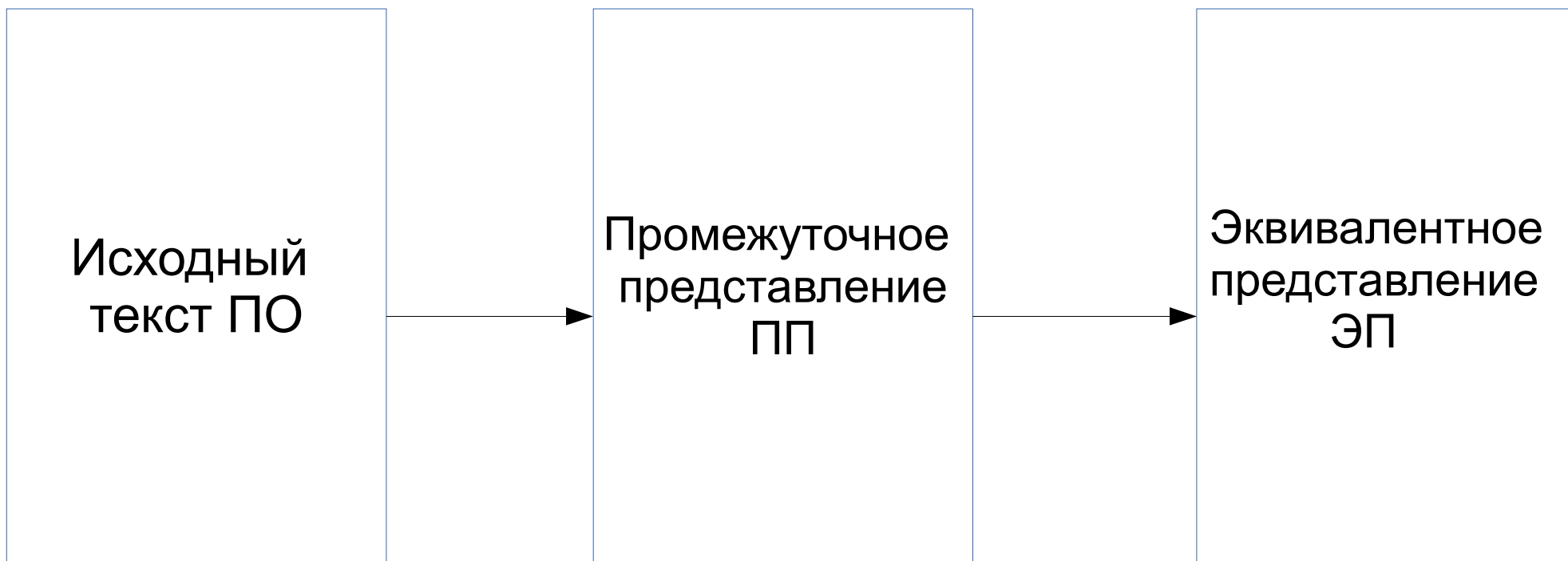


Экспериментальное изучение
количественных
закономерностей при анализе
исходных текстов ПО с открытым
кодом по эквивалентным
представлениям

Пустыгин А.Н., Зубов М.В., Старцев Е.В.
Челябинск

Челябинский государственный университет

Схема извлечения информации из ИСХОДНЫХ ТЕКСТОВ



Разработанные прототипы

1. Анализатор использования «жадных» методов,
2. Анализатор положения создаваемых объектов,
3. Построитель срезов представления потока управления,
4. Анализатор недостижимых функциональных блоков,
5. Построитель срезов диаграмм классов,
6. Построитель трасс между заданными функциональными блоками.

Тестовые проекты для исследования

Java-проекты:

1. java-генераторы представлений UCR,
2. (jclassgen) и UCFR, (jcfgen),
3. javac из пакета OpenJDK.

Python-проекты:

1. logilab,
2. pylint,
3. numpy,
4. Twisted,
5. bazaar[4].

