

Эффект биометрического зверинца в алгоритмах локального сопоставления

Андрей Михалев

науч. рук. С.Ю. Сартасов

СПбГУ

Математико-механический факультет
Кафедра системного программирования

13 июня 2017

Доступ к биометрическим системам аутентификации

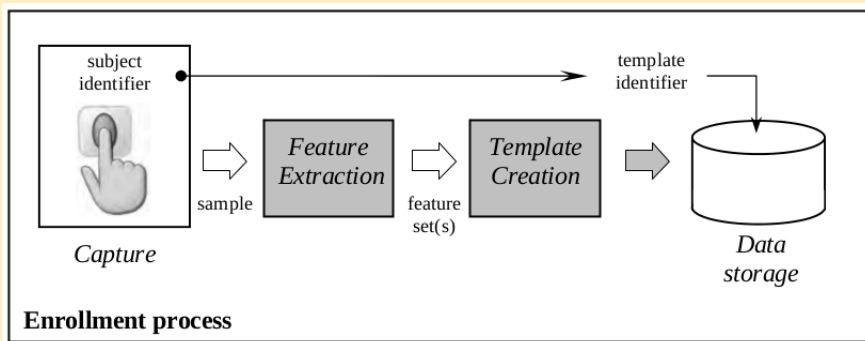
Преимущества отпечатка пальца:

- высокая точность
- невысокая стоимость систем распознавания

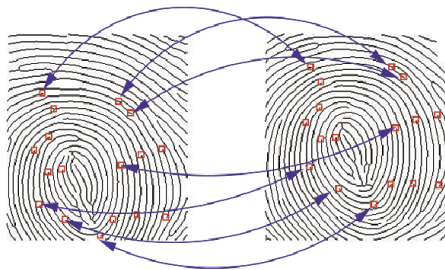
Предметная область

- *Минутция* - особые точки отпечатка пальца
- *Шаблон* - множество минутций отпечатка

Модель биометрической системы доступа



Алгоритмы локального сопоставления



Local matching

Global template matching score

Top local matching scores

1	$S(a, b)$
...	...
n_p	$S(c, d)$

Local Similarity Matrix

	1	2	...	j	...	m
1	$S(1, 1)$	$S(1, 2)$...	$S(1, j)$...	$S(1, m)$
2	$S(2, 1)$	$S(2, 2)$...	$S(2, j)$...	$S(2, m)$
...
i	$S(i, 1)$	$S(i, 2)$...	$S(i, j)$...	$S(i, m)$
...
n	$S(n, 1)$	$S(n, 2)$...	$S(n, j)$...	$S(n, m)$

Виды оценок сопоставления

- *Impostor score* – оценка сопоставления двух шаблонов отпечатков разных пользователей
- *Genuine score* – оценка сопоставления двух шаблонов отпечатков одного и того же пользователя

Термины *genuine score*, *impostor score* применимы к оценкам схожести произвольных локальных структур

Ошибки сопоставления

Ошибки сопоставления для порогового значения t :

- *False Match*
- *False Non-Match*

Вероятности для ошибок сопоставления

- *False Acceptance Rate (FAR)* – вероятность ошибки I рода, т.е. ложного совпадения отпечатков
- *False Rejection Rate (FRR)* – вероятность ошибки II рода, т.е. ложного несовпадения отпечатков

График FAR/FRR

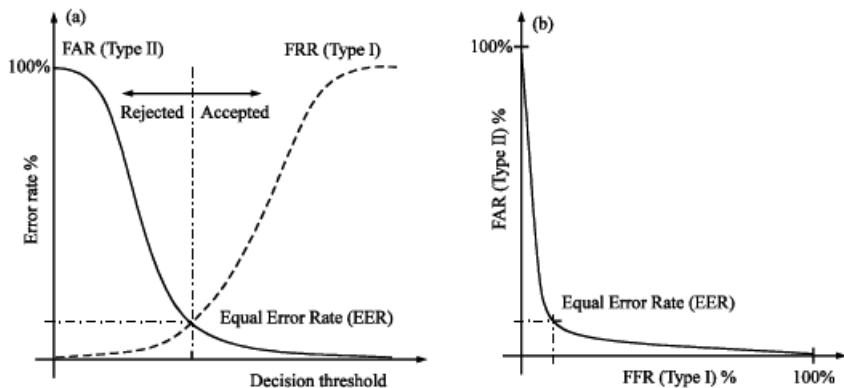


Рис.: Кривые FAR и FRR в зависимости от порогового значения и соответствующая ROC -кривая

