

ВИЯВЛЕННЯ НЕЯВНОЇ ІНФОРМАЦІЇ, ПРИХОВАНОЇ В НАВЧАЮЧІЙ ВИБОРЦІ.

Васильєв В.І., Кобець Н.М., Ланге Т.І.

В. І. Васильєв: Міжнародний науково-навчальний центр ЮНЕСКО/МПІ інформаційних технологій і систем НАН і МО України (МННЦ ІТІС), Україна, 252022, Київ, пр. Академіка Глушкова 40, т. (044)2664187, факс (044)2661570

Н. М. Кобець: Національний технічний університет України "Київський політехнічний інститут", пр. Перемоги 37, т. (044)4411892

Т. І. Ланге: Fachhochschule Merseburg, Fachbereich Elektrotechnik, FH Merseburg, Geusaer Straße, D-06217 Merseburg, Deutschland, т.(03461)462528, факс (03461)462919

Анотація - розглядається спосіб здобуття додаткової інформації з навчаючої вибірки, що дозволяє багаторазово збільшити число властивостей об'єктів. Це дозволяє істотно розширити можливості алгоритмів навчання розпізнаванню образів, побудованих на теорії редукції. Дається стислий опис алгоритму навчання, що синтезує простір, в якому досягається лінійне розділення образів, що задані навчаючою вибіркою.

ВСТУП

Пропонований підхід можна застосувати для поповнення інформації при розв'язанні задачі навчання розпізнаванню образів (задачі НРО) будь-якими відомими методами в межах геометричного підходу. Найбільш продуктивним цей підхід виявляється тоді, коли для розв'язання задачі НРО використовуються алгоритми, побудовані на теорії редукції [1], головною ідеєю якої є граничне спрощення вирішувальних правил, що діють у просторі мінімальної розмірності.

ОСНОВИ ТЕОРІЇ

В основу теорії редукції покладено фундаментальну теорему [2], яка стверджує, що якщо з N вирішувальних правил обирається одне, яке безпомилково розділяє випадкову і незалежну вибірку довжиною l , то імовірність невірної класифікації за допомогою цього правила з імовірністю більшою, ніж $1-\eta$, не перевищуватиме величини

$$\varepsilon \leq \frac{\ln N - \ln \eta}{l} \quad (1)$$

Показано [1], що якщо організувати послідовний синтез простору, в якому настане лінійне розділення образів, заданих на навчаючій вибірці довжиною l , то $\ln N \leq n_0 \ln m$, де m -кількість первинних властивостей, з яких обирається n_0 ознак. Використовуючи (1) можна синтезувати простір розмірності n_0 , в якому лінійне розділення випадкової та незалежної вибірки довжиною l гарантує потрібні значення ε та η .

Теорія редукції надає процедуру синтезу простору розмірністю n_0 , в якому досягається безпомилкове лінійне розділення навчаючої вибірки довжиною l . Величина n_0 , знайдена з (1), після підставлення $\ln N \leq n_0 \ln m$ визначає розмірність синтезованого простору, перевищення якої призводить до втрати гарантій досягнення заданих ε і η . Щоб розмірність простору не перевищувала n_0 , вводиться поняття мінімальної відокремлюючої сили ознаки, що визначається як частина "навантаження", яку під час розділення вибірки має взяти на себе кожна ознака [1].

Різновидом алгоритмів, побудованих на теорії редукції, є α -процедура [1]. Нехай навчаючу вибірку задано як навчаючу таблицю, в якій кожна властивість усіх об'єктів подана стовпцем, а кожному об'єкту відповідає певний рядок. Будь-який об'єкт належить до одного із образів V_1^* і V_2^* .

Y/X	x_1	x_2	...	x_m
y_1	x_{11}	x_{12}	...	x_{1m}
...
y_l	x_{l1}	x_{l2}	...	x_{lm}

Усі об'єкти, для яких $y_j \geq 0$, належать до образу V_1^* , а якщо $y_j < 0$, - до образу V_2^* . Розглянемо один стовпець, тобто одну неперервну властивість x_i . Встановимо на осі цієї властивості поріг x_i^0 , що розділяє всю вибірку на осі x_i на два класи еквівалентності V_{1i} і V_{2i} , які так чи інакше збігаються з V_1^* і V_2^* . Якщо обрати деякий критерій збігу множин V_1^* , V_2^* та V_{1i} , V_{2i} , можна знайти поріг x_i^0 , що оптимізує вибраний критерій.

