

Makkonnell, K. R., and Briu, S. L. *Ekonomiks: Printsipy, problemy i politika* [Economics: Principles, problems and politics]. Vol. 1. Moscow: Respublika, 1992.

Meskon, M. Kh., Albert, M., and Khedouri, F. *Osnovy menedzhmenta* [Fundamentals of management]. Moscow: Delo, 1997.

Pindayk, R., and Rubinfeld, D. *Mikroekonomika* [Microeconomics]. Moscow: Ekonomika; Delo, 1992.

Salyha, K. S. *Efektivnist hospodarskoi diialnosti pidpriemstv* [Efficiency of economic activity of enterprises]. Zaporizhzhia: ZTsNTEI, 2005.

Salyha, K. S. *Ekonomichne obgruntuvannia innovatsiinykh proektiv* [Economic justification of innovation projects]. Zaporizhzhia: KPU, 2010.

Savitskaya, G. V. *Analiz khozyaystvennoy deyatelnosti predpriyatiya* [Analysis of the economic activity of the enterprise]. Minsk: IP «Ekooperspektiva», 1997.

Sheremet, A. D., and Sayfulin, R. S. *Metodika finansovogo analiza* [Methods of financial analysis]. Moscow: INFRA-M, 1996.

УДК 519.86:336.717

ФОРМУВАННЯ АНСАМБЛЮ МОДЕЛЕЙ НА ОСНОВІ СПЕЦІАЛІЗАЦІЇ ЕКСПЕРТІВ ДЛЯ КЛАСИФІКАЦІЇ ПОЗИЧАЛЬНИКІВ – ФІЗИЧНИХ ОСІБ

©2018 БЕНЬ В. П.

УДК 519.86:336.717

Бень В. П. Формування ансамблю моделей на основі спеціалізації експертів для класифікації позичальників – фізичних осіб

Метою статті є дослідження ефективності застосування ансамблевих технологій для розв'язання задачі класифікації позичальників – фізичних осіб за рівнем кредитного ризику. Проаналізовано особливості інформаційного забезпечення процесу розв'язання задачі класифікації позичальників – фізичних осіб. Обґрунтовано необхідність застосування ансамблей моделей як одного із сучасних напрямків обробки великих масивів даних. Наведено авторський варіант алгоритму створення ансамблю на основі спеціалізації окремих моделей-експертів. Запропонований підхід реалізовано у двох варіантах за типами моделей, що використовуються як експерти. За першим варіантом ансамбль складається з логіт-регресій, у другому варіанті – з нейромереж. Аналіз отриманих розрахунків свідчить, що описана ансамблева структура дає змогу підвищити точність оцінювання кредитних ризиків позичальників, причому як окремі моделі-експерти доцільно використовувати нейромережі.

Ключові слова: ансамбль моделей, логіт-регресія, нейромережа.

Табл.: 3. **Бібл.:** 8.

Бень Владислав Петрович – провідний спеціаліст управління по корпоративних правах та інвестиційних проектах, АТ «Мотор Січ» (пр. Моторобудівників, 15, Запоріжжя, 69068, Україна)

E-mail: vlad.ben.1985@gmail.com

УДК 519.86:336.717

Бень В. П. Формирование ансамбля моделей на основе специализации экспертов для классификации заемщиков – физических лиц

Целью статьи является исследование эффективности применения ансамблевых технологий для решения задачи классификации заемщиков – физических лиц относительно уровня кредитного риска. Проанализированы особенности информационного обеспечения процесса решения задачи классификации заемщиков – физических лиц. Обоснована необходимость применения ансамблей моделей как одного из современных направлений обработки больших массивов данных. Приведен авторский вариант алгоритма создания ансамбля на основе специализации отдельных моделей-экспертов. Предложенный подход реализован в двух вариантах по типам моделей, использующихся в качестве экспертов. В первом варианте ансамбль состоит из логит-регрессий, во втором варианте – из нейросетей. Анализ полученных результатов свидетельствует, что описанная ансамблевая структура дает возможность повысить точность оценки кредитных рисков заемщиков, причем в качестве отдельных моделей-экспертов целесообразно использовать нейросети.

Ключевые слова: ансамбли моделей, логит-регрессия, нейросеть.

Табл.: 3. **Библ.:** 8.

