

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

ТЕХНОЛОГИИ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ

TECHNOLOGIES OF INFORMATION TRANSMISSION AND PROCESSING

МАТЕРИАЛЫ МЕЖДУНАРОДНОГО НАУЧНО-ТЕХНИЧЕСКОГО СЕМИНАРА
(Республика Беларусь, Минск, март – апрель 2022 г.)

Минск БГУИР 2022

УДК 621.391+004.04
ББК 32.811
Т38

Руководитель семинара В. Ю. Цветков

Научный программный комитет:

В. Ю. Цветков, А. А. Борискевич, В. К. Конопелько, В. В. Муравьев,
Н. И. Листопад, Б. В. Никульшин, И. С. Азаров, А. А. Дудкин, С. А. Золотой,
А. А. Пилюшко, С. Н. Касанин, Ю. Ю. Бобков, А. С. Войтенков,
С. М. Костромицкий, Ю. В. Бруттан, О. Дж. М. Аль-Фурайджи, Фам Хак Хоан

Т38 **Технологии** передачи и обработки информации : материалы
междунар. науч.-техн. семинара (Республика Беларусь, Минск,
март – апрель 2022 г.) = Technologies of Information Transmission and
Processing. – Минск : БГУИР, 2022. – 108 с. : ил.
ISBN 978-985-543-649-3.

Сборник содержит статьи, тематика которых посвящена научно-теоретическим разработкам в области сетей телекоммуникаций, информационной безопасности, технологий передачи и обработки информации.

Предназначен для научных сотрудников в области инфокоммуникаций, преподавателей, аспирантов, магистрантов и студентов технических вузов.

Научное издание

Корректор *В. В. Чепикова*

Ответственный за выпуск *В. Ю. Цветков*

Компьютерный дизайн и верстка *Е. Г. Макейчик*

Подписано в печать 04.05.2022. Формат 60×84 1/8. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 12,8. Уч.-изд. л. 10,9. Тираж 30 экз. Заказ 60.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».

Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,
№2/113 от 07.04.2014, №3/615 от 07.04.2014.

Ул. П. Бровки, 6, 220013, г. Минск

СОДЕРЖАНИЕ

С.И. Рудиков, В.Ю. Цветков, А.П. Шкадаревич Сжатие динамического диапазона инфракрасных изображений с адаптивным выравниванием и инверсией краев гистограммы	5
Х.А. Фам, С.Б. Саломатин Алгоритмы предварительного кодирования с учетом состояния канала связи для беспроводных систем ММО	12
В.Н. Кийко, Н.М. Наумович, М.В. Давыдов, В.И. Журавлев Влияние вибраций на фазовые шумы кварцевых генераторов	17
U.A. Vishnyakou, Hu Zhifeng, Fengshou Gu Development of IOT network for production quality control	21
Y. Xueying, I.A. Baryskievič Human activity recognition based on AdaBoost ensemble classifier	25
С.Х. Жэнь, И.М. Чэнь, В.К. Конопелько, В.Ю. Цветков Формирование образов ошибок для декодирования итеративных кодов	31
В.А. Аксенов, Е.М. Гнедько Технология IAB для гетерогенной радиосети 5G NR	35
Н.Д. Рудниченко, В.В. Вычужанин, А.А. Егошина, С.М. Вороной, Н.О. Шibaева Интеллектуальная система автоматизации оценивания тематик текстового контента	40
М.А. Алисиенко, А.П. Турлай, С.Б. Саломатин Вероятностное декодирование кодов с высокой плотностью проверок на четность с использованием алгоритма распространения доверия	46
В.В. Рабцевич, В.Ю. Цветков Уровневые модели и волновые алгоритмы сегментации АСМ-изображений	50
А.И. Наумович Распознавание патологий грудной клетки с помощью предварительного обучения на общедоступных наборах данных	57
Liu Zhenghua, U.A. Vishniakou Creating database IoT for product quality control	62
Zhao Di Sensor data frequency feature extraction	66
В.В. Мацкевич Возможности метода отжига в задаче обучения нейронных сетей	69
Y.P.Y. Qiu, J. Ma, Z.M. Liao Object recognition based on skeleton	74
Y. Xueying Human activity recognition based on random forest	77
Y.M. Chen, X.H. Ren Different applications and implements of the Hamming product codes	82
Zhao Di Multi-class classification SVM methods	85
D.V. Marchenko, Mbuyi Michelle Mulumba, Kuruma Mohammed, Koua Leonel Ignas Informational linear capacity of fiber-optic backbones of the countries of the west coast of Africa	88
Z.M. Liao, J. Ma, Y.P.Y. Qiu Research on skeleton matching classification	93
U.A. Vishniakou, D. Zongqi Optimizing smart home network protocol selection by hierarchy analysis method	96
Liu Zhenghua Structural design of the data base on the internet of things for product quality control	100

CONTENTS

S.I. Rudikov, V.Yu. Tsviatkou, A.P. Shkadarevich Dynamic range compression of infrared images with adaptive equalization and histogram edge inversion	5
H.A. Fam, S.B. Salomatin Pre-coding algorithms considering communication channel state for wireless MIMO systems	12
V.N. Kiyko, N.M. Naumovitch, M.V. Davydov, V.I. Zhuravlev Influence of vibrations on phase noises of quartz generators	17
U.A. Vishnyakou, Hu Zhifeng, Fengshou Gu Development of IoT network for production quality control	21
Y. Xueying, I.A. Baryskievič Human activity recognition based on AdaBoost ensemble classifier	25
X.H. Ren, Y.M. Chen, V.K. Konopelko, V.Yu. Tsviatkou Classify error patterns for decoding product codes	31
V.A. Aksyonov, E.M. Gnedzko IAB Technology for 5G NR hetnet	35
N.D. Rudnichenko, V.V. Vychuzhanin, A.A. Egoshina, S.M. Voronoy, N.O. Shibaeva Intellectual system of textual content subject estimation automatization	40
M.A. Alisienko, A.P. Turlay, S.B. Salomatin Probabilistic decoding of codes with a high density of parity checks using the belief propagation algorithm	46
V.V. Rabtsevich, V.Yu. Tsviatkou Level models and wave algorithms for segmentation of AFM images	50
A.I. Naumovich Chest pathologies recognition using pre-training on public datasets	57
Liu Zhenghua, U.A. Vishniakou Creating database IoT for product quality control	62
Zhao Di Sensor data frequency feature extraction	66
V.V. Matskevich Annealing method opportunities in neural networks training problem	69
Y.P.Y. Qiu, J. Ma, Z.M. Liao Object recognition based on skeleton	74
Y. Xueying Human activity recognition based on random forest	77
Y.M. Chen, X.H. Ren Different applications and implements of the Hamming product codes	82
Zhao Di Multi-class classification SVM methods	85
D.V. Marchenko, Mbuyi Michelle Mulumba, Kuruma Mohammed, Koua Leonel Ignas Informational linear capacity of fiber-optic backbones of the countries of the west coast of Africa	88
Z.M. Liao, J. Ma, Y.P.Y. Qiu Research on skeleton matching classification	93
U.A. Vishniakou, D. Zongqi Optimizing smart home network protocol selection by hierarchy analysis method	96
Liu Zhenghua Structural design of the data base on the internet of things for product quality control	100

УДК 004.932

СЖАТИЕ ДИНАМИЧЕСКОГО ДИАПАЗОНА ИНФРАКРАСНЫХ ИЗОБРАЖЕНИЙ С АДАПТИВНЫМ ВЫРАВНИВАНИЕМ И ИНВЕРСИЕЙ КРАЕВ ГИСТОГРАММЫ

С.И. РУДИКОВ¹, В.Ю. ЦВЕТКОВ², А.П. ШКАДАРЕВИЧ¹¹ Унитарное предприятие «НТЦ «ЛЭМТ» БелОМО», Республика Беларусь² Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 28 февраля 2022

Аннотация. Предложена модификация алгоритма адаптивной эквализации, растяжения и сжатия гистограммы, основанная на инверсии ее обрезанных краев. При сжатии гистограммы инверсия краев позволяет сохранить корреляцию значений и контраст для большей части смежных пикселей изображения. Показано, что при уменьшении динамического диапазона ландшафтных ИК-изображений предложенный алгоритм повышает блочный контраст по сравнению с алгоритмом адаптивной эквализации гистограммы.

Ключевые слова: сжатие динамического диапазона изображений, инфракрасные изображения, выравнивание гистограммы, инверсия гистограммы.

Введение

Для сжатия динамического диапазона и улучшения качества ландшафтных инфракрасных (ИК) изображений используются блочные алгоритмы, позволяющие адаптироваться к структуре изображения. Блочный алгоритм адаптивного выравнивания гистограммы (Adaptive Histogram Equalization, АНЕ) [1] на основе интегральной функции распределения яркости обеспечивает достаточно высокое качество изображений после преобразования, но не позволяет управлять формой глобальной гистограммы яркости. Это может приводить к недостаточному или чрезмерному локальному и глобальному контрасту преобразованного изображения. Данный недостаток свойственен многим модификациям алгоритма АНЕ [2–5]. Причина заключается в блочном выравнивании гистограммы, существенно ослабляющем результаты предкоррекции изображения. Поэтому эффективное управление формой глобальной гистограммы ИК-изображения при преобразовании динамического диапазона возможно только после блочного выравнивания гистограммы. Это реализовано в алгоритме адаптивной эквализации, растяжения и сжатия гистограммы (Histogram Equalization, Compression and Stretching, HECS) [6], который превосходит алгоритм АНЕ по контрасту за счет за счет обрезки краев, растяжения центральной части, растяжения (сжатия) и наложения обрезанных краев глобальной гистограммы. Недостатком алгоритма HECS является ухудшение контраста по краям динамического диапазона, что приводит к заметным артефактам на преобразованном ИК-изображении и дополнительной неравномерности его глобальной гистограммы.

Целью работы является повышение качества воспроизведения ИК-изображений при сжатии их динамического диапазона.

Постановка задачи

На рис. 1, *а* приведено ИК-изображение с широким динамическим диапазоном и его гистограмма. Эквализация (Histogram Equalization, HE) [7] выравнивает гистограмму ИК-изображения и позволяет получить ИК-изображение с узким динамическим диапазоном, но низким локальным контрастом (рис. 1, *б*). Для повышения локального контраста используется

алгоритм АНЕ [1], который делит исходное ИК-изображение на блоки пикселей, в пределах которых осуществляется выравнивание гистограмм. Блочный эффект уменьшается с помощью интерполяции. Разделение ИК-изображения на блоки в АНЕ позволяет адаптироваться к структуре изображения, но не обеспечивает равномерность глобальной гистограммы и возможность управления ее формой (рис. 1, *в*).

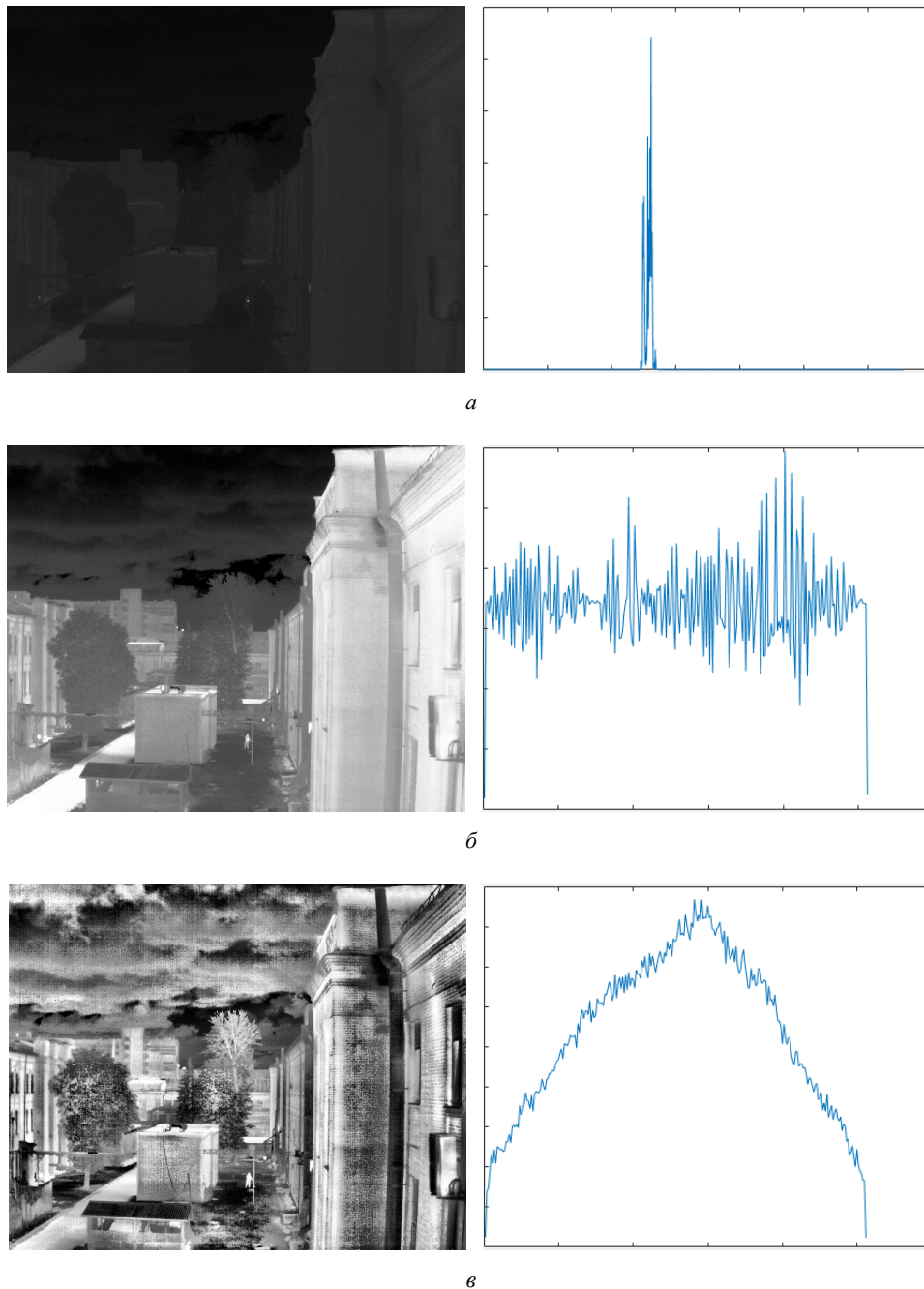


Рис. 1. ИК-изображения и их гистограммы: *а* – исходное; *б* – преобразованное с помощью НЕ; *в* – преобразованное с помощью АНЕ

Для повышения качества и расширения интервала управления характеристиками воспроизведения ИК-изображений при сжатии их динамического диапазона алгоритм HECS адаптивной эквализации, растяжения и сжатия гистограммы реализует двойное уменьшение динамического диапазона ИК-изображения с промежуточным управлением формой его гистограммы за счет ее частичного растяжения и сжатия [6]. Алгоритм реализуется в три этапа:

1. Предварительное уменьшение динамического диапазона $[0, L_{HDR} - 1]$ исходного ИК-изображения $I_{HDR} = \parallel i_{HDR}(y, x) \parallel_{(y=0, Y-1, x=0, X-1)}$ до $[0, L_{LDR1} - 1]$, $L_{HDR} < L_{LDR1}$ (этап 1) на основе адаптивной эквализации гистограммы – формируется ИК-изображение $I_{LDR1}(L_{LDR1}) = \parallel i_{LDR1}(y, x) \parallel_{(y=0, Y-1, x=0, X-1)}$ с динамическим диапазоном $[0, L_{LDR1} - 1]$.

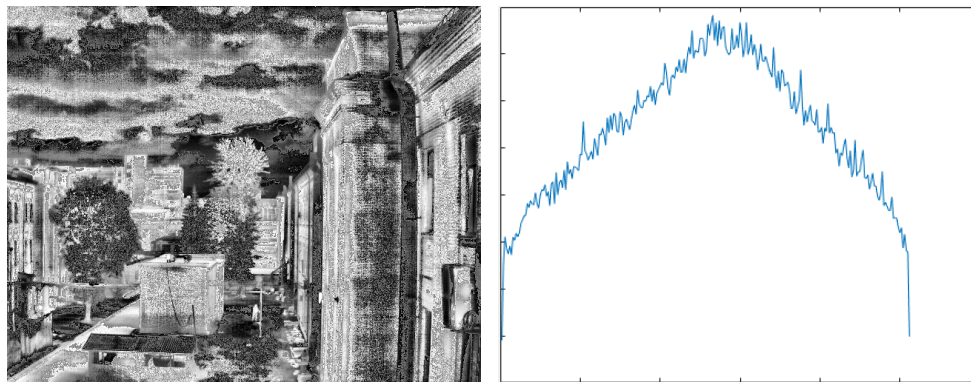
2. Управление формой гистограммы (этап 2):

2.1. Сжатие гистограммы на основе двух гамма-функций с коэффициентами G_L и G_H , применяемых к двум частям динамического диапазона $[0, T_G - 1]$ и $[T_G + 1, L_{LDR1} - 1]$, разделяемым на уровне $T_G = K_G(L_{LDR1} - 1)$ по коэффициенту K_G .

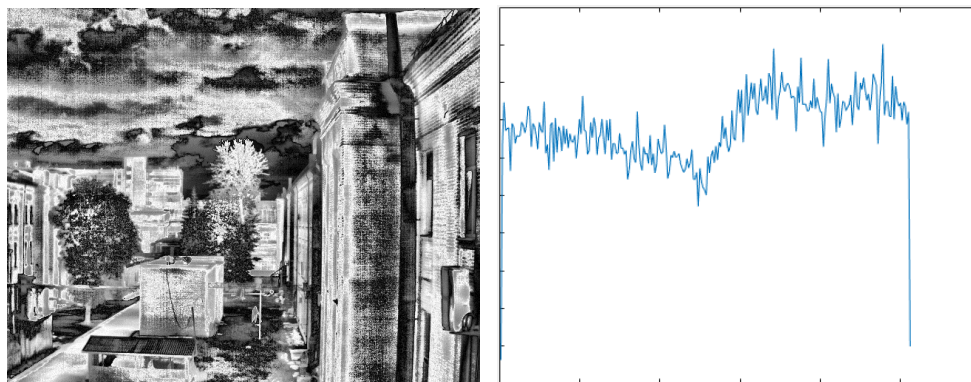
2.2. Линейное растяжение центральной части гистограммы, ограниченной динамическим диапазоном $[T, L_{LDR1} - 1 - T]$ по порогу T , с коэффициентом $(L_{LDR1} - 1)/(L_{LDR1} - 1 - 2T)$ и линейное растяжение (сжатие) боковых частей гистограммы, ограниченных диапазонами $[0, T - 1]$ и $[L_{LDR1} - T, L_{LDR1} - 1]$ по порогу T с коэффициентами T_L/T и T_R/T , где T_L, T_R – пороги.

3. Окончательное уменьшение динамического диапазона до $[0, L_{LDR2} - 1]$ на основе линейного сжатия гистограммы с коэффициентом L_{LDR2}/L_{LDR1} (этап 3).

Преобразованное с помощью алгоритма HECS ИК-изображение с узким динамическим диапазоном и его гистограмма представлены на рис. 2, а.



а



б

Рис. 2. Преобразованные ИК-изображения и их гистограммы: а – с помощью HECS; б – с помощью HECSm

В табл. 1 и 2 приведены выигрыши алгоритма HECS в сравнении с алгоритмом АНЕ в процентах по контрастности (стандартному отклонению) D_{ST} , среднему градиенту G_A , энтропии E_I и количеству локальных экстремумов N_{LE} , усредненные по 94 ИК-изображениям, разделенным на 6 типов в зависимости от формы гистограммы после адаптивной эквализации

(рис. 3): 11 ИК-изображений типа 1; 17 – типа 2; 31 – типа 3; 27 – типа 4; 5 – типа 5; 3 – типа 6. Характеристики D_{ST} , G_A , E_I являются средними по блокам 64×64 пикселей. Из табл. 1 и 2 следует, что алгоритм HECS превосходит алгоритм АНЕ по характеристикам D_{ST} , G_A , E_I и N_{LE} . Однако, как видно на рис. 1, *в* и рис. 2, *а* глобальная гистограмма ИК-изображения после преобразования с помощью алгоритма HECS выравнивается незначительно в сравнении с алгоритмом АНЕ. Таким образом, актуальной является задача модификации алгоритма HECS для повышения равномерности глобальной гистограммы.

Табл. 1. Выигрыши алгоритма HECS в сравнении с алгоритмом АНЕ при размере блока 32×32

Характеристика	Выигрыши алгоритма HECS в сравнении с АНЕ по типам ИК-изображений					
	Тип 1	Тип 2	Тип 3	Тип 4	Тип 5	Тип 6
D_{ST}	6,06	9,95	10,04	9,72	7,59	6,95
G_A	13,78	18,23	18,92	20,58	9,42	11,28
E_I	13,38	36,25	34,84	24,41	39,16	20,10
N_{LE}	13,38	9,97	9,17	17,22	4,26	12,97

Табл. 2. Выигрыши алгоритма HECS в сравнении с алгоритмом АНЕ при размере блока 64×64

Характеристика	Выигрыши алгоритма HECS в сравнении с АНЕ по типам ИК-изображений					
	Тип 1	Тип 2	Тип 3	Тип 4	Тип 5	Тип 6
D_{ST}	12,25	11,33	8,89	7,03	0,11	2,33
G_A	35,39	47,91	49,97	50,54	13,39	48,73
E_I	31,41	50,54	51,54	38,39	55,92	41,77
N_{LE}	14,68	14,89	14,11	25,94	6,88	18,11

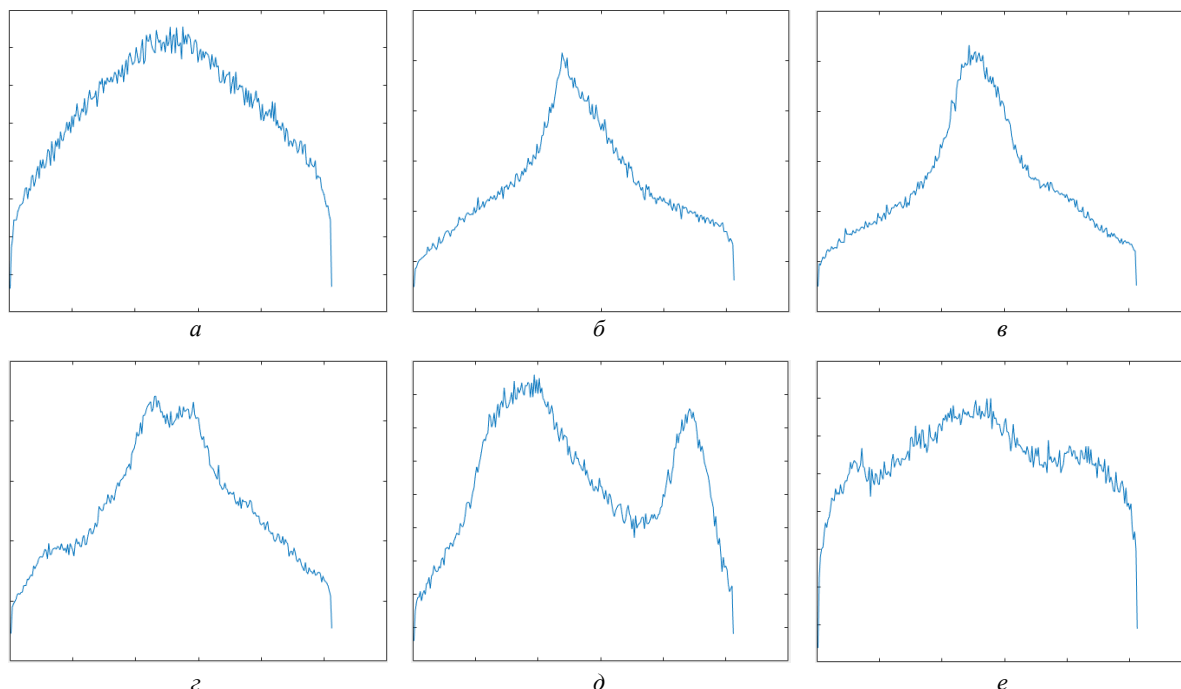


Рис. 3. Гистограммы ИК-изображений после адаптивного выравнивания: *а* – тип 1; *б* – тип 2; *в* – тип 3; *г* – тип 4; *д* – тип 5; *е* – тип 6

Адаптивная эквализация, растяжение и сжатие гистограммы с инверсией и наложением обрезанных краев

Для повышения глобального контраста ИК-изображений при уменьшении их динамического диапазона предлагается модификация алгоритма HECS адаптивной эквализации, растяжения и сжатия гистограммы, основанная на инверсии обрезанных краев глобальной

гистограммы. Сущность модификации состоит в преобразовании (линейном растяжении (сжатии) с коэффициентами T_L/T , T_R/T и инверсии) значений пикселей из диапазонов $[0, T-1]$ и $[L_{LDR1}-T, L_{LDR1}-1]$ с помощью выражения

$$\tilde{i}_{LDR1}(y, x) = \begin{cases} \left\lfloor \frac{T_L}{T} i_{LDR1}(y, x) - T_L \right\rfloor & \text{при } i_{LDR1}(y, x) < T, \\ (L_{LDR1}-1) - \left(T_R + \frac{T_R}{T} (i_{LDR1}(y, x) - (L_{LDR1}-1)) \right) & \text{при } i_{LDR1}(y, x) > L_{LDR1}-1-T \end{cases} \quad (1)$$

при $y = \overline{0, Y-1}$, $x = \overline{0, X-1}$.

На рис. 4 показан процесс преобразования гистограммы ИК-изображения в соответствии с предложенной модификацией алгоритма HECS. Сначала, с помощью порога T глобальная гистограмма ИК-изображения разделяется на три части: центральную и две крайних. Затем, центральная часть растягивается на весь динамический диапазон, а крайние – в соответствии с коэффициентами растяжения (сжатия) T_L/T и T_R/T . Кроме того, крайние фрагменты гистограммы инвертируются. Растянутая центральная часть, сжатые (растянутые) и инвертированные крайние фрагменты гистограммы складываются. Результирующая гистограмма выравнивается за счет увеличения вероятностей темных и светлых пикселей. Ее формой можно управлять с помощью порогов T, T_L, T_R .

Инверсия краев гистограммы позволяет сохранить корреляцию значений и контраст для большей части смежных пикселей изображения. На рис. 5 приведен фрагмент исходной гистограммы, разделенный на 4 диапазона, из которой формируются два фрагмента в результате обрезки краев, растяжения центральной части гистограммы и наложения обрезанных краев без инверсии и с инверсией. Из рис. 5 следует, что при наложении краев гистограммы без инверсии контрастные диапазоны 1–3 и 2–4 исходной гистограммы становятся неконтрастными после преобразования. При наложении краев гистограммы с инверсией контрастные диапазоны 1–3 и 2–4 исходной гистограммы сохраняют относительный контраст. Потеря контраста диапазонов 2–3 и 1–4 менее значима, поскольку диапазоны 2–3 соседние, а диапазон 1 является крайним в исходной гистограмме и включает относительно небольшое количество пикселей. Данное свойство предложенного преобразования подтверждает рис. 2, б, на котором приведены преобразованное с помощью предложенного модифицированного алгоритма ИК-изображение и его гистограмма.