На першому кроці α -процедури обирається будь-яка властивість, відокремлююча сила якої вище мінімально допустимої, і ця властивість розглядається як ознака x_1 . Так обирається перша координата простору, що синтезується. Тому для всіх властивостей x_i , що лишилися, для кожного об'єкту навчаючої вибірки обчислюються числа

$$\tilde{X}_i = \rho_i \cos(\beta_i + \alpha_{si}), \quad (2)$$

де $\rho_i = \sqrt{x_{i1}^2 + x_{i2}^2}$; $\beta_i = \arctg \frac{x_{i1}}{x_{i2}}$; α_{si} - змінний

параметр ($\alpha_{si} = 0^\circ \div 180^\circ$)

Обираємо таке α_{si}^* , для якого $\alpha_{si}^* = \arg \max_s \omega_{is}$,

де ω_{is} - кількість об'єктів, що вірно класифікуються за оптимального порога, встановлюваного на кожному напрямку, що визначається α_{si} .

Усі об'єкти навчаючої вибірки проектується на напрямок \tilde{X}_i , а самі ці напрямки визначатимуться величинами кутів α_{si} . Кожному α_{si} відповідає свій напрямок \tilde{X}_{si} , і на кожному напрямку встановлено свій оптимальний поріг \tilde{X}_{is}^* . Кінець кінцем, обирається такий напрямок $\tilde{X}_i(\alpha_{is}^*)$, на якому оптимальний поріг дає найкраще розділення множин V_{1i} і V_{2i} , якщо тепер відокремлююча сила на цьому напрямку перевищує мінімально припустиму, властивість, що розглядається, приймається за ознаку X_2 . Так обирається друга координата простору, що синтезується, яка відповідає напрямку $\tilde{X}_2(\alpha_{si}^*)$.

Далі обирається одна з властивостей X_k , що лишилися, і для неї в площині $\tilde{X}_2 - X_k$ обчислюється величина $\tilde{X}_k = \rho_k \cos(\beta_k + \alpha_{sk})$, де

$\rho_k = \sqrt{x_{k2}^2 + x_{k3}^2}$; $\beta_k = \arctg \frac{x_{k2}}{x_{k3}}$. Якщо при цьому

на оптимальному напрямку $\tilde{X}_3 = \tilde{X}_k(\alpha_{si}^*)$ оптимальному порогу відповідає відокремлююча сила, що перевищує мінімально припустиму, то властивість X_k береться за третю ознаку або третю координату простору, що синтезується. Така α -процедура повторюється, доки на одному із базових напрямків \tilde{X}_v оптимальний поріг не розділить об'єкти навчаючої вибірки повністю. Площина, що перпендикулярна цьому базовому напрямку і проходить через опти-мальний поріг цього напрямку, приймається за вирішувальне правило, що синтезоване α -процедурою.

ВИЯВЛЕННЯ ДОДАТКОВОЇ ІНФОРМАЦІЇ

Повернемося до навчаючої таблиці. Розглянемо будь-який рядок цієї таблиці, тобто один із об'єктів навчаючої вибірки. Для цього об'єкта кожна властивість $x_i, i = \overline{1, m}$ приймає своє конкретне значення x_{iv} . Оберемо деяку комбінацію властивостей (деякий фрагмент таблиці), зафіксуємо її на обраному об'єкті та позначимо булевим вектором $V(\beta_1, \dots, \beta_m)$. Розмірність цього вектора відповідає

розмірності навчаючої таблиці $i = \overline{1, m}$, а його складові визначаються співвідношенням

$$\beta_i = \begin{cases} 1, & \text{якщо } i \in k \\ 0, & \text{в іншому разі} \end{cases} \quad (3)$$

Тут k - множина стовпців навчаючої таблиці, що означає обраний фрагмент. Цей фрагмент описується вектором $X^* = \{x_1\beta_1; x_2\beta_2; \dots; x_m\beta_m\}$. Наприклад, якщо за фрагмент обрати стовпці ($i=1, 3, 5$ і $m=5$), то вектор V визначатиметься складовими $\beta_1=1, \beta_2=0, \beta_3=1, \beta_4=0, \beta_5=1$, тобто фрагмент визначається властивостями (стовпцями) x_1, x_3, x_5 : $X = \{x_1, x_3, x_5\}$

