Дипломна робота

Виконав: студент групи КА-35 Костюк В.І.

Керівник: Данилов В.Я.

22 червня 2017 р.

Тема: Визначення надійності функціонування об'єктів на основі штучних імунних мереж

- Об’єкт дослідження: система розпізнавання комп’ютерних вірусів.
- Предмет дослідження: штучні імунні системи, алгоритм негативного відбору для штучних імунних мереж.
- Мета: дослідити застосування імунних мереж для розв’язування задач комп’ютерної безпеки.
- Методи: алгоритм негативного відбору для штучних імунних мереж, сигнатурний метод сканування, еврістичний метод.
- Актуальність проекту мотивується швидкими темпами появи нових загроз комп’ютерам. Антивірусні програми можуть не встигати оновлювати бази даних, що в свою чергу може фатально відобразитись на надійності функціонування. Тому знаходження нових небезпек еврістичними методами за допомогою штучних імунних мереж може значно покращити ситуацію.

Постановка задачі

- Необхідно розробити алгоритм, який дозволяє ідентифікувати об’єкти, що можуть завадити функціонуванню системи, по наявній базі даних, а також дозволяє навчати систему та розпізнавати відсутні у наявній базі об’єкти.

Нехай X — множина описів об'єктів, Y —множина номерів (чи назв) класів. В нашому випадку лише два класи – «свій» та «чужий».

Існує відображення відображення y: $X \to Y$, значення якого відомі лише на елементах кінцевої навчальної вибірки $X^m = \{(x_1, y_1), \ldots, (x_m, y_m)\}$.

Потрібно побудувати алгоритм y: $X \to Y$, здатний класифікувати довільний об'єкт $x \in X$.