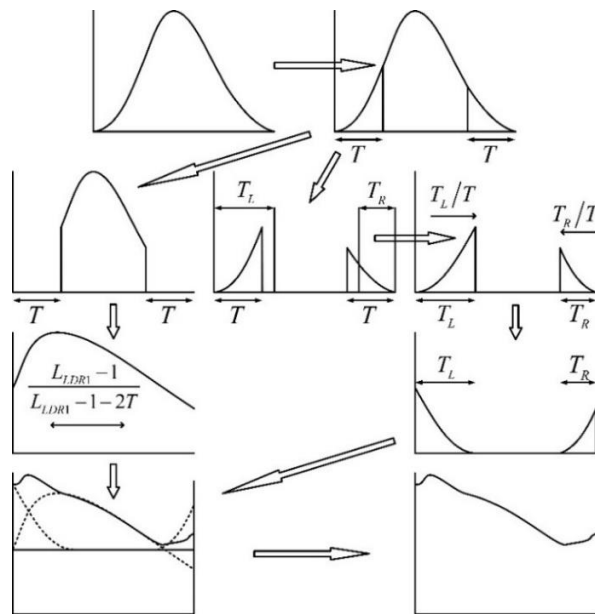


Рис. 4. Процесс растяжения и сжатия гистограммы с зеркальным отображением обрезанных краев глобальной гистограммы

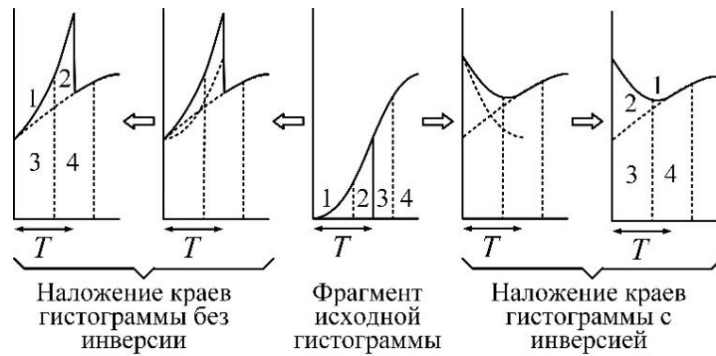


Рис. 5. Обрезка краев, растяжение центральной части гистограммы и наложение обрезанных краев без инверсии и с инверсией

Оценка эффективности алгоритмов сжатия гистограммы

В табл. 3 и 4 приведены выигрыши модифицированного алгоритма HECSm в сравнении с алгоритмом АНЕ в процентах по контрастности D_{ST} , среднему градиенту G_A , энтропии E_I и количеству локальных экстремумов N_{LE} , усредненные блокам 64×64 пикселей и 94 ИК-изображениям, разделенным на 6 типов в зависимости от формы гистограммы после адаптивной эквализации (см. рис. 3).

Табл. 3. Выигрыши алгоритма HECSm в сравнении с алгоритмом АНЕ при размере блока 32×32

Характеристика	Выигрыши алгоритма HECS в сравнении с АНЕ по типам ИК-изображений					
	Тип 1	Тип 2	Тип 3	Тип 4	Тип 5	Тип 6
D_{ST}	13,45	16,95	13,32	15,48	13,88	8,33
G_A	19,01	36,61	23,22	22,69	13,76	11,53
E_I	31,12	59,73	46,67	27,96	44,16	21,15
N_{LE}	7,65	10,60	7,98	12,99	3,05	10,04

Табл. 4. Выигрыши алгоритма HECSm в сравнении с алгоритмом АНЕ при размере блока 64×64

Характеристика	Выигрыши алгоритма HECS в сравнении с АНЕ по типам ИК-изображений					
	Тип 1	Тип 2	Тип 3	Тип 4	Тип 5	Тип 6
D_{ST}	23,12	24,02	19,86	16,62	13,21	12,57
G_A	41,71	46,22	46,01	42,52	19,03	35,80
E_I	41,35	57,97	59,24	41,06	59,23	40,91
N_{LE}	11,31	9,35	9,64	17,79	4,85	20,00

Из табл. 1–4 следует, что для ИК-изображений всех типов при любых размерах блоков алгоритм HECSm превосходит алгоритм АНЕ по характеристикам D_{ST} (на 13,6 % при размере блока 32×32 пикселей, на 18,2 % при размере блока 64×64 пикселей), G_A (на 21,1 % при размере блока 32×32 пикселей, на 38,6 % при размере блока 64×64 пикселей), E_I (на 38,5 % при размере блока 32×32 пикселей, на 50,0 % при размере блока 64×64 пикселей) и N_{LE} (на 8,7 % при размере блока 32×32 пикселей, на 12,2 % при размере блока 64×64 пикселей). При этом алгоритм HECSm превосходит алгоритм-прототип HECS по контрасту для ИК-изображений всех типов при всех размерах блоков (в 1,6 раза при размерах блоков 32×32 пикселей, в 2,6 раза при размерах блоков 64×64 пикселей); по характеристикам G_A (в 1,4 раза) и E_I (в 1,4 раза) при размерах блоков 32×32 пикселей; по характеристике G_A при размерах блоков 64×64 пикселей для изображений типов 1 и 5; по характеристике E_I при размерах блоков 64×64 пикселей для изображений типов 1–5.

Заключение

Предложен модифицированный алгоритм адаптивной эквализации, растяжения и сжатия гистограммы, основанный на инверсии ее обрезанных краев. При сжатии гистограммы инверсия ее краев позволяет сохранить корреляцию значений и контраст для большей части смежных пикселей изображения. При уменьшении динамического диапазона предложенный алгоритм повышает блочный контраст ландшафтных ИК-изображений на 13,6 % – 18,2 % при размерах блоков от 32×32 до 64×64 пикселей по сравнению с алгоритмом адаптивной эквализации гистограммы.

DYNAMIC RANGE COMPRESSION OF INFRARED IMAGES WITH ADAPTIVE EQUALIZATION AND HISTOGRAM EDGE INVERSION

S.I. RUDIKOV, V.Yu. TSVIATKOU, A.P. SHKADAREVICH

Abstract. A modification of the algorithm of adaptive equalization, stretching and compression of the histogram, based on the inversion of its clipped edges, is proposed. In histogram compression, edge inversion preserves value correlation and contrast for most adjacent pixels in an image. It is shown that with a decrease in the dynamic range of landscape IR images, the proposed algorithm increases the block contrast in comparison with the adaptive histogram equalization algorithm.

Keywords: image dynamic range compression, infrared images, histogram equalization, histogram inversion.

Список литературы

1. Kim T.K., Paik J.K., Kang B.S. // IEEE Transactions on Consumer Electronics. 1998. Vol. 44. No. 1. P. 82–87.
2. Reza A.M. // The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology. 2004. Vol. 38. No. 1. P. 35–44.
3. Kim J.-Y., Kim L.-S., Hwang S.-H. // IEEE Transactions on Circuits and Systems for Video Technology. 2001. Vol. 11. No. 4. P. 475–484.
4. Huang S.-C., Yeh C.-H. // Engineering Applications of Artificial Intelligence. 2013. Vol. 26. No. 5. P. 1487–1492.
5. Al-Sammaraiе M.F. // 10th International Conference on Computer Science and Education (ICCSE). Cambridge. 2015. P. 95–101.
6. Рудиков С.И., Цветков В.Ю., Шкадаревич А.П. // Весці Нацыянальнай акадэміі навук Беларусі. Серыя фізіка-матэматычных навук. 2021. Т. 66. № 4. С. 470–482.
7. Nithyananda C.R., Ramachandra A.C., Preethi. // International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). Chennai. 2016. P. 2512–2517.

АЛГОРИТМЫ ПРЕДВАРИТЕЛЬНОГО КОДИРОВАНИЯ С УЧЕТОМ СОСТОЯНИЯ КАНАЛА СВЯЗИ ДЛЯ БЕСПРОВОДНЫХ СИСТЕМ MIMO

Х.А. ФАМ, С.Б. САЛОМАТИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 28 февраля 2022

Аннотация. В статье рассматриваются алгоритмы линейного предварительного кодирования с использованием информации о состоянии канала на передающей стороне системы MIMO. Рассматривается синтез и эффективность алгоритмов предварительного кодирования, основанные на использовании следующих критериев: взаимной информации, минимума вероятности попарной ошибки, минимума среднеквадратической ошибки.

Ключевые слова: система MIMO, информация о состоянии канала – CSI, пространственно-временное кодирование (STC), идеальная CSI, корреляция замираний CSI, динамическая CSI.

Введение

В современную эпоху беспроводной связи неуклонный рост числа пользователей сотовой связи и их огромные потребности в скорости передачи данных требуют новых достижений в существующей сотовой инфраструктуре. Являясь ключевой технологией для систем беспроводной связи 5-го поколения (5G), технология массового ввода-вывода (MIMO) в настоящее время является одной из привлекательных технологий для будущего беспроводного доступа. Предпосылкой для использования массивного MIMO является то, что базовая станция должна получать информацию о состоянии канала (CSI) канала нисходящей линии связи.

Чтобы удовлетворить требования пользователей сотовой связи, многие методы передачи, такие как битовая загрузка, кодирование, методы предварительного кодирования, адаптивная модуляция, планирование с учетом канала и т.д., должны иметь точную информацию о состоянии канала (CSI) на стороне передатчика, чтобы добиться значительного выигрыша. Кроме того, точная CSI значительно улучшает производительность многих беспроводных технологий, например, множественных входов и множественных выходов (MIMO), сверхнадежных передач, ретрансляции и защиты физического уровня [1–7].

В режиме дуплекса с временным разделением (TDD) базовая станция может получать CSI нисходящей линии связи из канала восходящей линии связи на основе взаимности каналов [2]. В режиме дуплекса с частотным разделением каналов (FDD) взаимность больше не поддерживается, и CSI нисходящей линии связи необходимо оценивать в пользовательском оборудовании на основе пилот-сигнала и возвращать обратно в BS.

Более того, точная CSI позволяет упростить приемник за счет предварительного кодирования MIMO на передатчике, повысить надежность и более высокую пропускную способность канала и т.д. Для получения этих преимуществ без CSI на передатчике не обойтись. Алгоритмы точной CSI, являются альфой и омегой современной инфраструктуры беспроводной связи.

На передающей стороне, двумя основными компонентами обработки MIMO передаваемой информации являются пространственно-временное кодирование и предварительное кодирование (или «прекодирование»). Прекодирование представляет собой метод, использующий информацию о канале, доступную в передатчике. Использование на передающей стороне информации о состоянии канала связи позволяет увеличить скорость передачи

информации, улучшить зону покрытия и уменьшить сложность реализации приемника в беспроводных системах MIMO [1, 2].

В настоящей работе рассматриваются алгоритмы предварительного кодирования, основанные на использовании следующих критериев: взаимной информации, минимума вероятности попарной ошибки, минимума среднеквадратической ошибки. Оценивается эффективность применения предлагаемых алгоритмов.

Модели предварительного кодирования в беспроводных системах MIMO

Существует множество форм представления информации о состоянии канала на передающей стороне:

- точная информация о состоянии канала в каждый момент времени или идеальный информация о состоянии канала (идеальная CSI);
- часто бывает трудно получить в канале с замираниями – информация о корреляции замираний в канале (корреляционная CSI);
- информация о корреляции замираний и о среднем значении матрицы канал (динамическая CSI).

Разработка прекодирования для беспроводной связи MIMO в последние годы была активной областью исследований в настоящее время находит применение в новых беспроводных стандартах.

Структурная схема системы MIMO связи с прекодированием, приведена на рис. 1. На передающей стороне после кодера, сигнал подвергается линейному преобразованию с помощью матрицы F и затем поступает в канал связи MIMO с матрицей канала H . Матрица F линейного преобразования определяет алгоритм линейного прекодирования.

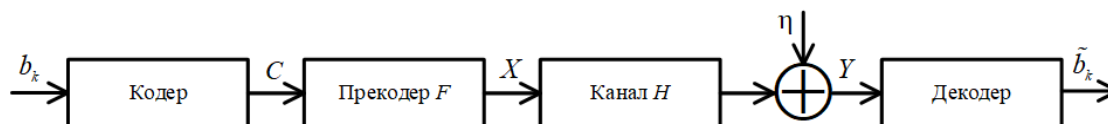


Рис. 1. Структурная схема системы MIMO связи с прекодированием

Кодер содержит блок канального кодирования и отображения символов, который доставляет векторные символы в прекодер. Модель принимаемого сигнала в системе связи MIMO с линейным прекодированием (приведенная на рис. 1) может быть записана в следующем выражении [3–7]:

$$Y = \sqrt{E_s} HFC + \eta, \quad (1)$$

где E_s – мощность передаваемого сигнала, η – аддитивный белый гауссовский шум.

Рассматривается синтез следующих алгоритмов прекодирования, основанный на использовании известных критериев: критерий взаимной информации, критерий минимума вероятности попарной ошибки, критерий минимума среднеквадратической ошибки.

1. Сингулярное разложение прекодирующей матрицы F описывается следующим выражением [3–7]:

$$F = U_F D V_F^H. \quad (2)$$

2. Прекодирование для случая, когда на передающей стороне имеется идеальная CSI. При идеальном CSI канал MIMO использует сингулярное разложение матрицы H [3–7]:

$$H = U_H \sum V_H^H, \quad (3)$$

оптимальные направления лучей – представление канала связи MIMO в виде совокупности параллельных каналов, путем левые сингулярные векторы прекодирующей матрицы F , определяющие направления лучей, должны соответствовать правым сингулярным векторам матрицы H , то есть $U_F = V_H$.

3. Прекодирование для случая, когда на передающей стороне имеется корреляция замираний CSI. Предполагаем, что в канале MIMO имеет корреляция на передающей стороне, а корреляция на приеме отсутствует. Матрица H в этом случае описывается как [3–7]:

$$H = H_{\omega} R_t^{1/2}, \quad (4)$$

оптимальные направления лучей, для всех критериев синтеза алгоритмов прекодирования, совпадают с собственными векторами корреляционной матрицы $R_t : U_F = U_t$, где матрица U_t входит в сингулярное разложение корреляционной матрицы $R_t : R_t = U_t \Lambda_t U_t^H$.

4. Прекодирование для случая, когда на передающей стороне имеется динамическая CSI. Матрица канала H данного случая описывается следующим выражением [3–7]:

$$H = \bar{H} + H_{\omega} R_t^{1/2}, \quad (5)$$

где $\bar{H} = E[H]$ – среднее значение матрицы H .

Решая квазиоптимальных направлений лучей получаем: $U_F = U_R$, где матрица U_R входит в сингулярное разложение матрицы $\bar{H}^H \bar{H} + M R_t$:

$$\bar{H}^H \bar{H} + M R_t = U_R \Lambda_R U_R^H. \quad (6)$$

Моделирование приведенных выше алгоритмов позволило получить следующие графики зависимостей вероятности ошибки на бит (BER) от отношения сигнал/шум (SNR).

Результаты моделирования

На рис. 2 продемонстрированы эффективности вероятности появления ошибки, когда на передающей стороне имеется идеальная CSI. Алгоритмы прекодирования обеспечивают значительный выигрыш, измеренный как в некодированном, так и в кодированном режиме. В некодированном режиме, при вероятности 10^{-6} выигрыш 4 дБ; в кодированном режиме, при вероятности 10^{-5} выигрыш 2 дБ.

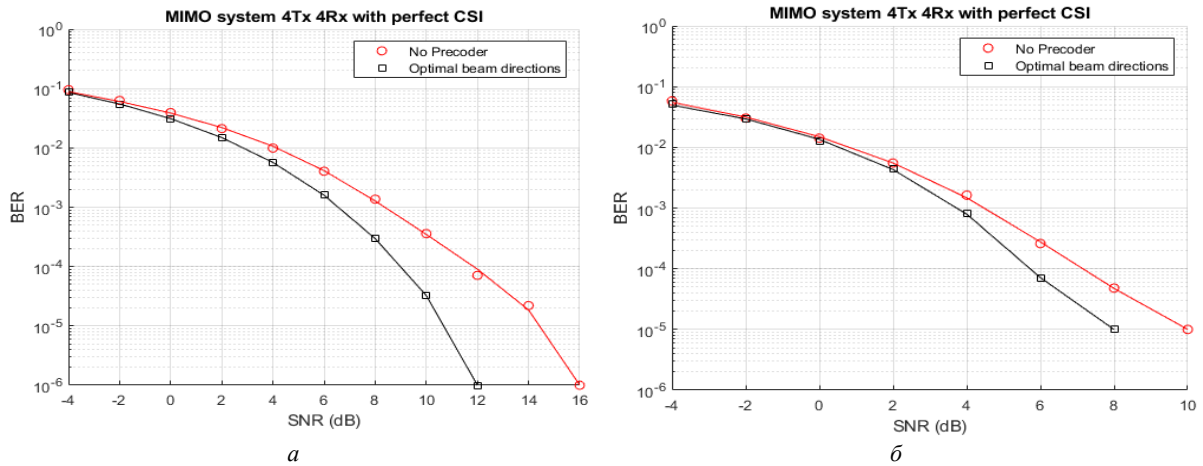


Рис. 2. Эффективность прекодирования в случае на передающей стороне имеет идеальную CSI:
 a – без кодирования; b – с пространственно-временным кодированием

На рис. 3 продемонстрированы эффективности вероятности появления ошибки, когда на передающей стороне имеется корреляционная CSI. Прекодирования обеспечивают значительный выигрыш, измеренный как в некодированном, так и в кодированном режиме. В некодированном режиме, при вероятности 10^{-5} выигрыш 4 дБ; в кодированном режиме, при вероятности 10^{-6} выигрыш 2 дБ.

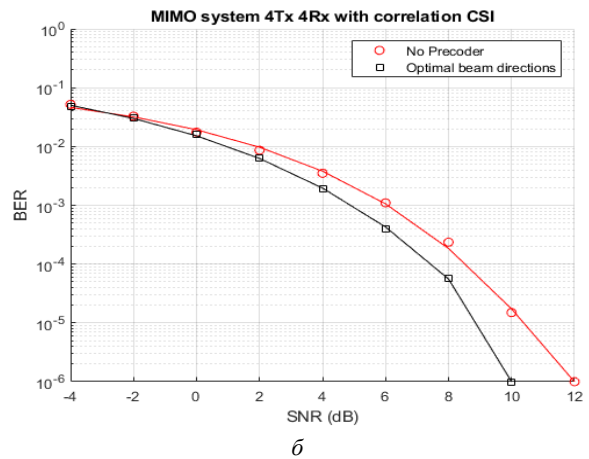
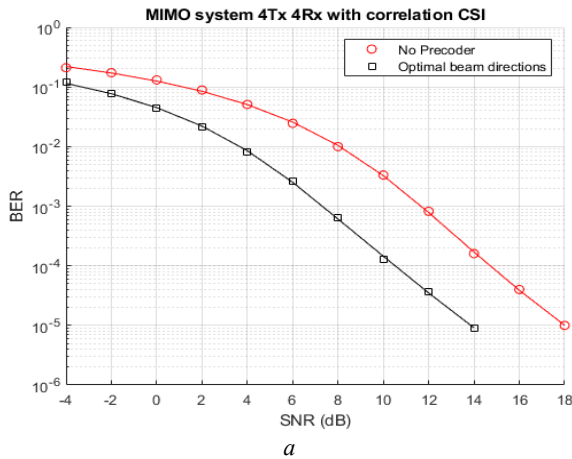


Рис. 3. Эффективность прекодирования в случае на передающей стороне имеет корреляционную CSI: a – без кодирования; b – с пространственно-временным кодированием

На рис. 4 продемонстрированы эффективности вероятности появления ошибки, когда на передающей стороне имеется динамическая CSI. Прекодирования обеспечивают выигрыш, измеренный как в некодированном, так и в кодированном режиме. В некодированном режиме, при вероятности 10^{-5} выигрыш 4 дБ; в кодированном режиме, при вероятности 10^{-5} выигрыш 2 дБ.

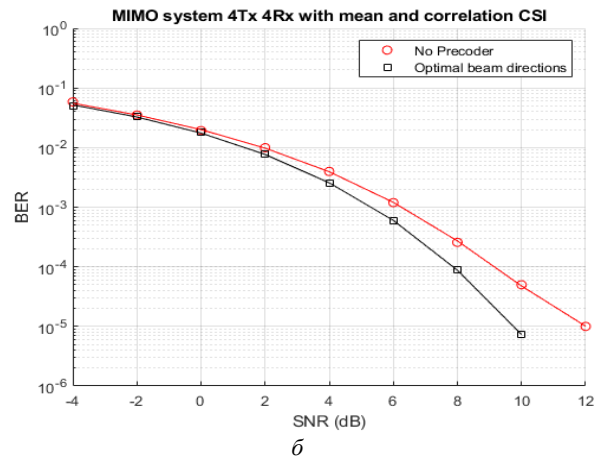
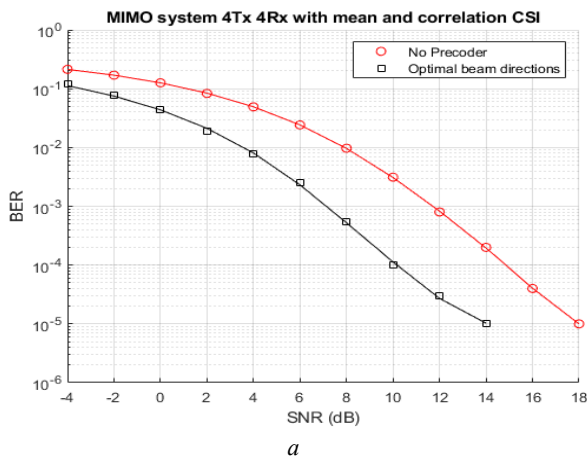


Рис. 4. Эффективность прекодирования в случае на передающей стороне имеет динамическую CSI: a – без кодирования; b – с пространственно-временным кодированием

Заключение

В работе рассмотрены алгоритмы предварительного кодирования в системе MIMO с использованием в схеме обратной связи информации о состоянии канала (CSI) канала.

Анализ моделирования алгоритмов прекодирования для различных форм CSI приводит к следующим результатам. Применение алгоритмов прекодирования позволяет получить достаточно большой выигрыш. Существует зависимость как от CSI, так и от количества антенн, конфигурации системы и отношения сигнал/шум. Выигрыш прекодирования обычно увеличивается с увеличением количества антенн. Когда CSI идеальна, прекодеры также обеспечивают дополнительный выигрыш от разнесения.

Выигрыш от прекодирования значителен как в некодированной, так и в кодированной областях. Выигрыш зависит от CSI, количества антенн, конфигурации системы и отношения сигнал/шум. Выигрыш прекодирования обычно увеличивается с увеличением количества антенн. Когда CSI идеальна, прекодеры обеспечивают дополнительный выигрыш от разнесения.

Прекодирующие матрицы, оптимальные по критериям взаимной информации и минимума среднеквадратической ошибки, имеют одни и те же правые сингулярные векторы. Эти векторы

определяют направления пространственных лучей, которые зависят от имеющейся на передающей стороне CSI.

Прекодирующие матрицы, оптимальные по всем описанным критериям и предназначенные для использования в условиях различного объема информации на передающей стороне о свойствах канала связи MIMO, имеют одни и те же правые сингулярные векторы. Из этих векторов составляется оптимальная прекодирующая матрица, которая согласована с корреляционной матрицей Q пространственно-временного кода.

PRE-CODING ALGORITHMS CONSIDERING COMMUNICATION CHANNEL STATE FOR WIRELESS MIMO SYSTEMS

H.A. FAM, S.B. SALOMATIN

Abstract. The article deals with linear precoding algorithms using channel state information on the transmitting side of a MIMO system. The synthesis and efficiency of precoding algorithms based on the use of the following criteria are considered: mutual information, minimum pairwise error probability, minimum mean square error.

Keywords: MIMO system, channel state information – CSI, space-time coding (STC), ideal CSI, CSI fading correlation, dynamic CSI.

Список литературы

1. Jankiraman M. Space-Time Codes and MIMO systems. USA, 2004.
2. Jafarkhani H. Space-Time Coding. Theory and practice. UK, 2005.
3. Tarokh V., Jafarkhani H., Calderbank A.R. // IEEE J. Select. Areas Commun. 1999. Vol. 17, No 3. P. 451–460.
4. Тыртышников Е.Е. Матричный анализ и линейная алгебра. М., 2007.
5. Visotsky E., Madhow U. // IEEE Trans. Inform. Theory. 2001. Vol. 47, No. 6. P. 2632–2639.
6. Kuo C.-C.J. [et al.] Precoding Techniques for Digital Communication Systems. USA, 2008.
7. Oesges C., Clerckx B. MIMO Wireless Communications. From real-world propagation to space-time code design. UK, 2007.

УДК 621.391.84

ВЛИЯНИЕ ВИБРАЦИЙ НА ФАЗОВЫЕ ШУМЫ КВАРЦЕВЫХ ГЕНЕРАТОРОВ

В.Н. КИЙКО, Н.М. НАУМОВИЧ, М.В. ДАВЫДОВ, В.И. ЖУРАВЛЕВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 18 февраля 2022

Аннотация. Разработана схема стенда для измерения фазовых шумов кварцевых генераторов. Приведены результаты измерений фазовых шумов при отсутствии и наличии вибраций кварцевых генераторов Морион и Wenzel.

Ключевые слова: фазовые шумы, вибрации.

Введение

Высокостабильные кварцевые генераторы (КГ) являются радиоэлектронными устройствами, вырабатывающими сигнал постоянной частоты с высокой температурной и временной устойчивостью, низким уровнем фазовых шумов. Однако при эксплуатации на подвижных объектах КГ подвергается воздействию механической вибрацией, которая влияет на частоту и фазу генерируемого сигнала.

Стабильность частоты и фазовые шумы КГ являются наиболее важной характеристикой.

Анализ дестабилизирующих вибрационных воздействий различных объектов

Современные радиоэлектронные комплексы оснащены радиоэлектронной аппаратурой и системами, которые выполняют функции управления, контроля, обнаружения, наведения, координации, связи и другие функции [1]. Эти системы размещаются на подвижных объектах – самолетах, вертолетах, кораблях, беспилотных летальных аппаратах, автомобильной и гусеничной технике, ракетах, спутниках – и при работе подвергаются воздействию целого комплекса дестабилизирующих факторов. Ориентировочные значения параметров вибрации, полученных при статистическом обобщении опытных данных для некоторых видов механических воздействий, приведены в табл. 1.

Табл. 1. Характеристики вибраций, действующих на аппаратуру, установленную на подвижных объектах

Источник вибрации	Характеристики вибрации
Транспортные средства колесного типа	Частоты возбуждающих колебаний подвески машины $2 \div 10$ Гц, кузова 8–15 Гц, ускорение 1 g, частоты возбуждения двигателя 20–60 Гц
Транспортные средства гусеничного типа	Ударновозбуждаемые вибрации в диапазоне частот 400–700 Гц, амплитуда колебаний на низких частотах $\pm 0,25$ мм
Корабль	Частоты возбуждающих колебаний и ускорения: кормовая часть 2–35 Гц, 0,05–0,5 g
Самолет	Частоты возбуждения 3–500 Гц, амплитуда колебаний $\pm 3,8$ мм на низких частотах Акустические вибрации с частотой до 130 КГц на уровне 150 дБ выше звукового порогового уровня
Управляемые снаряды	Частоты возбуждающих колебаний 30–5000 Гц, ускорения 5–30 g. На участке резонанса возможны ускорения до 40 g. Акустические вибрации с частотой до 10 КГц на уровне 130 дБ выше звукового порогового уровня

Воздействие вибрации на кварцевые генераторы

Для исследования характеристик КГ необходимо воздействие на него какого-либо внешнего фактора. Для данной цели выбран аддитивный белый гауссовский шум - это вид мешающего воздействия в канале передачи информации. Диапазон вибраций, которыми будет производиться воздействие на КГ, выберем от 10 Гц до 2 кГц. Аддитивный белый гауссовский шум в данном диапазоне будет иметь характеристики, наиболее приближенные к тем воздействиям, которые будут иметь место в реальных условиях. В качестве возмущающего воздействия при анализе случайных вибраций на КГ задана зависимость широкополосной случайной вибрации (ШСВ) от частот колебаний в виде, отображенном на рис. 1.