Тепер визначимо відмінності усіх об'єктів навчаючої вибірки від обраного об'єкта X_v за фрагментом k . Цю відмінність для кожного об'єкту X_s вибірки можна визначити співвідношенням:

$$L(X_v, X_s)_k = \sum_{i=1}^m |x_{iv} - x_{is}| \cdot \beta_i \quad (4)$$

Це співвідношення має всі властивості відстані у просторі обраного фрагмента, тобто $L(X_v, X_s) = 0$. Відмінність (4) зростає з віддаленням від опорного об'єкту в просторі обраного фрагменту і може розглядатися як похідна складна властивість усіх об'єктів, яка визначає, наскільки відмінний (схожий) будь-який об'єкт від опорного об'єкту в розумінні обраного фрагменту. Наприклад, співвідношення (4) може показати, наскільки дві людини відрізняються одна від одної зростом, вагою та розміром взуття. Тому кожна така похідна властивість може розглядатися α -процедурою як деяка нова властивість. Якщо рангом r нової складної властивості назвати число стовпців у фрагменті, а число стовпців вихідної навчальної таблиці дорівнює m , то така таблиця може породити m^* нових складних властивостей:

$$m^* = \sum_{r=1}^m C_m^r$$

Складні ознаки, що обираються із складних властивостей, дозволяють розв'язувати складні задачі, такі, наприклад, як розділення двох сфер, вкладених одна в одну. Таке розділення легко здійснити, застосовуючи тільки єдину ознаку, яка визначається опорним об'єктом, покладеним у центр цих сфер.

ВИСНОВКИ

Метод виявлення неявної інформації, що пропонується, дозволяє багаторазово збільшити кількість вихідних властивостей, з яких формуються ознаки, що сприяють розпізнаванню.

ЛІТЕРАТУРА

1. Васильев В.И. "Теория редукции в проблемах экстраполяции" // "Проблеми управління і інформатики"-1996.-№1-2.-С. 238-251.
2. Вапник В.Н., Червоненкис А.Я. "Теория распознавания образов (статистические проблемы обучения)".-М.:Наука, 1974.-416 с.

ПРО ЗАДАЧІ РОЗБИТТЯ

Надія ТИМОФІЄВА

МННЦ з ІТiС НАН України та Мiносвiти
252022, Київ, просп. Ак. Глушкова, 40, тел.: (044) 2664356

Nadija Tymofijeva. About partition problems

An analysis of classes of problems of partition is suggested, some property of set of partition is represented, method of generation of set of partition is described. On basis of this analysis mathematical model of problem of partition is built

Задачі розбиття відомі в найрізноманітніших постановках. Це розбиття множини, розсічення графів, мереж, кластеризація, таксономія, класифікація. З обчислювальної точки зору вони є одними з найбільш складних.

Для правильного вибору методу і алгоритму, за допомогою яких можна одержати оптимальний розв'язок, необхідно побудувати математичну модель, яка б адекватно описувала поставлену задачу. З цією метою в статті аналізуються типи задач розбиття, наведено деякі властивості множини розбиття, описано метод упорядкування усіляко можливих розбиття. На підставі цих досліджень побудовано математичну модель задачі розбиття.

Отже, задача розбиття полягає в упорядкуванні заданих об'єктів у порівняно однорідні групи, тобто за розробленими правилами проводиться розбиття заданої множини на підмножини. Задачі розбиття можна розділити за типом завдання об'єктів і підмножин, по яких розподіляються ці об'єкти, і за способом обчислення цільової функції.

За типом завдання об'єктів і підмножин виділимо такі задачі.

1. Задано n об'єктів, які належать до різних типів. Кількість підмножин, по яких розподіляються об'єкти окремого типу, і їх потужність задано.

2. Задано n об'єктів, які належать до різних типів. Кількість підмножин, по яких розподіляються об'єкти окремого типу, і їх потужність не задано.

3. Задано n об'єктів, які належать до одного типу. Кількість підмножин, по яких розподіляються об'єкти, і їх потужність задано.

4. Задано n об'єктів, які належать до одного типу. Кількість підмножин, по яких розподіляються об'єкти, і їх потужність не задано.