Эффект биометрического зверинца

1998 г., Дж. Доддингтон:

- *Biometric menagerie / biometric zoo*
- Классификация пользователей по *genuine* и *impostor* оценкам

2007 и 2010 гг., Н. Ягер и Т. Данстон:

- **Соотношения** между *genuine* и *impostor* оценками
- Расширение классификации Дж. Доддингтона

Эффект биометрического зверинца

Классы Доддингтона:

- *Sheeps* - большинство пользователей системы
- *Goats* - низкие *genuine* оценки
- *Lambs* - высокие *impostor* оценки
- *Wolves* - высокие *impostor* оценки

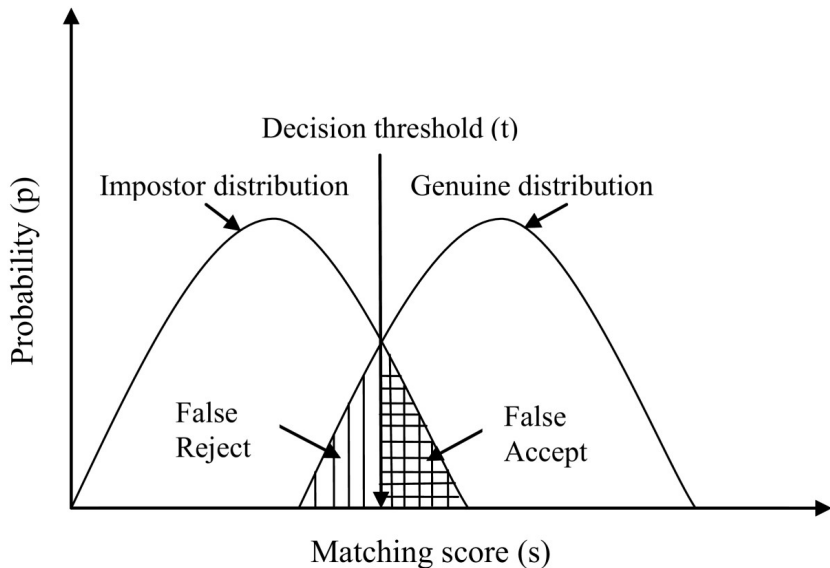
Классы Ягера-Данстона:

- *Chameleons* - $\mathcal{G}_H \cap \mathcal{I}_H$, увеличивают *FAR*
- *Phantoms* - $\mathcal{G}_L \cap \mathcal{I}_L$, увеличивают *FRR*
- *Doves* - $\mathcal{G}_H \cap \mathcal{I}_L$, идеальные пользователи
- *Worms* - $\mathcal{G}_L \cap \mathcal{I}_H$, увеличивают *FAR* или *FRR*

\mathcal{G}_H (\mathcal{G}_L) - высокие (низкие) *genuine* оценки

\mathcal{I}_H (\mathcal{I}_L) - высокие (низкие) *impostor* оценки

Распределение *genuine* и *impostor* оценок



Исследования биометрического зверинца

Исследования

- Н. По: 2008, 2009, 2010, 2012, 2013 и 2015 гг.
- Другие: 2008, 2009, 2011 - 2013, 2015 - 2016 гг.

Ограничение существующих подходов

Внимание уделяется исключительно оценкам сопоставления шаблонов

Вывод

Эффект биометрического зверинца напрямую связан с непропорциональным распределением *FAR* и *FRR* ошибок среди всех пользователей системы

Цель

Цель

Изучение эффекта биометрического зверинца в алгоритмах локального сопоставления и методов борьбы с ним

Вопросы

- 1 Наблюдается ли эффект биометрического зверинца на локальном уровне?
- 2 Лишены ли эффекта биометрического зверинца глобальные оценки, полученные путём объединения оценок сопоставления локальных структур шаблонов?
- 3 Можно ли использовать эффект биометрического зверинца в качестве критерия качества алгоритма сопоставления?

