

Санкт-Петербургский политехнический университет  
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

**Отчёт по курсовой работе  
по дисциплине «Анализ данных с интервальной  
неопределённостью»**

Выполнил студент:  
Габдрахманов Булат Маратович  
группа: 5040102/20201

Проверил:  
к.ф.-м.н., доцент  
Баженов Александр Николаевич

Санкт-Петербург  
2023 г.

## Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>2</b>
<b>2</b>	<b>Теория</b>	<b>2</b>
2.1	Бинарная мера совместности (покрытия) интервалов. . . .	2
2.2	Обобщение бинарной меры совместности (покрытия) интервалов. . . . .	2
<b>3</b>	<b>Изотопная подпись для некоторых типов органических соединений.</b>	<b>4</b>
<b>4</b>	<b>Пример</b>	<b>6</b>
<b>5</b>	<b>Заключение</b>	<b>8</b>

## Список иллюстраций

1	Вариации изотопного состава сульфидов в единицах $10^3 \cdot \delta^{34}S_{VCDT}$ . . . . .	5
2	Интервальная мода вариаций $10^3 \cdot \delta^{34}S_{VCDT}$ . . . . .	6

# 1 Постановка задачи

Необходимо провести анализ, используя несимметричные меры совместности интервалов.

## 2 Теория

### 2.1 Бинарная мера совместности (покрытия) интервалов.

По мере развития интервального анализа, были введены различные определения и конструкции оценки меры совместности интервальных объектов. Всем известно, что вопрос сравнения интервалов рассмотрен в публикации В.М. Левина. Однако в этой работе рассмотрены только абсолютные меры сравнения.

Вместе с тем, в практике обработки данных часто необходимо оперировать с относительными величинами. В частности, это нужно из-за необходимости соизмерения допусков и размеров деталей, потребности размерителей и значений измеряемых величин, и т.п.

В публикации S.Kabir et al, вводятся различные меры сходства или объектов нечётких множеств и интервалов, используются обозначения  $S(A, B)$ .

Кроме того, авторы вводят несимметричную меру *overlapping ratio* для интервала  $I_i$  относительно пары интервалов  $\{I_i, I_j\}$

$$OR(I_i, I_j) = \frac{|I_i \cap I_j|}{|I_i|} \quad (1)$$

$$s = \min(OR(I_i, I_j), OR(I_j, I_i)). \quad (2)$$

Мера *несходства*, *dissimilarity* соответственно имеет вид  $d = 1 - s$ .

Все меры, представленные в перечисленных работах, развивают идею Жаккара, и оперируют только с пересекающимися интервалами. В публикации обсуждаются также и непересекающиеся интервалы, но вычисления меры совместности не производится.

### 2.2 Обобщение бинарной меры совместности (покрытия) интервалов.

Приступим к обобщению меры совместности интервалов, имея целью описывать единообразно как накрывающие, так и не накрывающие вы-

борки. Для анализа данных необходимо проводить сравнение интервальных объектов универсальным образом, независимо от свойств накрытия. В первую очередь, введём базовую конструкцию совместности для двух интервалов. Для иллюстрации идеи, рассмотрим следующую числовую характеристику степени совместности двух интервалов  $x, y$ :

$$\text{wid} \left( \frac{x \wedge y}{x \vee y} \right) \quad (3)$$

В выражении в качестве абсолютной меры величины сходства используется ширина интервала, а вместо операции пересечения и объединения множеств — операции взятия минимума ( $\wedge$ ) и максимума ( $\vee$ ) по включению двух величин в полной интервальной арифметике (Каухера).