За способом обчислення цільової функції задачі розбиття діляться на такі, у яких взаємозв'язки між об'єктами задано явно, у кількісному значенні, і такі, у яких їх задано

неявно, у якісному значенні. Так, наприклад, у задачах компоновки, розбитті електричних схем на непересічні підмножини між об'єктами початкової множини задано кількість зв'язків. У цільовій функції, за якою здійснюється оцінка оптимального розв'язку, враховується сумарна кількість зв'язків між елементами, які входять до однієї підмножини, або сумарна кількість зв'язків між підмножинами. В задачах розпізнавання (таксономії), класифікації, кластеризації взаємозв'язок між об'єктами існує в якісному вимірі. Для встановлення цих зв'язків у кількісному вимірі вводяться міри подібності. Оскільки міра подібності - оцінка суб'єктивна, то використання для розв'язання однієї і тієї самої задачі різних мір подібності приводить до різних результатів. Отже, якщо взаємозв'язок між об'єктами визначається мірами подібності, то результат розв'язку задачі розбиття залежить і від способу визначення цих взаємозв'язків, і від завдання цільової функції, за якою проводиться пошук оптимального розв'язку.

Математична модель задач розбиття, як правило, задається однією множиною, між елементами якої існує взаємозв'язок. Початкові дані (взаємозв'язки між елементами цієї множини) подамо у вигляді матриці зв'язності. Якщо задача розбиття належить до першого типу за способом обчислення цільової функції, вважатимемо, що вхідні дані задано за умовою задачі. Якщо задача розбиття належить до другого типу (таксономія, кластеризація, класифікація), необхідно побудувати матрицю зв'язності, використовуючи залежно від класу задач такі міри подібності, як коефіцієнт кореляції, міри відстані, коефіцієнти асоціативності чи ймовірнісні коефіцієнти подібності [1-2].

Розглянемо деякі властивості розбиття множини і спосіб генерування усіляко можливих розбиття. Для їх упорядкування використаємо підхід, що враховує властивість періодичності, характерної для процесу генерування комбінаторних конфігурацій [3]. Згідно з ним усіляко можливі розбиття упорядковуються майже періодичними інтервалами, кожен із яких складається з основних майже періодів. Для цього необхідно:

а) визначити основний майже період множини розбиття і правило, за яким він формується;

б) визначити правило, за яким формується обмежувальне розбиття;

в) визначити правило, за яким генеруються усіляко можливі розбиття, які належать до майже періода σ -го порядку. Тим самим забезпечується генерування усіляко можливих розбиттів.

Розбиттям n -елементної множини A на η^k підмножин назвемо упорядковану множину підмножин $\rho^k = (\rho_1^k \dots \rho_{\eta^k}^k)$ таку,

$$\rho_1^k \cup \dots \cup \rho_{\eta^k}^k = A, \quad \rho_j^k \cap \rho_l^k = \emptyset, \\ j, l \in \{1, \dots, \eta^k\}, j \neq l, \quad \rho_j^k \neq \emptyset,$$

$$j \in \{1, \dots, \eta^k\}, \quad \eta^k \in \{1, \dots, n\},$$

$k \in \{1, \dots, q\}$, q - кількість усіляко можливих розбиттів. Упорядковану множину усіляко можливих розбиттів позначимо Θ .

Розбиття ρ^k характеризується

кількістю підмножин ρ_j^k і кількістю елементів, що входять до цієї підмножини. За цими ознаками у множині Θ виділено чотири типи розбиттів.

Означення 1. Розбиття $\rho^k \in \Theta$ основного майже періоду утворюється із попереднього ρ^{k-1} однією транспозицією j -го елемента t -ї підмножини ρ^{k-1} -го розбиття і першого елемента $t+1$ -ї підмножини ρ^{k-1} -го розбиття, $j = 1, \xi_i^k, t \in \{1, \dots, \eta^{k-1} - 1\}$,

ξ_i^k - кількість елементів у ρ_i^k .

Означення 2. Початкове розбиття ρ^k

t -го основного майже періоду назвемо обмежувальним. Воно або задається, або визначається за особливим правилом.

Означення 3. Майже періодом σ -го порядку упорядкованої множини Θ назвемо упорядковану її підмножину, яка містить усіляко можливі розбиття, утворені елементами $i_1^k, \dots, i_\sigma^k, \sigma \in \{\xi_1^k + 2, \dots, n\}$.

На підставі вищевикладеного подамо математичну модель у такому вигляді. Назвемо вагою величину, що визначає залежність між

елементом a_j і елементом a_l . Вагу між цими елементами задамо симетричною матрицею C . Введемо також $(0,1)$ -симетричну комбінаторну матрицю $Q(\rho^k)$, яка задає розподілення

елементів множини A по підмножинах ρ_j^k .