Огляд сигнатурного методу сканування

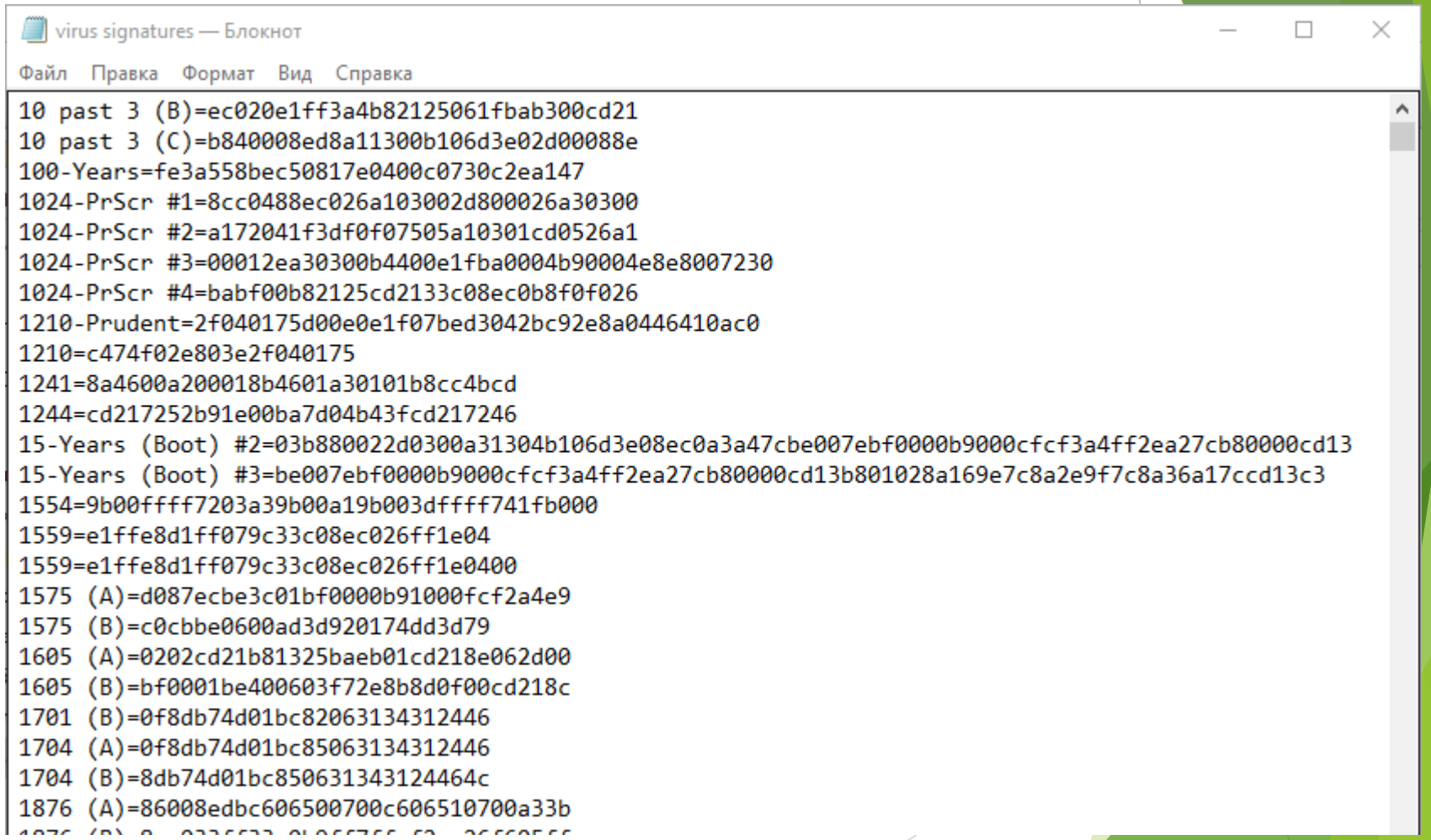
Сигнатури - класичний спосіб, що дозволяє виявляти і однозначно розпізнавати більшість вірусів. Сигнатура це - унікальний набір характеристик, однозначно характеризують якийсь об’єкт. Цей набір характеристик є меншою ніж загальної кількості характеристик, притаманних об'єкту. Це варіант хеш-функції, тобто неоднозначного відображення великого безлічі об'єктів на маленьке безліч комбінацій їх ознак. Набори «непрямих» ознак (тип файлу, розмір, час створення, взаємне розташування різних частин програми і т.п.) іноді називають «слабкими сигнатурами». Сигнатура повинна бути унікальною і характерною для вірусу, тобто такий, щоб порівнявши її з фрагментами програми, можна було однозначно сказати: це конкретний вірус чи ні. Для більшості вірусів вид сигнатури дуже простий - це послідовність розташованих один за одним байтів і адреса в файлі для початку цієї послідовності. Так само іноді потрібна довжина, якщо для різних вірусів використовуються різні за сигнатури.

Приклад набору сигнатур

Наведено фрагмент файлу із вхідними сигнатурами.