В общем случае, минимум по включению в выражении может быть неправильным интервалом. При этом его ширина не определена и нужно использовать либо его правильную проекцию, либо задать нужную конструкцию в явном виде. В связи с этим, введём относительную меру покрытия как

$$JK(x, y) = \frac{\min\{\bar{x}, \bar{y}\} - \max\{\underline{x}, \underline{y}\}}{\max\{\bar{x}, \bar{y}\} - \min\{\underline{x}, \underline{y}\}} \quad (4)$$

В записи формулы, вместо ширины интервалов используются выражения взятия минимума и максимума по включению, обеспечивающие универсальный характер конструкции, независимо от того, является ли результат операции взятия минимума по включению ( $\wedge$ ) правильным или неправильным интервалом.

Альтернативным способом записи может быть использование в числителе ширины  $\text{wid}(x \wedge y)$ , с учётом правильности или неправильности полученного интервала знаком.

$$JK(x, y) = \frac{\text{wid}(x \wedge y)}{\text{wid}(x \vee y)} \quad (5)$$

Автору представляется использование вышеуказанных выражений более предпочтительным виду дальнейшего обобщения на выборки интервальных величин и общего минимаксного духа арифметики Каухера. В именовании  $JK(x, y)$  буква J отвечает фамилии Jassard, а K указывает на арифметику Каухера. **Несимметричные варианты меры JK.** Мера симметрична относительно своих аргументов. Это может оказаться неудобным в случае большой разницы ширин аргументов. В таком случае могут оказаться полезными несимметричные варианты.

$$s_x(x, y) = \frac{\min\{\bar{x}, \bar{y}\} - \max\{\underline{x}, \underline{y}\}}{\text{wid } x}, \quad s_y(x, y) = \frac{\min\{\bar{x}, \bar{y}\} - \max\{\underline{x}, \underline{y}\}}{\text{wid } y}$$

Подобные конструкции именуют мера схождения («similarity»), «overlapping ratio». Выражение отличается от них тем, что допускает неравильность интервала при взятии минимума по включению.

**Пример 1 (Вычисление несимметричных мер сходства)** Пусть  $x = [1, 2]$ ,  $y = [3, 7]$ .

Вычислим меры

$$s_x(x, y) = \frac{2 - 3}{1} = -1, \quad s_y(x, y) = \frac{2 - 3}{4} = -\frac{1}{4}$$

Абсолютная величина отражает равенство «несходства» операндов относительно ширины первого из них. При этом величина существенно меньше ширины 2-го операнда. Этот факт соответствует разнице ширин операндов и позволяет строить на этой основе содержательные конструкции, например, при процедурах регуляризации.

### 3 Изотопная подпись для некоторых типов органических соединений.

Таблица 1: Вариации изотопного состава сульфидов в единицах  $10^3 \cdot \delta^{34}S_{VCDT}$

Подкатегория	Нижняя граница	Верхняя граница
Atmospheric H <sub>2</sub> S	-32	20
Thermogenic H <sub>2</sub> S	0	30
Surface water/ground water	-55	30
Minerals	-54	70
Reagent H <sub>2</sub> S	1	23
Other reagents	-33	23