Назвемо її комбінаторною матрицею розподілення. Елемент цієї матриці

$g_{jl}(\rho^k) = 1$, якщо елементи a_j і a_l належать до однієї підмножини, і $g_{jl}(\rho^k) = 0$ в іншому випадку. Матриця

$Q(\rho^k)$ утворюється із попередньої

$Q(\rho^{k-1})$ за правилами генерування

усіляко можливої множини розбиттів.

Подамо елементи правого поля матриці

$Q(\rho^k)$ комбінаторною функцією

$f_{\rho^k}(x)$, а елементи правого поля матриці

C - числовою функцією $\varphi(x)$. Функція цілі для задачі розбиття набуде вигляду

$$F(\rho^k) = \sum_{j=1}^m f_{\rho^k}(x_j) \varphi(x_j), \quad (1)$$

$$m = \frac{n(n-1)}{2}$$

Оскільки однорідні елементи мають між собою найбільші ваги і їх необхідно об'єднати в одну підмножину, то задача полягає в

знаходженні такого розбиття ρ^k , для якого

функція цілі (1) при виконанні заданих умов була б найбільшою. Сформульована математична модель дозволяє визначати множину значень цільової функції в задачах розбиття різних типів і тим самим обмежувати область пошуку оптимального розв'язку.

ЛІТЕРАТУРА

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов.- К.: Наук. думка, 1987. - 262 с.

2. Факторный, дискриминантный и кластерный анализ/ Дж.-О. Ким, Ч.У.Мьюллер, У.Р.Клекка, М.С.Олдендерфер, Р.К. Блэшфилдж; Под ред. И.С. Енюкова. -М.: Финансы и статистика, 1989.- 215 с.

3. Тимофієва Н.К. Про аргумент комбінаторних функцій// Проблеми оптимізації обчислень. Праці міжнародної конференції 6-8 жовтня 1997 р. м. Київ.- К.: 1997. - С. 303-307.

Support vector classifiers: Application to digital communications channel equalization

1. Felix Albu, 2. Dominique Martinez

1. Faculty of Electronics & Telecommunications, 1-3 Bd. Iuliu Maniu, Bucharest

felix@pns.comm.pub.ro, Tel: 401 410 54 00

2. CNRS-LAAS, 7 av. Col. Roche, 31077 Toulouse, France

dmartine@laas.fr, Tel: 33 561 33 64 43

ABSTRACT

The Support Vector (SV) method is a novel type of learning machine, based on statistical learning theory. In support vector classifiers, the input vectors are mapped to a high dimensional space and are then separated by the optimal linear hyperplane[1]. In the RBF case, the SV algorithm automatically determines centers, weights and threshold that minimize an upper bound on the expected test error. The classical approach in building a radial basis function classifier consists of choosing the centers by some clustering techniques. The weights are usually found by the pseudoinverse method or by LMS algorithm[2]. We applied these techniques to channel equalization, viewed as a classification problem, for reconstruction of binary signals. It is well-known that the radial basis function network has an identical structure to the optimal Bayesian symbol-decision equalizer solution[3]. But the computational requirement to implement can be very high, as the full RBF Bayesian solution usually requires a large number of centers. We compared the results of equalization task obtained with classical RBF training with those obtained by the SV algorithm. We showed the superiority of SV techniques, especially for high SNR ratio.

1. INTRODUCTION

To address a classification problem, one often would like to use as many features as possible to improve the classification result. However, most classifiers suffer from what is called the small size effect or peaking effect. There is a certain optimal number of features beyond which performance will only degrade. This is because, as the dimensionality of the feature space grows, estimating densities of the data becomes increasingly harder for a finite number of samples[4]. Vapnik proposed a method for finding a hyperplane

optimally dividing two classes, which does not depend on a probability estimation[1]. This optimal hyperplane is a linear decision boundary which separates the two classes and leaves the largest margin between the vectors of the two classes. It was observed that the optimal hyperplane is determined by only a small fraction of the data points, the so-called *support vectors*. To find a non-linear classifier, it is possible to transform the vectors in *input space* to some high-dimensional *feature space*. This allows for more difficult problems to be solved and increases the feature space dimensionality, but hardly influences the classifier complexity. Furthermore, it was shown that the probability of making an error depends only on the number of these support vectors (and therefore on the complexity of the classifier) and the number of training vectors. In the next sections we present the Support Vector machines (we refer mostly to the paper by Cortez and Vapnik [5]), the RBF equalizers and some computer simulations showing the performances of these techniques on digital communication channel equalization.