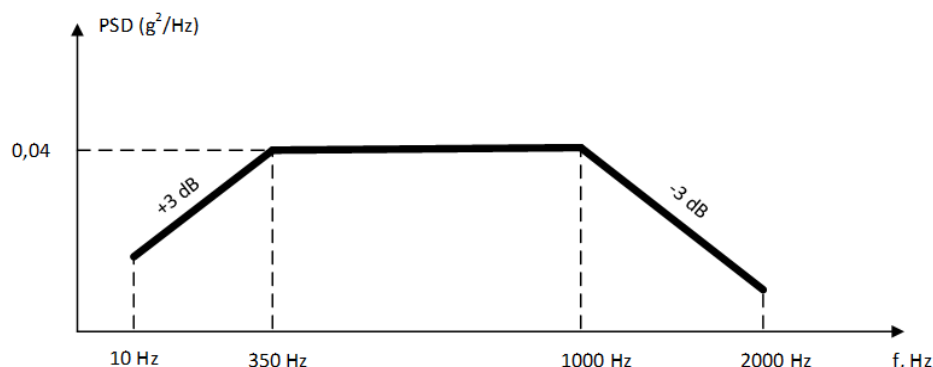


Рис. 1. График зависимости ШСВ от частот колебаний возмущающего воздействия

В диапазоне частот вибраций 300–1000 Гц, спектральная плотность вибраций постоянна и равна $0,04 \text{ g}^2/\text{Гц}$. На частотах менее 300 Гц и более 1000 Гц спектральная плотность вибраций уменьшается по закону 3 дБ/октаву.

Схема установки, с помощью которой измеряются фазовые шумы КГ (без воздействия и при наличии вибраций), приведена на рис. 2.

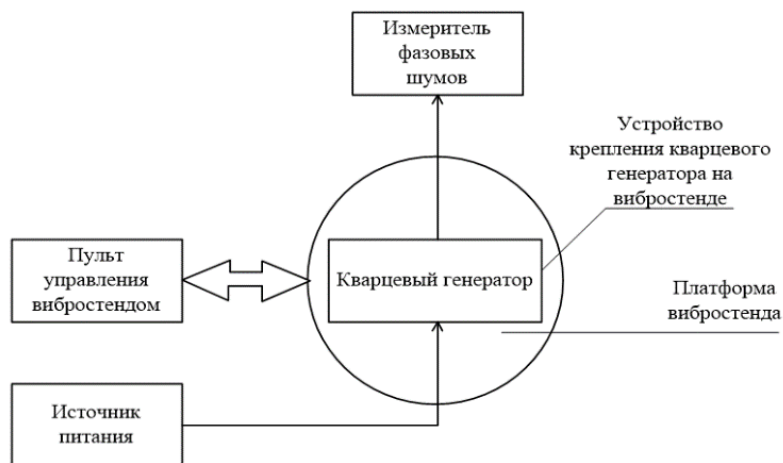


Рис. 2. Схема установки для измерения фазовых шумов

Для измерения уровня фазовых шумов используется анализатор FSWP26 фирмы Rohde&Schwarz. В качестве источника имитации воздействия внешних механических факторов применяется электродинамический стенд TV 5220/LS-120.

Воздействие ШСВ на кварцевые генераторы фирм Морион и Wenzel осуществлялось по оси z. Оба тестируемых КГ термостатированы и работают на частоте 100 МГц.

На рис. 3 приведены результаты экспериментальных исследований генератора ГК-213 фирмы Морион. Прямая наклонная линия показывает допустимый уровень фазовых шумов КГ. Нижняя кривая показывает уровень фазового шума при отсутствии вибраций. Верхняя экспериментальная зависимость приведена при наличии вибраций $0,04 \text{ g}^2/\text{Гц}$. Видно, что при наличии вибраций уровень фазовых шумов увеличивается от 30 до 50 дБ.

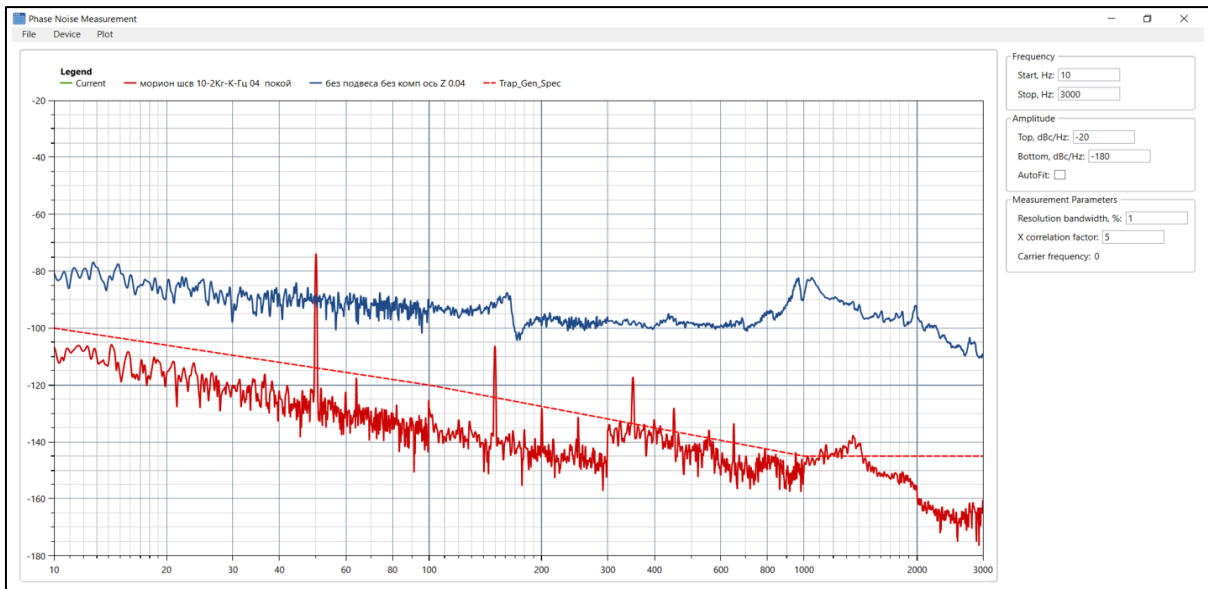


Рис. 3. Зависимость фазовых шумов генератора ГК-213 Морюн от частоты вибраций

На рис. 4 приведена зависимость фазовых шумов генератора Wenzel 100 MHz-SC Citrine от частоты вибраций.

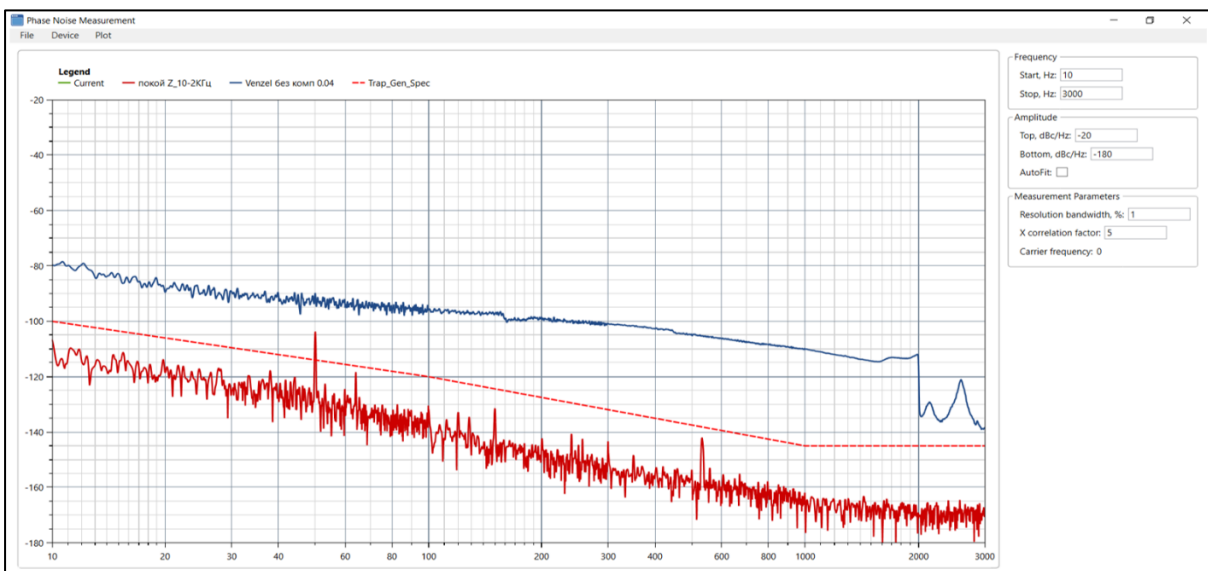


Рис. 4. Зависимость фазовых шумов генератора Wenzel 100 MHz-SC Citrine от частоты вибраций

Видно, что генератор Wenzel имеет значительно меньшее количество резонансных частот корпуса генератора. При этом уровень повышения фазовых шумов, также, как и генераторе ГК-213, составляет 30–50 дБ.

Заключение

Параметры современных радиоэлектронных систем в значительной мере определяются спектральной плотностью мощности фазовых шумов формируемых сигналов. Как видно из данных, полученных в результате эксперимента, вибрация существенно повышает уровень фазовых шумов КГ. Это ухудшает функции всей электронной системы, которая зависит от низкого фазового шума генератора. Для обеспечения необходимой защиты электронных устройств от интенсивных вибраций предлагается использование пассивных способов виброзащиты, к которым относятся виброизоляция (демпфирование), так и электронные методы компенсации фазового шума.

INFLUENCE OF VIBRATIONS ON PHASE NOISES OF QUARTZ GENERATORS

V.N. KIYKO, N.M. NAUMOVITCH, M.V. DAVYDOV, V.I. ZHURAVLEV

Abstract. A scheme of a stand for measuring the phase noise of quartz oscillators has been developed. The results of phase noise measurements in the absence and presence of vibrations of Morion and Wenzel quartz oscillators are presented.

Keywords: phase noise, vibration.

Список литературы

1. Карпушин В.Б. Вибрации и удары в радиоаппаратуре. М., 1971.

UDC 620.9:658.30

DEVELOPMENT OF IOT NETWORK FOR PRODUCTION QUALITY CONTROLU.A. VISHNYAKOU¹, HU ZHIFENG¹, FENGSHOU GU²¹ *Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*² *Centre for Efficiency and Performance Engineering, University of Huddersfield, UK**Received February 27, 2022*

Abstract. The network structure for achieving quality control of production based on the Internet of Things (IoT) technology is constructed. IoT network elements as analyzers, cloud platform, application port is described. Five criteria for selecting an IoT cloud platform are proposed. The mobile application for quality production is developed.

Keywords: IoT network, production quality control, IoT cloud platform.

Introduction

The Internet of Things is a collection of multiple technologies, which includes a variety of technologies such as sensors, network communication, device information security, device compatibility and data processing. Thus, for the multiple, large amount of data generated in the production process, IoT technology an effective technical means of quality control of production.

Currently, relevant researchers have used IoT technologies for quality control of production. For example, an IoT-based water quality monitoring system was developed in the reference [1]. An air quality monitoring platform based on IoT was created in the reference [2]. The reference [3] proposed a smart manufacturing system engineering (SMSE) approach to design a smart product quality monitoring system (SmartPQMS) with a practical application in steel production process as an example. In the paper [4], an IoT-based coffee quality monitoring and processing system was proposed.

However, in order to achieve quality control of production using IoT technology, a variety of IoT-based network elements need to be developed.

IoT structure network for milk quality control

For the specific production quality control, for instance, the production quality control of milk, a specific IoT network structure model needs to be constructed. The IoT network structure based on quality control of milk production was constructed. Figure 1 shows the basic structure of this network model. The model is composed of multiple milk analyzers, sensors, microcontrollers, gateway, IoT cloud platform and mobile applications. The data measured by the milk analyzer is transmitted through sensors and microcontrollers. By using the MQTT protocol to transfer the data to the gateway and then from the gateway to the cloud platform. Pre-processing and analysis of data in cloud platform by using the rule engine. In the cloud platform, the processing of data can be generally divided into two types, one is to directly send the data in the cloud platform to mobile devices for processing, and the other is to send the results of processing data in the cloud platform directly to mobile devices, so that the users of mobile devices can directly obtain the processing results. Which way to transfer data depends on different requirements.

Thus, an IoT network model based on quality control of milk production was constructed, but in order to better apply and implement the model, the details of the elements in the model need to be analyzed. The next part is the analysis of the elements in the model.

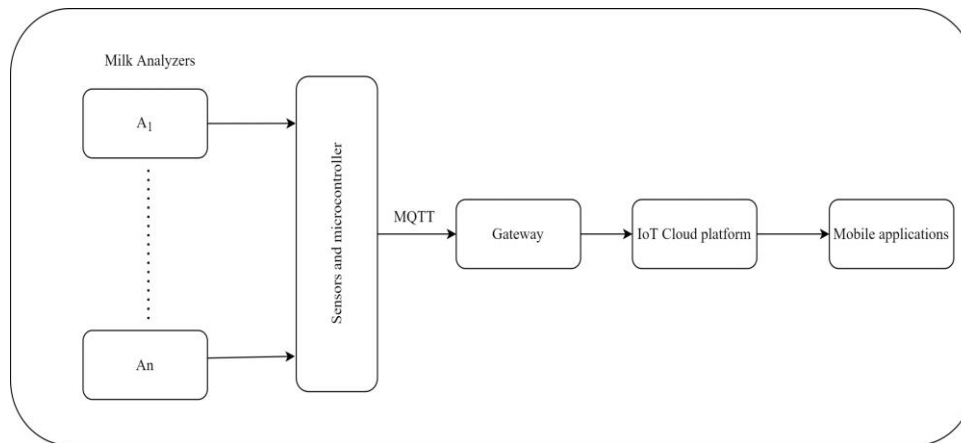


Figure 1. The IoT network structure for milk monitoring

Milk quality monitoring

The milk quality analyzer is a device for determining the quality characteristics of milk and products based on some indexes (fat content, density, sample acidity, lactose, sample temperature, etc.). A complete milk analyzer usually consists of hardware and software. For the hardware part, there are mainly included a multispectral sensor, a light source module, a quartz sample cell, a light-shielding cover, a collimator, a Raspberry Pi, a touch displays, a booster module, a battery management module, a Cooling fan, a Lithium battery, a power switch and so on.

For the software part, there are mainly two subsections, first subsection is data acquisition software, another is milk compositions analysis software. The first part is to collect data (the index of protein, fat, and density, etc.) from the milk, another part is for the analysis of milk compositions.

Port structure

With the development of sensing technology, communication technology and Internet of Things technology, the way to analyze milk quality using IoT technology corresponds to Automation 4.0.

In order to archive the Automation 4.0 on the milk quality assessment in dairy farms in different regions. In addition to communication technology, different communication ports are also required. However, different farms may use different type milk analyzers, the information from different milk analyzers can be received via parallel or serial ports. That means the milk analyzers can use the different ports and protocols. To make sure the received information to interact with the hardware of the cloud platform, it is necessary to provide the gateways-converters. A port structure known as MOXA has developed a solution that can make the connecting diverse devices and COM port simple [5].

Cloud environment platform

The IoT cloud platform is the central system of the entire IoT system and plays a vital role in the entire system. The IoT cloud platform is the technology fusion of the IoT platform and cloud computing. Accordingly, the IoT cloud platform can serve as a reliable gateway between the milk quality analyzer and the network, it can also provide cloud computing, data storage, data processing, etc., and can support the huge amount of data generated by the devices, as well as allow bilateral communication between devices. Of necessity, in addition to these capabilities, the IoT cloud platform needs to provide reliable and secure data transfer capabilities. In view of the IoT cloud platform has such powerful capabilities, there are many IoT platforms emerged, such as Microsoft Azure IoT cloud platform, Google Cloud's IoT Platform, IBM Watson IoT platform, AWS IoT platform, Ali cloud IoT cloud form, Cisco IoT Cloud Connect platform and Huawei IoT Cloud Platform, etc. However, every IoT cloud platform has its own pros and cons, how to choose the most suitable IoT cloud platform according to our requirements needs to be evaluated based on suitable assessment metrics. Here, some appropriate evaluation metrics are proposed for the selection of IoT cloud platforms.

1. Scalability. A reliable IoT cloud platform must support millions of devices connecting and communicating at the same time, and be capable of stable and normal operation, as well as supporting the scaling of device performance and number.

2. Equipment Control Capabilities. To reliably control end devices remotely through the IoT cloud platform, the IoT cloud platform can provide device management capabilities such as cloud application programming interfaces(APIs), developer tools, event logs, device command and control, diagnosis and prevention of device faults. In particular, the cloud API allows users to connect to and access back-end services such as databases, cloud services and cloud computing, the developer tools can be used to reprogram and manage IoT devices. The event logs provide access to the most relevant and valuable data to improve device management and operation by providing event recall capabilities. The function of diagnosis and prevention of equipment failure can provide the latest status of equipment operation, so that devices that are likely to fail can be detected and repaired, thus reducing property damage and safety accidents.

3. Over-the-Air Firmware Updates capability. This capability provides the solution to send new functionality to the device remotely and enables developers to optimize and enhance device functionality readily and reliably. However, it is worth noting that the implementation of this function requires the coordination of IoT hardware, device firmware, network connectivity and IoT cloud devices. The implementation of this function is an important indicator to examine the IoT cloud platform.

4. Security. The security performance of the IoT cloud platform is also an important evaluation indicator. Security does not only mean information and data security, but also hardware access security. The security features of IoT cloud platforms usually include information encryption, role access control, authentication, port control, etc.

5. Data Management. Device data can be seamlessly interfaced with the services provided by existing IoT cloud platforms, which will reduce the commissioning and access costs for devices to connect to the cloud platform.

These five metrics will provide a reliable evaluation system for selecting the right IoT cloud platform. Although some other indicators such as supported application protocols and cloud platform prices are also worth considering, users can make a reasonable choice of new indicators based on their own conditions relative to the five evaluation indicators mentioned above. According to the above five indicators, Aliyun IoT cloud platform will be a rational choice that can provide corresponding capabilities.

Application

The smartphone is simulated in android studio, which is a powerful tool for developing Android apps. As a result, the Milk Analyzer App was created in android studio. Figure 2 shows the structure of the system for collecting and analyzing milk quality indicators.

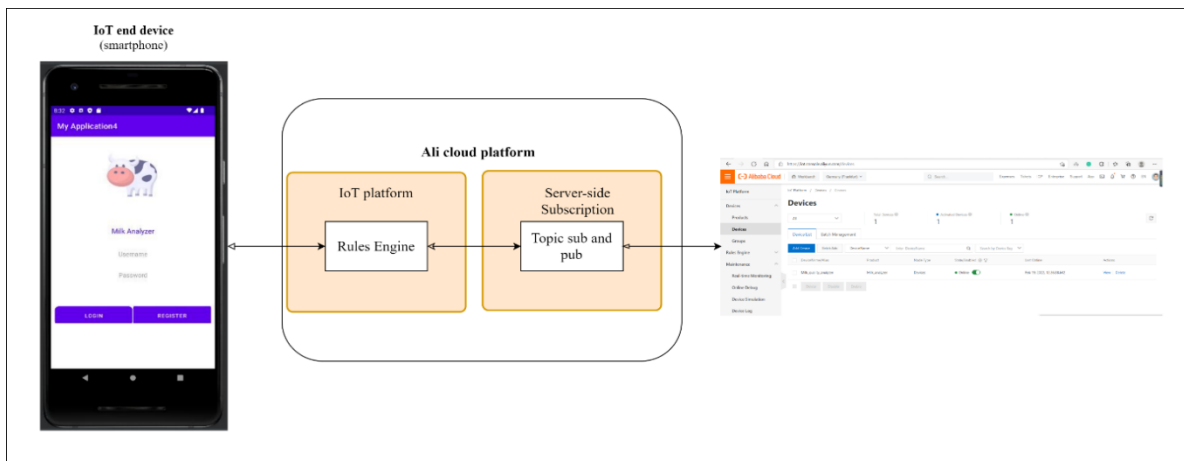


Figure 2. General structure of emulation of the IoT network on the Ali platform

On a mobile device, the application is divided into two sections: the login screen and the data processing screen (Figure 3). The user's work begins with login and registration (Figure 3, a), users enter

their username and password to access the data processing screen. In the data processing interface (Figure 3, *b*), four indicators are included in this interface: lactose, fat, chloride, and protein. Users can get data from the IoT cloud and test that data, for example, to distinguish whether it is normal quality or mastitis milk.

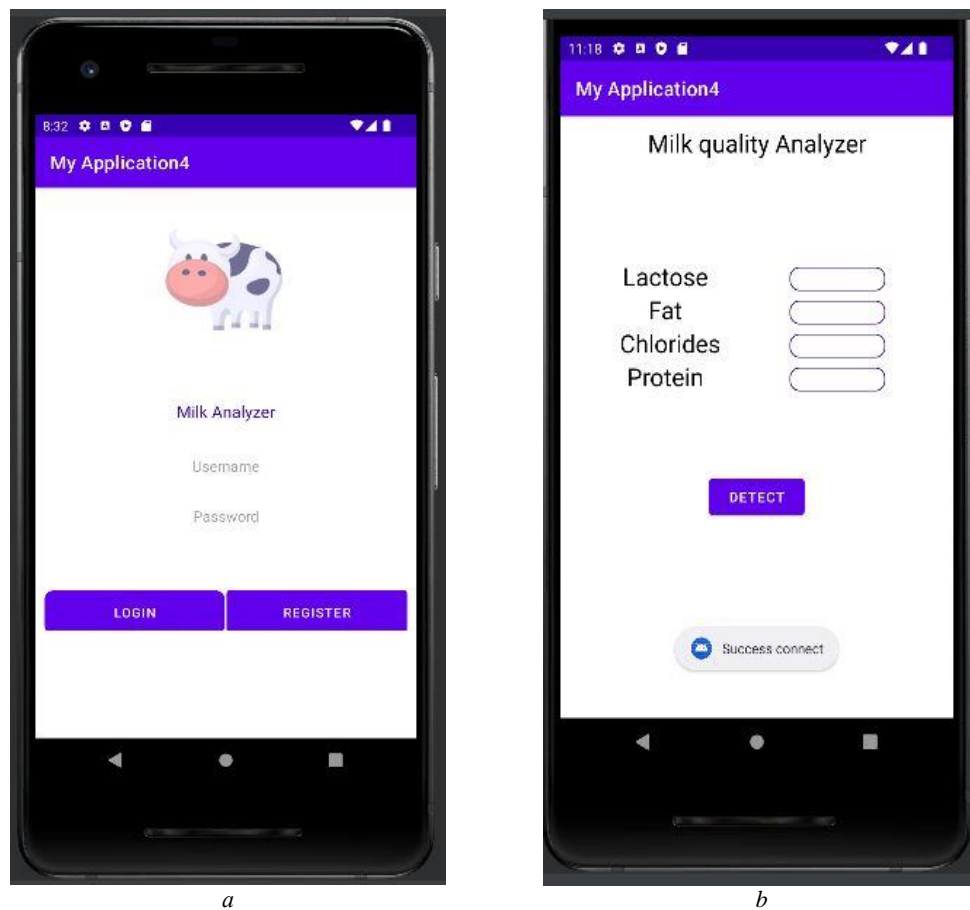


Figure 3. The section of application: *a* – The interface of login and registration; *b* – The interface of data processing

At that some elements of the IoT network based on quality control of milk production have been developed.

Conclusion

This paper considers some basic elements of IoT networks for milk production quality control, such as milk analyzers, port structures, IoT cloud platforms and applications on mobile devices. First, an IoT network structure for milk analysis is constructed, then the details of IoT network elements are described.

References

1. Moparthy N.R., Mukesh C., Sagar P.V. // Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). IEEE, 2018. P. 1–5.
2. Aamer H., Mumtaz R., Anwar H. [et al.] 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT). IEEE, 2018. P. 102–106.
3. Shin K.Y., Park H.C. // 19th International Conference on Control, Automation and Systems (ICCAS). IEEE, 2019, P. 1693–1698.
4. Rutayisire J., Markon S., Raymond N. // International Conference on Applied System Innovation (ICASI). IEEE, 2017, P. 1209–1212.
5. Connecting the COM port directly to the Cloud [Electronic resource]. URL: <https://moxa.pro/blogs/news/podklyuchaem-som-port-napryamuyu-v-oblako>.

HUMAN ACTIVITY RECOGNITION BASED ON ADABOOST ENSEMBLE CLASSIFIER

Y. XUEYING, I.A. BARYSKIEVIČ

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 18, 2022

Abstract. Human activity recognition (HAR) has been widely applied in the field and has good application prospects. Various classifiers in machine learning have shown excellent performance in their own fields. In this paper, AdaBoost ensemble classifier for human activity recognition is proposed to improve the performance of human activity recognition technology by using a weighted combination of multiple classifiers. The experimental results of HAR data were evaluated, and the total classification accuracy and receiver operating characteristic (ROC) area were calculated. The results show that the AdaBoost ensemble classifier framework proposed in this paper can accurately identify six kinds of human activities, and the AdaBoost ensemble classifier algorithm can significantly improve the HAR recognition accuracy.

Keywords: human activity recognition, AdaBoost, ensemble classifier.

Introduction

Human activity recognition (HAR) refers to the automatic detection of various physical activities that people perform in their daily lives. The system helps identify the activities that people perform and provides information feedback for intervention. Human activity recognition technology aims to perceive the external performance of human behavior and identify its categories according to the perception results. As a key technology that enables computers to provide services to people in a more active and natural way, human activity recognition has attracted the interest of researchers for its wide range of potential applications in recent years. Human activity recognition is widely used in surveillance [1], security [2], education [3], sports [4], medical [5] and other fields.

In recent years, with the intelligent mobile devices (such as smart phones and wearable devices) and related sensors (such as motion sensor, and skin conductance sensor) such as the rapid development of technology, user activity recognition technology research focus is from the method based on computer vision to the user to carry the intelligent recognition method based on other sensors on a mobile device [6]. These emerging user activity recognition technologies based on intelligent mobile devices do not rely on external devices and are more in line with the requirements of contemporary people for portable activity recognition. Machine learning classifiers are often used to evaluate the prediction accuracy of human activity recognition.

This research aims to introduce the basic process of human activity recognition and evaluate the performance of different classifiers. The smart phone-based accelerometer has a sampling frequency of 50 Hz and collects daily life data of human activities, including walking, walking upstairs, walking downstairs, sitting, standing, and lying. Use k-nearest neighbor (kNN) [7], Naïve Bayes (NB) [8], support vector machine (SVM) [9] and random forest (RF) [10] to evaluate the data set. The research results show that RF has the highest accuracy for human activity recognition.

This paper uses AdaBoost ensemble classifier to identify human activity data collected by human body sensor. The traditional classification method is to find a classifier closest to the actual classification function in a space composed of various possible functions, and the classification accuracy is often not ideal [11]. Ensemble classifier builds a group of base classifiers from training data (base classifiers mainly include RF, NB, SVM, kNN, etc.) and then classifies by voting the prediction of each base

classifier. By aggregating the prediction results of multiple classifiers, the classification accuracy of the classifier can be improved, and the weighted combination of multiple classifier models can achieve better performance, which has been well promoted and applied in practice.

In this paper, the AdaBoost ensemble classifier framework for human activity recognition is proposed, it shows in Figure 1.