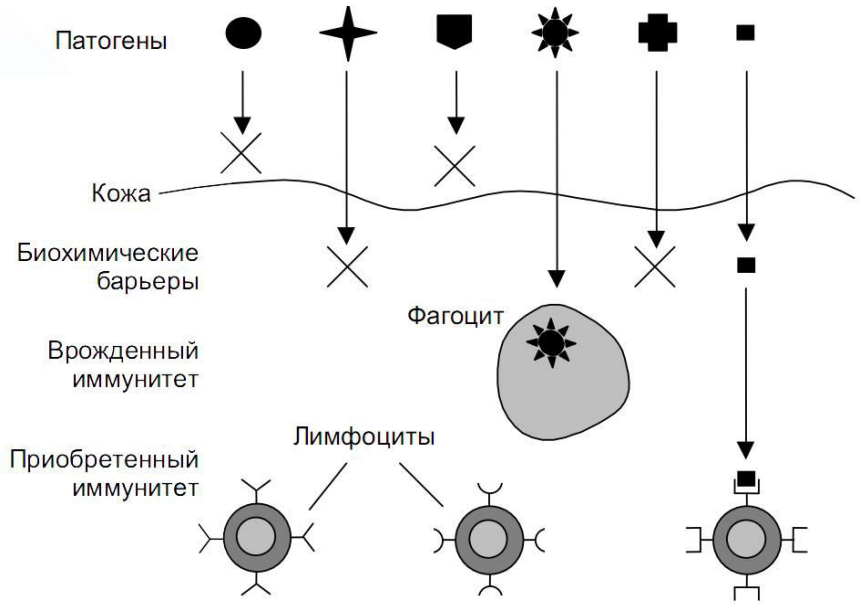
Дані взято за посиланням

http://www.nlnetlabs.nl/downloads/antivirus/antivirus/virussignatures.strings



Поняття імунної системи

Це система, основна функція якої полягає в захисті організму від постійних атак чужорідних мікроорганізмів (їх розпізнавання та знешкодження).

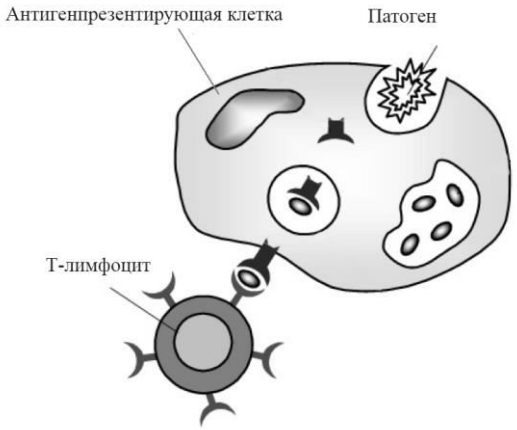


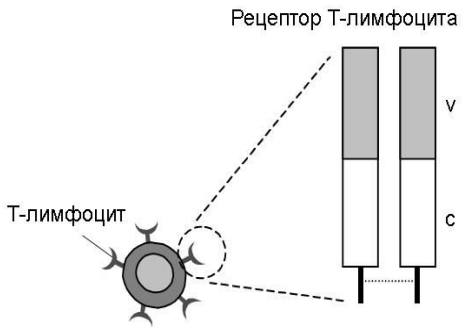
Лімфоцити

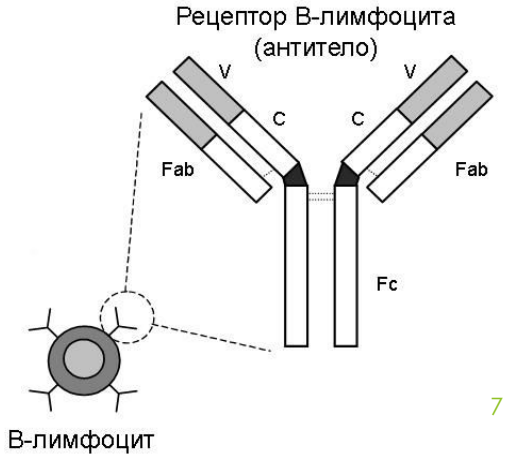
Це клітини імунної системи, які представлені двома великими популяціями Т-лімфоцитами і В-лімфоцитами.

Основна характеристика Т-лімфоцитів - здатність відрізняти клітини свого організму від будь-яких чужих та програмування на знищення аутоагрессивних клітин.

В-лімфоцит генетично запрограмований на синтез поверхневого рецептора специфічного до одного певного антигену.Зустрівши і розпізнавши цей антиген, В-лімфоцити розмножуються і диференціюються в плазматичні клітини, які утворюють велику кількість таких рецепторних молекул, званих антитілами.





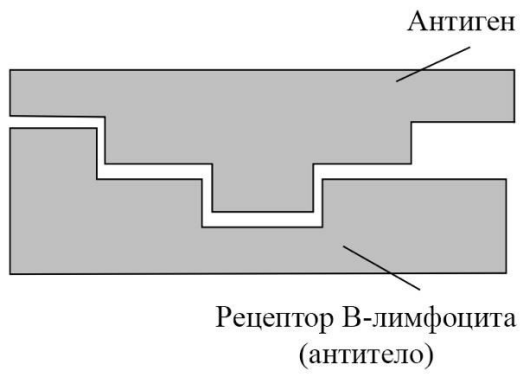


Антитіло, антиген, афінність

Антитіла - сполуки, які організм хребетних тварин виробляє у відповідь на антигени, чужорідні речовини. Вони розпізнають їх, після чого відбуваєється нейтралізація.