2. SUPPORT VECTOR MACHINES

2.1 THE OPTIMAL HYPERPLANE

The optimal hyperplane is defined as the plane separating two separable classes in such a way that its margin is as wide as possible. If we denote the training vectors by $\mathbf{x}_i, i = 1, \dots, l$ with the corresponding labels $y_i \in \{-1, 1\}$ then this yields

$$y_i [(\mathbf{w} \cdot \mathbf{x}_i) + b] \geq 1, \quad i = 1, \dots, l \quad (1)$$

if the optimal hyperplane is

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (2)$$

in which \mathbf{w} can be written as

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i, \quad \alpha_i \geq 0 \quad (3)$$

The discrimination function for a certain vector \mathbf{z} is therefore written as

$$f(\mathbf{z}) = \sum_{i=1}^l \alpha_i y_i (\mathbf{z} \cdot \mathbf{x}_i) + b \quad (4)$$

It can be shown that in order to find the optimal set of weights α_i , the following expression has to be maximized:

$$W(\Lambda) = \Lambda^T \mathbf{I} - \frac{1}{2} \Lambda^T \mathbf{D} \Lambda \quad (5)$$

w.r.t. $\Lambda = (\alpha_1, \dots, \alpha_l)$, subject to the following constraints:

$$\begin{aligned} \Lambda &\geq 0 \\ \mathbf{Y}^T \Lambda &= 0 \end{aligned} \quad (6)$$

\mathbf{D} is a matrix containing the dot products of the training vectors, multiplied by their labels:

$$\mathbf{D}_{ij} = y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (7)$$

Maximizing Eq. (5) is a quadratic programming problem with linear constraints.

2.2 SOFT MARGIN HYPERPLANE

The concept of the optimal hyperplane is only suitable for the separable case [4]. If classes overlap Eq. (1) is modified to allow for errors:

$$\begin{aligned} y_i [(\mathbf{w} \cdot \mathbf{x}_i) + b] &\geq 1 - \xi_i, \quad i = 1, \dots, l \\ \xi_i &\geq 0, \quad i = 1, \dots, l \end{aligned} \quad (8)$$

and the following function can be minimized:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (9)$$

The quadratic programming problem now becomes a dual one:

$$W(\Lambda, \delta) = \Lambda^T \mathbf{1} - \frac{1}{2} \left[\Lambda^T \mathbf{D} \Lambda + \frac{\delta^2}{C} \right] \quad (10)$$

w.r.t. $\Lambda = (\alpha_1, \dots, \alpha_l)$, and δ subject to constraints:

$$\begin{aligned} 0 &\leq \Lambda \leq \delta \mathbf{1} \\ \mathbf{Y}^T \Lambda &= 0 \end{aligned} \quad (11)$$

The dot product can be replaced by a generalized dot product that satisfies Mercer's theorem [2]. The only equation that changes is Eq. (7) which becomes

$$\mathbf{D}_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (12)$$

and the discrimination function

$$f(\mathbf{z}) = \sum_{i=1}^l \alpha_i y_i K(\mathbf{z} \cdot \mathbf{x}_i) + b \quad (13)$$

We used a radial basis function neural network with kernel width σ , whose the necessary number of kernels and their positions are found by the algorithm:

$$K(\mathbf{z}, \mathbf{x}_i) = \exp\left(-\frac{|\mathbf{z} - \mathbf{x}_i|^2}{\sigma^2}\right) \quad (14)$$

3. RBF EQUALIZERS

Channel equalization is a technique employed to combat the effects of intersymbol interference and noise which corrupt the transmission of signals across a communication channel (Fig. 1). The equalization aims to reconstruct the transmitted sequence with the minimum error probability, i. e. : $\tilde{s}(t) = s(t - d)$, where d is the delay.