Постановка задачи

Задачи:

- 1 Повторить эксперимент Ягера-Данстона для выявления эффекта биометрического зверинца
- 2 Выявить ограничения подхода Ягера-Данстона
- 3 Поставить эксперимент Ягера-Данстона на локальном уровне дактилоскопических алгоритмов
- 4 Экспериментально подтвердить, что алгоритмы консолидации приносят в систему эффект биометрического зверинца.
- 5 Предложить метод регистрации пользователей, учитывающий эффект биометрического зверинца

Эксперимент Ягера-Данстона и ограничение

Базовый метод:

- 1 Множество оценок сопоставления шаблонов
- 2 Строятся множества *genuine* и *impostor* оценок
- 3 Статистика по каждому шаблону
- 4 Проводится статистический тест

Ограничение базового метода

Выявление эффекта биометрического зверинца на глобальном уровне

Расширение ограничения подхода Ягера-Данстона

Предлагаемый метод

- 1 Множество оценок сопоставления локальных структур шаблонов.
- 2 Переопределить множества *genuine* и *impostor* оценок
- 3 Статистика по каждой локальной структуре каждого шаблона в системе
- 4 Проводится статистический тест

Не меняется

- Методология проведения эксперимента
- Формат входных данных

Эксперимент Ягера-Данстона на локальном уровне

Меняется

- Рассматриваемое множество оценок (множество локальных оценок)
- Концепция *genuine* и *impostor* оценок (сопоставление локальных структур)
- Способ определения множеств *genuine* и *impostor* оценок (разработан алгоритм *параллельных линий*)

Результаты работы алгоритма *параллельных линий*

- Пустое множество *genuine* оценок ($G_{set} = \emptyset$)
- Непустое множество *genuine* оценок ($G_{set} \neq \emptyset$)

Используемые технологии

Базы отпечатков:

- *FVC2000, FVC2002, FVC2004*
- *DB_1, DB_2, DB_3, DB_4*
- 80 изображений отпечатков для 10 разных пальцев

Всего 12 выборок и 960 отпечатков

ПО

- *Griaule Fingerprint SDK*
- *MCC_SDK*

Алгоритмы

- *Feng (minutiae-based версия)*
- *LSS, LSA, LSS-R, LSA-R*

Эксперимент на глобальном уровне

- 1 Рассматриваем множество глобальных оценок сопоставления
- 2 Выделяем множества *genuine* и *impostor* оценки в терминах сопоставления шаблонов
- 3 Выявляем классы биометрического зверинца среди шаблонов пользователей для двух статистик (математического ожидания и медианы)

MCC								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	2	1	2	3	1	2	1	2
Phantoms	0	0	0	0	0	0	0	0
Doves	0	0	0	0	0	0	1	1
Worms	0	0	0	0	0	0	0	0

Таблица: Количество выборок, в которых найден зверинец на глобальном уровне

<i>minutia-based Feng</i>								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	10	8	11	10	7	3	2	4
Phantoms	5	6	7	7	5	6	5	4
Doves	1	1	0	0	1	1	0	0
Worms	0	0	0	0	0	0	0	0

Таблица: Количество выборок, в которых найден зверинец на глобальном уровне

Эксперимент на локальном уровне

- 1 Рассматриваем множество локальных оценок сопоставления
- 2 Выделяем множества *genuine* и *impostor* оценки в терминах сопоставления локальных структур с помощью алгоритма *параллельных линий*
- 3 Выявляем классы биометрического зверинца среди локальных структур шаблонов для двух статистик (математического ожидания и медианы)

Зверинец можно рассматривать:

- с учётом локальных структур с $G_{set} = \emptyset$ (с $G_{set} = \emptyset$)
- без учёта локальных структур с $G_{set} = \emptyset$ (без $G_{set} = \emptyset$)

MCC								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	0	0	0	0	0	0	0	0
Phantoms	12	12	12	12	12	12	12	12
Doves	0	0	0	0	0	0	0	0
Worms	12	12	12	12	12	12	12	12

Таблица: Количество выборок, в которых найден зверинец на локальном уровне с $G_{set} = \emptyset$