Антигени - молекули патогена, які ініціюють реакцію набутого імунітету (імунна відповідь).

Афінність описує взаємодію між антитілом і антигеном, це відображення міцності хімічного зв'язку молекули антитіла з окремим епітопом антигену. Чим більше антитіло специфічно до даного антигену, тим міцніше буде хімічний зв'язок і тим вище буде афінність.



Штучна імунна система

Штучна імунна система (ШІС) - це алгоритм навчання, що ґрунтується на принципах функціонування природної імунної системи. Штучна імунна система об’єднує *апріорне* знання з адаптивними можливостями біологічної імунної системи що створює могутню альтернативу наявним сьогодні методам розпізнавання образів, навчання, пошуку об’єктів, що можуть завадити оптимізації.

Штучна імунна система наслідує поняття природньої імунної системи, що були пояснені на попередніх слайдах.

Антиген являє собою задачу, яку необхідно вирішити, а антитіло - можливе рішення задачі. Популяція антитіл являє собою безліч можливих рішень задачі.

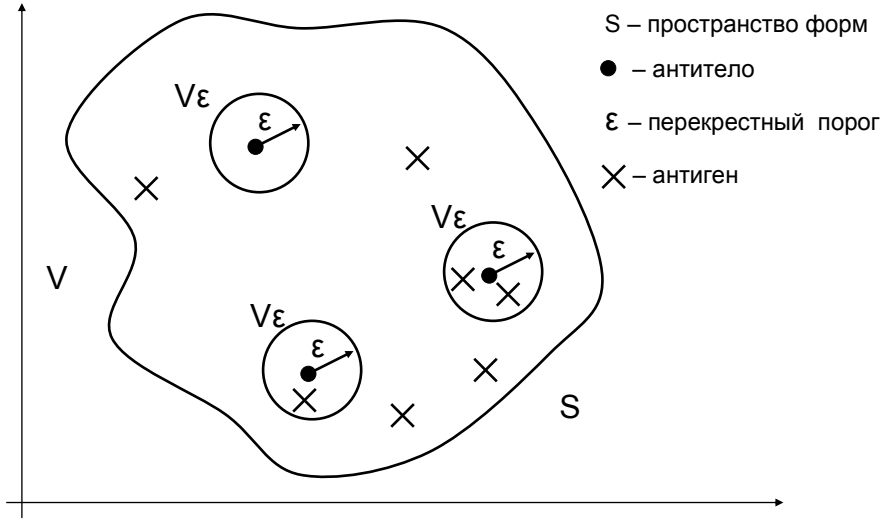
Математично узагальнена форма антитіла або антигену, може бути представлена у вигляді вектора атрибутів (набору координат), які можуть розглядатися як точка в L-вимірному матеріальному просторі форм:

$$Ab = \langle Ab_1, Ab_2, \dots, Ab_L \rangle;$$

$$Ag = \langle Ag_1, Ag_2, \dots, Ag_L \rangle.$$

Простір форм - це L-мірний простір, в якому вимірювання відповідають набору атрибутів, залучених у взаємодію антитіл і антигенів, що дозволяє представляти антитіла і антигени точками в цьому просторі. Як правило, антитіло і антиген мають одну і ту ж довжину L.

Афінність - характеристика, кількісно описує силу взаємодії антигену і антитіла.



Символьний простір форм.

Вектор атрибутів в символьному просторі форм складається з різних типів атрибутів, серед яких хоча б один є символьним.

Афінність між антитілом і антигеном в символьному просторі форм обчислюється аналогічно хеммінгову простору форм.