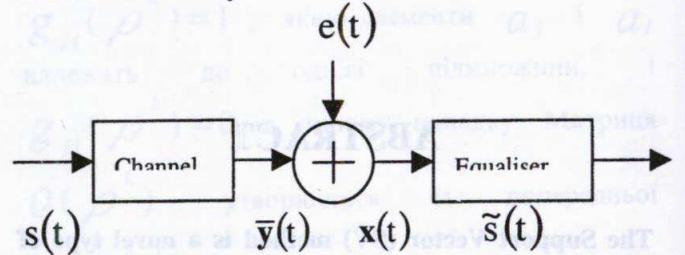


Fig.1 Discrete-time model of data transmission system

The channel is usually modeled by a FIR filter with the following transfer function :

$$H(z) = \sum_{i=0}^{n_h-1} h_i z^{-i}, \quad (15)$$

where h_i is the channel impulse response components and n_h its length. In our study the transmitted symbol $s(t)$ is taken from the data set $\{\pm 1\}$; it forms an i.i.d. sequence, and $e(t)$ is an additive white Gaussian noise with zero mean and variance σ_e^2 [3].

A RBF network is a two layer network comprising a hidden layer and an output layer. The hidden layer contains n neurons which compute the Euclidian distance between a center vector \mathbf{c}_i and an input vector $\mathbf{x} = [x(t)x(t-1)\dots x(t-m+1)]^T$. The result is passed through a nonlinear function Φ_i to generate the hidden node output. Functions Φ_i are

chosen to be Gaussian: $\Phi_i = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{\sigma_i^2}\right)$,

where σ_i is called the width. The output layer is computed by a weighted linear combination of the n neurons of the hidden layer. The overall response is a mapping:

$$f(\mathbf{x}) = \sum_{i=1}^n w_i \Phi_i, \quad (16)$$

where w_i are the weights.

It has been shown [3] that RBFN achieves an implementation of the optimal Bayesian equalizer if the channel is known and the parameters of the network are well chosen (i. e., the number of hidden neurons n is equal to the number of desired channel states: $n = 2^{m+n_h-1}$; the RBFN centers are placed at desired channel state vectors:

$\mathbf{c}_i = [\bar{y}_i(t) \bar{y}_i(t-1) \dots \bar{y}_i(t-m+1)]^T$; the weights are chosen in the data set: $\{\pm 1\}$, and $\sigma_i^2 = 2\sigma_e^2 \gamma_i^2 = 2\sigma_e^2$, $i \in [1;n]$. Note that the RBFN structure can be complex when m and n_h are large.

In our work, the training of RBFN was done using a two-steps approach: in the first step a classical k-means clustering procedure was used to find the location of the centers and the widths were set at $\sigma_i = \frac{d_m}{\sqrt{2 \cdot n}}$, where $i \in [1;n]$ and d_m is the maximum distance between the chosen centers[6]. The k-means algorithm is an unsupervised learning method based only on input training samples. It partitions the input data set into n cluster centers. In the second step, the weights were trained using the least mean squares (LMS) algorithm.

4. SIMULATION STUDY

For the purpose of graphical display, the equalizer order is chosen equal to 2. Let the channel transfer function be $H_1(z) = 0.5 + z^{-1}$. The delay was 1.

Figure 2 shows 50 samples which form the data clusters and the SV centers(indicated by extra circles).

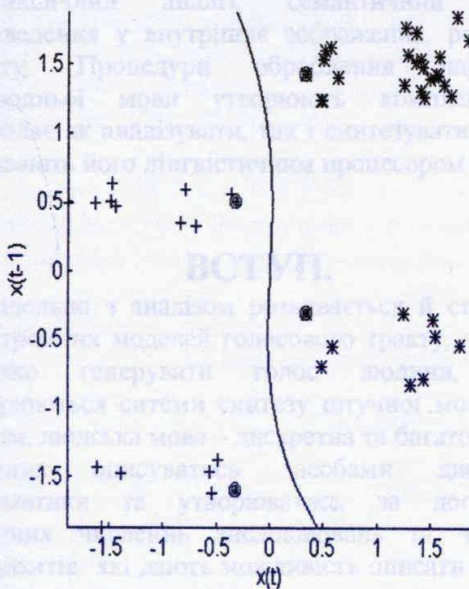


Fig.2 RBF centers automatically found by the SV algorithm (indicated by extra circles), and the decision boundary, $H_1(z) = 0.5 + z^{-1}$, SNR=17 dB, $\sigma^2 = 2$

The other channel used for various SNR's was:

$$H_2(z) = -0.2052 - 0.5131 \cdot z^{-1} + 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4} \quad (17)$$