***Нарушение гладкости зависимостей
при анализе динамической типизации
по исходному тексту***

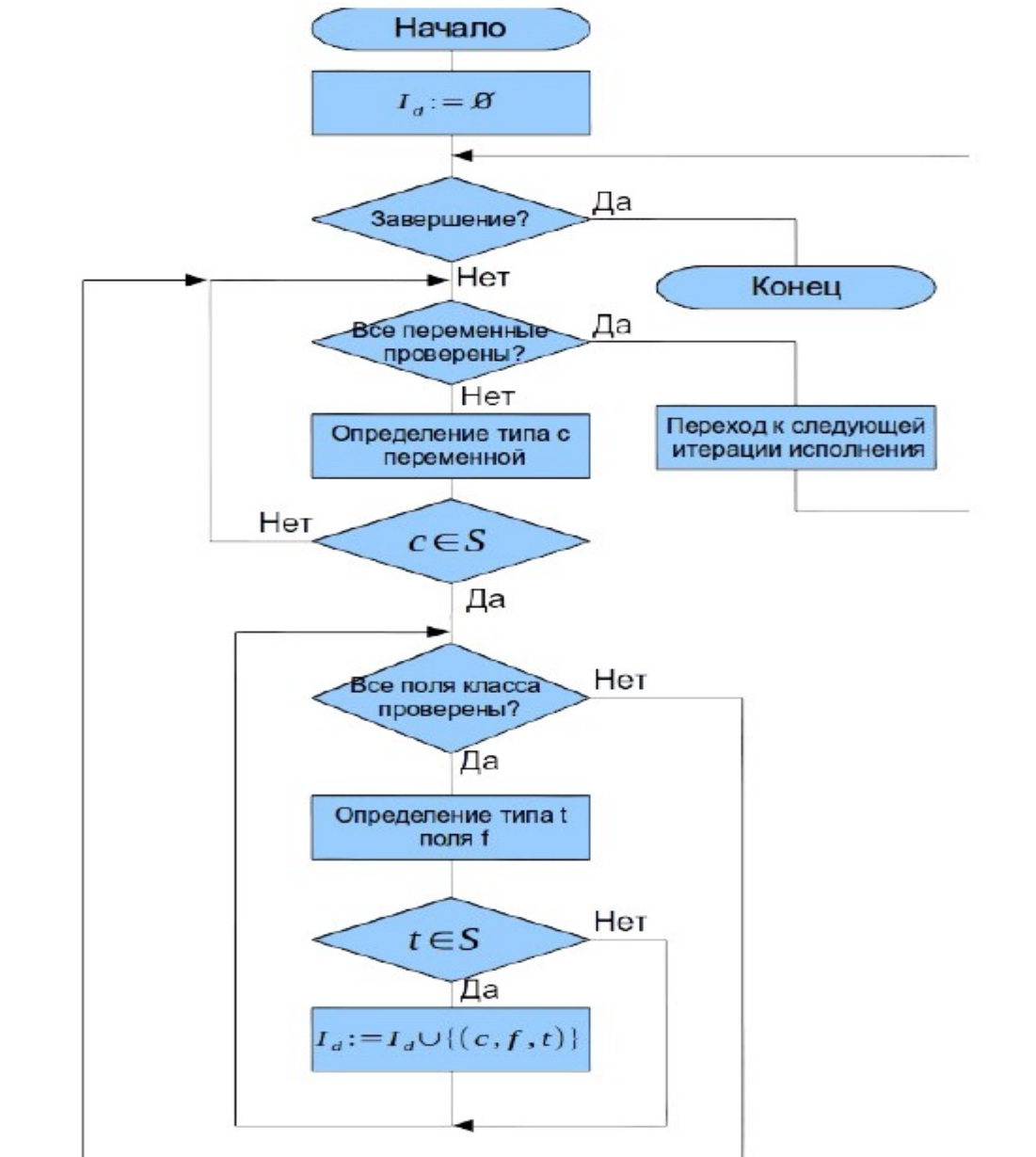
Процедуры, исследовавшиеся в текстах на python

«Утиная типизация» - метод подбора при статическом анализе классов-кандидатов для полей классов, называемых в методике утиными.

