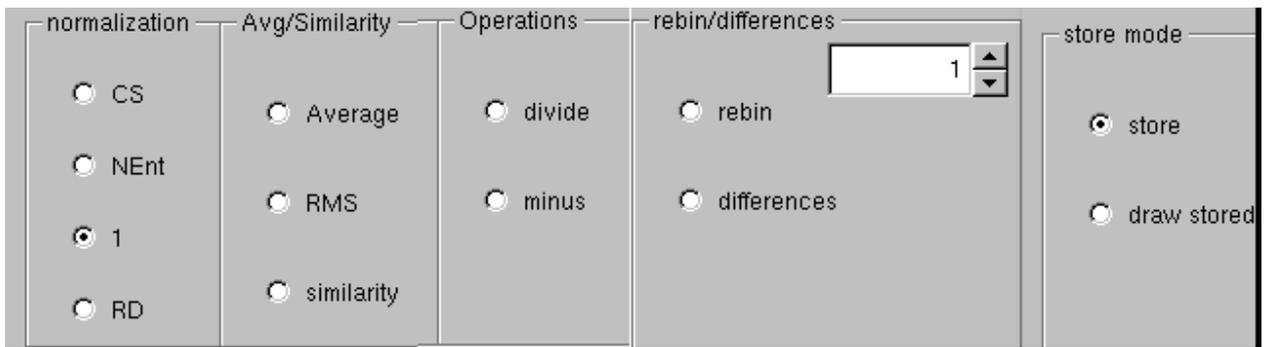


Рис. 3. Панель задания операций с гистограммами

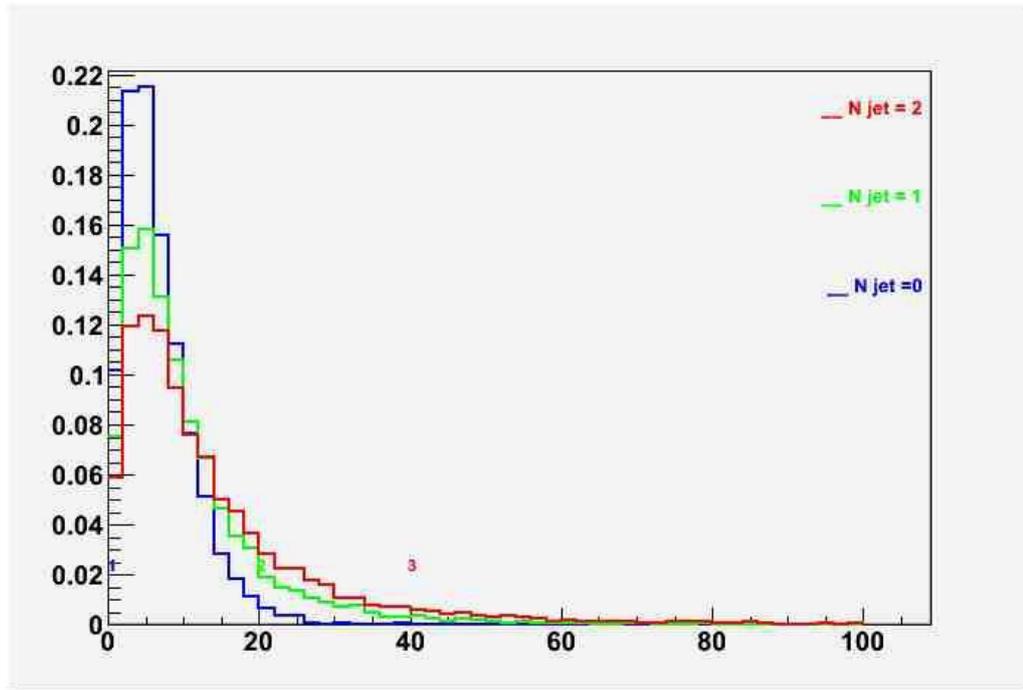


Рис. 4. Пример одновременного построения гистограмм из нескольких входных файлов

Во второй колонке панели управления находятся кнопки, позволяющие проводить сравнение некоторых характеристик гистограмм. Можно вывести средние по гистограмме (кнопка *Average*), либо среднеквадратичные отклонения (кнопка *RMS*), либо величину, характеризующую “похожесть” гистограмм, заданную по какому-либо критерию. На рис. 5 приводится график среднеквадратичных отклонений для выбранных гистограмм.

В третьей колонке панели управления определяются операции деления и вычитания гистограмм, в четвёртой — возможность объединения бин. При этом в специальном окошке надо ввести количество объединяемых бин.

Самая правая колонка панели управления позволяет отобразить гистограммы для построения.

Таким образом реализован пакет для анализа и сравнения большого количества гистограмм, обладающий солидной функциональностью. Графический интерфейс облегчает понимание инструментария и выбора режимов работы.

Благодаря простому и функциональному инструментарию пользователь легко и быстро может производить анализ большого объема данных.

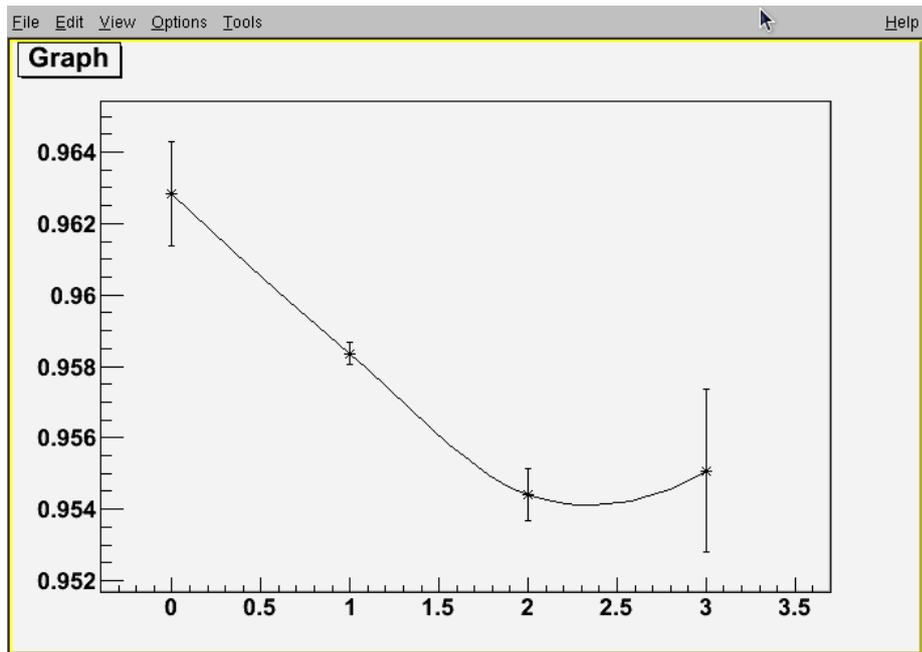


Рис. 5. График сренеквадратичных отклонений для выбранных гистограмм

Библиографический список

<http://root.cern.ch>