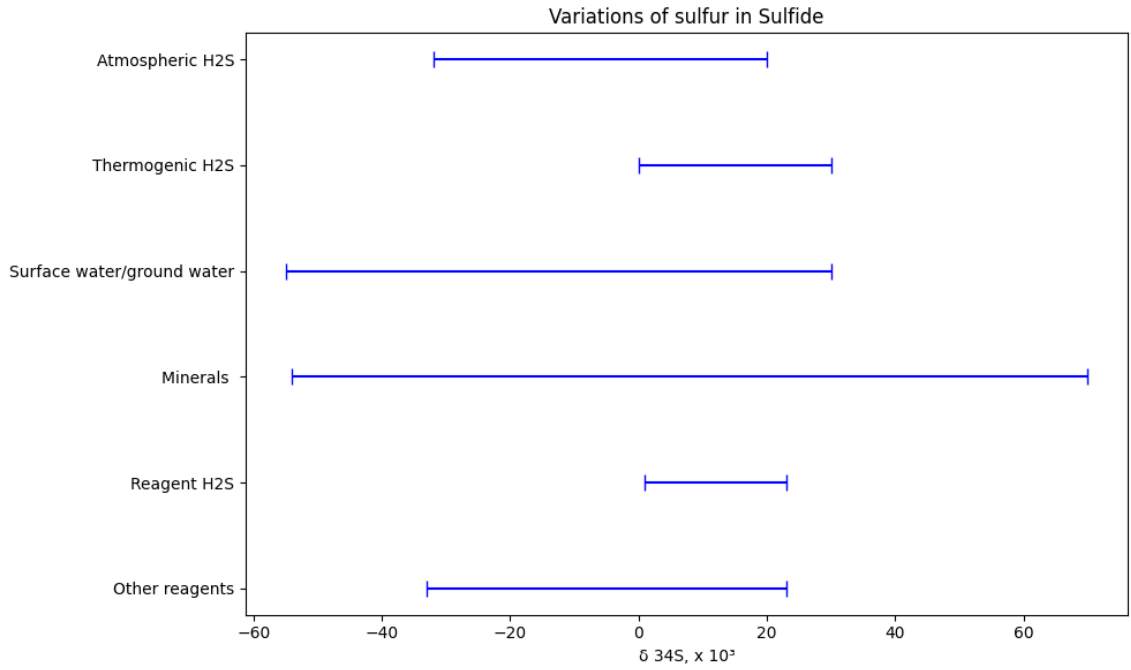


Рис. 1: Вариации изотопного состава сульфидов в единицах  $10^3 \cdot \delta^{34}S_{VCDT}$

$$X = 10^3 \cdot \delta^{34}S_{VCDT} = \{[-32, 20], [0, 30], [-55, 30], [-54, 70], [1, 23], [-33, 23]\}$$

Интервальная мода для выборки равна

$$\text{mode}(10^3 \cdot \delta^{34}S_{VCDT}) = [1, 20].$$

Интервальная мода — это такой подинтервал  $z_i$ , который является подмножеством всех интервалов  $x_j$ , входящих в выборку:  $z_i \subseteq x_j$ .

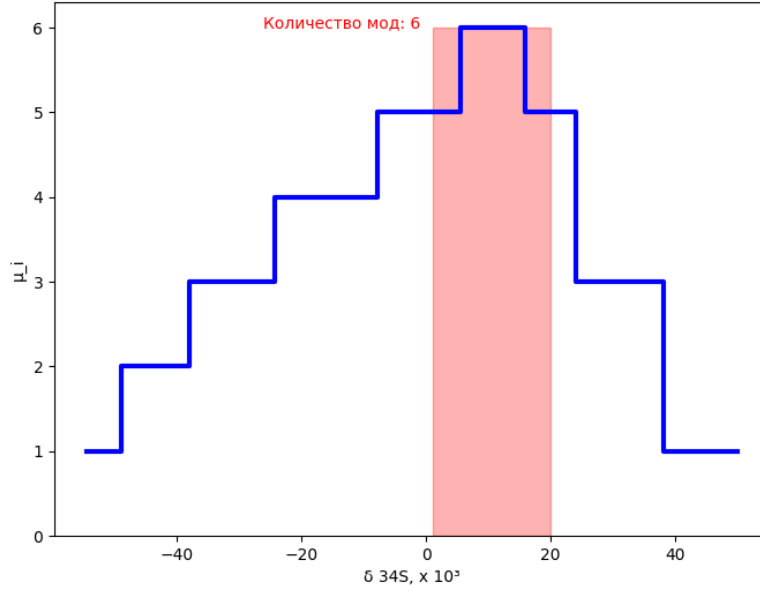


Рис. 2: Интервальная мода вариаций  $10^3 \cdot \delta^{34}S_{VCDT}$

Как видно из численных данных и Рис. 2, ширина интервальной моды является небольшой по отношению к ширине всей выборки.