We used a 2-64-1 RBF network, and the delay was chosen $\tau = 2$ [7]. The number of samples for training was 250 (wee need all at once for the SV method). The performance of our structures (RBF or SV equalizers) were measured in terms of \log_{10} BER (BER= Bit Error Rate), for 200000 samples. We can see the superiority of SV method over a RBF network using this k-means clustering method especially for high SNR ratio (Fig. 3). The SV method doesn't work well for low SNR's because the errors were too much punished. We used $C = 10^8$ and $\sigma^2 = 0.5$. Better results could be obtained by lowering C and using higher values for σ^2 , directly related to SNR ratio. A cross-validation technique could be used, although other heuristic methods would be indicated. It should be noted that with the clustering method presented in [3] the RBF structure could achieve the optimal Bayesian performance for a symbol by symbol equalizer. The convergence of this procedures requires for our particular case more than 600 samples. We also need a correct estimate of the channel order. The SV method provides a rigorous way of choosing the number and locations of RBF centers specifically selected for our classification problem.

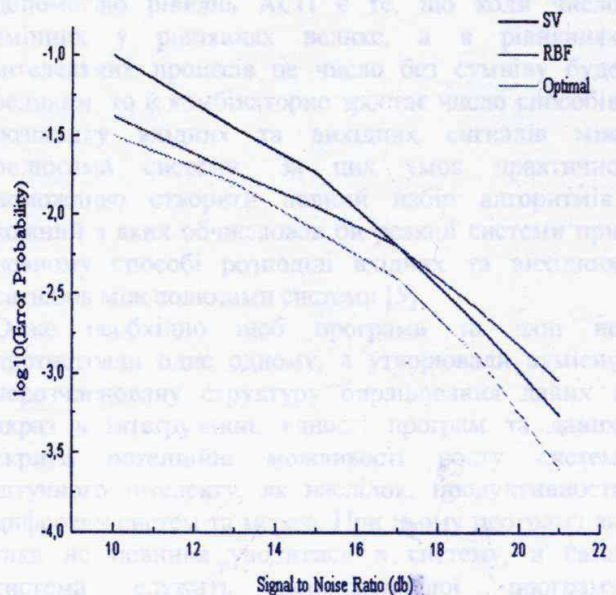


Fig.3 Comparison of performance. Channel H2, delay=2, m=2

5. CONCLUSION

The traditional view of RBF equalizers has been one where the clustering heuristic used for training them was considered very important. In contrast, the SV algorithm provides a way of choosing the number and the locations of RBF centers, where the centers are those examples that are critical for the classification task. The classical k-means clustering algorithm can only achieve a local optimal solution, which depends on the initial locations of cluster centers. A consequence of this local optimality is that some initial centers can become stuck in regions of the input domain with few or no input patterns, failing to move where they are needed [8]. Our simulation showed the superiority of SV method over a RBF equalizer using a k-means clustering procedures for high SNR's ratio. Whether a RBF network can realize optimal equalizer solution depends crucially on positioning the centers at the desired channel states. This requires a great number of samples for training using a supervised or unsupervised k-means procedures and a correct estimate of the channel order. With a small number of training samples and no knowledge about the channel order, the SV method could become a suitable choice for digital communications channel equalization. In the future, our work will focus on attempting to enhance the speed in training and test phase while looking closely at their application to nonlinear channels.

REFERENCES

- [1] Vapnik, V. "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995
- [2] B. Mulgrew : "Applying Radial Basis Functions", IEEE Signal Processing Magazine, March 1996.
- [3] S. Chen, B. Mulgrew, P. M. Grant, "A clustering technique for digital communications channel equalization using Radial Basis Function networks", IEEE Transactions on Neural Networks, Vol 4, July 1993.
- [4] D. Tax, D. Ridder and R.P.W. Duin, "Support vectors classifiers: a first look", Proceedings of the Third Annual Conference of the Advanced School for Computing and Imaging, ASCI, DELFT, June 1997
- [5] Cortes, C. and Vapnik, V. "Support Vector Networks", Machine Learning 20:273-297, 1995
- [6] B. Dorizzi, J. C. M. Mota, F. Albu, "A step towards equalisation for radiomobile channel: Neural Networks and Variable Selection", Workshop on "Emerging Techniques for Communication Terminals", Toulouse, France, July 1997
- [7] S. Chen, B. Mulgrew, S. McLaughlin, "Adaptive Bayesian Equalizer with Decision Feedback", IEEE Trans. Signal Processing, 41 (9), pp. 2918-2927, 1993
- [8] R. O. Duda and P.E. Hart, "Pattern Classification and Scene Analysis", Wiley, New York, 1973