Бень Владислав Петрович – ведущий специалист управления по корпоративным правам и инвестиционным проектам, АО «Мотор Сич» (пр. Моторостроителей, 15, Запорожье, 69068, Украина)

E-mail: vlad.ben.1985@gmail.com

UDC 519.86:336.717

Ben V. P. The Formation of Model Assembly on the Basis of Specialization of Experts for Classification of Borrowers – Natural Entities

The article is aimed at researching the efficiency of the use of assembly technologies to solve the task of classifying borrowers – natural entities in relation to the level of credit risk. Features of information support of the process of solving the task of classifying borrowers – natural entities were analyzed. The necessity of application of model assemblies as one of contemporary directions of processing of big data has been substantiated. The author's own variant of algorithm of creation of an assembly on the basis of specialization of separate models-experts has been provided. The proposed approach is implemented in two versions according to the types of models used as experts. In the first variant the assembly consists of logit-regressions, in the second variant – of neural networks. An analysis of the obtained results has proved that the described assembly structure gives an opportunity to increase the accuracy of assessment of the credit risks of borrowers, and it is expedient to use neural networks as separate models-experts.

Keywords: model assemblies, logit-regression, neural network.

Tbl.: 3. **Bibl.:** 8.

Ben Vladyslav P. – Leading Specialist of the Department for Corporate Rights and Investment Projects, Motor Sich JSC (15 Motorobudivnyk Ave., Zaporizhzhia, 69068, Ukraine)

E-mail: vlad.ben.1985@gmail.com

Ідея створення ансамблів (комітетів) моделей пов'язана з проблемою розв'язування задач, що вимагають використання значних масивів даних. У сучасній аналітиці проблеми збереження та оброб-

ки надвеликих обсягів інформації узагальнюються терміном *BigData*. У більшості публікацій, присвячених *BigData*, прийнято визначати основні характеристики такого типу даних принципом «трьохV»: *volume*

(обсяг), *velocity* (швидкість надходження) та *variety* (різноманітність) [1]. Перша характеристика стосується фізичних обсягів інформації, друга розуміється у двох сенсах: як швидкість постійного надходження нової інформації та як необхідність швидкої її обробки. Третя характеристика вказує на необхідність одночасної обробки суттєво різнорідної інформації, яка складається зі структурованих, слабо структурованих або взагалі неструктурованих даних.

Діяльність банківських установ має чітку клієнтоорієнтовану спрямованість. Банки зберігають по кожному клієнту анкетні дані, кредитну історію, історію спілкування з клієнтами, внутрішню інформацію тощо. Крім власної інформації, банківські установи можуть звертатися до зовнішніх бюро кредитних історій, операторів мобільного зв'язку, проводити моніторинг інтернет-активності клієнтів. Накопичення такої кількості різнорідної інформації неодмінно призводить до проблем, пов'язаних з *BigData*. Таку інформацію зазвичай дуже складно ефективно обробити за допомогою однієї моделі. У тих випадках, коли одна модель дає високий рівень якості (прогнозу чи класифікації), мова йде про сильну модель. Під силою моделі розуміють її точність чи ефективність [2]. У більшості випадків спроби описати складні взаємозв'язки у масивах *BigData* різнорідної інформації за допомогою єдиної моделі не дають прийнятної рівня якості результатів. Тоді доводиться звертатися до використання так званих слабких моделей. Одним із підходів підвищення ефективності (підсилення) моделей у такій ситуації є утворення ансамблів моделей.

У задачах класифікації на основі великих масивів даних доцільно застосовувати ансамблі моделей, оскільки таким чином можна одночасно з однією глобальною задачею розв'язати кілька більш простих. Досліджувана проблема або поетапно, або одразу поділяється на сегменти, для кожного з яких розробляється окрема модель, а результати розрахунків моделей об'єднуються певним способом.

Підходи до створення та застосування ансамблевих структур почали розвиватися в кінці минулого століття з роботи Р. Шепайра [3], де було вперше запропоновано ідею бустінгу (підсилення). Успішність подальших модифікацій початкової ідеї бустінгу та створення нових алгоритмів роботи комітетів моделей дали поштовх до використання ансамблів у різних сферах досліджень. Однак лише останнім часом комітети моделей почали використовувати також і для розв'язання задачі класифікації позичальників банку. Тому кількість публікацій за даною темою досить обмежена.