<i>minutia-based</i> Feng								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	0	4	0	3	0	6	0	4
Phantoms	12	12	12	12	12	12	12	12
Doves	0	0	0	0	0	0	0	0
Worms	12	12	12	12	12	12	12	12

Таблица: Количество выборок, в которых найден зверинец на локальном уровне с $G_{set} = \emptyset$

MCC								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	1	2	2	4	1	2	3	5
Phantoms	0	0	5	5	0	0	8	10
Doves	6	6	2	2	7	6	2	1
Worms	2	1	1	1	2	1	0	0

Таблица: Количество выборок, в которых найден зверинец на локальном уровне без $G_{set} = \emptyset$

<i>minutia-based</i> Feng								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	10	11	12	11	11	12	12	12
Phantoms	7	10	9	12	8	12	11	12
Doves	0	4	0	6	0	4	0	5
Worms	0	5	1	6	0	3	0	12

Таблица: Количество выборок, в которых найден зверинец на локальном уровне без $G_{set} = \emptyset$

Преимущества локального уровня

Преимущества:

- Найден эффект биометрического зверинца
- Можно отфильтровать *плохие* локальные структуры
- Фильтрация может избавить систему от некоторого класса

Недостатки:

- Алгоритм *параллельных линий*

Итеративный алгоритм фильтрации

Эксперимент Ягера-Данстона на локальном уровне

Основные шаги алгоритма

Пока в выборке найден класс зверинца:

- 1 Рассматриваем список локальных структур
- 2 Ищем первую локальную структуру принадлежащую этому классу
- 3 Если не нашли, то алгоритм заканчивает работу
- 4 Проверяем, что после её удаления в шаблоне остаётся не менее 75% локальных структур
- 5 Если проверка пройдена, то удаляем и повторяем эксперимент Ягера-Данстона для оставшихся локальных структур
- 6 Иначе алгоритм заканчивает работу

MCC								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	1	2	2	4	1	2	3	5
Phantoms	0	0	5	5	0	0	8	10
Doves	6	6	2	2	7	6	2	1
Worms	2	1	1	1	2	1	0	0

Таблица: Количество выборок со зверинцем до фильтрации на локальном уровне без $G_{set} = \emptyset$

<i>minutia-based</i> Feng								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	10	11	12	11	11	12	12	12
Phantoms	7	10	9	12	8	12	11	12
Doves	0	4	0	6	0	4	0	5
Worms	0	5	1	6	0	3	0	12

Таблица: Количество выборок со зверинцем до фильтрации на локальном уровне без $G_{set} = \emptyset$

MCC								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	0	0	0	0	0	0	0	0
Phantoms	0	0	5	5	0	0	8	10
Doves	6	6	2	2	7	6	2	1
Worms	0	0	0	0	0	0	0	0

Таблица: Количество выборок со зверинцем после фильтрации классов *Worms* и *Chameleons* на локальном уровне без $G_{set} = \emptyset$

<i>minutia-based Feng</i>								
Consolidation	LSS		LSS-R		LSA		LSA-R	
Statistic	E-val	Med	E-val	Med	E-val	Med	E-val	Med
Chameleons	3	10	1	7	1	7	0	3
Phantoms	6	10	8	12	6	12	11	12
Doves	0	5	0	7	0	7	0	7
Worms	0	0	0	0	0	0	0	0

Таблица: Количество выборок со зверинцем после фильтрации классов *Worms* и *Chameleons* на локальном уровне без $G_{set} = \emptyset$