Под утиным полем понимается поле класса, об агрегируемом классе которого (объект какого класса хранится в поле) делается предположение.

Агрегация - отношение «часть-целое» между двумя равноправными объектами, когда один объект имеет ссылку на другой объект.

Алгоритм выполнения эксперимента по тестированию утиной типизации



Испытание анализатора путем сравнения с результатами динамического анализа

1. С целью тестирования типизации на этапе статического анализа параллельно выполняется практически стандартная процедура динамического исследования на модельных сценариях исполнения для тестовых проектов.

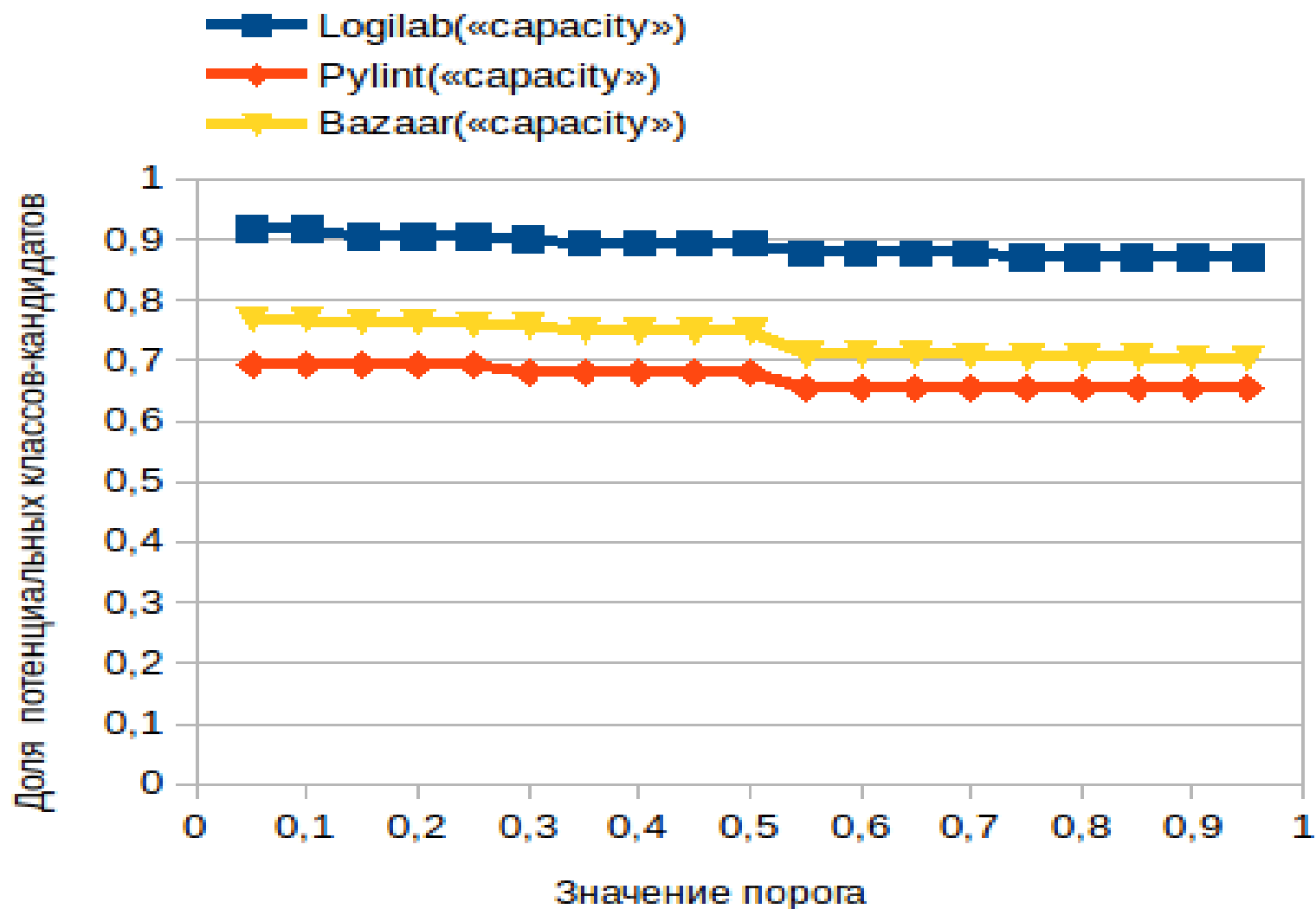
2. Результаты динамического исследования считаются эталонной информацией о типах данных в полях объектов.