Цікавою є робота [4], в якій описано процедуру розробки ансамблю для розв'язання задачі проведення поведінкового скорингу. У роботі наведено основні аспекти, що впливають на підвищення точності роботи ансамблевих структур, і розглянуто реалізацію

одного з них, а саме – метод узагальнення результатів окремих моделей ансамблю. Отже, залишаються відкритими для дослідження інші питання, зокрема підготовка даних, що будуть навчальними масивами для окремих моделей ансамблю.

На сьогодні розроблено та описано значну кількість різнорідних видів ансамблів, які різняться за алгоритмами реалізації [5–7]. Деякі з них, наприклад бустінг, мають кілька модифікацій процесу побудови ансамблів та вже перетворилися в окремі сімейства алгоритмів. Основні технології формування ансамблів та їх характеристики наведено в *табл. 1*.

Як можна бачити з *табл. 1*, загалом виділяють дві категорії ансамблів: зі статичною та динамічною інтеграцією моделей. При статичній інтеграції дані прикладу, для якого проводиться класифікація, не впливають на процедуру об'єднання результатів розрахунків моделей ансамблю. А при динамічній інтеграції процедура об'єднання результатів кожен раз коригується для проведення класифікації кожного конкретного прикладу. Крім того, алгоритми побудови ансамблів розрізняють також за способами формування навчальних вибірок для окремих моделей та різними схемами, що використовуються у процедурі узагальнення результатів їх роботи.

Технології, розроблені для категорії статичних ансамблів, доцільно використовувати при дослідженні однорідних даних. До цих технологій належать усереднення, стекінг, бустінг, беггінг.

Технологія усереднення результатів роботи моделей ансамблю є найпростішою. Вона передбачає, що всі моделі було побудовано на одній навчальній вибірці, а результат роботи комітету визначається простим голосуванням, тобто більшістю голосів.

Розташування алгоритму стекінгу в *табл. 1* слід вважати дещо умовним, оскільки в [7] зазначається, що загальна концепція використання даного методу відсутня, а його головна ідея знаходить свою реалізацію в різнорідних варіантах. Сутність даного алгоритму полягає в такому:

1) для комбінування в ансамбль застосовуються моделі не одного, а різних типів – наприклад, нейромережа, дерево рішень та логістична регресія;

2) замість алгоритму голосування вводиться концепція метанавчання. На вхід деякої метамоделі, яку називають моделлю першого рівня, подаються результати класифікації кожним експертом, які відносять до нульового рівня.

Далі результати розрахунків моделі першого рівня передаються на модель другого рівня і так далі, доки не буде досягнуто певної умови зупинки процесу.

Для масивів неоднорідних даних окремі їх підмножини можуть краще описуватись різними моделями, тобто побудову моделей-експертів слід здійснювати не для всієї вибірки в цілому, а для окремих

її частин. У таких випадках мова йде про спеціалізацію експертів. Для врахування спеціалізації при формуванні ансамблів використовуються технології динамічної інтеграції моделей.

При розв'язанні задачі класифікації позичальників – фізичних осіб важливо враховувати суттєву неоднорідність досліджуваних даних. Так, клієнти кредитних установ мають значні відмінності за показниками, які використовуються для оцінювання кредитоспроможності. Наприклад, вікова категорія позичальників може змінюватися в межах від 20 до 60 і більше років, рівень освіти – від середньої до наявності двох чи більше вищих освіт і т. п. І різні категорії позичальників характеризуються різним рівнем кредитного ризику. Тому виділення з усього наявного масиву даних більш однорідної вибірки, наприклад із позичальників лише з вищою освітою, дасть змогу досліджувати таку їх групу, яка характеризується більш-менш подібними соціально-економічними умовами існування, що дозволить ефективніше виявляти закономірності поведінки позичальників.

Для обробки таких даних доцільно застосовувати підходи до утворення ансамблів, що враховують спеціалізацію експертів.

Так, у нашому дослідженні [8] для оцінювання рівня кредитоспроможності фізичної особи запропоновано методу побудови ансамблю, що враховує вплив таких якісних показників, як рівень освіти та статус працюючого. За кожним з цих показників із початкового масиву даних утворюється по декілька підгруп. Наприклад, показник «рівень освіти» поділяється на два підрівні: «наявність однієї чи більше вищих освіт» та «наявність середньої та середньої спеціальної освіти». З усього масиву даних обираються лише ті записи, які відповідають позичальникам із певним підрівнем даного якісного показника, утворюючи таким чином два окремі масиви більш однорідних даних. Аналогічна процедура здійснюється за іншим показником (статусом працюючого). Додатково було здійснено поділ початкової вибірки на три підгрупи за такою важливою з точки зору оцінки кредитоспроможності фізичних осіб характеристикою, як наявність утриманців. Для сформованих таким чином масивів даних стає можливою реалізація ідеї спеціалізації експертів.