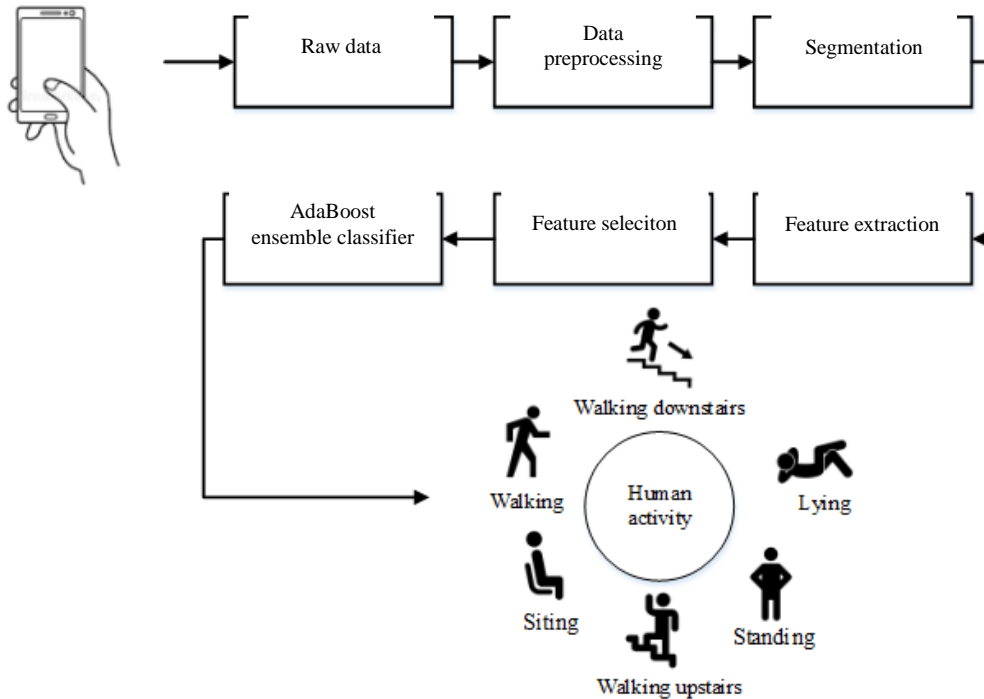


Figure 1. Human activity recognition based on AdaBoost ensemble classifier

Different base classifiers (RF, NB, SVM, and kNN) are compared to study the performance evaluation of base classifier and AdaBoost ensemble classifier. Experimental results show that the ensemble classifier based on AdaBoost algorithm is feasible in practical problems. Significantly improved automatic HAR performance. The search for local extrema is a basic operation for many image processing tasks.

Classification methods

AdaBoost algorithm can be applied to a variety of base classifiers to form a variety of combined classifiers. This paper mainly selects RF, NB, SVM, KNN and decision tree C4.5 as base classifiers for research.

Random forest is an ensemble classifier based on multiple base classifiers [10]. The construction process of the random forest is as follows: Firstly, bootstrap sampling is used to extract K samples from the original training set, and the sample size of each sample is required to be consistent with the original training set. Secondly, a decision tree model is constructed for each sub-sample, and K decision trees are trained. The last K decision trees are voted for the final classification. The random forest algorithm constructs a decision tree for each training subset and generates N decision trees. Node splitting is the core of the algorithm. Only through node splitting can a complete decision tree be generated. Each tree selects attributes based on the Gini index.

Naïve Bayes classifier is based on the assumption that each attribute of the sample is independent from each other, requiring fewer parameters to be estimated and smaller error rate than other classifiers [8]. When the correlation between attributes is small, NB classifier can achieve better performance.

Support vector machine (SVM) is a classifier for linear and nonlinear data, which is characterized by simultaneously minimizing empirical error and maximizing geometric edge regions [9]. It uses nonlinear mapping to transform original training data into high dimensional data. In this new dimension,

the linear optimal separation hyperplane is studied. By appropriately nonlinear mapping to an efficient high-dimensional hyperplane, the maximum spacing hyperplane is established, which is discovered by support vector machines using support vectors and boundaries. It is assumed that the larger the distance or gap between parallel hyperplanes, the smaller the total error of the classifier. Then the covering theorem can be used to achieve linear separation in the result feature space.

k-Nearest-Neighbor (kNN), as one of the classical classification methods, is a non-parametric classification method based on comparison learning, which has the characteristics of simple implementation and high robustness [7]. It is used to store all available cases into multiple categories and predict the classification of new cases based on the nearest k-nearest neighbor. kNN algorithm uses distance measure functions (such as Euclidean distance) to find k-nearest neighbors. Most of k nearest texts belong to a certain category, so the samples also belong to this category.

C4.5 is an extension of ID3 [8] and classifies samples by generating decision trees. Decision tree is an inductive learning algorithm, which extrapolates classifiers in the form of decision tree from training sample sets and uses top-down recursive method. C4.5 uses the information gain rate function as the classification standard, and uses the value of "classification information" to standardize the information gain, avoiding the disadvantage of using the information gain to select attributes with more values. Compared with ID3, it can discretize continuous attributes, process incomplete data, and prune trees in the process of tree construction.

AdaBoost is an iterative algorithm [11]. Firstly, different training subsets are sampled from the same training sample set, and then different base classifiers are trained with these different training subsets. Finally, these base classifiers are combined to form a strong classifier. In AdaBoost algorithm, each sample in the training sample set is assigned a weight, which represents the probability that this sample is selected into the training subset by a base classifier. The AdaBoost process is shown in Figure 2 and Figure 3.

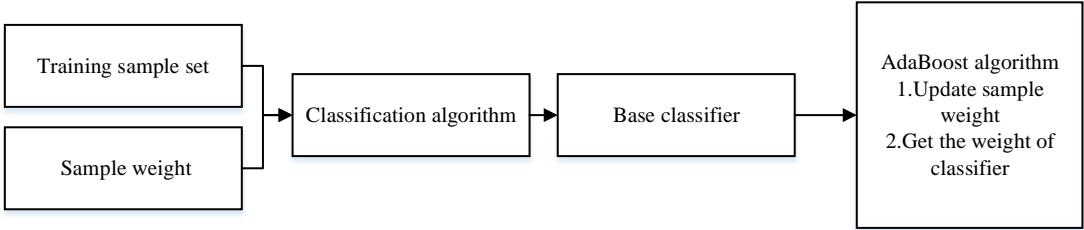


Figure 2. AdaBoost iterative training process

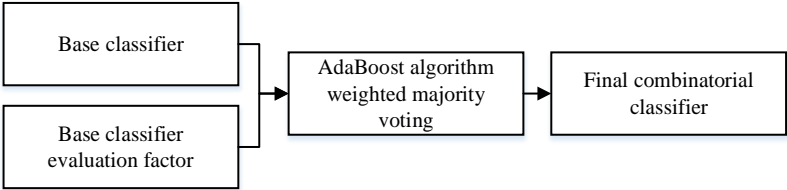


Figure 3. The block diagram of the generation of combined classifier (AdaBoost with single algorithms)

In each iteration, the weight of the sample is updated according to whether the classification of the sample is correct or not and the classification accuracy of the whole training sample set last time. Then, according to these weights, sample points are selected to obtain a new training subset and train the next base classifier. If a sample is correctly classified, its weight will increase and its probability of being selected into the next training subset will increase, while if a sample is incorrectly classified, its probability of being selected into the next training subset will decrease, thus making the AdaBoost method focus on those relatively difficult samples.

In the concrete implementation process, first of all, the weight of all samples in the initial seasonal training sample set is equal. Secondly, in the t iteration, sample points are selected according to the sample weight to form a training subset, and the training base classifier H_t is trained. All samples are classified to improve the weight of incorrectly classified samples and reduce the weight of correctly classified samples; the updated weighted sample set is used to generate the next training subset and train

the next base classifier H_{i+1} . Finally, multiple base classifiers generated by iteration are used to classify the samples, and weighted majority voting is carried out on the classification results to obtain the final result.

Final classifier: As shown in Figure 3, a set of base classifier h_1, h_2, \dots, h_l generated by iteration is used first sample X is classified to get the classification result $h_1(x), h_2(x), \dots, h_l(x)$, and then they are voted with weight $\alpha_1, \alpha_2, \dots, \alpha_l$ to get the classification result.

Experimental results

Through accelerometer and gyroscope embedded in smart phone to collect speed and gyroscope sensor data, sensor data collection of mobile phone data collection application matlab@mobile is opened, and placed in the pants pocket of the activity, when the activity is completed, take out the mobile phone to close the application. In the process of data collection, the scene mode of normal life was simulated. Six basic physical activities were completed: 1. Walking, 2. Walking upstairs, 3. Walking downstairs, 4. Sitting, 5. Standing and 6. Lying. The 66 time features extracted from ACC data and related to 6 physical activities are used in the analyzed HAR system.

Figure 4 shows the original ACC data distribution point diagram based on two features in the time domain (TotalAccYMean and TotalAccXMean). It follows from the figure recognizing 6 classes of activities requires design of a new feature space by mean of the classifier proposed in the paper.

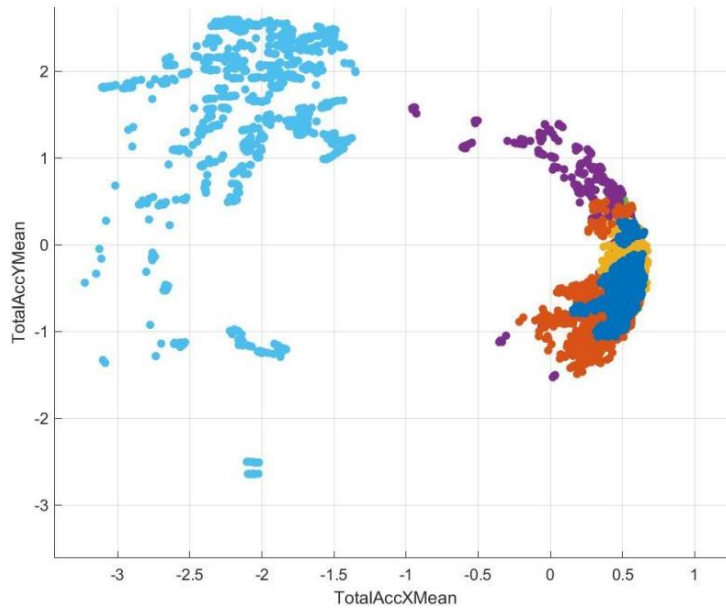


Figure 4. Distribution of ACC data activity classes in two dimensional feature space

Performance evaluation measures

In this paper, experiments are set up to evaluate the classifier to achieve consistency evaluation of model quality and quantitative study of prediction performance of the model. Accuracy measures are often used to evaluate the performance of classifiers. In fact, this metric measures the percentage of correctly classified examples. In the case of binary classification, the classification accuracy can be expressed as:

$$Accuracy = \frac{T_N + T_P}{F_P + F_N + T_N + T_P}, \quad (1)$$

where T_N (True negatives) represents the correct classifications of negative examples, T_P (True positives) represents the correct classifications of positive examples, F_N (False negatives) and

F_p (False positives) represent, respectively the positive examples incorrectly classified into the negative classes and the negative examples incorrectly classified into the positive classes.

In the process of training, the training model can match the training data, but cannot match the test set well. In order to better deal with the evaluation method of bias and variance tradeoff, this study adopts the 10-fold cross-validation method to randomly divide the data set into 10 parts, ensuring that each part takes turns as a test set, and the remaining 9 parts as a training set, so as to select more appropriate parameters for the model.

Area under curve (AUC) is the area under the ROC curve. We often use AUC value as the evaluation standard of the model, because in many cases, the ROC curve cannot clearly explain which classifier has a better effect. As a value, the classifier with a larger AUC has a better effect. Due to the complexity of human movement and the changeable environment, there are still many problems in human movement pattern recognition. For example, detailed classification of different motion modes requires more effective feature extraction and more efficient classification algorithms. In this study, we focus on how to obtain a more efficient classification algorithm, and propose the ensemble classifier AdaBoost algorithm. In order to suggest the feasibility of the algorithm, different base classifiers (RF, NB, SVM, kNN, C4.5) were adopted, with the same parameters for each classifier.

The experiment divided the data set into training set and test set, and adopted the method of 10-fold cross-validation to prevent over-fitting. The tables 1 and 2 show the Classification results of single classifier and AdaBoost ensemble classifier on activities. The total accuracy of single classifier and AdaBoost classifier was compared.

Table 1. Multiclass classification accuracy estimation of single classifier and AdaBoost ensemble classifier

Algorithm	Accuracy (%)						Roc Area
	Walking	Walking upstairs	Walking downstairs	Sitting	Standing	Laying	AUC
RF (single)	93,4	93,5	92,9	91,3	93,8	96,8	0,99
RF (AdaBoost)	96,6	97,0	93,0	93,7	90,0	99,9	1,00
NB (single)	90,2	72,1	82,7	42,6	66,1	89,9	0,92
NB (AdaBoost)	91,7	72,9	82,5	80,0	71,6	97,2	0,96
SVM (single)	94,7	93,4	91,2	83,2	88,5	100,0	0,95
SVM (AdaBoost)	96,1	95,5	96,8	85,7	83,6	99,9	1,00
KNN (single)	84,9	86,8	93,9	82,5	65,9	96,7	0,99
KNN (AdaBoost)	93,8	94,1	94,9	66,4	68,9	98,4	0,98
C4.5(single)	91,3	88,3	87,8	91,6	89,9	99,7	0,97
C4.5 (AdaBoost)	91,8	91,1	89,8	91,2	92,5	100,0	0,97

Table 2. The total accuracy of single classifier and AdaBoost classifier

Classification Method	Single accuracy (%)	AdaBoost accuracy (%)
RF	93,60	95,03
NB	73,90	82,65
SVM	91,80	92,90
kNN	85,10	86,08
C4.5	91,40	92,07

Table 1 shows the classification results of single classifier and AdaBoost ensemble classifier, and Table 2 shows the total accuracy of single classifier and AdaBoost ensemble classifier. According to the analysis in Table 1 and Table 2, among the single classifiers, the random forest classifier performed best, with the accuracy rate of all six kinds of activities higher than 90 % and the total classification accuracy rate of 93,6 %. SVM and C4.5 also performed well, and the total classification accuracy of single classifier reached 91,8 % and 91,4. However, their shortcoming lies in the uneven prediction accuracy and the large difference of recognition effect among different activities. NB is the worst performer overall.

As can be seen from the tables, the AdaBoost ensemble classifier has improved the recognition of each activity and the whole model to a certain extent. After the AdaBoost integrated classifier, RF, NB, SVM, KNN and C4.5 were increased by 1,7 %, 8,75 %, 1,1 %, 0,98 % and 1,3 % respectively. The data show that all the classifiers mentioned in this paper have improved the classification accuracy,

especially the accuracy of weak classifiers has been significantly improved. After using the AdaBoost ensemble learning method, the area under the ROC curve of RF and SVM was increased to 1.

Conclusion

Human activity recognition has developed rapidly in recent years. The main steps include data preprocessing and feature extraction, feature selection, training of classifier and implementation of classification algorithm. Starting from classifier training and classification algorithm, this paper proposes AdaBoost ensemble classifier for human activity recognition. In HAR, AdaBoost ensemble classifier is combined with k-nearest neighbor, naive Bayes network, C4.5, SVM and random forest. This class uses a weighted combination of multiple classifiers to improve the performance of human activity recognition technology. The experimental results show that the proposed AdaBoost ensemble classifier framework can enough accurately (95 % total accuracy) identify simple human activities, including walking, walking upstairs, walking downstairs, and site selection, standing and lying down. The AdaBoost ensemble classifier algorithm significantly improves HAR recognition accuracy.

References

1. Hsu Y., Yang S. // IEEE access. 2008. Vol. 1. P. 11–15.
2. Tao D., Guo Y. // IEEE Trans Image Process .2016. Vol. 11. P. 25–31.
3. Kim Y., Son Y. // Classification of Children’s sitting postures using machine learning algorithms. 2018. Vol. 2. P. 8–12.
4. Ermes M., Pärkkä J. // IEEE transactions on information technology in biomedicine. 2008. Vol. 13. P. 20–26.
5. Nazabal A., Ghahramani Z. // IEEE journal of biomedical and health informatics .2015. Vol. 20. P. 1342–1351.
6. Ahad M. // Intelligent Systems. 2017. P. 119–128.
7. Theodoridis S., Pikrakis A. Introduction to Pattern Recognition. Academic Press. 2010.
8. Maswadi K. // Multimedia Tools and Applications. 2021. Vol. 14. P. 9–26.
9. Vapnik V. The Nature of Statistical Learning Theory. Springer. 2000.
10. Breiman L., Friedman J. Classification and Regression Trees. CRC press. 1984.
11. Kondratieff Y. Introduction to machine learning. Morgan Kaufmann Press. 2018.

ФОРМИРОВАНИЕ ОБРАЗОВ ОШИБОК ДЛЯ ДЕКОДИРОВАНИЯ ИТЕРАТИВНЫХ КОДОВ

С.Х. ЖЭНЬ, И.М. ЧЭНЬ, В.К. КОНОПЕЛЬКО, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 28 февраля 2022

Аннотация. Рассматривается формирование библиотек образов ошибок при двумерном кодировании информации на основе анализа известных библиотек образов ошибок. В статье показано, что образы ошибок можно разделять на три подмножества: нетранспонированные, транспонированные и транспонированные сами в себя.

Ключевые слова: формирование образов ошибок, транспонированные образы, нетранспонированные образы, транспонированные образы сами в себя.

Введение

Двумерное (итеративное) кодирование применяется для борьбы с ошибками, возникающими при передаче, обработке и хранении информации в радиотехнических системах, инфокоммуникационных сетях [1, 2]. При двумерном кодировании кодовое слово представляется в виде таблицы и кодируется по строкам и столбцам соответствующими кодами $C_1(n_1, k_1, d_1)$ и $C_2(n_2, k_2, d_2)$, где n – длина кода, k – число информационных символов, $d = d_1 \cdot d_2 \geq 2t + 1$ – кодовое расстояние кода произведения, где t – кратность корректируемых ошибок в таблице данных [3]. Местоположение ошибок в кодовой матрице представляется образом ошибок – бинарной матрицей, в которой единичные элементы соответствуют ошибочным символам [4].

Для формирования библиотеки образов ошибок декодировании итеративных кодов в [5] предложен ранговый метод, основанный на синтезе исходного множества всех возможных матриц с заданным числом единиц, определяющим кратность ошибок, и прореживании исходного множества с помощью четырех параметров минимизации, учитывающих ранг и распределение единиц в строках и столбцах исходной матрицы. Из-за большого объема исходного множества и числа операций по его минимизации ранговый метод имеет высокую временную и пространственную сложность. Позиционный подход к формированию библиотеки образов ошибок [6], использующий меньший размер исходного множества и упрощенные условия минимизации, не позволяет существенно снизить вычислительную сложность.

Ниже рассматривается формирование библиотек образов ошибок на основе анализа известных библиотек.

Анализ известных библиотек образов ошибок

В [5, 6] исследованы методы формирования библиотек образов ошибок, основанные на обработке таблиц размерности $t \times t$, преобразованных в одномерные векторы длины, распределение этих векторов по группам, классам, типичным и нетипичным образам. Это позволяет проводить поиск образов ошибок, устранять избыточные, формировать отличающиеся друг от друга образы. Благодаря этому существенно сокращается время необходимое на проведение соответствующего анализа. Однако уже для $t \geq 4 \div 8$ время вычисления образов составляет десятки минут и часов и стремительно растет с дальнейшим увеличением кратности

ошибок. При применении позиционного и рангового методов наблюдаются высокие временные затраты для нахождения библиотек ошибок (табл. 1).

Табл. 1. Преобразованная библиотека образов для $t=4$ из [5]

Нетранспонированная (A) ↔ Транспонированная (A ^T)																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-01</p>	1	1	1	1													↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> </table> <p>T4-12</p>	1				1				1				1				<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td>1</td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-03</p>	1	1	1		1												↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-07</p>	1	1			1				1							
1	1	1	1																																																																		
1																																																																					
1																																																																					
1																																																																					
1																																																																					
1	1	1																																																																			
1																																																																					
1	1																																																																				
1																																																																					
1																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td>1</td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-04</p>	1	1				1	1										↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-08</p>	1	1			1					1							<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-05</p>	1	1	1					1									↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-13</p>	1				1				1							
1	1																																																																				
	1	1																																																																			
1	1																																																																				
1																																																																					
	1																																																																				
1	1	1																																																																			
			1																																																																		
1																																																																					
1																																																																					
1																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-06</p>	1	1					1	1									↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> </table> <p>T4-14</p>	1				1					1				1			<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-11</p>	1	1					1					1					↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> </table> <p>T4-12</p>	1				1					1					1	
1	1																																																																				
		1	1																																																																		
1																																																																					
1																																																																					
	1																																																																				
	1																																																																				
1	1																																																																				
		1																																																																			
			1																																																																		
1																																																																					
1																																																																					
	1																																																																				
		1																																																																			
Транспонированная сама в себя (A = A ^T)																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-02</p>	1	1			1	1											<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-09</p>	1	1			1						1						<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td></td><td>1</td><td>1</td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-10</p>		1	1		1				1								<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td>1</td></tr> </table> <p>T4-16</p>	1					1					1					1		
1	1																																																																				
1	1																																																																				
1	1																																																																				
1																																																																					
		1																																																																			
	1	1																																																																			
1																																																																					
1																																																																					
1																																																																					
	1																																																																				
		1																																																																			
			1																																																																		

В табл. 2 приведен анализ библиотек образов ошибок для $t=4$ [6].

Табл. 2. Преобразованная библиотека образов для $t=4$ из [6]

Нетранспонированная (A) ↔ Транспонированная (A ^T)																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-01</p>	1	1	1	1													↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> </table> <p>T4-03</p>	1				1				1				1				<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td>1</td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-04</p>	1	1	1		1												↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-08</p>	1	1			1				1							
1	1	1	1																																																																		
1																																																																					
1																																																																					
1																																																																					
1																																																																					
1	1	1																																																																			
1																																																																					
1	1																																																																				
1																																																																					
1																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td></td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-05</p>		1	1	1	1												↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> </table> <p>T4-11</p>		1			1				1				1				<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-06</p>	1	1			1		1										↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-10</p>	1	1			1					1						
	1	1	1																																																																		
1																																																																					
	1																																																																				
1																																																																					
1																																																																					
1																																																																					
1	1																																																																				
1		1																																																																			
1	1																																																																				
1																																																																					
	1																																																																				
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-07</p>	1	1					1	1									↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> </table> <p>T4-12</p>	1				1					1				1			<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td>1</td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-14</p>	1	1					1					1					↔	<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> </table> <p>T4-15</p>	1				1					1					1	
1	1																																																																				
		1	1																																																																		
1																																																																					
1																																																																					
	1																																																																				
	1																																																																				
1	1																																																																				
		1																																																																			
			1																																																																		
1																																																																					
1																																																																					
	1																																																																				
		1																																																																			
Транспонированная сама в себя (A = A ^T)																																																																					
<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-02</p>	1	1			1	1											<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td></td><td>1</td><td>1</td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-09</p>		1	1		1				1								<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td>1</td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <p>T4-13</p>	1		1			1			1								<table border="1" style="width: 40px; height: 40px; border-collapse: collapse;"> <tr><td>1</td><td></td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td></td></tr> <tr><td></td><td></td><td>1</td><td></td></tr> <tr><td></td><td></td><td></td><td>1</td></tr> </table> <p>T4-16</p>	1					1					1					1		
1	1																																																																				
1	1																																																																				
	1	1																																																																			
1																																																																					
1																																																																					
1		1																																																																			
	1																																																																				
1																																																																					
1																																																																					
	1																																																																				
		1																																																																			
			1																																																																		

Анализ библиотек [5, 6] образов ошибок кратностью $t \geq 2$ показал, что все эти образы можно разделить на три соответствующие подмножества: транспонированное, нетранспонированное и транспонированное сами в себя. В табл. 1 и 2 представлены библиотеки образов для $t = 4$ разбитые на три подмножества.

Установлено, что при создании библиотек образов ошибок не требуется производить поиск транспонированного подмножества (достаточно протранспонировать вычисленное нетранспонированное подмножество). Это позволит снизить вычислительные затраты на формирование библиотек образов ошибок в два раза. Благодаря использованию образов, которые переходят сами в себя при транспонировании для ошибок кратности $t = 2$ в качестве базового образа достаточно использовать только матрицы $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ и $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Для уменьшения вычислительной сложности формирования образов ошибок предлагается метод, основанный на итерационном расширении матриц базовых образов ошибок [7]. На рис. 1 приведена структура синтезированного генератора, в которой формирование подмножества образов ошибок производится для каждого предыдущего образа t путем последовательного увеличения размера матрицы образа и кратности ошибок до $t+1$, исключения составляют совпадающие матрицы для $t+1$. В качестве базовой матрицы, не исключенной из рассмотрения, можно выбирать любую из подмножества, которая и будет последующим образом ошибок. Таким образом, выбранное подмножество образов будет базовым для формирования последующих образов ошибок кратности $t+1$ [7].

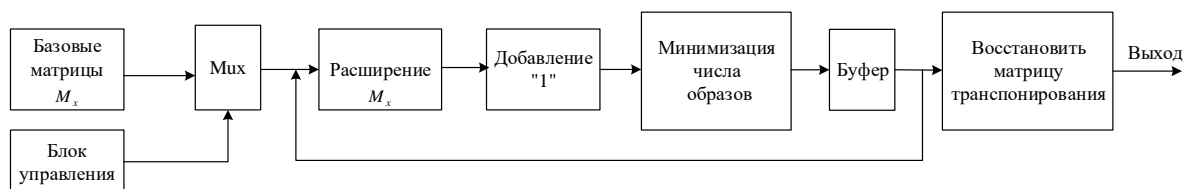


Рис. 1. Структурная схема генератора образов ошибок для итерационного метода

Из этого базового образа можно вычислять образы для $t = 3$ которые в свою очередь будут базовыми при синтезе образов ошибок кратности $t = 4$ и т.д. Это можно представить в виде графа (рис. 2).

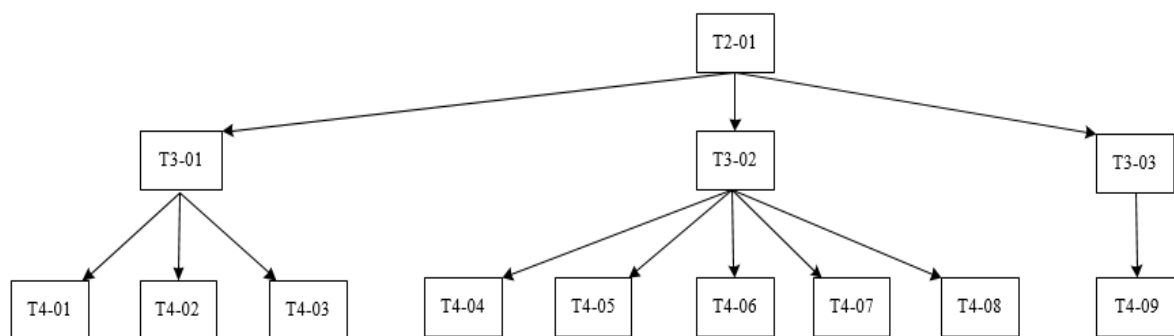


Рис. 2. Граф вычисления нетранспонированных подмножеств библиотек образов ошибок для $t = 2...4$

Предложенный метод отличается от известных рангового и позиционного методов итерационным формированием библиотеки образов ошибок в результате увеличения на каждой итерации размера предыдущей матрицы образа ошибок на единицу и деления образов ошибок на подмножества. В табл. 3 приведены библиотеки образов ошибок для $t = 4$ и их транспонированные множества при использовании предлагаемого решения.

