

# КОПУЛЫ КАК ИНСТРУМЕНТ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Литвинчук Д.В. Королев К.Ф.

Волорова Н.А. – к.т.н., доцент

В настоящее время все более актуальным становится вопрос о получении, подготовке и анализе данных. Копула-функция представляет собой инструмент для выполнения последнего. Метод копула-функции определяет характер взаимодействия между случайными величинами, что позволяет сформировать более адекватную модель описания совместного распределения нескольких переменных, чем многомерный нормальный закон. Важным преимуществом функции является то, что при моделировании многомерного распределения, помимо частных распределений она учитывает характер их взаимодействия. Копула-функция может быть применена для описания совместного распределения переменных имеющих разные частные распределения.

Общеизвестно, что лишь по маргинальным распределениям невозможно достоверно восстановить  $n$ -мерное совместное распределение. Всё дело в том, что в таком случае нам не хватает информации о взаимозависимости одномерных распределений. Копула, в данном случае, является недостающим звеном, содержащим всю информацию о взаимосвязи маргиналов.

Копула — это распределение, определённое на  $n$ -мерном единичном кубе, все маргиналы которого распределены равномерно на  $[0;1]$ .

Теорема Склара (Sklar's theorem) утверждает, что какими бы ни были  $n$ -мерное распределение  $F(x) = F(x_1, x_2, \dots, x_n)$  и его маргинальные распределения  $F_k(x_k), k = 1 \dots n$  существует такая копула  $C(u_1, u_2, \dots, u_n)$ , что для всех  $x \in R^n$  выполняется равенство:

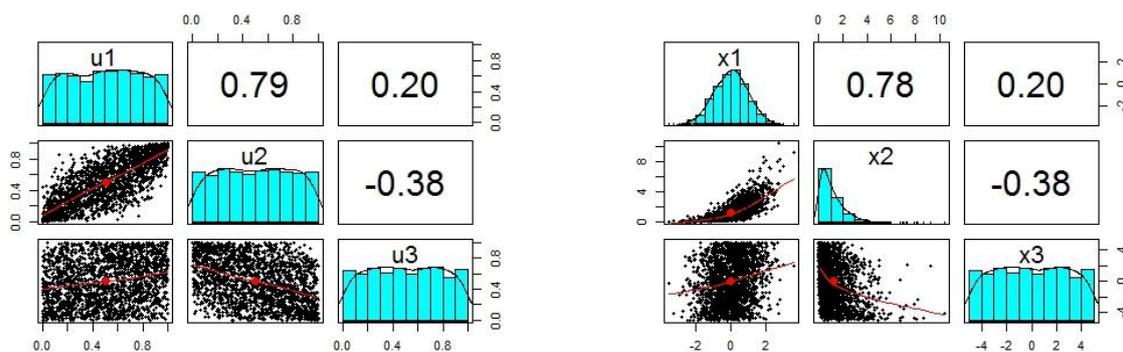
$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$$

Копула является обыкновенной функцией распределения, потому для копул можно определить плотность как неотрицательную функцию  $c(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$ , для которой выполняется:

$$C(u_1, u_2, \dots, u_n) = \int_0^{u_1} \int_0^{u_2} \dots \int_0^{u_n} c(s_1, s_2, \dots, s_n) ds_1 ds_2 \dots ds_n$$

Следствие Теоремы Шкляра устанавливает связь между плотностями совместного и маргинальных распределений с плотностью копулы, описывающей данное распределение.

$$c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = \frac{f(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n f_i(x_i)}$$



Выборка на основе копулы

Результирующая выборка

Копулы предлагают элегантный способ моделирования сложных распределений. Например, с помощью копулы Гаусса можно построить  $n$ -мерное распределение по заданной корреляционной матрице и маргинальным распределениям.

$$C_{GA}(u_1, u_2, \dots, u_n) = \Phi_n(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n))$$

Где  $\Phi$  — функция нормального распределения. Благодаря копулам, данный способ позволяет отдельно задавать корреляционные зависимости маргиналов и сами маргиналы. Для примера, сгенерируем выборку на 2000 элементов, средствами языка R, из распределения с маргиналами: нормальным, экспоненциальным, равномерным; и попарными коэффициентами корреляции равными 0,8, 0,2, -0,4. Для этого сгенерируем выборку из трёхмерного нормального распределения с соответствующей корреляционной

матрицей. Затем, к полученным данным, по координатам применим сначала  $\Phi$ , получив тем самым выборку на основе копулы, а затем  $F_i^{-1}$  — обратную функцию для данной координаты.

