электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51038.

УДК 621.391

Двумерное распознавание сигнала на основе метода *k* ближайших соседей

Якубов Р.Ж., *студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Информационная безопасность»

Научный руководитель: Троицкий И.И., к.т.н., доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Информационная безопасность»
v.a.matveev@bmstu.ru

Одной из наиболее значимых угроз безопасности информации с ограниченным уровнем доступа можно назвать угрозу утечки данных при их передаче, возникающую вследствие невозможности контроля распространения информативного сигнала при его прохождении через физическую среду, что может привести к относительно простому перехвату данных злоумышленниками.

При перехвате данных актуальной является задача классификации передаваемого сигнала, а именно, определения того, что передается по каналу: «ноль» или «единица». Одним из способов воспрепятствования получения злоумышленниками информации с ограниченным доступом является зашумление каналов передачи данных. Это приводит к усложнению распознавания сигнала.

В данной статье рассматривается возможность распознавания двумерного сигнала методом k ближайших соседей. Критерием эффективности использования данного метода будем считать P^* - вероятность правильного распознавания сигнала, определяемую опытным путем. На основании [1] и [2] представим описание метода k ближайших соседей.

Пусть $\mathcal{T} = \{t^\alpha\}$, $\alpha = \{0,1\} -$ множество, состоящее из двух помеченных выборок. В таком случае, правило классификации входного элемента t заключается в том, что ему присваивается метка α, наиболее часто встречающаяся среди k ближайших к нему элементов множества $\mathcal{T}$.

В случае использования данного метода для распознавания сигнала, передаваемого по 2 каналам, в качестве выборок t_l^α, где $l-$ номер элемента во множестве $\mathcal{T}$, используются вектора вида $t_l^\alpha = \left(t_l^\alpha{}_1, t_l^\alpha{}_2\right)$, а входной элемент $t = \left(t_{l_1}, t_{l_2}\right)$.

Определение расстояния от входного элемента t до остальных элементов множества $\mathcal{T}$ зависит от выбора метрики. В рамках данной работы предлагается несколько метрик:

Евклидово расстояние:

$$d(t_l^\alpha, t) = \sqrt{\sum_{i=1}^{n+1}\left(t_l^\alpha{}_i - t_{l_i}\right)^2} \quad (1)$$

Манхэттенское расстояние:

$$d(t_l^\alpha, t) = \sum_{i=1}^{n+1}\left|t_l^\alpha{}_i - t_{l_i}\right| \quad (2)$$

Расстояние Чебышева:

$$d(t_l^\alpha, t) = \max\left|\left(t_l^\alpha{}_i - t_{l_i}\right)\right| \quad (3)$$

Степенное расстояние:

$$d(t_l^\alpha, t) = \left(\sum_{i=1}^{n+1}\left|t_l^\alpha{}_i - t_{l_i}\right|^p\right)^{1/r} \quad (4)$$

где параметры $p = 2, r = 1$.

Для получения значений P^* будут рассматриваться 2 случая:

1) В первом канале передаются сигнал+шум, во втором передается шум.
2) В двух каналах передается пара: сигнал и шум.

Дадим описание каналов первого случая.

$$y = \frac{a}{2}\eta + \varepsilon, x = \delta$$

где ε, η, δ – случайные величины, при этом $\eta -$ информативный сигнал, и представлен в виде

$$\eta = \begin{cases} 1, & p = \frac{1}{2} \\ -1, & p = \frac{1}{2} \end{cases}$$

Шумы имеют нормальные распределения и задаются следующим образом:

$\varepsilon \sim N(0, \sigma_\varepsilon^2)$ - шум в канале с сигналом y

$\delta \sim N(0, \sigma_\delta^2)$ - шум в канале x.

$a = |m_1 - m_{-1}|$ - сигнал в канале.

Условные плотности распределения вероятностей $P\left({}^{y}/_{\eta = 1}\right)$ и $P\left({}^{y}/_{\eta = -1}\right)$ имеют нормальный закон распределения со значениями математического ожидания m_1, m_{-1} и $\sigma_\varepsilon^2, \sigma_\varepsilon^2$, соответственно.