Табл. 3. Результат формирования библиотек образов ошибок при $t=4$

Нетранспонированная (A)				Транспонированная (A ^T)																																																															
<table border="1"><tr><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-01	1	1	1	1													<table border="1"><tr><td>1</td><td>1</td><td>1</td><td></td></tr><tr><td></td><td></td><td></td><td>1</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-02	1	1	1					1									<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr></table> T4-01 ^T	1				1				1				1				<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr></table> T4-02 ^T	1				1				1					1		
1	1	1	1																																																																
1	1	1																																																																	
			1																																																																
1																																																																			
1																																																																			
1																																																																			
1																																																																			
1																																																																			
1																																																																			
1																																																																			
	1																																																																		
<table border="1"><tr><td>1</td><td>1</td><td>1</td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-03	1	1	1				1										<table border="1"><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td>1</td><td>1</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-04	1	1					1	1									<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-03 ^T	1				1				1	1							<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr></table> T4-04 ^T	1				1					1				1		
1	1	1																																																																	
		1																																																																	
1	1																																																																		
		1	1																																																																
1																																																																			
1																																																																			
1	1																																																																		
1																																																																			
1																																																																			
	1																																																																		
	1																																																																		
<table border="1"><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr><tr><td></td><td></td><td></td><td>1</td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-05	1	1					1					1					<table border="1"><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td>1</td><td>1</td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-06	1	1				1	1										<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td></td><td></td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr></table> T4-05 ^T	1				1					1					1		<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-06 ^T	1				1	1				1						
1	1																																																																		
		1																																																																	
			1																																																																
1	1																																																																		
	1	1																																																																	
1																																																																			
1																																																																			
	1																																																																		
		1																																																																	
1																																																																			
1	1																																																																		
	1																																																																		
Транспонированная сама в себя (A = A ^T)																																																																			
<table border="1"><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-07	1	1					1				1						<table border="1"><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-08	1	1					1			1							<table border="1"><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td>1</td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table> T4-09	1	1			1	1											<table border="1"><tr><td>1</td><td></td><td></td><td></td></tr><tr><td></td><td>1</td><td></td><td></td></tr><tr><td></td><td></td><td>1</td><td></td></tr><tr><td></td><td></td><td></td><td>1</td></tr></table> T4-10	1					1					1					1
1	1																																																																		
		1																																																																	
		1																																																																	
1	1																																																																		
		1																																																																	
	1																																																																		
1	1																																																																		
1	1																																																																		
1																																																																			
	1																																																																		
		1																																																																	
			1																																																																

Заключение

Анализ библиотек образов показал, что все образы ошибок известных библиотек можно разделить на три подмножества: нетранспонированные, транспонированные и транспонированные сами в себя. Это позволяет при создании библиотек образов ошибок уменьшить число анализируемых образов в 2 раза.

CLASSIFY ERROR PATTERNS FOR DECODING PRODUCT CODES

X.H. REN, Y.M. CHEN, V.K. KONOPELKO, V.Yu. TSVIATKOU

Abstract. The formation of libraries of error patterns in two-dimensional encoding of information based on the analysis of known libraries of error patterns is considered. The article shows that error patterns can be divided into asymmetric and symmetric, meanwhile the asymmetric patterns conclude two subsets: untransposed and transposed patterns.

Keywords: classification of error patterns, symmetric patterns, asymmetric patterns, untransposed patterns, transposed patterns.

Список литературы

1. Elias P. // In Transactions of the IRE Professional Group on Information Theory. 1954. Vol. 4. P. 29–37.
2. Конопелько В.К. [и др.]. Теория прикладного кодирования. Минск, БГУИР, 2004.
3. Lin. S., Costello D.J. Error Control Coding, 2nd edition. USA, 2004.
4. Смолякова О.Г., Конопелько В.К. // Докл. БГУИР. 2008. С. 19–28.
5. Смолякова О.Г., Конопелько В.К., Аль-Алем Ахмед Саид. // Материалы науч.-техн. конф. «ТСЗИ-2009». 2009. С. 71–72.
6. Конопелько В.К., Липницкий В.А., Спичекова Н.В. // Докл. БГУИР. 2010. №6. С. 127–131.
7. Ren X H., Konopelko V.K. // Материалы науч.-техн. конф. «Кодирование и цифровая обработка сигналов в инфокоммуникациях». Минск: БГУИР. 2020. С. 12–14.

УДК 654.16

ТЕХНОЛОГИЯ IAB ДЛЯ ГЕТЕРОГЕННОЙ РАДИОСЕТИ 5G NR

В.А. АКСЕНОВ, Е.М. ГНЕДЬКО

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь**Поступила в редакцию 28 февраля 2022*

Аннотация. Представлен обзор технологии Integrated Access and Backhaul (IAB), предназначенной для соединения большого количества пикосотовых базовых станций, работающих в рамках гетерогенной радиосети 5G. Представлен краткий анализ зависимости производительности пикосотовой сети при использовании указанной технологии.

Ключевые слова: 5G NR, диапазон FR2, формирование лучей ДН, гетерогенная топология сети, сеть доступа, транспортная сеть, OFDM, ресурсный блок.

Введение

Освоение частотного диапазона FR2 (24,250–52,600 ГГц), предусмотренного в рамках технологии 5G, крайне привлекательно возможностью использования действительно широкополосных каналов до 400 МГц и, соответственно, достижением ранее невиданных скоростей передачи в радиоинтерфейсе. Однако, на столь высоких частотах тяжело получить большие размеры сот в радиопокрытии. Для решения этой проблемы разработчики стандартов 5G предусмотрели, по сути, только два технических решения:

- использование антенных решеток, формирующих узкий лепесток диаграммы направленности, периодически сканирующий секторную соту (технология beamforming);
- использование концепции гетерогенной сети, HetNet, с организацией радиопокрытия территории посредством множества «быстрых» пико- и фемтосот, работающих внутри «медленных» макро- и микросот, использующих классические низкочастотные диапазоны (диапазон FR1 410–7125 МГц).

Имеются оценки [1], что на площади современной городской микросоты (радиус до 1000 метров), обслуживаемой одной базовой станцией, потребуется развернуть около 50 пикосот outdoor (вне зданий, радиус 50–150 метров) и еще больше фемтосот indoor (внутри зданий). Столь существенное повышение плотности базовых станций на территории невозможно без модифицирования транспортной сети (называемой в мировой практике Backhaul), посредством которой все базы соединяются с подсистемой коммутации (называемой core network).

Integrated Access and Backhaul

В настоящее время транспортная сеть Backhaul физически реализуется с использованием оптоволоконных линий, прокладываемых в городской канализации связи, или с использованием радиолинков (радиорелейных линий). Только второй вариант без ограничений подходит для быстрого и гибкого развертывания большого количества пикосотовых базовых станций. Однако, классическая архитектура сотового сайта с транспортным радиолинком, показанная на рис. 1, а, неприемлема из-за высокой цены на транспортное оборудование и необходимости платить за дополнительный радиоспектр, который этот радиолинк использует. Кажется, вполне логичным, что при использовании в 5G цифровых антенных решеток с синтезируемыми узкими лучами ДН, предложили выделить один такой луч для транспортного обмена в то время, как остальные лучи будут по-прежнему обеспечивать беспроводной доступ (access) терминалов в соте. Такое решение получило название интегрированного доступа и транспорта – Integrated Access and

Backhaul, IAB [2]. Более точно, в случае использования одной и той же антенной решетки для доступа и транспорта, говорят о «virtual IAB Mobile Termination» (vIAB-MT). Вариант архитектуры сайта при vIAB-MT показан на рис. 1, б. На рис. 1 сокращенно обозначены: RRU – remote radio unit; BBU – baseband unit; IDU/ODU – indoor/outdoor unit; CPRI – common public interface; IF – intermediate frequency; DU – distributed unit.

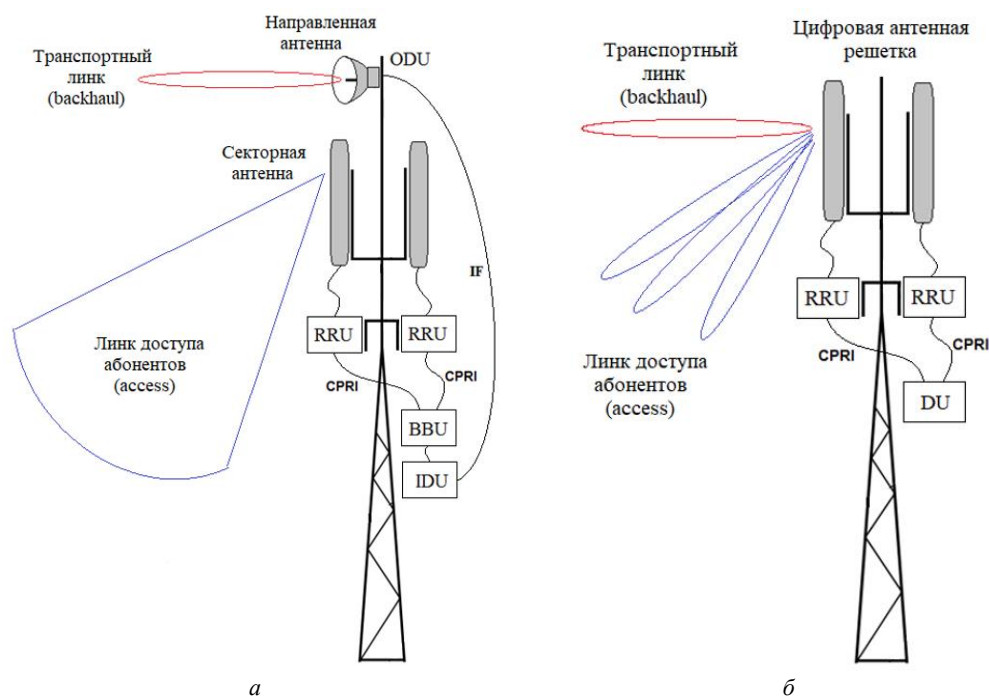


Рис. 1. Архитектура сайтов с беспроводным транспортным линком:
 а – сайт для стандартов 3G–4G; б – сайт 5G NR с технологией IAB

Следует отметить, что в общем случае для транспортного линка в рамках технологии IAB может использоваться и отдельная направленная антенна, так как трансиверы RRU предусматривают подключение к ним нескольких антенн. Такое решение сайта называется IAB-MT. При этом сохраняется главное принципиальное отличие IAB: транспортный линк и линк доступа используют одинаковую технологию передачи на основе OFDMA и одинаковый диапазон частот.

Топология соединения пикосотовых базовых станций по технологии IAB естественным образом распадается на два основных типа [3, 4]. Во-первых, это цепочечное соединение баз, расположенных вдоль проспектов, улиц, иных магистралей. В этом случае установка оборудования будет осуществляться на осветительные столбы. Такой симбиоз получил специальное название Pole Station (англ. «Станция на столбе»). Pole Station получает электропитание от коммунальной осветительной электросети и обслуживает трафик с заполненных публикой проспектов. Именно по этому сценарию уже развернуты пикосотовые кластеры в ряде городов США [5].

Во-вторых, это древоподобная ветвящаяся топология, которая более характерна для соединения пикосотовых баз, развернутых в зонах радиотени от зданий, иных «темных точках». Или в «горячих точках», случайным образом разбросанных на улице и/или внутри зданий и сооружений. Понятно, что в реальной сети используются оба из описанных вариантов топологии.

Система названий, введенная для обозначения элементов топологии транспортной сети IAB [2–4], представлена на рис. 2. Первая базовая станция в цепочке или корневая станция в «дереве» подключаются по высокопроизводительному оптическому каналу к Core Network.

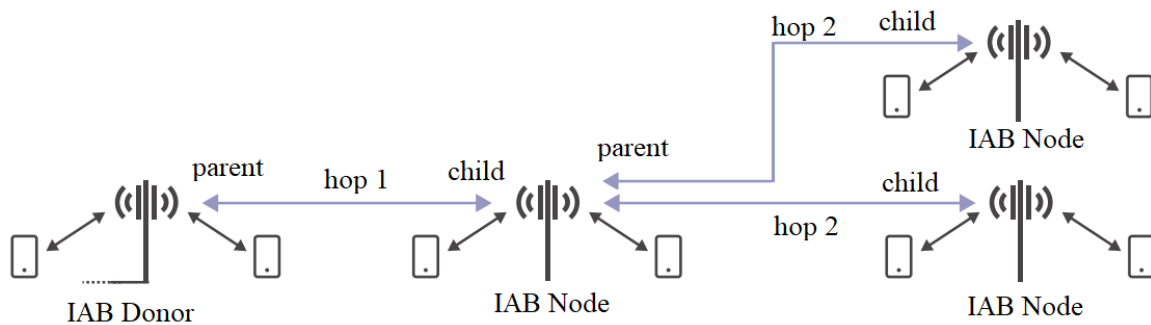


Рис. 2. Система названий для обозначения элементов топологии транспортной сети IAB

Такие базы играют роль донора – IAB Donor – всей вышестоящей транспортной сети. Остальные базовые станции в сети именуется просто узлами – IAB Node. Передача транспортной информации от донора к ближайшему узлу (узлам) и от него далее называется переходом (англ. hop – перескок). Соответственно, появляется понятие номера или порядка перехода. Все узлы в цепочке, начиная от донора и далее, вступают между собой в иерархические отношения по схеме «parent – child», или «родительский узел – дочерний узел».

Влияние IAB на производительность сотовой сети

Передача информации в радиointерфейсе 5G NR, как и в LTE, носит гранулированный характер [6]. В качестве минимальной гранулы выступает ресурсный блок Resource Block (RB), состоящий из 12 поднесущих. Для частотного диапазона FR2 предусмотрены нумерологии, равные 3 и 4. Количество передаваемых RB для этих нумерологий представлены в табл. 1. Около 10 % процентов ресурсных блоков используются под служебные цели (синхронизация, управление антеннами и лучами, опорные сигналы, измерения и подобное). Так же важно отметить, что для FR2 стандартизирован только временной дуплекс TDD.

Табл. 1. Количество ресурсных блоков RB в радиointерфейсе 5G NR для диапазона FR2

Нумерология	Разнос поднесущих, кГц	Минимальное количество RB	Максимальное количество RB	Минимальная ширина канала, МГц	Максимальная ширина канала, МГц
3	120	24	275	34,56	396
4	240	24	138	69,12	397,44

В первом приближении можно считать, что по транспортной сети backhaul будут доставляться RB, необходимые для обслуживания нагрузки в сотах, в линиях доступа. Но как транспортные RB, так и RB доступа будут черпаться из одного и того же ресурса. Положим, что используется нумерология 3 и минимальная ширина канала 34,56 МГц с 20-ю ресурсными блоками для трафика. Например, если двум оконечным базам на рис. 2 потребуется по 10 RB для обслуживания нагрузки, то донорская база и база первого перехода останутся совершенно без возможности обслуживать свою собственную нагрузку в данный квант времени.

Очень опасны для IAB сети проблемы в радиоканалах транспортных линков, поскольку во всех системах с OFDMA в этом случае выполняется переход к модуляции поднесущих в RB от более эффективного вида модуляции на менее эффективный по скорости, но более помехоустойчивый. Представим себе, что в нашем примере для сети на рис. 2 крайние базы ждут ресурсные блоки с битной емкостью, соответствующей модуляции 16QAM, а на участке первого перехода из-за интерференции, например, возможна передача лишь с модуляцией QPSK, которая в четыре раза «медленнее». В оконечных сотах произойдет блокировка в реализации востребованных абонентами сервисов.

Однако наиболее интересное и не столь очевидное влияние на производительность работы сотовой сети с технологией IAB оказывает топология самой этой сети. Более точно, тот критерий, которым руководствуются при соединении донора и узлов IAB друг с другом. В [7] было выполнено моделирование производительности сети, состоящей из трех доноров и 54 пикосот, распределенных на некоторой территории. В первом случае доноров и другие узлы соединили

по критерию высокого качества транспортного радиоканала (максимизации в линиях RSRP – Reference Signal Received Power). Получилась топология, показанная на рис. 2, *а*. Во втором случае использовали критерий «не более трех дочерних узлов от каждого родительского», или иначе, не более трех прямых соединений для каждого узла. Топология под этот критерий показана на рис. 2, *б*. Каждая цветная линия на рисунках указывает переход своего порядка.

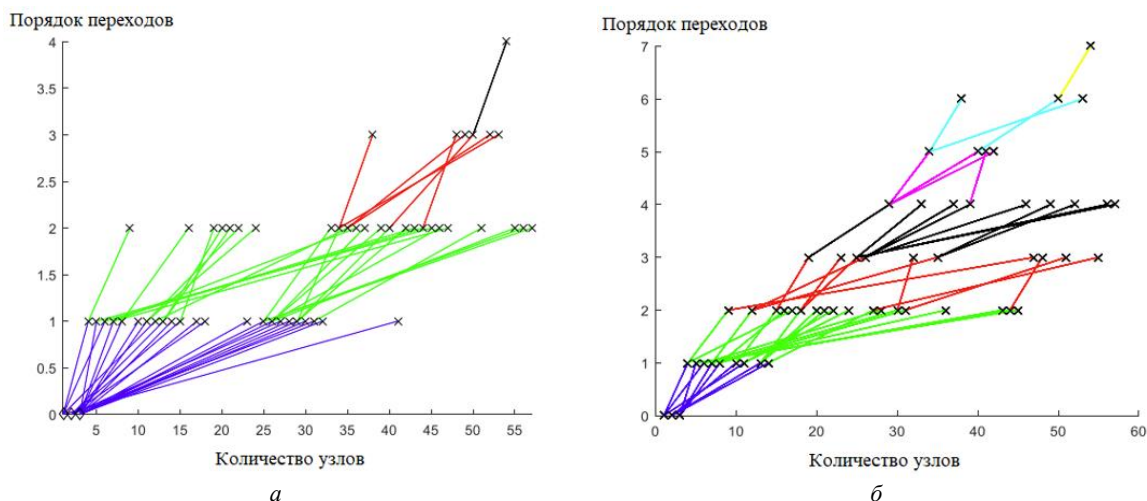


Рис. 2. Варианты топологии IAB сети для разных критериев соединения узлов: *а* – критерий максимизации RSRP; *б* – критерий «не более трех дочерних узлов от каждого родительского»

Как видно из рис. 2, эти два метода создают очень разные топологии с точки зрения порядка переходов. Критерий максимизации RSRP дает меньший порядок переходов в целом (до 4-х), но приводит к множественным переходам от одного родительского узла IAB. В случае ограничения количества прямых соединений до 3-х наблюдается рост порядка переходов до 7.

Результаты моделирования показаны в табл. 2. Хотя среднее число переходов увеличено, топология с дочерним ограничением в 3 узла имеет в 10 раз лучшую воспринимаемую пользователем пропускную способность (user perceived throughput, UPT) 5 %-го перцентиль. Он также имеет на 75 % лучшую среднюю пропускную способность DL (50 %-ый перцентиль) по сравнению с методом максимизации RSRP. В [7] полагается, что столь существенная разница связана с меньшей перегрузкой начальных транзитных переходов от трех узлов-доноров, а нагрузка лучше сбалансирована и распределена по топологии сети IAB.

Табл. 2. Сравнение воспринимаемой пользователем пропускной способности для различных топологий IAB

Метод формирования топологии транспортной сети	5 %-ый перцентиль UPT	50 %-ый перцентиль UPT
Максимизация RSRP	8 Мб/с	160 Мб/с
Три прямых соединения для узла 4	80 Мб/с	280 Мб/с

Заключение

Приведенный обзор технологии IAB показывает, что она определенно дает большие преимущества операторам сотовой связи в быстром и недорогом развертывании пикосотовых гетерогенных сетей, обеспечивающих сверхскоростной доступ мобильных пользователей в диапазоне FR2. Однако потенциальные возможности пикосотового покрытия могут оказаться не достигнутыми из-за нюансов распределения частотно-временного ресурса и конкретной топологии транспортной сети IAB. Требуется дальнейшие теоретические и экспериментальные исследования для выработки оптимальных решений при проектировании и реализации сетей с технологией IAB.

IAB TECHNOLOGY FOR 5G NR HETNET

V.A. AKSYONOV, E.M. GNEDZKO

Abstract. An overview of the Integrated Access and Backhaul (IAB) technology is presented, designed to connect a large number of pico-cell base stations operating within a heterogeneous 5G radio network. A brief analysis of the dependence of the performance of a picocellular network using this technology is presented.

Keywords: 5G NR, FR2 frequency range, beamforming, HetNet, access network, backhaul network, OFDM, resource block.

Список литературы

1. IAB—the cost-effective solution to quickly expand 5G mmWave coverage. [Электронный ресурс]. URL: <https://www.rcrwireless.com/20200727/5g/iab-the-cost-effective-solution-to-quickly-expand-5g-mmwave-coverage-analyst-angle>.
2. 3GPP TS 38.174 V16.5.0 (2021-12). NR; Integrated access and backhaul radio transmission and reception (Release 16). [Электронный ресурс]. URL: <https://www.3gpp.org/dynareport/38174.htm>.
3. Henrik Ronkainen [et al.] // Ericsson technology review. June 23, 2020. [Электронный ресурс]. URL: <https://www.ericsson.com/49e6f6/assets/local/reports-papers/ericsson-technology-review/docs/2020/introducing-integrated-access-and-backhaul.pdf>.
4. Integrated Access and Backhaul for 5G. [Электронный ресурс]. URL: <https://www.mpirical.com/blog/integrated-access-and-backhaul-for-5g>.
5. Draft Small Cell Design Guidelines. Third version. March 21, 2019. [Электронный ресурс]. URL: https://ddot.dc.gov/sites/default/files/dc/sites/ddot/page_content/attachments/Final%20Third%20Version%20of%20the%20Small%20Cell%20Guidelines.pdf.
6. 5G NR Resource Block Definition and RBs Calculation. [Электронный ресурс]. URL: <https://www.techplayon.com/nr-resource-block-definition-and-rbs-calculation>.
7. 5G Americas White paper. Innovations in 5G Backhaul Technologies: IAB, HFC & Fiber. Copyright 2020 5G Americas. [Электронный ресурс]. URL: <https://www.5gamericas.org/wp-content/uploads/2020/06/Innovations-in-5G-Backhaul-Technologies-WP-PDF.pdf>.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АВТОМАТИЗАЦИИ ОЦЕНИВАНИЯ ТЕМАТИК ТЕКСТОВОГО КОНТЕНТА

Н.Д. РУДНИЧЕНКО, В.В. ВЫЧУЖАНИН, А.А. ЕГОШИНА,
С.М. ВОРОНОЙ, Н.О. ШИБАЕВА

Государственный университет «Одесская политехника», Украина

Поступила в редакцию 25 февраля 2022

Аннотация. В работе приведены результаты разработки и исследования интеллектуальной системы автоматизации оценивания тематик текстового контента на базе применения алгоритмов машинного обучения. Проведен анализ проблематики в задачах обработки естественного языка, обоснована актуальность рассматриваемых задач. Разработана концепция работы системы, описан ее компонентный состав, приведена характеристика собранных данных для анализа, проведены численные эксперименты по обучению тестированию созданных моделей машинного обучения, подтверждающие эффективность использования бустинга и опорных векторов для решения поставленной задачи. Предложены пути дальнейшего совершенствования разработанной системы.

Ключевые слова: обработка естественного языка, машинное обучение, интеллектуальные информационные системы.

Введение

В связи с постоянным ростом информационных источников и технических устройств поддержки процессов ознакомления человека с интересующими данным задача эффективного анализа разнородных текстов большого объема является важной практически в любой современной области человеческой деятельности, где его экспертиза, опыт и знания могут быть изложены в текстовом виде [1]. Появление широкого доступа к сети Интернет привело к бурному росту объемов доступной текстовой информации, находящей в открытом доступе, что значительно ускорило развитие научных направлений в области обработки текста, известной как обработка естественного языка (ОЕЯ) [2]. Спецификой данной сферы научной деятельности является автоматизация процессов обработки и анализа текстовых данных на базе использования методов и подходов искусственного интеллекта (ИИ) и машинного обучения (МО) [3–5].

Важными задачами в сфере ОЕЯ являются анализ тональности текстового контента, соотнесения его к заданным категориям, оценка уникальности концепций и идей контента, уровня согласованности и полноты текста, статистических показателей и характеристик, а также сжатия объемов текста с минимизацией потери содержательной части [6]. Отсутствие универсального и системного подхода в методиках решения поставленных задач приводит к тому, что для существующих программных решений характерным является низкий уровень эффективности и гибкости при решении обозначенных задач [7]. Это связано в том числе с тем, что полноценная автоматизация всех процессов обработки и анализа текстовых данных возможна только при реализации логики понимания системой текста на уровне человека, с учетом определения сложно формализуемых или нечетких метаданных, что требует создания полноценного ИИ, не являющегося пока возможным и осуществимым технически из-за нехватки знаний о специфике процесса восприятия информации (в частности, в текстовом виде) человеком [8–13].

В связи с этим возникает необходимость разработки и внедрения в процессы анализа текстовых данных гибких, функциональных и удобных интеллектуальных систем (ИС) и

моделей, обеспечивающих как возможности решения бизнес-процессов, так и проведения экспериментальных аналитических исследований, в частности, мониторинга эффективности каждого из имплементированных подходов с возможностью внесения корректировок в его работу, расширение функционала благодаря модульному составу [14–16].

Цель данной работы является исследование возможностей решения задачи оценки и соотнесения текстового контента к сформированному множеству тематических категорий с целью автоматизации обработки и сегментации информационных материалов, размещаемых на различных ресурсах сети Интернет и в системах электронного документа.

Концепция интеллектуальной системы

Концептуально, ИС состоит из 3 отдельных модулей: импорта и предобработки данных, интеллектуального анализа данных, визуализации и интерпретации результатов вычислительных экспериментов.

Выбранными алгоритмами МО для проведения сравнительного исследования работы системы являются: логистическая регрессия, наивный классификатор Байеса, метод опорных векторов, метод ближайшего соседа, случайный лес и метод бустинга.

Ключевым функционалом ИС является: загрузка «сырого» набора данных в виде набора текстовых файлов, содержащих в себе текст статьи и название раздела, к которому она будет принадлежать; преобразования составленного набора данных в структурированный вид, имеющий табличный вид, более удобный для дальнейшего анализа и работы; первичный анализ данных, которая включает в себя подсчет и определение метаданных, в том числе количества текстовых статей в каждой из категорий, а также расчет процента текстов по категориям, размер статьи в формате txt; сохранение набора данных, прошедших первичный анализ в формате *.csv для его удобной обработки программным обеспечением; очистка текстов, включающий в себя удаление специальных символов и знаков препинания, которые не нужны для дальнейшей классификации текстов, объединение слов, написанных с большой и маленькой буквы, несущих одинаковый смысл; лемматизации и стемминг; расчет TF-IDF вектора; создание словарей наиболее значимых слов для проведения оценки текстов; инициирование процесса обучения на базе выбранных алгоритмов МО. Обученные модели сохраняются в формате *.pickle; выбор лучшей модели созданных классификаторов путем сравнения их точности и эффективности по метрикам; представление результатов в виде графиков с использованием двух методов снижения размерности: PCA и t-SNE график; визуализация результатов классификации текстов с указанием точности работы каждой модели МО.

После запуска ИС пользователь осуществляет загрузку данных, которые будут обрабатываться. Выбор наборов данных для обучения возможен путем подключения разных источников (форматы csv, txt, doc, не защищенный pdf, файлы баз данных) или из сети Интернет (ссылки на веб-ресурсы или на прямые текстовые файлы). Все данные хранятся в промежуточном формате txt для упрощения процесса дальнейшей обработки.

Далее модуль программы, отвечающий за обработку данных получает dataset, загруженный в систему, формируя на выходе из всех загруженных наборов текста сводную таблицу, содержащей колонки с названием файла, его содержанием, категорией (тематикой), к которой относится текст. Последний столбец является выходной переменной, может быть размечен вручную в рамках данного модуля, в случае наличия этого набора в датасете он может быть отмечен с помощью элементов интерфейса.

На следующем этапе данные, представленные в виде таблицы, проходят первичный анализ, чтобы выяснить, насколько такая выборка является полной для обучения моделей ML и насколько точно набор данных описывает предметную область.

Во время первичного анализа рассчитываются такие показатели как подсчет количества текстов в каждой из выбранных категорий, процент, который составляют тексты с каждой рубрики в соотношении к общему количеству текстов и средний объем текста в каждой из выбранных тематик. В результате длинные тексты удаляются из выборки, что позволяет сбалансировать данные исходя из указанных пользователей параметров и ограничений.

Далее проводится этап очистки данных для перехода к дальнейшему анализу его состава. В это время проводится уменьшение объема текстов, выделяются наиболее значимые

семантические элементы. Очистка текстов выполняется путем изъятия специальных символов и знаков препинания, удаляются часто повторяющиеся слова, не несущие существенного смыслового содержания (союзы, наречия) и притяжательные местоимения, формы единственного и множественного числа в одинаковых словах объединяются в одну, проводится приведение слов к единому регистру. Отдельно выполняются процессы лемматизации и стемминга.

Следующим этапом является создание пакета словарей, состоящих из слов, которые чаще всего встречаются в текстах отдельных рубрик, что позволяет обеспечить семантическую связь между контекстом статей. Обработанные данные и словари сохраняются отдельными файлами для дальнейшего использования в обучении.