	Описание	Дата	Номер рейса	Страна	Пункт отправ.	Пункт приб.	Цена
Ab1	Business	2011	212	Brazil	Campinas	Greece	546
Ab2	Holiday	2010	312	U.K.	London	Paris	102
Ag	Holiday	2010	212	U.K.	London	Greece	546
Ag-Ab1	1	1	0	1	1	0	0
Ag-Ab2	0	0	1	0	0	1	1

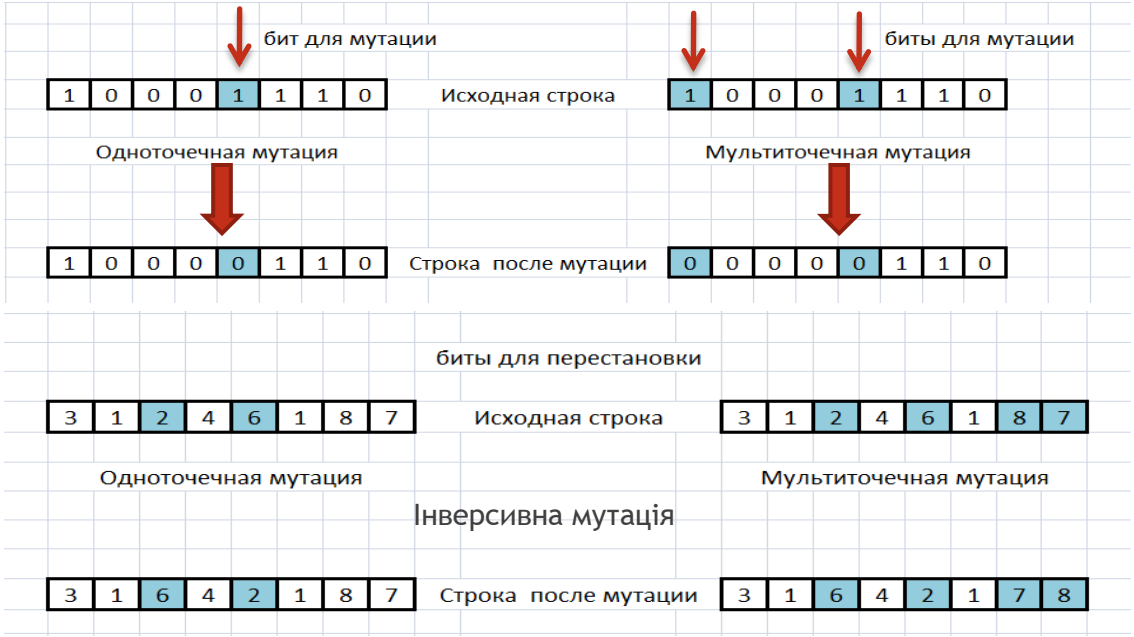
Оператор мутації

Оператор мутації призначений для внесення змін в рядок атрибутів, що представляє антитіло.

Мутація має дві важливі функції:

- внесення різноманітності в популяцію антитіл;

- поліпшення афінності відібраних антитіл.



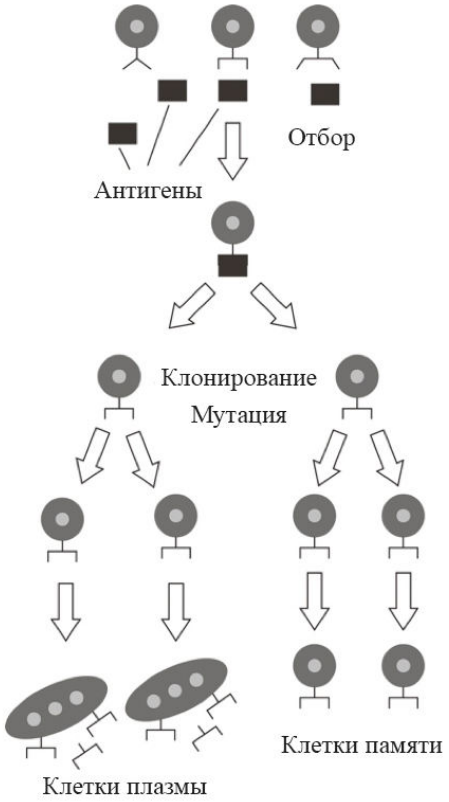
Індуктивна мутація:

$$Ab' = Ab + \alpha \cdot N(0, \sigma)$$

Моделі штучних імунних систем

Обчислювальні моделі ШІС розроблені на основі спостережуваних властивостей природної імунної системи і включають в себе найбільш перспективні з точки зору обчислень особливості. Найбільш часто в літературі використовуються такі обчислювальні моделі:

- модель негативного відбору
- модель позитивного відбору
- модель клонального відбору
- модель імунної мережі



Модель клонального відбору

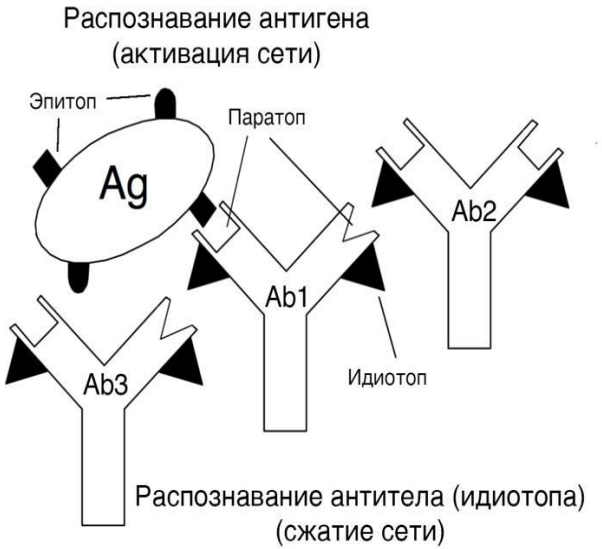
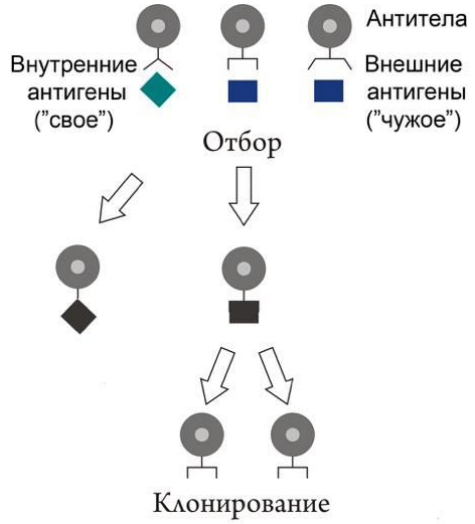


Схема взаємодії в імунній мережі



Розпізнавання свій/чужий при негативному і позитивному відборі

Негативний відбір

У штучному негативному відборі основними використовуваними елементами, які виконують розпізнавання "свого" і "чужого", є детектори (аналоги Т-лімфоцитів в ЄІС) - бітові рядки фіксованої довжини L.

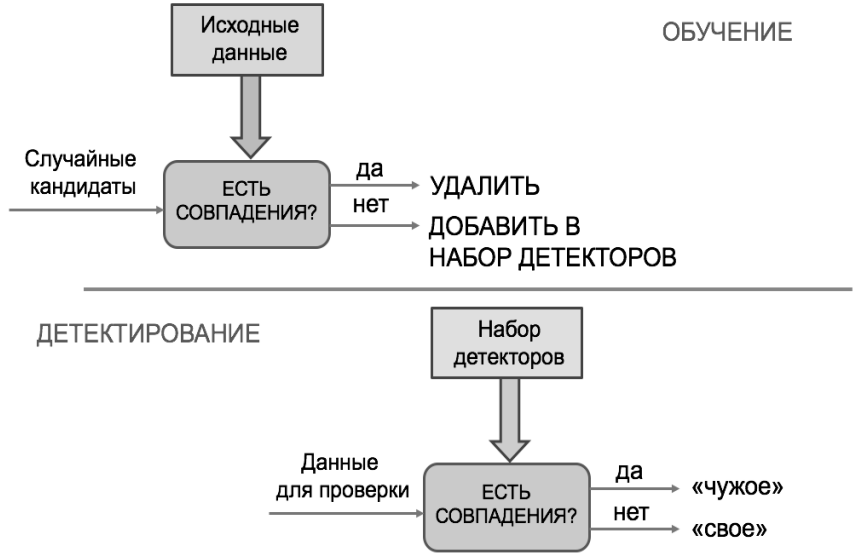
Випадковим чином генеруються рядки - кандидати в детектори. Ці випадкові рядки проходять етап відбору шляхом порівняння зі "своїм". Ті випадкові рядки, які збігаються зі "своїм", видаляються, інші стають детекторами