Список использованных источников:

1. Благовещенский Ю.Н. Основные элементы теории копул. Прикладная эконометрика, 2012.
2. Фантаццини Д. Моделирование многомерных распределений с использованием копула-функций. Прикладная эконометрика 2011.

## ПРОГРАММНАЯ СИСТЕМА СБОРА НОВОСТЕЙ ДЛЯ ГОРОДСКИХ ИНФОРМАЦИОННЫХ ПОРТАЛОВ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Королёва П.Б.

Жвакина А.В. – к.т. тех. наук, доцент

Информационный портал представляет собой универсальный сайт, основной контент которого – информационные тематические статьи. Существует множество классификаций таких сайтов. Блоги, словари и справочники, объявления и вакансии, социальные сети – это лишь некоторые из них. Данная работа связана с ещё одним типом – с информационными сайтами городов и районов.

Можно с уверенностью сказать, что почти каждый город имеет собственный сайт. Обычно он содержит не только историю города, основную информацию о существующих организациях и различные справочные данные, но и новости. Новости группируются по их содержанию, и почти всегда на сайте выделяются такие разделы, как «В Беларуси» или «В мире».

Проработав пару месяцев с контентом одного городского информационно-новостного сайта, я узнала, что содержание вышеупомянутых разделов формируется с помощью более крупных новостных сайтов-источников. Это происходит вручную путём изучения контента стороннего сайта, отбора новостей и стандартной их публикации с ссылками на оригинальные источники.

Данный процесс требует достаточно много усилий от контент-менеджера. Это приводит к необходимости решения задачи оптимизации рабочего времени и упрощения агрегации информации. Ручной «копи-паст» новостей в настоящее время, когда появляется всё больше и больше различных агрегаторов и грабберов является проявлением нерационального подхода. В таких случаях целесообразно использовать автоматизированную настраиваемую систему сбора новостей.

Разработанная мною программа представляет собой именно такую систему. В перспективе её можно интегрировать на любой сайт для агрегации любых видов данных, но в данный момент акцент ставится именно на сбор новостей. Принцип работы указан на рисунке 1:

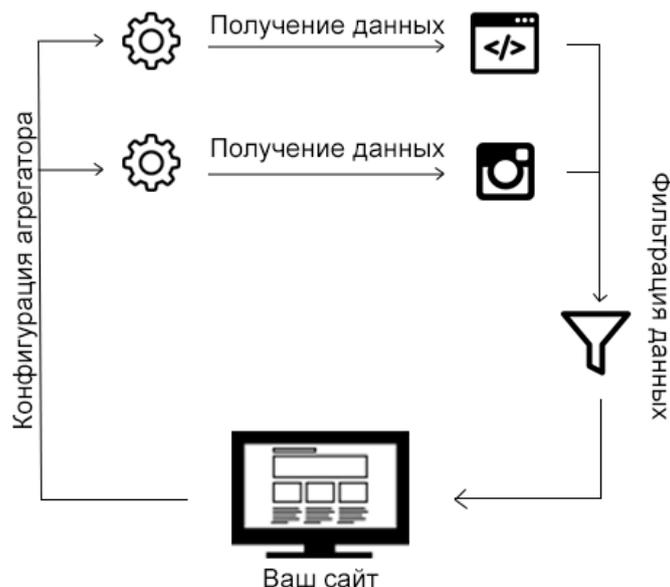


Рис. 1– Схема работы системы сбора новостей

Данный программный продукт обеспечивает возможность выбора имеющихся источников информации или создания и настройки новых.