Пользователь задает тематики, по которым необходимо провести процесс оценки принадлежности текстов. Далее иницируются вычислительные процессы по разбиению выборок текстовых данных на обучающие и тестовые, создаются объекты классов по каждому из 6 используемых методов ML. Для проведения вычислительных процессов в параллельном режиме на нескольких вычислительных узлах предусмотрена опция подключения Apache Mesos и библиотеки Numba для задействования ресурсов графических аппаратных ускорителей.

На базе полученных результатов вычислений производится расчет метрик оценки качества классификации, главной из которых является точность (отношение правильных ответов к общему числу экземпляров текстов). Еще одной метрикой является полнота классификации, как соотношение правильно положительных ответов ко всем положительным ответам.

После этого полученные данные оценки метрик сохраняются в текстовый файл, агрегируются и передаются на вход графического модуля, на котором осуществляется визуализация графиков анализа точности классификации. Для большей степени интерпретируемости результатов необходимым является использование методов снижения размерности данных. Используются две методики: анализ основных компонентов, сущность которой заключается в расчете собственных значений и собственных векторов матрицы данных для формирования минимального количества переменных, которые обеспечивают максимальное значение дисперсии в данных. Используется в случае, когда объемы текста не превышают 10 000 знаков; t-SNE, в своей основе содержит вероятностный подход, что удобно при визуализации данных больших размеров (более 10 000 знаков). Позволяет минимизировать расхождение между распределением, измеряющим попарно сходство входящих объектов, и распределением, измеряющим попарно сходство соответствующих низкомерных точек встраивания.

Компонентный состав интеллектуальной системы

Программная имплементация ИС выполнена на основе использования языка программирования Python 3.8, среды разработки PyCharm и ряда функциональных библиотек по обработке и анализу данных, созданию моделей МО. Компонентный состав ИС в виде набора пакетов приведен на рис. 1. Главный компонент программы (MainApp) выполняет функции вызова других компонентов ИС для обработки запросов приложения по обработке данных, выполнению вычислительных процессов по созданию моделей ML, построения графиков и общей бизнес-логики работы системы.

Для удобства проект ИС, представляющий собой модульную структуру, в среде разработки разделен на отдельные пакеты, в каждом из которых содержатся классы, реализующие необходимый функционал исходя из их логической связности между собой. Каждый из которых содержит свой набор программных компонентов.

«Raw dataset» включает в себя необработанные данные в виде файлов и логику реализации импорта.

«Dataset Creation» обеспечивает логику создания структурированного набора данных.

«ExploratoryDataAnalysis» содержит классы, отвечающие за первичный анализ данных.

«Feature Engineering» обеспечивает очистку данных для дальнейшей обработки и выделение признаков.

«Model Training» содержит соответствующие классы, реализующие вызов и создание объектов моделей классификаторов, а также их сериализованные структуры, загружаемые в процессе анализа данных через интерфейс пользователя. Также в данном пакете размещен класс,

содержащий методы для сравнения всех классификаторов и выбора наиболее эффективного для поставленной задачи с учетом расчета метрик.

«ScrapingData» имплементирует логику сбора текстовых данных с указанных источников для дальнейшего анализа.

«UI» реализует функционал графического интерфейса пользователя и вывода визуализаций с элементами управления.

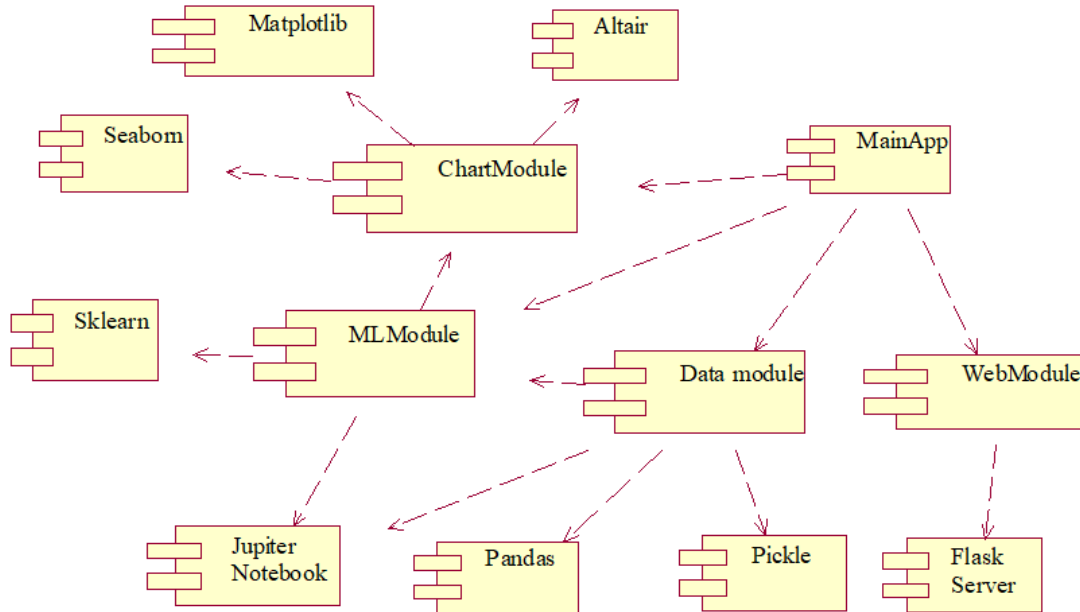


Рис. 1. Диаграмма основных компонентов ИС

Исследование результатов экспериментов

В качестве исходных данных взяты наборы статей из веб-сайтов qz, theguardian, Feedly. Этапами исследования стали сбор и предобработка данных, отбор статей по наиболее популярным рубрикам для обучения моделей МО и оценки их точности с подбором наиболее подходящих значений гиперпараметров для лучшего из выявленных классификаторов.

На базе анализа собранной метаинформации по статьям было подобрано пять основных категорий статей: бизнес, развлечения, политика, IT и здоровье. По результатам подсчета количество текстов в каждой из категорий построена диаграмма (рис. 2), позволяющая нагляднее представить данные (процент каждой из категорий в выборке). Согласно данной визуализации наибольшее число текстов в категории IT, а меньше всего в развлечениях, но несмотря на это данные являются сбалансированными, общая выборка репрезентативна, что позволяет избежать дополнительных манипуляций по их обработке.

В процессе проведения исследования выбрано случайное разбиение набора данных: 75% наблюдений, составляющих тренировочный набор и 25 % наблюдений, составляющих тестовый набор. Процесс настройки гиперпараметров моделей был выполнен с перекрестной валидацией тренировочного набора, чтобы каждая модель МО стала более обобщающей.

Следующим этапом был проведен расчет TF-IDF метрик для определения насколько важно каждое слово в тексте. Формализация процесса заключается в следующем: если слово встречается в любом документе часто, при этом встречаясь редко во всех других документах, то это слово имеет большую значимость для того же документа. Следует отметить, что всего выделено около 500 слов для каждой из тематик.

Формирование словарей из отдельных слов для каждой тематики текстов позволила обеспечить лучшие (чем при обычных словосочетания) соответствие отдельных слов семантическому смыслу в рамках каждой категории.

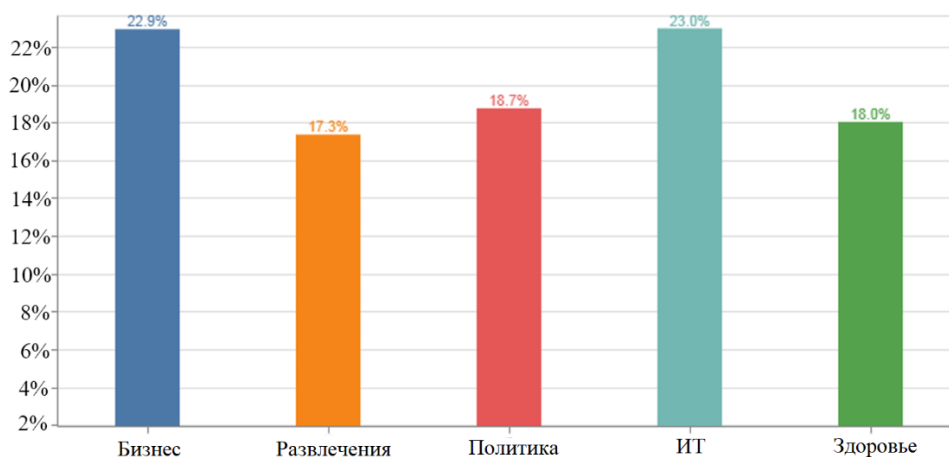


Рис. 2. Диаграмма распределения данных по классам выборки

Результаты оценки созданных моделей МО в рамках ИС приведены в табл. 1.

Табл. 1. Результаты оценки созданных моделей МО

№	Название модели	Обучающая выборка			Тестовая выборка		
		Точность	Полнота	Временные затраты, с	Точность	Полнота	Временные затраты, с
1	Бустинг	0,991	0,908	150	0,912	0,813	40
2	Опорные вектора	0,956	0,888	112	0,893	0,811	31
3	Наивный байес	0,713	0,663	43	0,65	0,558	12
4	Ближайший сосед	0,739	0,552	48	0,701	0,615	11
5	Случайный лес	0,885	0,782	140	0,798	0,745	33
6	Логистическая регрессия	0,771	0,699	38	0,743	0,703	9

Согласно табл. 1 следует отметить, что модель бустинга является наиболее точной. Наивный байесовский классификатор и метод ближайших соседей показывают менее значительные результаты, однако их скорость работы достаточно высока.

Высокие показатели точности и полноты характерны для модели опорных векторов, скорость ее обучения и тестирования также выше модели бустинга почти на 30 %, минимальный разброс значений между тестовой и тренировочной выборками также наблюдается для данной модели и для логистической регрессии, являющейся наиболее быстрой, но существенно менее точной.

Выводы

В результате проведения данного исследования была разработана интеллектуальная система автоматизации процесса оценки тематик текстового контента базе использования предложенной концепции предобработки данных и создания разных моделей машинного обучения с успешной апробацией ее использования на реальных наборах данных. Установлено, что наиболее эффективными моделями для соотнесения текстового контента к заданным тематикам являются модели бустинга и опорных векторов. В дальнейшей работе в данном направлении предлагаемыми путями расширения возможностей системы являются: внедрения возможностей итеративного прогона моделей в выбранном диапазоне во многопоточном режиме, что может обеспечить существенное снижение временных затрат на проведение вычислительных экспериментов; расширение методики предобработки данных путем выделения дополнительной метаинформации по статистическим и семантическим параметрам текстов.

INTELLECTUAL SYSTEM OF TEXTUAL CONTENT SUBJECT ESTIMATION AUTOMATIZATION

N.D. RUDNICHENKO, V.V. VYCHUZHANIN, A.A. EGOSHINA,
S.M. VORONOV, N.O. SHIBAEVA

Abstract. This article presents the results of the development and research of an intellectual system of textual content subject estimation automatization based on the application of machine learning algorithms. The analysis of problems in the tasks of natural language processing is carried out, the relevance of the tasks under consideration is substantiated. The concept of system was developed, its component structure was described, the characteristics of the data collected for analysis were described and numerical experiments on training and testing of created models of machine learning were conducted. These experiments confirmed the effectiveness of using boosting and support vectors for solving the task. Ways to further improve the developed system are proposed.

Keywords: natural language processing, machine learning, intelligent information systems.

Список литературы

1. Вычужанин В.В., Рудниченко Н.Д. Методы информационных технологий в диагностике состояния сложных технических систем. Одесса, 2019.
2. Рудниченко Н. и др. // Информационные управляющие системы и технологии. Проблемы и решения. 2019. С. 31–46.
3. Rudnichenko N. [et al.] // Proceedings of the 9th International Conference «Information Control Systems & Technologies». 2020. P. 371–385.
4. Lidong W., Cheryl A. // International Journal of Mathematical, Engineering and Management Sciences. 2016. Vol. 1. P. 52–61.
5. Junfei Q. [et al.] // EURASIP Journal on Advances in Signal Processing. 2016.
7. Bliznyuk B.O. [et al.] // Visnik of Kharkiv National University of the Name of V.N. Karazin. 2017.
8. Yuskov V.S., Barannikova I.V. // Mining information and analytical bulletin. 2017. Vol. 3. P. 272–278.
9. Widiastuti N.I. // IOP Conference Series: Materials Science and Engineering. 2018. Vol. 407.
10. Li H. // National Science Review. 2018. Vol. 5.
11. Noskov D.V. // Bulletin of Science and Education. 2018. Vol. 40. P. 39–41.
12. Jain A., Mandowara J. // International Journal of Computer Application. 2016. Vol. 2.
13. Krasnyansky M.N. [et al.] // Voronezh State University Bulletin. Series: Systems Analysis and Information Technology. 2018. Vol. 3. P. 173–182.
14. Степанов П.А. // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2013. № 12. С. 106–112.
15. Юсков В.С., Баранникова И.В. // Горный информационно-аналитический бюллетень. 2017. № 3. P. 272–278.
16. Вахтина О.П. // Вестник магистратуры. 2019. № 98. С. 21–24.

ВЕРОЯТНОСТНОЕ ДЕКОДИРОВАНИЕ КОДОВ С ВЫСОКОЙ ПЛОТНОСТЬЮ ПРОВЕРОК НА ЧЕТНОСТЬ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА РАСПРОСТРАНЕНИЯ ДОВЕРИЯ

М.А. АЛИСИЕНКО, А.П. ТУРЛАЙ, С.Б. САЛОМАТИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 18 февраля 2022

Аннотация. Рассмотрены подходы вероятностного декодирования кодов с высокой плотностью контроля четности на основе алгоритма распространения доверия с жесткими и адаптивными мягкими решениями. Рассмотрены примеры схем декодирования с использованием нейронной сети. Предложена методика оценки сложности декодирования.

Ключевые слова: алгоритм распространения доверия, коды с высокой плотностью проверок на четность, проверочная матрица, нейронная сеть, адаптивный алгоритм декодирования.

Введение

Современные коды, такие как турбокоды, коды LDPC и полярные коды, обычно достигают хорошей производительности только при относительно больших длинах блоков. При коротких и средних длинах блоков коды с высокой плотностью проверок на четность (HDPC), такие как коды ВСН и коды Рида-Соломона, могут быть лучшим выбором из-за их свойства высокой корректирующей способности [1–3]. Хотя для таких кодов существуют алгоритмы жесткого решения низкой сложности, эти алгоритмы не достигают полной мощности исправления ошибок, которую имеют короткие коды HDPC при оптимальном декодировании. С другой стороны, по своей природе классические коды имеют очень плотную проверочную матрицу с небольшим обхватом и большим количеством коротких циклов. Поэтому выполнение на них стандартного итеративного декодирования обычно приводит к плохим результатам.

В данной работе рассматриваются алгоритмы построения эффективных итеративных декодеров кодов с высокой плотностью контроля четности.

Вероятностный алгоритм распространения доверия

Алгоритм распространения доверия выполняется на основе векторов, состоящих из вещественных величин, полученных на выходе канала путем пересчета вероятностей через алгоритм распространения доверия (belief propagation decoding). На основе принятого из канала вектора \mathbf{C} формируются два вектора, для двоичного случая, вероятностей того, что в принятом векторе на данной позиции находился заданный символ [3–5].

Каждому ненулевому элементу проверочной матрицы кода присваивают две величины: $q_{i,j}^x$ и $r_{i,j}^x$. Величина $q_{i,j}^x$ является вероятностью того, что j -ый символ принятого вектора имеет значение x по информации, полученной из всех проверок, кроме i -ой. Величина $r_{i,j}^x$ является вероятностью того, что проверка i выполняется, если j -ый символ принятого вектора равен x , а все остальные символы проверок имеют распределение вероятностей, заданное величинами $\{q_{i,j'}^x : j' \in N(i) \setminus j\}$ где $N(i)$ – множество символов, входящих в i -ую проверку.

Перед началом работы алгоритму требуется инициализация, далее работа идет по принципу пересчета вероятностей символов принятого вектора, используя для пересчета

вероятностей правило Байеса для апостериорной вероятности события. Одна итерация алгоритма представляет собой следующую последовательность действий:

1. Для всех проверок вычисляются величины $\Delta r_{i,j}$ и пересчитываются вероятности $r_{i,j}^x$ для $x = \{0,1\}$.
2. Для всех символов принятого вектора пересчитываются вероятности $q_{i,j}^x$.
3. Формируются векторы псевдоапостериорной вероятности q_j^0 и q_j^1 .
4. Формируется вектор решения \mathbf{C}' по следующему правилу: элемент вектора $c_j = 1$, если q_j^1 больше заданного порогового значения, иначе $c_j = 0$.
5. Если вектор \mathbf{C}' является кодовым словом, декодирование заканчивается, в противном случае выполняется следующая итерация алгоритма.

Существует несколько алгоритмов мягкого декодирования для кодов HDPC с высокой производительностью одним из которых является адаптивный алгоритм вероятностного декодирования.

Сложность «мягких» алгоритмов выше, чем сложность алгоритма «жесткого» декодирования путем инвертирования битов, но качество декодирования повышается за счет использования дополнительной информации на выходе канала. Однако, точность работы «мягкого» алгоритма зависит от инициализации. Чем точнее инициализация произведена, тем точнее будет конечный результат. Для канала с гауссовским шумом инициализация может быть произведена при помощи информации о дисперсии шума в канале. Для других распределений шума в канале или при неизвестных характеристиках шума точная инициализация алгоритма может оказаться сложной задачей.

Адаптивный алгоритм «мягкого» вероятностного декодирования

Известно несколько алгоритмов мягкого декодирования для кодов HDPC с высокой производительностью, одним из которых является адаптивный алгоритм вероятностного декодирования ВР [1, 4].

Модификация алгоритма ВР путем использования нейронной сети, значительно улучшает его производительность для кодов HDPC. В нейронном алгоритме ВР, сообщения вычисляются так же, как и в традиционном ВР на основе LLR, а затем умножаются на набор мультипликативных весов. Веса уникальны для каждой итерации; следовательно, в качестве композиции аффинных преобразований и нелинейностей нейронный декодер ВР можно рассматривать как своего рода не полностью связанную нейронную сеть с прямой связью, как показано на рис. 1. Каждый уровень сети состоит из обучаемой операции передачи сообщений.

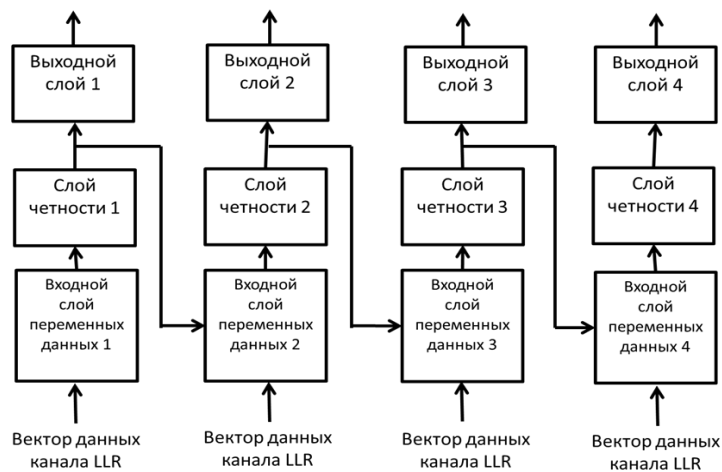


Рис. 1. Декодер на основе не полностью связанной нейронной сети с прямой связью

Нейронный алгоритм ВР является примером «глубокого развертывания» [6–8]. Методологически итерационный алгоритм (который может быть или не быть алгоритмом обучения) разворачивается, и каждой итерации назначаются обучаемые параметры, чтобы получить глубокую архитектуру.

Для обучения нейронных декодеров ВР может быть использован алгоритм минимизации средней перекрестной энтропии между каждым программным выходом декодера (пропущенным через сигмовидную функцию для получения значений в диапазоне [0, 1]) и каждым соответствующим битом переданного кодового слова.

Возможна модификация алгоритма декодирования «сумма-произведение» (SP). В основе модификации лежит алгоритм адаптации матриц проверки четности от итерации к итерации на основе логарифмического отношения правдоподобия (LLR) входящего сигнала. Аппаратная реализация декодера требует применения алгоритма исключения Гаусса, который используется в каждой итерации SP.

Декодер на основе линейного программирования

Метод интерпретирует проверочную матрицу как набор ограничений и использовали методы линейного программирования для декодирования полученного кодового слова [9].

Декодер линейного программирования (ДЛП) действует на фундаментальном многограннике, созданном ограничениями матрицы проверки четности. Многогранник имеет вершины, которые являются допустимыми кодовыми словами, и вершины, которые являются недопустимыми кодовыми словами. Декодирование линейным программированием в конечном итоге останавливается на одной из вершин, которое называют псевдокодовым словом (ПКС) [10]. Сходство в производительности и аналогичное поведение декодирования ДЛП и распространения доверия (ВР) привели к использованию веса и распределения РСВ для оценки и улучшения производительности декодирования.

Обогащение структуры матрицы контроля четности путем добавления избыточных строк улучшает функцию распределения веса ПКС и производительность декодирования. Здесь необходимо учитывать, что добавление большего количества строк в матрицу проверки на четность может увеличить количество коротких циклов в графе и снизит производительность его декодирования.

Алгоритмы декодирования используют избыточные матрицы проверки на четность. Один из вариантов декодирования определяет матрицы проверки на четность ($n \times n$), путем получения их из кодовых слов минимального веса двойственного кода [1]. Возможно использование «временной» избыточной матрицы проверки на четность путем изменения базовой матрицы проверки на четность на протяжении всего процесса декодирования. Различная структура избыточных матриц декодеров влияет на вычислительную сложность декодирования [11, 12].

«Временная» избыточность, достигается за счет использования группы изоморфизмов перестановок $\text{Per}(C)$, которая определяется как набор перестановок координатных мест, которые переводят код C в себя. Случайным образом выбирая элементы из $\text{Per}(C)$, декодер изменяет набор ограничений, которые используются во времени. Этот процесс также можно рассматривать как смену многогранника при очередной попытке приблизиться к решению или как декодирование с переставленным мягким входным вектором, что эквивалентно декодированию с помощью переставленной матрицы контроля четности.

Методика оценки сложности декодирования

Все выражения алгоритма задаются как функции числа ребер двудольного графа, индуцированного кодом. Таким образом, сравнение двух разных декодеров, которые используют разные матрицы проверки на четность, достигается путем нормализации количества ребер их графов.

Если алгоритмы используют одну и ту же матрицу проверки на четность, то сложность декодера оценивается путем усреднения количества итераций суммы-произведения, выполняемых до тех пор, пока не будет достигнуто допустимое кодовое слово или пока

количество итераций не достигнет определенного предела в случае, если декодер не сходится к кодовому слову. Одна итерация определяется как действие по отправке сообщений из узлов-переменных в узлы-переменные, обработка этих сообщений в узлах-проверках и отправка новых сообщений из узлов-проверок в узлы-переменные.

Матрица проверки четности \mathbf{H} для некоторого кода $C(n, k)$ имеет размер $(n \times n)$. Нормируя относительную сложность, можно установить все декодеры в одном масштабе. Количество ребер в графе равно nd_{\min} . Относительная сложность – это отношение количества ребер в графах. Сложность алгоритмов можно сравнить, умножив среднее количество итераций суммы-произведения на показатель относительной сложности.

Так, для кода Голея (24, 12, 8) относительную сложность можно оценить, как 1,7. Для кода БЧХ (31, 16, 7) относительная сложность равна 1,5; для кода (63, 45, 7) – относительная сложность равна 3,2. Здесь относительная сложность пропорциональна длине кода.

Заключение

Рассмотрены различные алгоритмы итеративного декодирования кодов с высокой плотностью проверок на четность. Показано, что итерационное адаптивное декодирование позволяет повысить эффективность «мягкого» декодирования таких кодов. Приведена методика и результаты оценки сложности декодирования некоторых кодов.

PROBABILISTIC DECODING OF CODES WITH A HIGH DENSITY OF PARITY CHECKS USING THE BELIEF PROPAGATION ALGORITHM

M.A. ALISIENKO, A.P. TURLAY, S.B. SALOMATIN

Abstract. Approaches for probabilistic decoding of high-density parity codes based on a belief propagation algorithm with hard and adaptive soft decisions are considered. Examples of decoding schemes using a neural network are considered. A technique for estimating the complexity of decoding is proposed algorithm for searching for single-pixel extremes of halftone images based on centrally symmetric scanning is proposed. It is shown that the algorithm works much faster than the best known algorithms for detecting key points of images.

Keywords: belief propagation algorithm, high-density parity-check codes, parity check matrix, neural network, adaptive decoding algorithm.

Список литературы

1. Richardson T., Urbanke R. Modern Coding Theory. Cambridge, UK: Cambridge University Press, 2008.
2. Koetter R., Vardy A. // IEEE Transaction on Information Theory. 2003. Vol. 49. P. 2809–2825.
3. Doan N. [et al.] // arXiv preprint arXiv: 1811.00124. 2018.
4. Luo F. Machine learning for future wireless communication. Wiley TEEE Press, 2020
5. Tseng H., Wu J. // The International Journal of Electronics.1993. Vol. 75, No. 4. P. 589–594.
6. Liva G. [et al.] // arXiv preprint arXiv: 1610.00873. 2016.
7. O’Shea T., Hoydis J. // IEEE Transactions on Cognitive Communications and Networking. 2017. Vol. 3, No. 4. P. 563–575.
8. Hershey J.R., Le Roux J., Weninger F. // Mitsubishi Electric Research Laboratories. Tech. Rep. 2014. TR2014–117.
9. Feldman J. Decoding Error-Correcting Codes via Linear Programming. PhD dissertation. Massachusetts Institute of Technology. 2003.
10. Kelley C.A., Sridhara D. // IEEE Trans. Inform. Theory. 2007. Vol. 53, No. 11, P. 4013–4038.
11. Fossorier M. P. C., Mihaljević M., Imai H. // IEEE Trans. Commun. 1999. Vol. 47. P. 673–680.
12. Halford T.R. The Extraction and Complexity Limits of Graphical Models for Linear Codes. PhD dissertation. University of Southern California, 2007.

УРОВНЕВЫЕ МОДЕЛИ И ВОЛНОВЫЕ АЛГОРИТМЫ СЕГМЕНТАЦИИ АСМ-ИЗОБРАЖЕНИЙ

В.В. РАБЦЕВИЧ, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 28 февраля 2022

Аннотация. Предложены уровневые модели и алгоритмы волновой сегментации полутоновых изображений поверхностей материалов, получаемых с помощью атомного силового микроскопа. Показано, что комбинация результатов нескольких алгоритмов позволяет повысить точность сегментации и определить соответствие фрагментов изображений предложенным моделям.

Ключевые слова: сегментация изображений, модели изображений, атомная силовая микроскопия.

Введение

Для исследования поверхностей материалов широко используется атомная силовая микроскопия (АСМ) [1–3]. Она основана на оценке интенсивности силового воздействия со стороны поверхности на микронзонд и позволяет преобразовать информацию о топологии поверхности в полутоновое АСМ-изображение. Одна из задач, решаемых при анализе АСМ-изображений, заключается в автоматической (без участия оператора) оценке количества структурных составляющих (частиц, фаз, зерен) на поверхностях материалов. Для решения данной задачи используется сегментация, выделяющая области пикселей АСМ-изображений, соответствующие данным объектам.

Для сегментации АСМ-изображений широко используются градиентные и маркерные алгоритмы морфологического водораздела [4–6]. Эти алгоритмы выделяют области без предварительной информации об изображении, но имеют значительные ошибки сегментации и высокую вычислительную сложность. Меньшую вычислительную сложность имеют алгоритмы, основанные на волновом выращивании областей [7–9], но они требуют определения начальных точек роста. Это осуществляется под контролем оператора т.к. автоматическое выделение начальных точек [10–11] приводит к недостаточной или избыточной сегментации. Алгоритмы выращивания областей без предварительного выбора начальных точек роста [12] имеют наименьшую вычислительную сложность, но сегментируют АСМ-изображения с большой ошибкой.