За запропонованим підходом початковий масив розбивається на підгрупи, які перетинаються між собою. У такому вигляді формування масивів даних для навчання окремих моделей-експертів схоже на відповідну процедуру за технологією беггінгу. Однак за алгоритмом беггінгу вибірки утворюються випадковим чином, а в запропонованому варіанті формування вибірок є логічно обумовленим і приводить до отримання більш однорідних масивів даних.

Очевидно, що можна будувати моделі, використовуючи будь-яку потрібну кількість якісних по-

казників. Узагальнення результатів розрахунків усіх моделей доцільно здійснювати шляхом простого або зваженого голосування. Ваги для моделей відповідатимуть рівню точності класифікації експерта (чим він є вищим, тим більшою є його вага).

Сформована таким чином технологія побудови ансамблю, на нашу думку, найкращим чином пристосована саме для специфіки задачі оцінювання кредитоспроможності позичальників – фізичних осіб. Вона поєднує в собі й ідею спеціалізації експертів, і логічний спосіб формування масивів даних для навчальних вибірок, і класичний спосіб поєднання результатів розрахунків окремих моделей ансамблю.

У даній роботі наведено результати досліджень за даним підходом у двох варіантах. За першим як окремі моделі-експерти застосовуються нейромережі, а за другим – використано *logit*-моделі.

База даних містить інформацію за шістьма характеристиками позичальників (вік, стаж на останньому місці роботи, загальний стаж, кількість депозитів, наявність виплачених у минулому кредитів і кількість дітей у сім'ї) та нараховує 2175 спостережень. Для проведення обчислень використано пакет STATISTICA.

Враховуючи особливості розподілу позичальників між класами та підгрупами в досліджуваній базі даних, у кожній підгрупі виділялося близько 80% записів для навчальної вибірки та 20% – для тестової. Для вибору експертів за кожною підгрупою розглядалися три кращі мережі за показниками точності класифікації на навчальній та тестовій вибірках, а також по узагальнених даних. Результати проведених обчислень наведено в *табл. 2*.

Аналіз даних *табл. 2* вказує, що більш ефективну класифікацію за представленими даними здійснюють мережі з радіально-базисною архітектурою, і лише в одному випадку кращий результат демонструє тришаровий перцептрон (він виявився ефективнішим для масивів меншої розмірності).

Результати розрахунків усіх моделей ансамблю узагальнювалися простим голосуванням. Висновок про точність класифікації, що проведена ансамблем моделей за запропонованою технологією, можна зробити на основі перевірки за усім початковим масивом даних, проводячи порівняння отриманого результату з результатами розрахунків інших комітетів або окремих нейромереж.

За всім масивом даних комітетом було правильно класифіковано 63,5% надійних позичальників та 60% дефолтних.

Реалізований алгоритм побудови ансамблю було також застосовано для випадку використання як експертів *logit*-моделей. По кожній підгрупі даних проведено розрахунки параметрів *logit*-регресій на основі лише тієї частини даних, яка була використана як навчальні вибірки для нейронних мереж. У резуль-

Показники ефективності нейромереж різної архітектури для вибору окремих моделей-експертів при формуванні ансамблю