3. Далее вычисляется отношение количества типизированных полей по статическому к количеству по динамическому анализу.

Доля классов, подобранных в качестве кандидатов на разных порогах.

Порог	Logilab	Pylint	Bazaar
0,1	0,92	0,69	0,77
0,2	0,90	0,69	0,76
0,3	0,90	0,68	0,76
0,4	0,89	0,68	0,75
0,5	0,89	0,68	0,75
0,6	0,88	0,65	0,71
0,7	0,88	0,65	0,71
0,8	0,87	0,65	0,71
0,9	0,87	0,65	0,70

Зависимость доли типизированных утиных полей от порога



О подходящих значениях порога для использования при подборе классов кандидатов при выполнении утиной типизации.

Значительные изменения результатов срабатывания происходят на участке значений порога от 0.5 до 0.7.

Значения порога менее 0.5 брать не имеет смысла, т. к. на 2 проектах это вообще никак не отражается на результатах

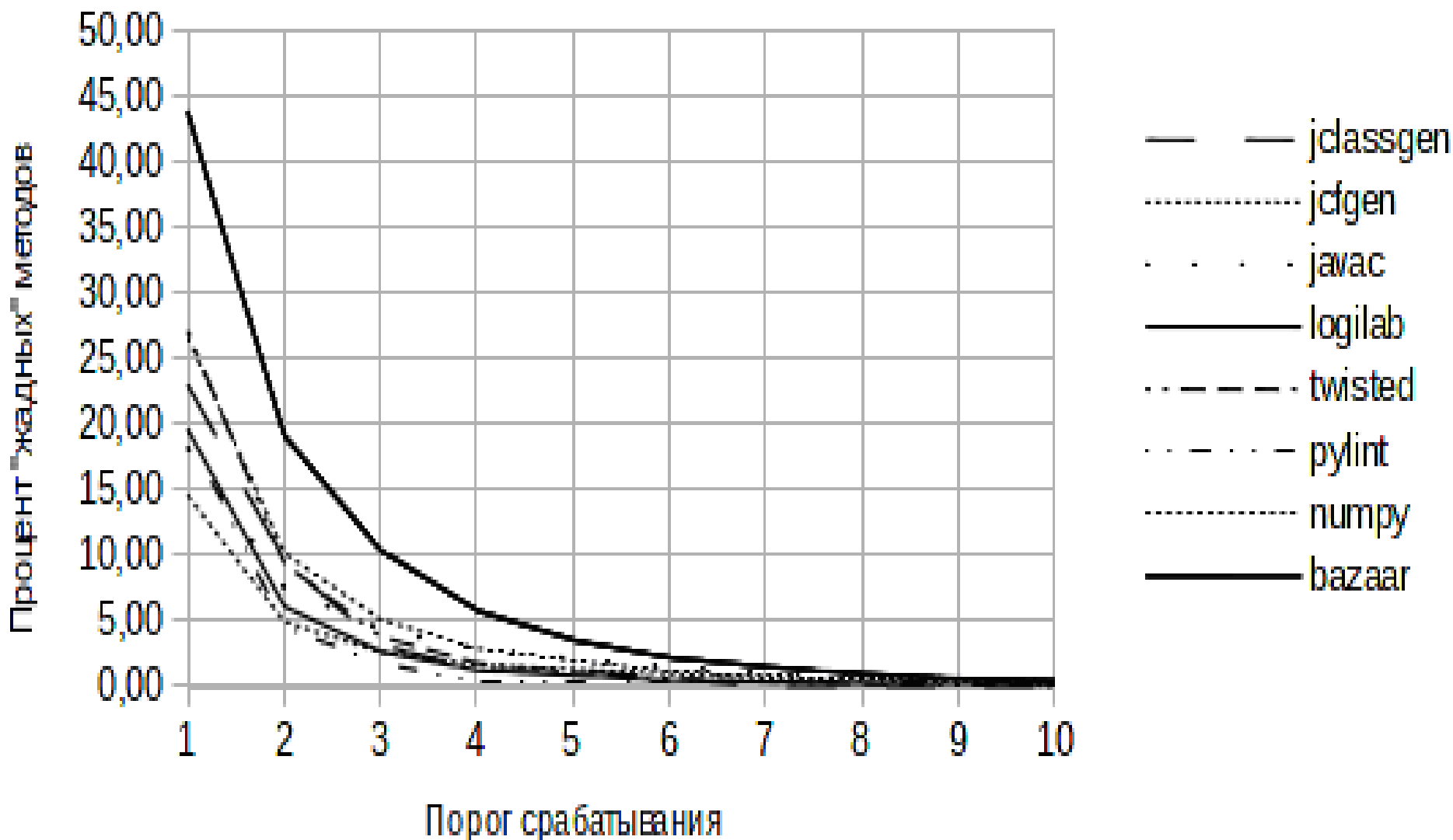
На значениях порога выше 0.7 результаты довольно близки к случаю полного соответствия сигнатуры класса-кандидата и сигнатуры утинового поля.