Целью работы является повышение точности сегментации АСМ-изображений.

Модели АСМ-изображений

Предлагаются три модели АСМ-изображений, представляющих различные варианты расположения объектов относительно друг друга на поверхности материала.

Одноуровневая модель $\{M_p, M_s, M_e, F_1\}$ АСМ-изображения представляет плотно расположенные объекты простой формы (рис. 1, а) в виде матрицы пикселей $M_p = \|m_p(y, x)\|_{(y=0, Y-1, x=0, X-1)}$ размером $Y \times X$ пикселей по вертикали и горизонтали ($m_p(y, x) \in [0, 2^B - 1]$), где B – разрядность пикселей, определяющая ширину динамического

диапазона $[0, 2^B - 1]$), разделенной на N_S областей $S = \|s(n_S)\|_{(n_S=1, N_S)}$ локальных максимумов $E = \|e(n_S)\|_{(n_S=1, N_S)}$ с монотонным уменьшением яркости; где M_S – матрица сегментации, $M_S = \|m_S(y, x)\|_{(y=0, Y-1, x=0, X-1)}$; M_E – матрица локальных максимумов, $M_E = \|m_E(y, x)\|_{(y=0, Y-1, x=0, X-1)}$; F_1 – специфичные для одноуровневой модели ограничения на распределение яркостей пикселей в областях локальных максимумов. Похожая модель предложена в [13, 14], но для поиска локальных максимумов и минимумов на полутоновых изображениях с широким динамическим диапазоном.

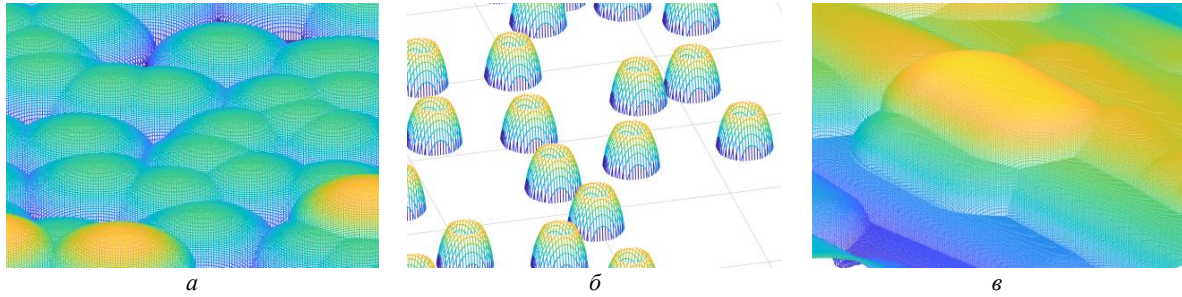


Рис. 1. Варианты расположения объектов на поверхности материала: *a* – для одноуровневой модели; *б* – для двухуровневой модели; *в* – для многоуровневой модели

Значения элементов $m_S(y, x)$ матрицы M_S сегментации указывают на номера n_S областей $s(n_S)$, которым они принадлежат ($m_S(y, x) = n_S$, $n_S \in [1, N_S]$) (рис. 2). Значения элементов $m_E(y, x)$ матрицы M_E локальных максимумов указывают на соответствие пикселей изображения n_S -му локальному максимуму ($m_E(y, x) = n_S$) или несоответствие локальным максимумам ($m_E(y, x) = 0$). Локальный максимум $e(n_S)$ может состоять из одного или более элементов: $\forall y \forall x (m_E(y, x) = n_S) \Rightarrow (m_E(y, x) \in e(n_S))$. Количество локальных максимумов $e(n_S)$ совпадает с количеством областей $s(n_S)$ изображения и каждой области изображения соответствует один локальный максимум: $\forall n_1 \forall n_2 (s(n_1) \leftrightarrow e(n_2)) \Rightarrow (n_1 = n_2)$.

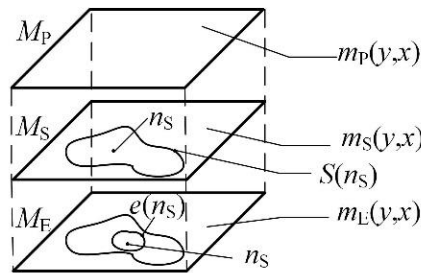


Рис. 2. Связь между матрицами в моделях АСМ-изображений

Для каждой области $s(n_S)$ АСМ-изображения, описываемого одноуровневой моделью, существует только один локальный максимум $e(n_S)$. Следовательно, между каждым пикселем $m_P(y_B, x_B)$ области $s(n_S)$ ($m_S(y_B, x_B) = n_S$) и соответствующим локальным максимумом $m_P(y_E, x_E)$ ($m_E(y_E, x_E) = n_S$, $m_S(y_E, x_E) = n_S$) области (если максимум многопиксельный, то любым его пикселем) существует множество $R(y_B, x_B)$ последовательностей пикселей (маршрутов) $r(y_B, x_B, n_R)$, в которых значения пикселей монотонно убывают, начиная с локального максимума $m_P(y_E, x_E)$, $R(y_B, x_B) = \{r(y_B, x_B, n_R)\}_{(n_R=1, N_R)}$, где n_R – номер маршрута; N_R –

количество маршрутов. Маршрут $r(y_B, x_B, n_R)$ монотонного изменения яркости между пикселями $m_P(y_B, x_B)$ и $m_P(y_E, x_E)$ образуется последовательностью из $N_P(n_R)$ пикселей $r(y_B, x_B, n_R) = \{m_P(y_R(y_B, x_B, n_R, n_P), x_R(y_B, x_B, n_R, n_P))\}_{(n_P=1, N_P(n_R))}$ таких, что выполняется $m_S(y_R(y_B, x_B, n_R, n_P), x_R(y_B, x_B, n_R, n_P)) = n_S$, где n_P – порядковый номер пикселя в маршруте n_R ; наименьший номер $n_P = 1$ соответствует пикселю с координатами (y_B, x_B) : $m_P(y_R(y_B, x_B, n_R, 1), x_R(y_B, x_B, n_R, 1)) = m_P(y_B, x_B)$, а наибольший номер $n_P = N_P(n_R)$ соответствует пикселю с координатами (y_E, x_E) : $m_P(y_R(y_B, x_B, n_R, N_P(n_R)), x_R(y_B, x_B, n_R, N_P(n_R))) = m_P(y_E, x_E)$.

Для одноуровневой модели $\{M_P, M_S, M_E, F_1\}$ распределение яркостей пикселей в области $s(n_S)$ локального максимума $m_P(y_E, x_E)$ определяется существованием для каждого пикселя $m_P(y_B, x_B)$ из области $s(n_S)$ хотя бы одного маршрута $r(y_B, x_B, n_R)$, все $N_P(n_R)$ элементов которого удовлетворяют ограничению F_1 , определяемому с помощью выражения

$$\begin{aligned} & \exists n_R (m_P(y_R(y_B, x_B, n_R, n_P), x_R(y_B, x_B, n_R, n_P))) \leq \\ & \leq m_P(y_R(y_B, x_B, n_R, n_P + 1), x_R(y_B, x_B, n_R, n_P + 1)) \wedge \\ & \wedge (m_S(y_R(y_B, x_B, n_R, n_P + 1), x_R(y_B, x_B, n_R, n_P + 1)) = n_S) \Rightarrow \\ & \Rightarrow (m_S(y_B, x_B) \leftarrow n_S) \end{aligned} \quad (1)$$

при $n_P = 1, N_P(n_R) - 1$.

Пример профиля яркости АСМ-изображения для одноуровневой модели приведен на рис. 3, а.

Двухуровневая модель $\{M_P, M_S, M_E, F_2\}$ АСМ-изображения представляет отдельно расположенные объекты простой формы (рис. 1, б), где F_2 – специфичные для двухуровневой модели ограничения на распределение яркостей пикселей в областях локальных максимумов. Для этой модели распределение яркостей пикселей в области $s(n_S)$ локального максимума $m_P(y_E, x_E)$ определяется существованием для каждого пикселя $m_P(y_B, x_B)$ из области $s(n_S)$ хотя бы одного маршрута $r(y_B, x_B, n_R)$, все $N_P(n_R)$ элементов которого удовлетворяют ограничению F_2 , определяемому с помощью выражения

$$\begin{aligned} & \exists n_R ((d_P(n_P + 1, n_P + 2) \leq d_P(n_P, n_P + 1)) \vee \\ & \vee (d_P(n_P + 1, n_P + 2) > d_P(n_P, n_P + 1)) \wedge d_P(n_P, n_P + 1) \geq T_P) \wedge \\ & \wedge (m_S(y_R(y_B, x_B, n_R, n_P + 1), x_R(y_B, x_B, n_R, n_P + 1)) = n_S) \Rightarrow \\ & \Rightarrow (m_S(y_B, x_B) \leftarrow n_S) \end{aligned} \quad (2)$$

при $n_P = 1, N_P(n_R) - 1$,

где $d_P(j, i) = m_P(y_R(y_B, x_B, n_R, i), x_R(y_B, x_B, n_R, i)) - m_P(y_R(y_B, x_B, n_R, j), x_R(y_B, x_B, n_R, j))$; T_P – минимальное значение разности значений смежных пикселей, определяющее принадлежность пикселя объекту, а не подложке.

Нулевые элементы матрицы сегментации модели соответствуют подложке. Пример профиля яркости АСМ-изображения для двухуровневой модели приведен на рис. 3, б.

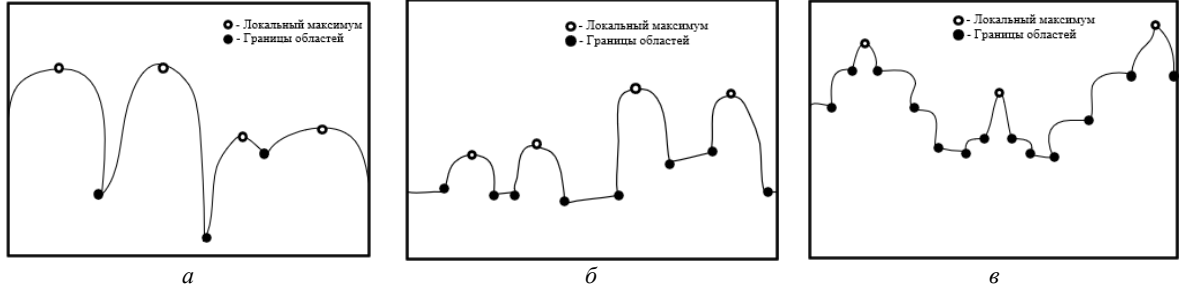


Рис. 3. Примеры профилей яркости АСМ-изображений: *a* – для одноуровневой модели; *б* – для двухуровневой модели; *в* – для многоуровневой модели

Многоуровневая модель $\{M_P, M_S, M_E, M_L, F_L\}$ АСМ-изображения представляет плотно или отдельно расположенные объекты сложной формы (рис. 1, *в*), где F_L – специфичные для многоуровневой модели ограничения на распределение яркостей пикселей в областях локальных максимумов; M_L – матрица декомпозиции АСМ-изображения, $M_L = \|m_L(y, x)\|_{(y=0, Y-l, x=0, X-l)}$; $m_L(y, x)$ – элемент матрицы уровней, $m_L(y, x) \in [0, L_S]$; L_S – количество уровней декомпозиции. Для этой модели распределение яркостей пикселей в области $s(n_S)$ локального максимума $m_P(y_E, x_E)$ определяется существованием для каждого пикселя $m_P(y_B, x_B)$ из области $s(n_S)$ хотя бы одного маршрута $r(y_B, x_B, n_R)$, все $N_P(n_R)$ элементов которого удовлетворяют ограничениям F_L , определяемым с помощью выражений

$$\begin{aligned} & \exists n_R \left((d_P(n_P + 1, n_P + 2) \leq d_P(n_P, n_P + 1)) \vee \right. \\ & \left. \vee (d_P(n_P + 1, n_P + 2) > d_P(n_P, n_P + 1)) \wedge d_P(n_P, n_P + 1) \geq T_P \right) \wedge \\ & \wedge (m_S(y_R(y_B, x_B, n_R, n_P + 1), x_R(y_B, x_B, n_R, n_P + 1)) = n_S) \Rightarrow \\ & \Rightarrow (m_S(y_B, x_B) \leftarrow n_S), m_L(y_B, x_B) \leftarrow l_S \end{aligned} \quad (3)$$

при $n_P = 1, \overline{N_P(n_R) - 1}$,

$$\begin{aligned} & \exists n_R (d_P(n_P, n_P + 1) < T_P) \wedge \\ & \wedge (m_P(y_R(y_B, x_B, n_R, n_P), x_R(y_B, x_B, n_R, n_P)) \leq \\ & \leq m_P(y_R(y_B, x_B, n_R, n_P + 1), x_R(y_B, x_B, n_R, n_P + 1))) \wedge \\ & \wedge (m_S(y_R(y_B, x_B, n_R, n_P + 1), x_R(y_B, x_B, n_R, n_P + 1)) = n_S) \Rightarrow \\ & \Rightarrow (m_S(y_B, x_B) \leftarrow n_S), m_L(y_B, x_B) \leftarrow l_S \end{aligned} \quad (4)$$

при $n_P = 1, \overline{N_P(n_R) - 1}$,

где l_S – номер уровня декомпозиции, $l_S \in [1, L_S]$.

Условия (3) и (4) описывают области АСМ-изображений соответственно с высокой и низкой скоростью уменьшения яркости пикселей в направлении от локального максимума. В случае сложной формы объектов необходимо учитывать несколько уровней декомпозиции для того, чтобы выделить все объекты по условиям (3) и (4). Каждый уровень декомпозиции рассматривается независимо от других. Сегменты, соответствующие одному уровню декомпозиции имеют одинаковые номера в матрице M_L декомпозиции. Пример профиля яркости АСМ-изображения, соответствующего многоуровневой модели приведен на рис. 3, *в*.

Алгоритмы выделения объектов на АСМ-изображениях

На основе моделей (1) – (4) разработаны соответствующие алгоритмы выделения на АСМ-изображениях объектов различных типов. Алгоритмы основаны на волновом выращивании областей локальных максимумов с их выбором в порядке убывания значений [15], сущность которого состоит в использовании изменяющегося от максимума к минимуму порога q яркости ($q = \overline{2^B - 1, 0}$) для выбора пикселей роста областей (локальных максимумов) или пикселей, присоединяемых к пикселям смежных существующих областей, яркость которых удовлетворяет ограничениям F_1 , F_2 или F_L .

Алгоритм L1 сегментации АСМ-изображений, описываемых одноуровневой моделью, состоит из следующих шагов [15].

1. Загрузка изображения M_p ; формирование матриц M_S , M_E ; $q \leftarrow 2^B - 1$.
2. Цикл по порогу q (пока $q \neq 0$ и существуют пиксели, удовлетворяющие условию F_1).
 - 2.1. Поиск локальных максимумов, соответствующих порогу q .
 - 2.2. Присоединение к существующим областям пикселей, соответствующих порогу q и удовлетворяющих условию F_1 .
 - 2.3. Уменьшение порога яркости: $q \leftarrow q - 1$.
 - 2.4. Окончание цикла по порогу q яркости.
3. Вывод матрицы сегментации M_S .

Алгоритм L2 сегментации АСМ-изображений, описываемых двухуровневой моделью, отличается от алгоритма сегментации АСМ-изображений, описываемых одноуровневой моделью, только условием F_2 присоединения пикселей, соответствующих порогу, к существующим областям.

Алгоритм LM сегментации АСМ-изображений, описываемых многоуровневой моделью, состоит из следующих шагов.

1. Загрузка изображения M_p ; формирование матриц M_S , M_E , M_L ; $l_S \leftarrow 1$.
2. Цикл по уровням l_S декомпозиции (пока существуют несегментированные пиксели).
3. Инициализация порога яркости: $q \leftarrow 2^B - 1$.
4. Цикл по порогу q (пока $q \neq 0$ и существуют пиксели, удовлетворяющие условиям F_L).
 - 4.1. Поиск локальных максимумов, соответствующих порогу q .
 - 4.2. Присоединение к существующим областям пикселей, соответствующих порогу q и удовлетворяющих условию F_L .
 - 4.3. Уменьшение порога яркости: $q \leftarrow q - 1$.
 - 4.4. Окончание цикла по порогу q яркости.
5. Обнуление всех сегментированных пикселей: $\forall y \forall x (m_S(y, x) > 0) \Rightarrow (m_p(y, x) \leftarrow 0)$ при $y = \overline{0, Y - 1}$, $x = \overline{0, X - 1}$.
6. $l_S \leftarrow l_S + 1$.
7. $d \leftarrow d + 1$.
8. Окончание цикла по уровням l_S декомпозиции.
9. Вывод матриц сегментации M_S и декомпозиции M_L .

Описание топологий поверхностей материалов на основе комбинаций моделей АСМ-изображений

Фрагменты реального АСМ-изображения могут соответствовать различным моделям: $\{M_p, M_S, M_E, F_1\}$, $\{M_p, M_S, M_E, F_2\}$, $\{M_p, M_S, M_E, M_L, F_L\}$. Применение несоответствующего алгоритма (L1, L2, LM) для сегментации может привести к ошибкам. В этой связи предлагается комплексный подход к сегментации АСМ-изображений, основанный на формировании и анализе результатов различных комбинаций трех предложенных алгоритмов.

В табл. 1 приведено описание результатов применения алгоритмов сегментации на основе предложенных моделей к АСМ-изображениям различных типов. Символом «+» отмечены результаты сегментации без ошибок. Символом «-» отмечены результаты недостаточной сегментации.

Табл. 1. Результаты применения алгоритмов L1, L2, LM сегментации к АСМ-изображениям, описываемым одно-, двух- и многоуровневой моделями

Алгоритм сегментации	Результаты сегментации АСМ-изображений, соответствующих моделям		
	Одноуровневая модель	Двухуровневая модель	Многоуровневая модель
L1	+	-	-
L2	+	+	-
LM	+	+	+
$L1 - L2$	0 ($L1 = L2$)	Сегментация простой подложки	Недостаточная сегментация сложной подложки
$L1 \cap LM$	+	Соотнесение областей и прилегающих фрагментов простой подложки	Соотнесение фрагментов объектов сложной формы
$\overline{L2} \cap LM$	0 ($L2 = 0$)	Сегментация простой подложки	Сегментация сложной подложки

Из табл. 1 следует, что только алгоритм LM сегментирует АСМ-изображения любого типа без ошибок. Однако он имеет высокую вычислительную сложность, обусловленную многократным повтором сканирования изображения в поисках начальных точек роста и присоединяемых к существующим областям пикселей, и не позволяет классифицировать фрагменты АСМ-изображений. Из табл. 1 следует также, что комбинации результатов нескольких алгоритмов позволяют получить больше информации о структуре поверхности материала, чем результаты отдельных алгоритмов. Комбинации результатов позволяют также определить тип модели для каждого фрагмента АСМ-изображения.

Заключение

Для повышения точности сегментации АСМ-изображений поверхностей материалов предложены одно-, двух- и многоуровневая модели, определяющие условия присоединения пикселей к существующим сегментам. На основе данных моделей разработаны три волновых алгоритма сегментации АСМ-изображений, имеющих различную вычислительную сложность и точность выделения областей на АСМ-изображениях, описываемых различными уровнями моделями. Показано, что комбинирование результатов этих алгоритмов позволяет получить больше информации о структуре поверхности материала, чем результаты отдельных алгоритмов.

LEVEL MODELS AND WAVE ALGORITHMS FOR SEGMENTATION OF AFM IMAGES

V.V. RABTSEVICH, V.Yu. TSVIATKOU

Abstract. Level models and algorithms for wave segmentation of halftone images of material surfaces obtained with an atomic force microscope are proposed. It is shown that the combination of the results of several algorithms can improve the accuracy of segmentation and determine the correspondence of image fragments to the proposed models.

Keywords: image segmentation, image models, atomic force microscopy.

Список литературы

1. Ulyanova T.M. [et al.] // Crystallography Reports. 2006. Vol. 51. P. 144–149.
2. Kuznetsova T. [et al.] // Journal of Materials Engineering and Performance. 2016. Vol. 25 (12). P. 5450–5459.
3. Geisse N.A. // Materials Today. 2011. Vol. 12 (7–8). P. 40–45.
4. Vincent L. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1991. Vol. 13 (6). P. 583–598.
5. Gonzalez R. C., Woods R.E. Digital Image Processing, Third Edition. 2008.
6. Beucher S., Lantuéjoul C. Use of watersheds in contour detection. International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation. 1979.
7. Adams R., Bischof L. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1994. Vol. 16 (6). P. 641–647.
8. Fan M., Lee T. // Image Processing IET. 2015. Vol. 9 (6). P. 478–485.
9. Fan J. [et al.] // Pattern Recognition Letters. 2005. Vol. 26 (8). P. 1139–1156.
10. Van Herk M. // Pattern Recognition Letters. 1992. Vol. 13. P. 517–521.
11. Gil J., Werman M. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1993. Vol. 15. P. 504–507.
12. Pratt W.K. Digital Image Processing, Third Edition. 2001.
13. Нгуен А.Т., Цветков В.Ю. // Информатика. 2019. № 16 (3). С. 23–36.
14. Нгуен А.Т., Цветков В.Ю. // Докл. БГУИР. 2021. № 19 (4). С. 61–69.
15. Рабцевич В.В. [и др.] // Известия НАН Беларуси. 2019. № 2 (121). С. 215–231.

УДК 519.686

РАСПОЗНАВАНИЕ ПАТОЛОГИЙ ГРУДНОЙ КЛЕТКИ С ПОМОЩЬЮ ПРЕДВАРИТЕЛЬНОГО ОБУЧЕНИЯ НА ОБЩЕДОСТУПНЫХ НАБОРАХ ДАННЫХ

А.И. НАУМОВИЧ

Белорусский государственный университет, Республика Беларусь

Поступила в редакцию 28 февраля 2022

Аннотация. Рассмотрена проблема распознавания патологий грудной клетки в условиях малых объемов обучающей выборки. Предложен подход к решению задачи, основанный на нейросетевой технологии. Разработана архитектура и способ обучения сети, а также алгоритм преобразования набора данных для обучения.

Ключевые слова: сверточная нейронная сеть, распознавание образов, обработка рентгенографических изображений.

Введение

В условиях пандемии важнейшей задачей врачей является выявление заболеваний легких на ранней стадии их проявления [1–3]. В настоящее время для получения данных о состоянии пациентов с заболеваниями легких используют широкий спектр различных стационарных и мобильных флюорографических аппаратов. Из-за резкого увеличения числа пациентов возрастает нагрузка на врачей-специалистов, поэтому актуальной в такой ситуации становится проблема автоматизации процесса обработки гетерогенных данных.

Для эффективного решения обозначенной проблемы требуется разработать соответствующий набор алгоритмов и на их основе программное обеспечение, автоматизирующее трудоемкие этапы обработки данных [4–5].

Формализация и постановка задачи

Прежде, чем перейти к постановке задачи, рассмотрим процесс принятия решения при традиционном подходе к проведению скрининга. Как отмечалось выше, он проводится среди широких слоев населения с использованием различного рода стационарного и портативного медицинского оборудования. Полученные в результате разнородные рентгенографические снимки в кратчайшие сроки должны быть обработаны врачами рентгенологами. Поэтому актуальной является задача автоматизации процесса анализа разнородных рентгенографических изображений. Необходимо разработать и реализовать на практике алгоритмы, которые позволят специалисту-рентгенологу уделить максимальное внимание неоднозначным (сложным) случаям течения болезни. Такие алгоритмы должны гарантировать уверенное (предварительное) решение в максимальном количестве случаев. Когда же степень «уверенности» алгоритма в поставленном диагнозе мала, необходимо предоставить ему интерпретацию работы алгоритма. Интерпретация должна дать возможность специалисту быстро определить справедливость или ложность поставленного алгоритмом предварительного диагноза. Наконец, разрабатываемый подход к решению задачи должен быть достаточно производительным применено к большому объему гетерогенных данных.

Для решения обозначенной проблемы широко используются нейросетевые технологии обработки данных. Важным этапом применения таких технологий является обучение нейронной сети. Процесс, как правило, реализуют на общедоступном наборе данных. Затем выполняют так

называемую калибровку нейронной сети на целевом наборе. Наличие размеченного (целевого) набора данных является в некотором смысле недостатком рассматриваемого подхода. Потому что на практике сбор и подготовка размеченных данных (для каждого типа оборудования) является достаточно дорогостоящим процессом. Поэтому актуальной является задача разработки алгоритмов, требующих минимальный объем данных из целевого домена для обучения нейронной сети.

Проведем формализацию и постановку задачи.

Под рентгенографическим изображением (или просто изображением) будем понимать тензор $x \in [0,1]^{w \times h \times c}$, где $[0,1]$ – допустимый диапазон яркости пикселей изображения, w – ширина изображения, h – высота изображения, а c – количество каналов изображения.

Пусть $X = \{x\}$ – некоторое множество рентгенографических изображений. Пусть также $P = \{p_1, p_2, \dots, p_k\}$ – множество патологий, которые можно обнаружить на снимках из X (включая также отсутствие патологии на снимке). Пусть задано множество $X_0 \subset X$, такое, что для $\forall x \in X_0, x \rightarrow y \in P$, т.е. каждому объекту множества X_0 ставится в соответствие метка патологии из множества P .

Необходимо построить алгоритм $A: X_0 \times X \rightarrow P$, который по входному рентгенографическому изображению распознал бы на нем патологию из множества P .

Следует заметить, что постановка задачи не предполагает наличие целевого домена данных.

Пусть $X_D \subset X$ – множество изображений из целевого домена. Наличие меток патологий P_D для этого набора изображений не уточняется. В таком случае необходимо построить алгоритм $A: X_0 \times X_D \times X \rightarrow P$.

По результатам теоретических исследований необходимо разработать архитектуру и состав программного комплекса по распознаванию патологий грудной клетки по рентгенографическим изображениям.

Алгоритмическое решение задачи

На практике распространены ситуации, когда множество X_D достаточно мало. В данной статье рассматривается задача, в которой $|X_D| = 12$, $Y = \{\text{туберкулез, пневмония}\}$. Рассмотрим решения практической задачи, когда в реальных условиях заданы три снимка с туберкулезом, три снимка с пневмонией, а на остальных снимках патологий не выявлено.

С учетом специфики конкретной прикладной задачи, предлагается подход, основанный на использовании нейросетевой технологии обработки данных [6]. Он состоит из следующих основных шагов:

1. Предварительная обработка снимка.
2. Распознавание наличия любой патологии на снимке x .
3. Автоматизированное уточнение диагноза для снимков, имеющих патологию.
4. Оценка степени уверенности алгоритма в поставленном предварительном диагнозе.
5. Ручной анализ неоднозначных ситуаций специалистом-рентгенологом.

В силу ограниченности объема статьи предлагается подробнее остановиться на этапах распознавания наличия патологии и уточнения диагноза. Предлагается использовать две различные нейронные сети сверточного типа стандартной архитектуры.

Наибольший интерес представляет подход к обучению этих нейронных сетей. Поскольку выборка X_D не позволяет обучить их эффективно (из-за малого объема), предлагается использовать общедоступные наборы рентгенографических изображений для обучения нейронных сетей. Статистические свойства общедоступных данных (средняя яркость и дисперсия пикселей) отличаются от целевых. Предлагается процедура обучения и архитектура нейронной сети, позволяющие решить эту проблему.

Процедура обучения нейронной сети представляет собой процесс, состоящий из двух основных этапов.

На первом этапе к общедоступному набору данных применяется преобразование $G(x)$. В результате получается новая обучающая выборка, характеристики которой соответствуют целевому набору данных.

На втором этапе, на полученном наборе данных выполняется обучение нейронной сети методом градиентного спуска.

На практике основная сложность заключается в выборе преобразования $G(x)$. На основании анализа данных и визуального сравнения изображений предлагается алгоритм построения $G(x)$.