Архітектура мережі, кількість входів, кількість нейронів проміжного шару	Відсоток правильно класифікованих спостережень у навчальній вибірці, %	Відсоток правильно класифікованих спостережень у тестовій вибірці, %	Узагальнені дані по всьому масиву (без поділу на навчальну та тестову вибірки)		Правильно класифікованих, %
			Клас	Усього	
<i>Підгрупа позичальників із «наявністю однієї чи більше вищих освіт», відібраних за показником «рівень освіти»</i>					
Тришаровий перцептрон, 6, 6	64	58	0	439	61,73
			1	297	60,27
<i>Підгрупа позичальників із «наявністю середньої та спеціальної середньої освіти», відібраних за показником «рівень освіти»</i>					
Радіально-базисна, 6, 7	62	61	0	692	63,29
			1	747	60,37
<i>Підгрупа позичальників із «власною справою», відібраних за показником «статус працівника»</i>					
Радіально-базисна, 6, 6	63	58	0	76	64,47
			1	69	56,52
<i>Підгрупа «найманих працівників», відібраних за показником «статус працівника»</i>					
Радіально-базисна, 6, 16	61	59	0	809	58,96
			1	738	60,98
<i>Підгрупа позичальників із «іншим статусом», відібраних за показником «статус працівника»</i>					
Радіально-базисна, 6, 3	60	56	0	246	58,65
			1	237	54,88
<i>Підгрупа даних «відсутні», відібраних за показником «наявність утриманців»</i>					
Радіально-базисна, 5, 15	63	59	0	491	60,29
			1	489	61,55
<i>Підгрупа даних «одна особа», відібраних за показником «наявність утриманців»</i>					
Радіально-базисна, 5, 5	59	58	0	435	57,93
			1	355	58,59
<i>Підгрупа даних «дві та більше особи», відібраних за показником «наявність утриманців»</i>					
Радіально-базисна, 5, 6	68	61	0	205	68,29
			1	200	60,5

таті отримано вісім логістичних регресій, показники точності яких наведено в табл. 3.

Аналіз даних табл. 3 свідчить, що всі моделі, крім побудованої на даних по статусу працівника «власна справа», є статистично значущими на навчальних вибірках (масив навчальних даних для вказаної моделі є найменшим у дослідженні та складає лише 110 прикладів). Точність класифікації на навчальних вибірках для всіх моделей майже однакова із дещо вищою ефективністю визначення ненадійних позичальників. На тестових вибірках у більшості моделей спостерігається значне зниження точності класифікації. Отримані результати свідчать про недоцільність застосування logit-регресій для розв'язання задачі класифікації за відсутності достатньої кількості навчальних даних. Узагальнений результат розрахунків цих моделей за описаним вище алгоритмом побудо-

ви ансамблю продемонстрував точність передбачення надійних позичальників на рівні 56,6%, а дефолтних – 61,8%.

ВИСНОВКИ

Популярні сьогодні алгоритми бустінгу та бегінгу доцільно застосовувати для однорідних масивів початкових даних. Тоді як при класифікації позичальників – фізичних осіб початкові дані є суттєво неоднорідними. Для таких масивів кращі результати дають ансамблі зі спеціалізацією експертів. Підсумовуючи результати експериментальних розрахунків, можна зробити висновок, що комітет моделей, який був утворений на основі використання logit-регресій, недоцільно використовувати для розв'язання поставленої задачі класифікації позичальників – фізичних осіб за рівнем кредитного ризику. Вищу точ-

Показники точності logit-регресій

Показник, що використо- вувався для фор- мування підгрупи даних	Навчальна вибірка		Тестова вибірка		Значення χ^2
	Відсоток правильно класифікованих значень «0»	Відсоток правильно класифікованих значень «1»	Відсоток правильно класифікованих значень «0»	Відсоток правильно класифікованих значень «1»	
Рівень освіти – «наявність одної чи більше вищих освіт»	60,8	66,3	31	67	47,561
Рівень освіти – «наявність середньої та спеціальної середньої освіти»	59,6	66,8	62	48	98,497
Статус працівника – «власна справа»	60	58,2	45	48	16,164
Статус працівника – «найманий працівник»	62,5	67,0	45	56	124,11
Статус працівника – «інший статус»	56,3	62,6	56	65	29,389
Наявність утриман- ців – «відсутні»	58,8	66,5	52	30	67,646
Наявність утриман- ців – «одна особа»	58,5	62,0	40	30	37,573
Наявність утриман- ців – «дві та більше особи»	59,4	69,4	11	78	36,753

ність класифікації дає комітет моделей, отриманий на основі нейромереж. Запропонований підхід, крім розв'язування задачі класифікації позичальників – фізичних осіб, також може бути застосований для випадків, коли початковий масив даних має невелику розмірність і більшість алгоритмів формування ансамблів не даватимуть задовільних результатів. Такі задачі виникатимуть, зокрема, при дослідженнях нових банківських продуктів, які ще не мають широкого розповсюдження, а, отже, обсяги статистичної інформації за ними недостатні для застосування класичних технологій формування ансамблів. ■