Сравнение оригинальных и отфильтрованных данных

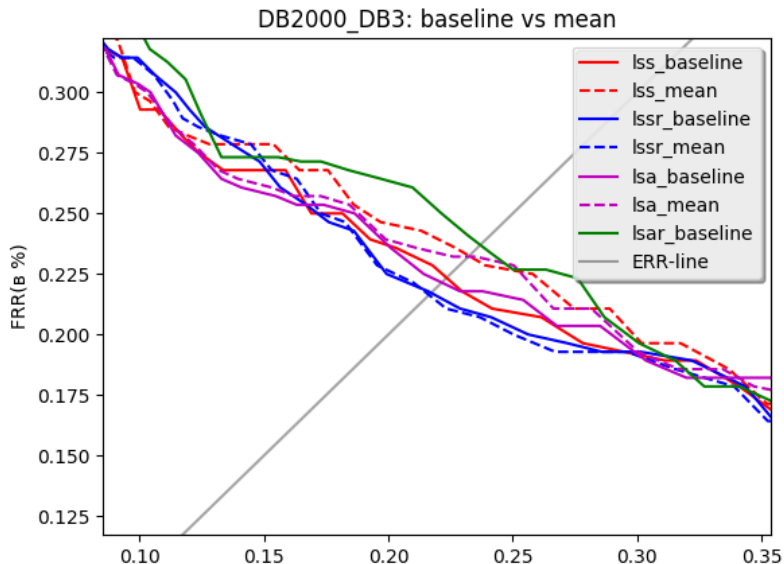
Оригинальные данные (baseline)

Рассматриваются оценки сопоставления шаблонов из некоторой выборки

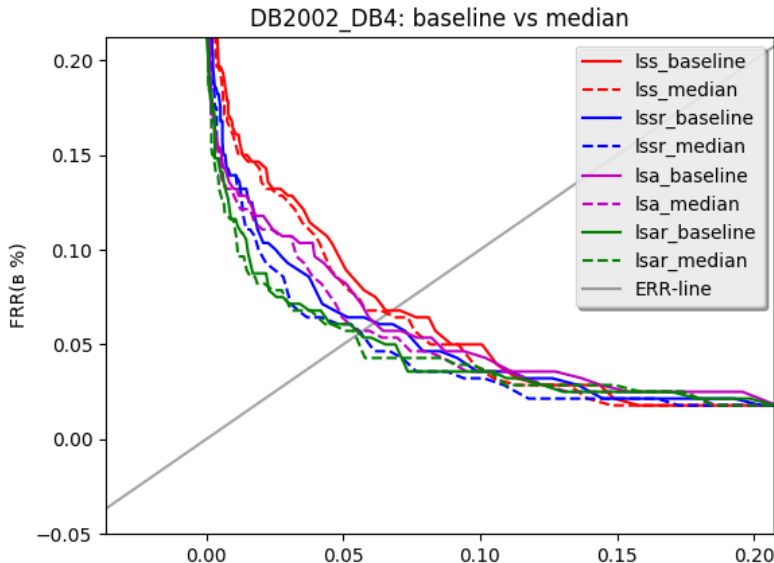
Отфильтрованные данные

- 1 Эксперимент Ягера-Данстона для локальных структур шаблонов исходных выборок
- 2 Выявление эффекта биометрического зверинца на локальном уровне
- 3 Фильтрация плохих локальных структур
- 4 Получение урезанных шаблонов
- 5 Формирование оценок сопоставления урезанных шаблонов

После фильтрации *Worms* без $G_{set} = \emptyset$



После фильтрации *Chameleons* без $G_{set} = \emptyset$



Результаты работы

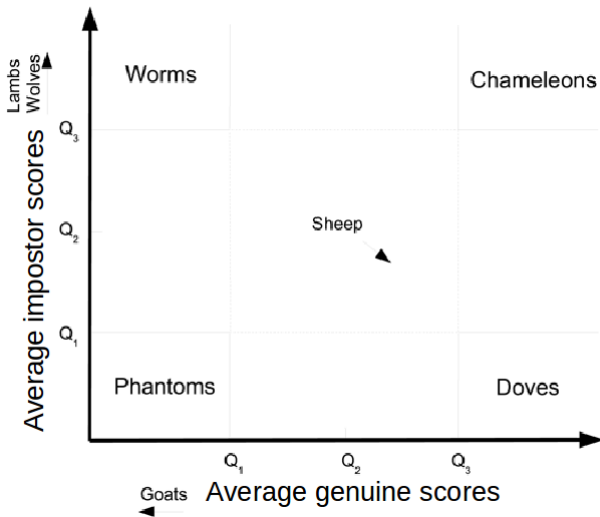
- Проведен эксперимент на глобальном уровне
- Предложено расширение базового метода до локального уровня
- Проведен эксперимент для локальных оценок сопоставления
- Эффект биометрического зверинца выявлен на глобальном и локальном уровнях
- Показано, что алгоритмы консолидации приносят эффект в глобальный уровень
- Предложен метод фильтрации *плохих* локальных структур
- Выступление на конференции СПИСОК-2017