На етапі детектування набір детекторів використовується для перевірки, чи є вхідні дані "своїм" або "чужим". Якщо вони збігаються з одним з детекторів, вони оголошуються "чужим" або аномалією.

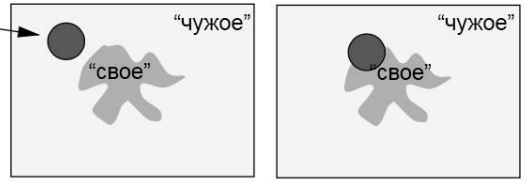
1. Визначити своє як сукупність S рядків довжини L над кінцевим алфавітом, яку необхідно захищати або контролювати.

2. Утворити набір детекторів D, кожен з яких не повинен відповідати будь-якому рядку в S. Як правила зіставлення двох бінарних рядків використовується правило r безперервних біт (два рядки співпадають, якщо існує таке вікно розміром r, в межах якого всі біти обох рядків збігаються).

3. Перевірити S на предмет змін шляхом безперервного порівняння детекторів з D з елементами S. Якщо хоча б один з детекторів виявиться відповідним, значить, відбулася зміна, оскільки детектори по визначенню відібрані так, щоб не відповідати будь-якому рядку з S.

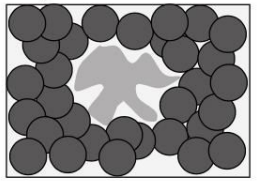


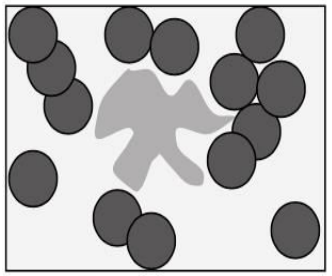
Случайные строки



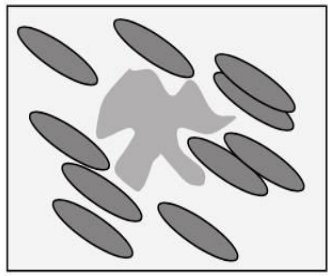
ДОБАВИТЬ

УДАЛИТЬ

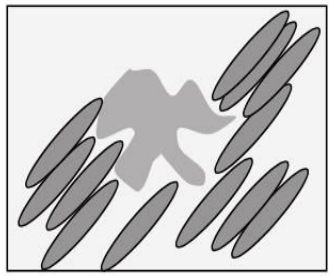




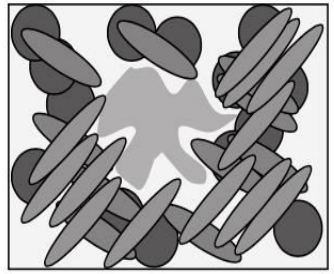
набор локальных детекторов 1



набор локальных детекторов 2

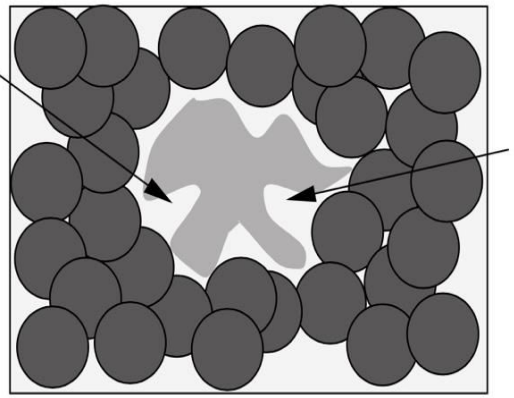


набор локальных детекторов 3



детектирование с использованием всех наборов детекторов

дыры



дыры

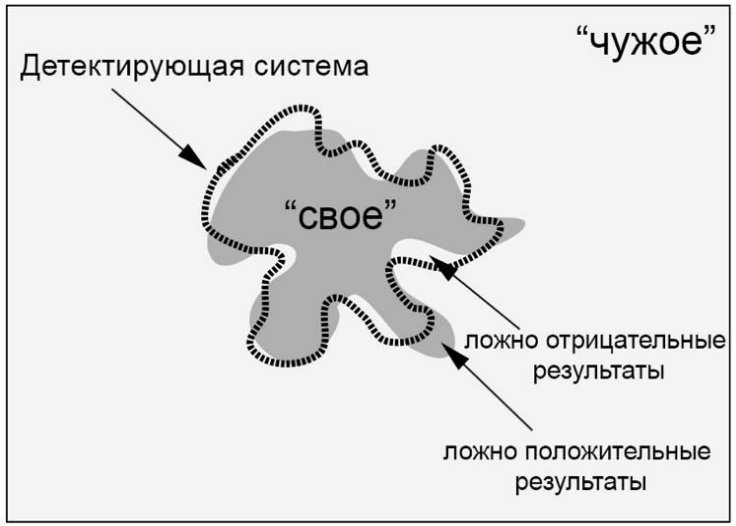
Недоліки

Хоча метою є розділити два класи, реально зразки тільки одного класу ("своє") доступні для навчання системи.

Негативний відбір буває неефективним і неминучі помилково позитивні результати

Межі "свій" - "чужий" нечіткі, тому що найчастіше "своє" і "чуже" поділяють спільні області.

"Своє" з часом змінюється, тому очікуються проблеми з клітинами пам'яті, які з часом стануть неефективними.



Псевдокод

```