Нарушение гладкости зависимости при анализе «жадных» методов ПО ИСХОДНОМУ ТЕКСТУ

Жадными считаются функции, которые часто обращаются к методам объекта другого класса. Считается, что следует переносить такие методы в класс, часто их использующий.

Для классификации жадных методов количество обращений к «чужим» методам сравнивают с некоторым пороговым уровнем, устанавливающим границу между «жадным» и допустимым.

Графическая зависимость доли количества жадных методов от общего количества методов для тестовых проектов



Процент найденных методов в зависимости от порога, %

Проект	1	2	3	4	5	6	7	8	9	10
jclassgen	23,08	9,62	3,21	1,28	1,28	0,64	0,00	0,00	0,00	0,00
jcfgen	14,62	4,98	2,66	1,66	1,33	1,00	0,66	0,66	0,66	0,00
pylint	^{19,45}	4,56	1,82	0,30	0,30	0,30	0,30	0,30	0,30	0,30
logilab	19,67	6,14	2,69	1,25	0,86	0,48	0,38	0,29	0,29	0,19
numpy	26,72	10,23	5,15	2,93	1,98	1,19	0,87	0,56	0,40	0,32
javac	18,22	7,57	4,42	2,88	2,13	1,49	1,11	0,87	0,62	0,49
twisted	27,16	9,42	3,84	1,80	1,09	0,62	0,35	0,19	0,14	0,14
bazaar	43,89	19,21	10,44	5,83	3,57	2,21	1,51	1,00	0,63	0,48

Статистика зависимости доли «жадных» методов как функции порога срабатывания

Оптимальным значением является диапазон 5...8. На нем наблюдается практически линейная зависимость доли жадных методов от величины порога.

При значении порога выше 8 наблюдается отсутствие зависимости или же 0 значение для некоторых проектов.

При значениях порога от 1 до 4 наблюдается резкое падение процента искомых функций (практически в 10 раз).

Вывод

Изучение количественных характеристик эквивалентных представлений в процессе анализа исходных текстов обнаруживает нарушение гладкости графиков зависимостей

В координатах «число срабатываний — значение аргумента классификации»