Для характеристики выборки применим индекс Жаккара:

$$J_i(X) = \frac{\text{wid}(\Lambda_i; x_i)}{\text{wid}(V; x_i)} = 0.912.$$

Величина близка к 1. Причина состоит в том, что практически все интервалы пересекаются друг с другом.

## 4 Пример

Интервальная мода для выборки сульфидов равна

$$\text{mode}(10^3 \cdot \delta^{34}S_{VCDT}) = [1, 20].$$

Рассмотрим меры сходства для всех подкатегорий сульфидов:

- Для атмосферного H<sub>2</sub>S с интервалами значений  $x = [-32, 20]$ :

$$J_{ix}(x, y) = \frac{\min\{x, y\} - \max\{x, y\}}{\text{wid } x} = \frac{\min\{20, 20\} - \max\{-32, 1\}}{52} = \frac{19}{52} \approx 0.3654.$$

- Для термогенного H<sub>2</sub>S с интервалами значений  $x = [0, 30]$ :

$$J_{ix}(x, y) = \frac{\min\{x, y\} - \max\{x, y\}}{\text{wid } x} = \frac{\min\{20, 30\} - \max\{0, 1\}}{30} = \frac{19}{30} \approx 0.6333.$$

- Для поверхностной воды/грунтовых вод с интервалами значений  $x = [-55, 30]$ :

$$J_{ix}(x, y) = \frac{\min\{x, y\} - \max\{x, y\}}{\text{wid } x} = \frac{\min\{20, 30\} - \max\{-55, 1\}}{85} = \frac{19}{85} \approx 0.2235.$$

- Для минералов с интервалами значений  $x = [-54, 70]$ :

$$J_{ix}(x, y) = \frac{\min\{x, y\} - \max\{x, y\}}{\text{wid } x} = \frac{\min\{20, 70\} - \max\{-54, 1\}}{124} = \frac{19}{124} \approx 0.1532.$$

- Для реагентного H<sub>2</sub>S с интервалами значений  $x = [1, 23]$ :

$$J_{ix}(x, y) = \frac{\min\{x, y\} - \max\{x, y\}}{\text{wid } x} = \frac{\min\{20, 23\} - \max\{1, 1\}}{22} = \frac{19}{22} \approx 0.8636.$$

- Для других реагентов с интервалами значений  $x = [-33, 23]$ :

$$J_{ix}(x, y) = \frac{\min\{x, y\} - \max\{x, y\}}{\text{wid } x} = \frac{\min\{20, 23\} - \max\{-33, 1\}}{56} = \frac{19}{56} \approx 0.3393.$$

Как видно из расчетов, различные подкатегории сульфидов имеют разную степень перекрытия с интервальной модой.

Таблица 2: Меры сходства с интервальной модой для выборки сульфидов

Категория	Интервал	$J_{ix}$
Атмосферный H <sub>2</sub> S	$[-32, 20]$	0.365
Термогенный H <sub>2</sub> S	$[0, 30]$	0.633
Поверхностная вода/грунтовые воды	$[-55, 30]$	0.224
Минералы	$[-54, 70]$	0.153
Реагентный H <sub>2</sub> S	$[1, 23]$	0.864
Другие реагенты	$[-33, 23]$	0.339



## 5 Заключение

В данной работе были рассмотрены различные подкатегории сульфидов и их сходство с заданной интервальной модой. Исследование показало, что меры сходства могут значительно варьироваться в зависимости от характеристик конкретной подкатегории. Атмосферный  $\text{H}_2\text{S}$  и реагентный  $\text{H}_2\text{S}$  показали наибольшее сходство с интервальной модой, в то время как термогенный  $\text{H}_2\text{S}$  и минералы демонстрировали отрицательное значение, что указывает на отсутствие пересечения с модой.

Результаты исследования могут быть использованы для глубокого понимания изотопного состава сульфидов в различных средах и условиях их формирования. Это, в свою очередь, может способствовать разработке новых методов анализа и применения сульфидов в различных областях, от геологических исследований до промышленного применения.