input   : S = set of patterns to be recognised, n the number of worst elements to select for removal
output  : M = set of memory detectors capable of classifying unseen patterns
begin
     Create an initial random set of antibodies, A

  forall  patterns in S   do
     Determine the affinity with each antibody in A
     Generate clones of a subset of the antibodies in A  with the highest affinity.
        The number of clones for an antibody is proportional to its affinity
     Mutate attributes of these clones to the set A , and place a copy of the highest
        affinity antibodies in A  into the memory set, M
     Replace the n lowest affinity antibodies in A with new randomly generated antibodies
  end

end
```

Робота програмного продукту

```
"C:\Program Files\Java\jdk1.8.0_121\bin\java"
Results:
(Real = 0, Researched = 0) = 37852
(Real = 1, Researched = 0) = 14
(Real = 0, Researched = 1) = 33
(Real = 1, Researched = 1) = 56
```

Спрогнозовані значення\Істинні значення	0	1
0	37852	14
1	33	56

- Похибка: 0,00124

Висновки

В роботі розглянуто використання штучних імунних систем в задачах визначення надійності функціонування об’єктів. За поданими сигнатурами було проведено сканування сигнатур та виведено статистику, що відображає ефективність та правильність роботи методу.

Рекомендації щодо подальших досліджень

В якості подальших дослідження можуть виступати наступні:

- Застосування нових методів сканування вірусів, не тільки за сигнатурою. Відслідковування поведінки об’єктів та контроль доступу.
- Вдосконалення точності та швидкості роботи програми підбиранням відстані, за якою розраховується афінність.
- Точна оцінка методу у порівнянні з існуючими інструментами
- Вдосконалення коду програми та створення інтерфейсу, що дозволить проводити повноцінні дослідження

Дякую за увагу!