ЛІТЕРАТУРА

1. Big Data в Райффайзенбанке. URL: <https://habrahabr.ru/company/raiffeisenbank/blog/332496/>
2. Вапник В. Н., Червоненкіс А. Я. Теория распознавания образов (статистические проблемы обучения). М.: Наука, 1974. 416 с.
3. Schapire, R. E. Theoretical Views of Boosting and Applications // Algorithmic Learning Theory, 10th International Conference (Tokyo, Japan). December 1999. Proceedings of Algorithmic Learning Theory, 1999. URL: <http://rob.schapire.net/papers/Schapire99d.pdf>
4. Кузнецов И. А., Киреев В. С. Разработка ансамбля алгоритмов классификации с использованием энтропийного показателя качества для решения задачи поведенче-

ского скоринга // Труды XVIII Международной конференции DAMDID/RCDL'2016 «Аналитика и управление данными в областях с интенсивным использованием данных» (Ершово, 11–14 октября 2016). URL: <http://ceur-ws.org/Vol-1752/paper07.pdf>

5. Хайкин С. Нейронные сети: полный курс. 2-е изд./ пер. с англ. М.: Издательский дом «Вильямс», 2006. 1104 с.

6. Научная сессия МИФИ-2007. IX Всероссийская научно-техническая конференция «Нейроинформатика-2007»: Лекции по нейроинформатике. Часть 2. М.: МИФИ, 2007. 148 с.

7. Паклин Н. Б., Орешков В. П. Бизнес-аналитика: от данных к знаниям: учебное пособие. 2-е изд. испр. СПб: Питер. 2013. 704 с.

8. Савіна С. С., Бень В. П. Об'єднання моделей logit-регресій як комітету експертів для оцінки кредитоспроможності позичальника. *Нейро-нечіткі технології моделювання в економіці*. 2015. № 4. С. 154–188.

Науковий керівник – Матвійчук А. В., доктор економічних наук, професор, професор кафедри економіко-математичного моделювання ДВНЗ «КНЕУ імені Вадима Гетьмана» (м. Київ)

REFERENCES

- “Big Data v Rayffayzenbanke” [Big Data in Raiffeisenbank]. <https://habrahabr.ru/company/raiffeisenbank/blog/332496/>
Khaykin, S. *Neironnyye seti* [Neural networks]. Moscow: Vilyams, 2006.

Kuznetsov, I. A., and Kireyev, V. S. "Razrabotka ansamblya algoritmov klassifikatsii s ispolzovaniyem entropiynogo pokazatelya kachestva dlya resheniya zadachi povedencheskogo skoringa" [Development of an ensemble of classification algorithms using the entropy quality index for solving the problem of behavioral scoring]. DAMDID/RCDL'2016 «Analitika i upravleniye dannymi v oblastiakh s intensivnym ispolzovaniyem dannykh». <http://ceur-ws.org/Vol-1752/paper07.pdf>

Nauchnaya sessiya MIFI-2007. IX Vserossiyskaya nauchno-tekhnicheskaya konferentsiya «Neyroinformatika-2007»: Lektsii po neyroinformatike [Scientific session of MEPI-2007. IX All-Russian Scientific and Technical Conference "Neuroinformatics-2007": Lectures on Neuroinformatics]. Part 2. Moscow: MIFI, 2007.

Paklin, N. B., and Oreshkov, V. P. *Biznes-analitika: ot dannykh k znaniyam* [Business Intelligence: from data to knowledge]. St. Petersburg: Piter, 2013.

Schapire, R. E. "Theoretical Views of Boosting and Applications". Proceedings of Algorithmic Learning Theory, 1999. <http://rob.schapire.net/papers/Schapire99d.pdf>

Savina, S. S., and Ben, V. P. "Obiednannia modelei logit-rehresii yak komitetu ekspertiv dlja otsinky kredytopromozhnosti pozychalnyka" [Combining logit regression models as a committee of experts to assess the borrower's creditworthiness]. *Neiro-nechitki tekhnolohii modeliuвання v ekonomitsi*, no. 4 (2015): 154-188.

Vapnik, V. N., and Chervonenkis, A. Ya. *Teoriya raspoznavaniya obrazov (statisticheskiye problemy obucheniya)* [Theory of pattern recognition (statistical learning problems)]. Moscow: Nauka, 1974.