Полученные результаты говорят о том, что:

- Фильтрация может быть применена на промежуточном этапе при регистрации пользователей
- Исследуемый эффект может быть использован в качестве критерия качества алгоритма сопоставления
- Нет однозначного способа борьбы с эффектом биометрического зверинца
- Отрываются возможности для более детального анализа этого эффекта

- Модификация алгоритма параллельных линий
- Рассмотрение других алгоритмов локального сопоставления
- Более детально изучить алгоритмы консолидации
- Изменение критериев фильтрации или другой порядок фильтрации классов биометрического зверинца

Биометрический зверинец на плоскости



Эксперимент Ягера-Данстона

- Множество пользователей системы \mathcal{P}
- Множество оценок \mathcal{S}
- Множество $S(j, k) \subset \mathcal{S}$ для пары пользователей $j, k \in \mathcal{P}$

Определить для каждого k -го пользователя:

- 1 $G_k = S(k, k)$ - множество *genuine* оценок
- 2 $I_k = S(j, k) \cup S(k, j), \forall j \neq k$ - множество *impostor* оценок
- 3 Математическое ожидание: $g_k = \overline{G}_k, i_k = \overline{I}_k$

Определить для всех пользователей системы:

- 1 Упорядоченные по возрастанию множества:

$$\mathcal{G} = \bigcup_{k \in \mathcal{P}} g_k, \mathcal{I} = \bigcup_{k \in \mathcal{P}} i_k$$

- 2 Множества элементов:

- $\mathcal{G}_H \subset \mathcal{P}$ - первые 25% элементов множества \mathcal{G}
- $\mathcal{G}_L \subset \mathcal{P}$ - последние 25% элементов множества \mathcal{G}
- $\mathcal{I}_H \subset \mathcal{P}, \mathcal{I}_L \subset \mathcal{P}$ - аналогично

Эксперимент Ягера-Данстона

Рассмотрим класс *chameleons*: $\mathcal{G}_H \cap \mathcal{I}_H$

- Пусть $c = |C|$ - ожидаемое количество хамелеонов в системе.
- Гипотеза H_0 : вероятность того, что пользователь является хамелеоном, есть $p = 1/16$
- *Эксперимент*:

$$f(c, n, p) = \sum_{i=c}^n \binom{n}{c} p^i (1-p)^{n-i},$$

где $n = |\mathcal{P}|$ - количество испытаний в эксперименте

- H_0 отклоняется, если $f(c, n, p) < \alpha$, где $\alpha = 0.05$, тогда возможны два случая:
 - хамелеонов очень мало
 - хамелеонов очень много

Локальный уровень

Алгоритм локального сопоставления

- Входные данные: два шаблона A и B
- Результат: *Local Similarity Matrix (LSM)* $\Gamma_{A,B}$

Матрица локальных схожестей $\Gamma_{j,k}$

- Множество *локальных оценок*
- Содержит оценки сопоставления локальных структур из шаблонов j -го и k -го пользователей
- Аналог множества $S(j,k)$ на локальном уровне

Эксперимент Ягера-Данстона на локальном уровне

Для i -й локальной структуры шаблона A :

$$G_i = \bigcup_{k=1}^m \bigcup_{l=1}^{n_k} \Gamma_{\{A,k\}}(i,l) \mid \Gamma_{\{A,k\}}(i,l) - \textit{genuine},$$

$$I_i = \bigcup_{k=1}^m \bigcup_{l=1}^{n_k} \Gamma_{\{A,k\}}(i,l) \mid \Gamma_{\{A,k\}}(i,l) - \textit{not genuine}, A \neq k$$

где:

- m - количество сопоставлений шаблона A с другими шаблонами
- n_k - количество локальных локальных структур в k -м шаблоне
- $\Gamma_{\{A,k\}}(i,l)$ - оценка сопоставления i -й локальной структуры шаблона A с l -й локальной структурой шаблона k