Во втором случае, описанным в [3], рассматриваются сигналы $y_i = \frac{a_i}{2}\eta + \varepsilon_i$, где $i = \{1,2\}$, где η – не зависящая от ε_i случайная величина:

$$\eta = \begin{cases} 1, p = \frac{1}{2} \\ -1, p = \frac{1}{2} \end{cases}$$

$\varepsilon_i \sim N\left(0, \sigma_{\varepsilon_i}^2\right)$ - шум в i - ом канале, распределенный по нормальному закону,

$a_i = \left|m_1^{(i)} - m_{-1}^{(i)}\right|$ - сигнал в i - ом канале.

Параметры условных плотностей распределения вероятностей $P\left({}^{y_i}/_{\eta = 1}\right)$, $P\left({}^{y_i}/_{\eta = -1}\right)$ $N\left(m_1^{(i)}, \sigma_i^2\right)$ и $N\ (m_{-1}^{(i)}, \sigma_i^2)$, где $i = \{1,2\}$.

При моделировании передачи сигнала считается, что при $\eta = 1$ передается значение "1", при $\eta = 0$ передается "0".

Представленные выше каналы передачи данных и процедура распознавания сигнала методом k ближайших соседей были промоделированы в среде Matlab R2011. При этом отношение сигнал/шум было принято равным 0.02. В ходе работы менялись следующие параметры:

1. Коэффициенты корреляции шумов R в каналах менялись от 1 до -1 с шагом $h = 0.1$.
2. Количество ближайших соседей k, используемых при классификации, изменялось от 1 до 511. Были перебраны все нечетные значения k от 1 до 11, далее использовался более значимы шаг перебора.
3. Для оценки расстояний были использованные перечисленные выше метрики(1 - 4).
4. Объем тестовых выборок $V = 10000$.

В ходе работы были получены следующие результаты. При уменьшении модуля коэффициента корреляции вероятность успешного распознавания сигнала резко уменьшалась. Далее графически представлена зависимость вероятности P^* от количества ближайших соседей, используемых при классификации для различных метрик.

Результаты работы метода k ближайших соседей для каналов передачи первого случая в координатах: P^* - вероятность успешного распознавания сигнала, k - количество ближайших соседей, учитываемых при распознавании, представлены на рисунках 1-4:

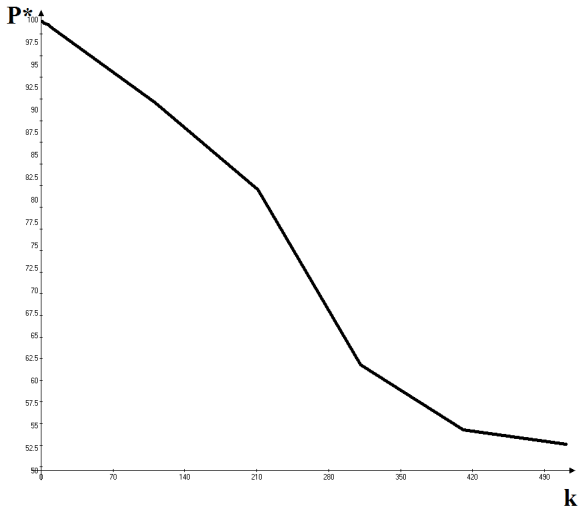

Рис. 1. Полученные результаты. Евклидова метрика. Сигнал+шум и шум

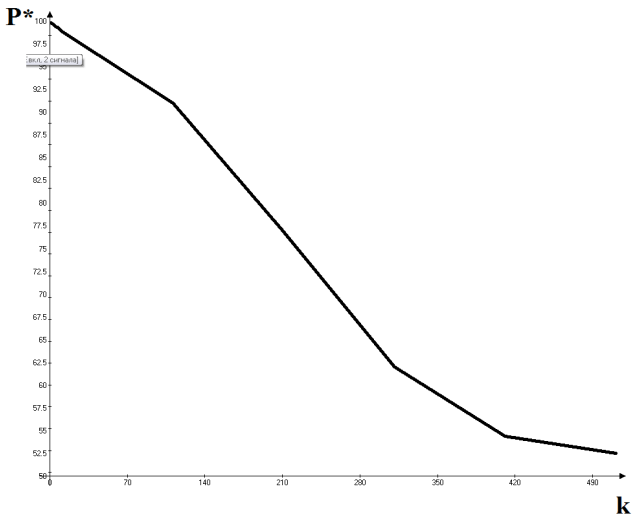

Рис. 2. Полученные результаты. Степенная метрика. Сигнал+шум и шум

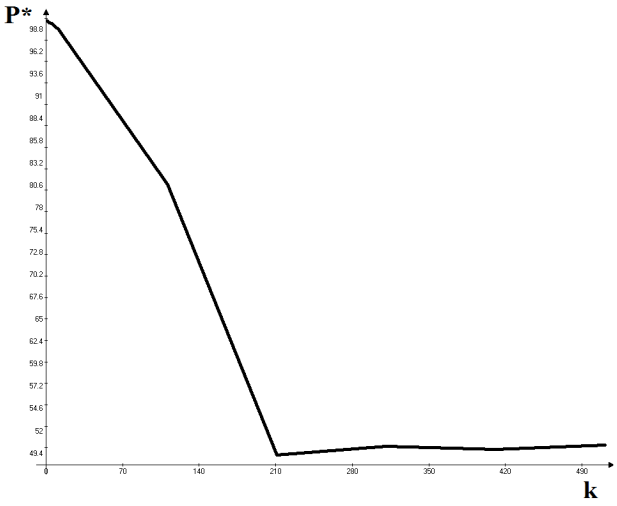

Рис. 3. Полученные результаты. Манхэттенская метрика. Сигнал+шум и шум

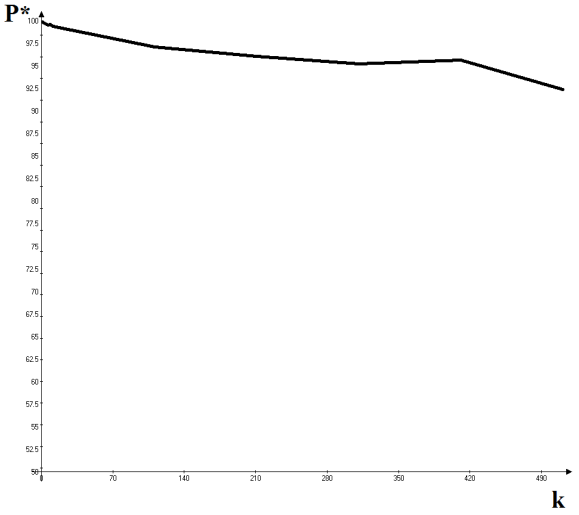

Рис. 4. Полученные результаты. Метрика Чебышева. Сигнал+шум и шум

Результаты работы метода k ближайших соседей для каналов передачи данных второго случая в координатах: P^* - вероятность успешного распознавания сигнала, k - количество ближайших соседей, учитываемых при распознавании, представлены на рисунках 5-8:

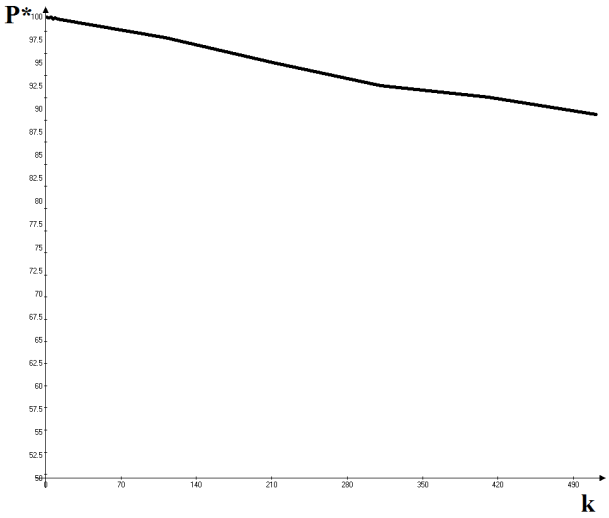

Рис. 5. Полученные результаты. Евклидова метрика. 2 канала сигнал+шум

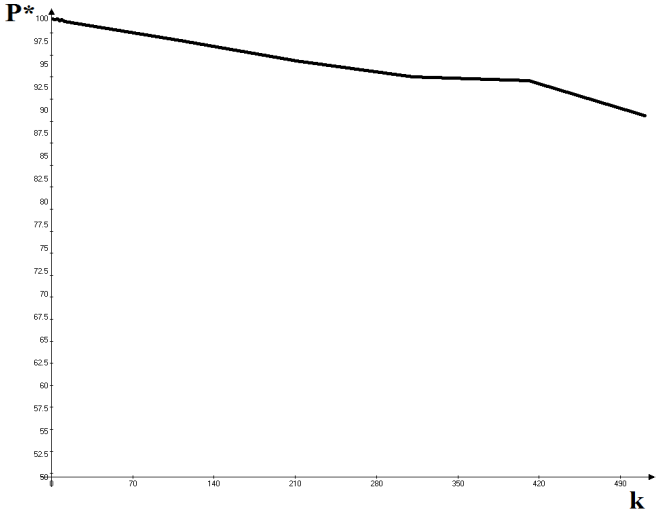

Рис. 6. Полученные результаты. Степенная метрика. 2 канала сигнал+шум

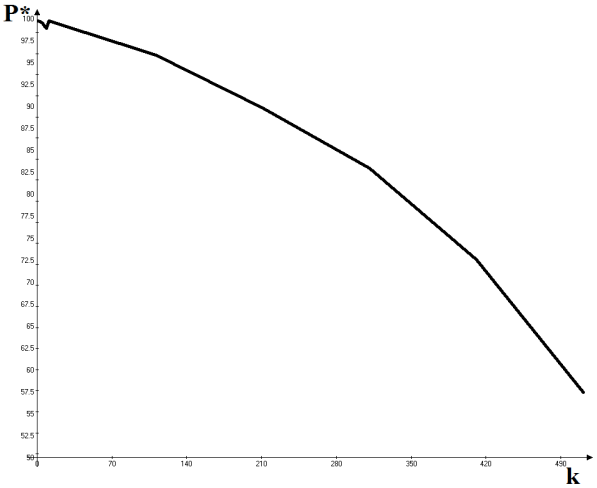

Рис. 4. Полученные результаты. Манхэттенская метрика. 2 канала сигнал+шум

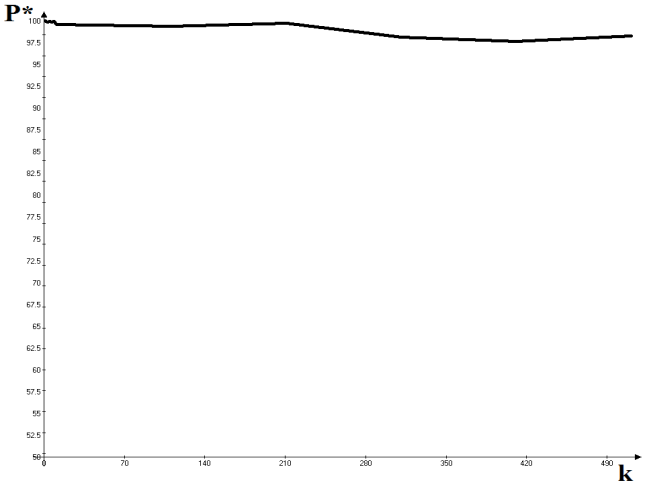

Рис. 5. Полученные результаты. Метрика Чебышева. 2 канала сигнал+шум

Исходя из полученных результатов, можно сказать, что все перечисленные метрики подходят для распознавания передаваемого сигнала. При этом наиболее эффективно метод k ближайших соседей работает при меньших значениях k. В случае, если необходимо использовать большие значения k, целесообразно применение метрики Чебышева.

Следует заметить, что зашумлении каналов передачи данных необходимо обеспечивать независимость генерируемых шумов (т.е. значение коэффициента корреляции между шумами каналах должно стремиться к 0), так как при увеличении значения коэффициента корреляции вероятность перехвата данных возрастает.

Список литературы

1. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. 510 с.
2. Мерков А.Б. Введение в методы статистического обучения. Режим доступа: http://www.recognition.mccme.ru/pub/RecognitionLab.html/slbook.pdf (дата обращения 27.04.2015).
3. Троицкий И.И., Басараб М.А., Матвеев В.А. Определение вероятности распознавания дискретного сигнала в аддитивном шуме для двух каналов передачи информации // Электромагнитные волны и электронные системы. 2014. №11, т.19. С. 40-44.