Пусть $x \in X_0, y \in X_D$. Пусть преобразование G является композицией преобразований сдвига яркости $g_1(x, \delta) = x + \delta$ и изменения контраста $g_2(x, \gamma) = (x - \bar{x}) \cdot \gamma + \bar{x}$, где \bar{x} – средняя яркость изображения, γ, δ – параметры преобразований. Необходимо подобрать в некотором смысле оптимальные значения γ^* и δ^* параметров γ, δ , чтобы статистические характеристики обоих выборок были схожи.

В работе рассматриваются 2 статистики. Средней яркостью выборки X назовем величину

$$\bar{X} = \frac{1}{Nwh} \sum_{x \in X, i=1, j=1}^{w,h} x_{ij}. \quad \text{Дисперсией выборки } X \text{ будем называть величину}$$

$$DX = \frac{1}{Nwh} \sum_{x \in X} \sum_{i=1, j=1}^{w,h} (x_{ij} - \bar{X})^2. \quad \text{Подобрать параметры } \gamma, \delta \text{ необходимо так, чтобы}$$

$|\bar{X} - \bar{X}_D| \rightarrow \min$ и $|DX - DX_D| \rightarrow \min$. Поскольку задача многокритериальная, введем дополнительный параметр $\alpha, 0 \leq \alpha \leq 1$ и запишем целевую функцию в виде $\alpha |\bar{X} - \bar{X}_D| + (1 - \alpha) |DX - DX_D| \rightarrow \min$.

Предлагается алгоритм нахождения значений γ, δ , основанный на переборе допустимых значений аргументов по сетке. Пусть $\{(\gamma_i, \delta_i)\}$ – множество допустимых значений параметров, а гиперпараметр α зафиксирован. Пусть $G(X) = \{y \mid y = G(x), \forall x \in X\}$.

Шаг 0. Для множества X_D вычислить среднее \bar{X}_D с дисперсией DX_D .

Шаг 1. Для каждого набора (γ_i, δ_i) выполнить шаги 2–4.

Шаг 2. Построить выборку $G(X)$. Вычислить ее характеристики \bar{X}_G и DX_G .

Шаг 3. Вычислить значение функции $\alpha |\bar{X}_G - \bar{X}_D| + (1 - \alpha) |DX_G - DX_D|$.

Шаг 4. Если значение функции – наименьшее среди рассмотренных, то пометить пару (γ_i, δ_i) как наилучшую на текущий момент. Если все параметры рассмотрены, то закончить выполнение. Иначе перейти к шагу 1, используя следующую пару параметров $(\gamma_{i+1}, \delta_{i+1})$.

Основной недостаток алгоритма заключается в его трудоемкости. В ходе выполнения необходимо провести перебор γ, δ . Кроме того, из-за необходимости дискретизации пространства допустимых значений аргументов возникает возможность пропуска значений параметров, дающих хорошее качество преобразования.

Для преодоления обозначенных проблем использован подход, основанный на методе стохастического градиентного спуска по параметрам γ, δ . Для простоты вычисления производных функции потерь преобразуем ее к виду $L(\delta, \gamma) = \alpha (\bar{x} - \bar{y})^2 + (1 - \alpha) (Dx - Dy)^2$.

После упрощения функция потерь представляется в виде $L(\delta, \gamma) = \alpha \delta^2 + (1 - \alpha) (1 - \gamma^2)^2 (Dx)^2$.

Рассмотрим шаги алгоритма градиентного спуска.

Шаг 0. Для начальной инициализации положить $\gamma^0 = 0, \delta^0 = 0$.

Шаг 1. Зафиксировать объект $x \in X, t = t + 1$.

Шаг 2. Применить преобразование к изображению: $y = \gamma(x - \bar{x}) + \bar{x} + \delta$.

Шаг 3. Вычислить частные производные функции потерь по формулам $\frac{\partial L}{\partial \gamma} = 2\alpha\delta$,

$$\frac{\partial L}{\partial \gamma} = -4\gamma(1-\alpha)(1-\gamma^2)(Dx)^2.$$

Шаг 4. Выполнить шаг алгоритма градиентного спуска по формулам: $\gamma^{t+1} = \gamma^t - \beta \frac{\partial L}{\partial \gamma}$, $\delta^{t+1} = \delta^t - \beta \frac{\partial L}{\partial \delta}$, где β – шаг обучения.

Шаг 5. Если $|\gamma^{t+1} - \gamma^t| < \epsilon$ и $|\delta^{t+1} - \delta^t| < \epsilon$, то прекратить выполнение и принять γ^t и δ^t как искомые значения. Иначе перейти к Шагу 1.

Таким образом, предложен алгоритм построения преобразования G как композиции преобразований изменения яркости и контраста. Рассмотрим теперь изменения, которые необходимо внести в стандартную архитектуру нейронной сети.

Пусть x – объект обучающей выборки, $F: X \rightarrow R^n$ – сверточная нейронная сеть. Пусть $\bar{x} = G(x)$ – преобразованный снимок обучающей выборки. Применим нейронную сеть к обоим снимкам: $y = F(x), \bar{y} = F(\bar{x})$. Для достижения инвариантности к условиям съемки и параметрам медицинского оборудования, необходимо, чтобы $L_1 = \|\bar{y} - y\| \rightarrow \min$. Пусть L_2 – стандартная функция потерь при решении задачи классификации над вектором x . Тогда для обучения нейронной сети необходимо минимизировать функцию потерь $L = \alpha L_1 + (1-\alpha)L_2, \alpha \geq 0$.

Результаты экспериментов и обсуждение

Работоспособность предложенного подхода была проверена в ходе экспериментальных исследований. В качестве общедоступных наборов данных использованы NIH [7], RSNA [8], VinBigData [9], Shenzhen Tuberculosis Dataset [10]. Для проверки качества обученных моделей архитектуры DenseNet использовалось 12 снимков, полученных с помощью портативного медицинского оборудования.

Результаты тестирования бинарной модели на общедоступных данных представлены в табл. 1.

Табл. 1. Результаты тестирования бинарной модели

Pathology Label	F1 Score
Sick	0,7312
No finding	0,8392

Результаты тестирования уточняющей диагностической модели на общедоступных данных представлены в табл. 2.

Табл. 2. Результаты тестирования уточняющей диагностической модели

Pathology label	F1 Score
Tuberculosis	0,9652
Pneumonia	0,9712
Other	0,9808

Матрица ошибок модели для выборки X_0 представлена в табл. 3.

Табл. 3. Матрица ошибок для целевой выборки

Pathology label	No Finding	Tuberculosis	Pneumonia
No Finding	5	1	0
Tuberculosis	0	3	0
Pneumonia	1	0	2

Заключение

В работе предложен подход к распознаванию патологий легких, который представлен в виде последовательности шагов предварительной обработки снимка, распознавания наличия патологии и уточнения диагноза. Предложен подход к обучению нейронной сети (при малых объемах целевой выборки), основанный на подборе параметров композиции преобразований изображений.

На основе предложенных подходов и алгоритмов разработан соответствующий программный комплекс по распознаванию болезней легких. Результаты проведенных экспериментов свидетельствуют о работоспособности предложенного подхода.

CHEST PATHOLOGIES RECOGNITION USING PRE-TRAINING ON PUBLIC DATASETS

A.I. NAUMOVICH

Abstract. The problem of chest pathologies recognition in then conditions of insufficient size of the training dataset is considered. An approach to solving the problem based on neural network technology is proposed. The architecture and method of network training have been developed, as well as an algorithm for converting the dataset for training purposes.

Keywords: convolucional neural network, pattern recognition, X-Ray image processing.

Список литературы

1. Yao J [et. al.] Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. Academic Radiology, 2011.
2. Park S. [et. al.] // arXiv preprint arXiv, 2021.
3. CheXpert: A Large Chest X-Ray Dataset and Competition [Электронный ресурс]. URL: <https://stanfordmlgroup.github.io/competitions/chexpert/>.
4. Andrew Naumovich /Andrew Naumovich // Research Papers Collection "Open Semantic Technologies for Intelligent Systems ", Issue 5, 2021, P. 233–237.
5. Krasnoproshin V., Naumovich A. // Proceedings of the 15-th International Conference "Pattern Recognition and Information Processing". 2021. P 43–46.
6. Краснопрошин В.В., Головки В.А. Нейросетевые технологии обработки данных. Минск, 2017.
7. NIH Chest X-ray dataset [Электронный ресурс]. URL: <https://cloud.google.com/healthcare-api/docs/resources/public-datasets/nih-chest>.
8. RSNA Pneumonia Detection Challenge [Электронный ресурс]. URL: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>.
9. VinBigData Chest X-ray Abnomalities Detection [Электронный ресурс]. URL: <https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection>.
10. Tuberculosis Chest X-rays (Shenzhen) [Электронный ресурс]. URL: <https://www.kaggle.com/raddar/tuberculosis-chest-xrays-shenzhen>.

CREATING DATABASE IOT FOR PRODUCT QUALITY CONTROL

LIU ZHENGHUA, U.A. VISHNIAKOU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 5, 2022

Abstract. The structure of the IoT network for product quality control is presented. In the process of comparing database servers, the implementation of the SQL was selected to represent product quality indicators. The process of configuring the SQL server for use in such IoT network on a cloud platform is shown.

Keywords: product quality control, database, MySQL, cloud server.

Introduction

Product quality database programming, taking milk quality control as an example, mainly analyzes the advantages and disadvantages of several commonly used relational databases, such as MySQL, SQL Server, Oracle database [1], respectively, ease of use, flexibility, openness, security, and the use performance and other aspects are analyzed.

After the comparative analysis, the MySQL database is used to implement database programming. Next, it is introduced how to install the MySQL database on the server and how to use SQL statement to create a database, create data tables.

Comparative analysis of commonly used relational databases

1. MySQL.

The advantages of MySQL:

- small size, fast speed, low total cost of ownership, open source;
- support multiple operating systems;
- it is an open source database, and the provided interface supports connection operations in multiple languages.

The disadvantages of MySQL:

- does not support hot backup;
- the biggest disadvantage of MySQL is its security system, which is mainly complex rather than standard, and changes only when MySQL admin is called to reread user permissions;
- there is no Stored Procedure language, which is the biggest limitation for programmers accustomed to enterprise-level databases;
- the price of MySQL varies by platform and installation method. MySQL for Linux is free if installed by the user or a system administrator rather than a third party, and third-party options require a license fee. Unix or Linux self-installation is free; Unix or Linux third-party installation is charged.

2. SQL Server.

The advantages of SQL Server:

- ease of use, scalability suitable for distributed organizations, data warehouse functions for decision support, integration with many other server software, good price-performance ratio, etc.;
- brings flexibility to data management and analysis, allowing units to respond calmly in a rapidly changing environment to gain a competitive advantage.

The disadvantages of SQL Server:

- SQL Server can only run on Windows without any openness. Operating system stability is very important to databases. Windows9X series products are more focused on desktop applications. NT server is only suitable for small enterprises and Windows platform reliability, security and scalability are very limited. Unix-like is tried and tested especially for handling large databases;
- the parallel implementation and coexistence model of SQL server is immature, and it is difficult to deal with the increasing number of users and the limited scalability of data volumes;
- no security certificate obtained;
- performance: SQL Server performs well with multiple users;
- client support and application mode. Only C/S mode is supported, and the SQL Server C/S structure only supports ADO, DAO, OLEDB, and ODBC connections for Windows clients;
- SQL Server's complete rewrite of the code has undergone long-term testing, constant delays, and many features take time to prove not fully compatible with earlier products. Use requires certain risks.

3. Oracle.

The advantages of Oracle:

- Oracle can run on all mainstream platforms (including Windows), fully supports all industry standards, adopts a completely open strategy to enable customers to choose suitable solutions, and fully supports developers;
- Oracle's parallel server extends the capabilities of Windows NT by allowing a group of nodes to share work in the same cluster, providing a high-availability and high-scalability cluster solution. If Windows NT cannot meet the needs, users can move the database to UNIX. Oracle's parallel server has a high degree of integration for the clustering mechanisms of various UNIX platforms;
- Oracle is certified to the ISO standard at the highest level of certification. It provides multiple layers of security, including controls for assessing risk, preventing unauthorized data breaches, detecting, and reporting database activity, and implementing data access control in databases through data-driven security;
- Oracle is almost the highest performance relational database, maintaining the world record of TPC-D and TPC-C under the open platform;
- Oracle's multi-level network computing supports a variety of industrial standard network client connections such as ODBC, JDBC, and OCI;
- Oracle has a long development experience and is fully backward compatible. It has been widely recognized and applied, and the use risk is low.

The disadvantages of Oracle:

- high requirements for hardware;
- the price is relatively expensive;
- it is more troublesome to manage and maintain;
- the operation is more complicated and requires higher technical content.

Through the above analysis, MySQL is widely used in small and medium websites on the Internet. Because of its small size, high speed, low total cost of ownership, and especially open source, many companies use MySQL databases to reduce costs. The MySQL database can be regarded as one of the fastest running SQL language databases. In addition to having many features that other databases do not have, MySQL database is a completely free product, and users can download MySQL databases directly from the Internet without paying anything. In this paper, MySQL database is used for the management of milk quality database.

Database implementation

This project is aimed at the implementation of the Internet of Things product quality control database. The IoT architecture of this project is shown in Figure 1. This database is built on a cloud server. The implementation of the database can be performed through the command window of the cloud server through the creation of the database and the design of the corresponding table through SQL statements [2]. The process includes some steps.

Step 1. We configure the MySQL server on the cloud platform. Before connecting to the MySQL server remotely, we need to configure the MySQL server to listen on its internal IP address. Then, we create a non-root user account for the MySQL client to connect to the server [3].

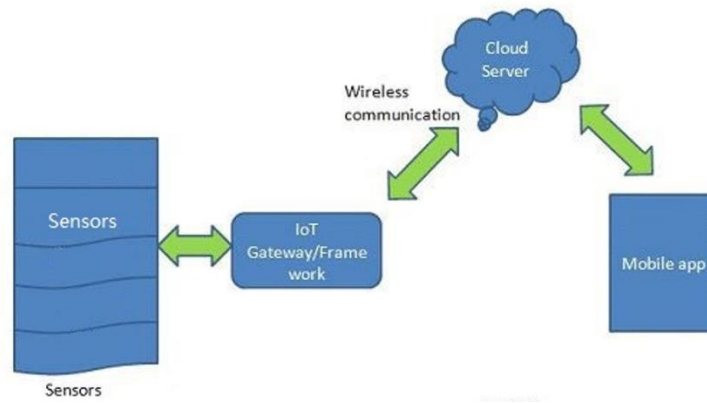


Figure 1. Product quality control IoT structure

Step 2. In Cloud Shell, we use SSH to connect to the my-server instance (see Figure 2).

```

TX packets 13388 bytes 1209435 (1.2 MB)
TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

hu1612240@my-server:~$ ping 34.125.192.234
PING 34.125.192.234 (34.125.192.234) 56(84) bytes of data.
64 bytes from 34.125.192.234: icmp_seq=1 ttl=64 time=0.931 ms
64 bytes from 34.125.192.234: icmp_seq=2 ttl=64 time=0.479 ms
64 bytes from 34.125.192.234: icmp_seq=3 ttl=64 time=0.479 ms
64 bytes from 34.125.192.234: icmp_seq=4 ttl=64 time=0.580 ms
64 bytes from 34.125.192.234: icmp_seq=5 ttl=64 time=0.411 ms
64 bytes from 34.125.192.234: icmp_seq=6 ttl=64 time=0.366 ms
64 bytes from 34.125.192.234: icmp_seq=7 ttl=64 time=0.351 ms
64 bytes from 34.125.192.234: icmp_seq=8 ttl=64 time=0.568 ms
64 bytes from 34.125.192.234: icmp_seq=9 ttl=64 time=0.434 ms
64 bytes from 34.125.192.234: icmp_seq=10 ttl=64 time=0.353 ms
64 bytes from 34.125.192.234: icmp_seq=11 ttl=64 time=0.387 ms
^C
[1]+  Stopped                  ping 34.125.192.234
hu1612240@my-server:~$ sudo -i
sudo: -: command not found
hu1612240@my-server:~$ sudo -i
root@my-server:~# mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 13
Server version: 5.7.37-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
  
```

Figure 2. Connect to database on server

Step 3. We update the /etc/mysql/mysql.conf.d/mysqld.cnf configuration file with the following information:

```

LOCAL_IP=$(curl http://metadata.google.internal/computeMetadata/v1/instance/network-interfaces/0/ip \
-H "Metadata-Flavor: Google") sudo sed -i "s|bind-address.*|bind-address = $LOCAL_IP|" /etc/mysql/mysql.conf.d/mysqld.cnf
  
```

Step 4. We restart the MySQL service to apply the changes to the running server:

```

sudo mysql --user=root -p[ROOT_PASSWORD] -e "show databases"
  
```

Step 5. We verify that the server is running locally. As shown in Figure 2 (connection to DataBase on Server) and Figure 3 (show databases) connect to the database and view the instance.

```

mysql> show dateabases;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server ve
rsion for the right syntax to use near 'dateabases' at line 1
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| milkquality |
| mysql |
| performance_schema |
| sys |
+-----+
5 rows in set (0.00 sec)
  
```

Figure 3. Show Database

Step 6. Create a database instance.

Step 7. Create product_information table.

Through the creation statement of the data table (see Figure 4) the milk quality indicates the database to which the table belongs, "milk_information" is the table name, the data type of MID (Milk

ID) is int, the data length is 255, NOT NULL is not empty, AUTO_INCREMENT is incremented by 1, the data type of CID (Cow ID) is int, "Fat", "Protein", "Lactose", "MUN" (Milk Urea Nitrogen), "SCC" (Somatic Cell Content), "BC" (Bacteria Content) data types are decimal, where the first value in parentheses indicates the total number of digits of the data, the second value indicates the number of digits after the decimal point, "GettingDate" is a date time type in the format (yyyy-MM-dd HH:mm:ss), PRIMARY KEY('MID'), indicating that "MID" is the primary key in the milk_information table [4].

```
mysql> use milkquality;
Database changed
mysql> show tables;
Empty set (0.00 sec)

mysql> CREATE TABLE `milkquality`.`milk_information` (
-> `MID` int(255) ZEROFILL NOT NULL AUTO_INCREMENT,
-> `CID` int(255) NOT NULL,
-> `Fat` decimal(3, 2) NULL,
-> `Protein` decimal(3, 2) NULL,
-> `Lactose` decimal(3, 2) NULL,
-> `MUN` decimal(3, 1) NULL,
-> `SCC` decimal(3, 1) NULL,
-> `BC` decimal(3, 1) NULL,
-> `GettingDate` datetime NULL,
-> PRIMARY KEY (`MID`)
-> );
Query OK, 0 rows affected (0.02 sec)
```

Figure 4. Create table milk_information

Conclusion

By comparing the advantages and disadvantages of MySQL, Oracle, and SQL server databases, we concluded that MySQL, due to its small size, fast speed, low total cost of ownership, especially open source, makes the using MySQL databases to reduce costs. In addition to having many features those other databases do not have, MySQL database is a completely free product, and users can download MySQL databases directly from the Internet without paying anything. So, we chose MySQL database as our target database.

We introduce how to create our MySQL database on the server through the development documentation provided by the Google Cloud Platform. We create a database instance and related data tables by connecting to the database and using SQL statements.

References

1. Zhang Yaqi // Tianjin Science and Technology Journal. 2019. Vol. 46, Iss. 9. P. 52–54.
2. Hao Ziyuan., Zhang Xu., Yang Wei. // Journal of Agricultural Machinery. 2017. Vol. 48, Iss. 1. P. 315–320.
3. Set up client access to MySQL on Compute Engine (via private IP address). Google Cloud [Electronic resource]. URL: https://cloud.google.com/architecture/mysql-remote-access#add_a_new_firewall_rule
4. Tang Hanming., Zhai Zhenxing., Guan Baojun. In-depth and easy-to-understand MySQL. M., 2019.

SENSOR DATA FREQUENCY FEATURE EXTRACTION

ZHAO DI

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received March 8, 2022*

Abstract. Frequency domain features are important for human activity recognition. However, the raw signal needs to be converted to frequency domain features using a fast Fourier transform. In the frequency domain, the time series data of each component is converted by using the Fast Fourier Transform (FFT) algorithm.

Keywords: Fast Fourier algorithm, Feature extraction.

Introduction

In a human activity recognition (HAR) system, to build a good performance recognition system, feature extraction is an important aspect. Some studies suggest that the use of efficient feature extraction methods may improve the final performance better than using an effective classifier [1]. Frequency domain analysis can observe the characteristics of the signal based on the frequency. Generally speaking, the frequency domain analysis is more concise, and observing the signal on the frequency domain makes the analysis of the problem more profound and convenient. The frequency domain feature extraction algorithm adopted in this article is a fast Fourier algorithm, which can transform the sample signal to the frequency domain and extract the frequency domain features of the sample signal to facilitate the frequency domain analysis.

The frequency domain characteristic parameters

Spectrum analysis includes amplitude spectrum and phase spectrum, and amplitude spectrum is the most commonly used. When the health state of the key components of the reducer changes, the frequency component in the signal spectrum of the sample data changes accordingly [2]. Therefore, we can first accurately characterize the signal spectrum information by analyzing the frequency domain characteristics of the signal, and then obtain the human behavior state under different situations. The frequency domain characteristic parameters used in this thesis include center of gravity frequency, mean frequency, root mean square frequency and frequency standard deviation.

1. Center of gravity frequency.

The center of gravity frequency (FC) can describe the frequency of large signal components in the spectrum, reflecting the distribution of the signal power spectrum. In other words, for a given band range, the energy contained below the center of gravity frequency is half the total energy of the signal. In other words, the center of gravity frequency is a weighted average of the amplitude of the power spectrum as the weights, so the center of gravity frequency moves to locations where the power spectrum amplitude is larger. The parameters defined by using the expression:

$$FC = \frac{\int_0^{+\infty} f \bar{p}(f) df}{\int_0^{+\infty} p(f) df}, \quad (1)$$

where f is frequency of the power spectrum $p(f)$ of the signal, $\bar{p}(f)$ is the power spectrum of the signal.

2. Mean square frequency.

The mean square frequency is the square of the RMS frequency. Unlike the center of gravity frequency, the mean square frequency is the weighted average of the square of the signal frequency, weighted by the amplitude of the power spectrum. The mean square frequency (MSF) is describing the change in the position of the main band of the power spectrum. The parameters defined by using the expression:

$$\text{MSF} = \frac{\int_0^{+\infty} f^2 p(f) df}{\int_0^{+\infty} p(f) df}, \quad (2)$$

where f is frequency of the power spectrum $p(f)$ of the signal, $p(f)$ is the power spectrum of the signal.

3. Center of gravity frequency.

The root mean square frequency is the calculated square root of the mean square frequency. It also describes the position changes of the main frequency band of the power spectrum. The parameters defined by using the expression:

$$\text{RMSF} = \sqrt{\text{MSF}}. \quad (3)$$

4. Frequency variance.

The frequency variance is the square of the frequency standard deviation, another metric measure of the energy dispersion of the power spectrum. Frequency standard deviations Frequency standard deviation is the radius of inertia centered on the center of gravity frequency. If the amplitude of the spectrum near the center of gravity is large, the frequency standard deviation is small; if the spectrum near the center of gravity is small. The frequency standard deviation describes the degree of dispersion of the power spectrum energy distribution. The parameters defined by using the expression:

$$\text{VF} = \frac{\int_0^{+\infty} (f - \text{FC})^2 p(f) df}{\int_0^{+\infty} p(f) df}, \quad (4)$$

where f is frequency of the power spectrum $p(f)$ of the signal, $p(f)$ is the power spectrum of the signal.

In order to quickly calculate the above parameters quickly the following formula based on the discrete Fourier transform properties is used:

$$F_{\text{FC}} = \frac{\sum_{n=1}^N p(f) p(f)}{2\pi \sum_{n=1}^N p(f)^2}, \quad (5)$$

$$F_{\text{FC}} = \frac{1}{N} \sum_{n=1}^N p(f), \quad (6)$$

$$F_{\text{RMSF}} = \sqrt{\frac{\sum_{n=1}^N p(f)^2}{4\pi^2 p(f)^2}}, \quad (7)$$

$$F_{\text{RVF}} = \sqrt{\frac{\sum_{n=1}^N (-F_{\text{FC}})^2 p(f)}{\sum_{n=1}^N p(f)}}, \quad (8)$$

where $\bar{p}(f) = \frac{p(f) - p(f-1)}{\Delta}$, Δ is the frequency resolution interval, and N is the number of frequency sample points.

Pros and cons of frequency feature extraction

1. Pros of frequency feature extraction:

- frequency domain analysis can observe the signal features according to the frequency, and the frequency domain representation is more concise;
- the physical significance of frequency domain analysis is clear and it can provide more intuitive feature information than the time domain.

2. Cons of frequency feature extraction:

- frequency feature extraction can only be conducted in the effective interval;
- because the time domain truncation causes energy leakage, the amplitude of the discrete spectrum, phase and frequency can produce large errors.

Program code implementation

In Matlab, firstly we need to fast Fourier transform the signal, and then use the four features of the center of gravity frequency, the mean square frequency, root mean square frequency and the Frequency variance to calculate the eigenvalue of the frequency domain. The program code implementation of the article in Figure 1.

```
1 function [ frequencystruct ] = frequencyDomainFeatures(src ,fs)% src is the raw data
2 %Compute frequency-domain statistics
3 %*****FFT transform the signal*****
4 FS=fs;
5 N=length(src);n=0:N-1;
6 freq=n*fs/N;
7 f=abs(fft(src,N)*2/N);
8
9 x=f(1:N/2); % Ordinate   Frequency amplitude
10 freq=freq(1:N/2)'; % Abscissa   Frequency value
11
12 % plot(freq,x);
13 % title('Original signal frequency domain waveform');
14 % xlabel('Frequency/hz');
15 % ylabel('Amplitude/v');
16
17 %*****Calculate frequency domain eigenvalues*****
18 frequencystruct.MF=mean(x); %average frequency
19 frequencystruct.FC=sum(freq.*x)/sum(x);%Center of gravity frequency
20 frequencystruct.RMSF=sqrt(sum([freq.^2].*x)/sum(x));%Root mean square (RMS) frequency
21 frequencystruct.RVF=sqrt(sum([(freq-frequencystruct.FC).^2].*x)/sum(x));%frequency standard deviation
22
23
24
25 end
```

Figure 1. Program code implementation

Conclusion

Frequency domain analysis can observe the signal characteristics according to the frequency. Generally, the time domain analysis is more intuitive, but the frequency domain analysis is more concise, and observing the signals in the frequency domain makes the analysis of the problem more profound and convenient. At present, frequency domain analysis has become a trend of signal analysis. Through feature extraction, we can extract the center of gravity frequency, mean frequency, root mean square frequency and frequency standard deviation in three directions of x, y, and z, a total of twelve feature vectors, we can carry out the next step of data analysis work through these twelve feature vectors, that will become convenient and concise.

References

1. Wu X. [et al.] // IEEE Trans. On KDE. 2014. Vol. 26. P. 97.
2. Manual J. [et al.] // IJCV. 2022. Vol. 130. P. 199–200.

УДК 621.391

ВОЗМОЖНОСТИ МЕТОДА ОТЖИГА В ЗАДАЧЕ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ

В.В. МАЦКЕВИЧ

Белорусский государственный университет, Республика Беларусь

Поступила в редакцию 22 февраля 2022

Аннотация. В работе рассматривается проблема обучения нейронных сетей. Предложен алгоритм обучения на основе метода отжига. Показано, что разработанный алгоритм обладает гораздо большей эффективностью, чем существующие градиентные методы.

Ключевые слова: метод отжига, метод градиентного спуска, обучение, нейронная сеть.

Введение

В настоящее время происходит стремительное развитие цифровых технологий. Объемы получаемой информации из различных источников стремительно растут [1]. Возникает множество проблем, связанных с большими объемами данных: передача, хранение и обработка. Для решения данных проблем были разработаны различные алгоритмы.

Из-за широкого разнообразия обрабатываемой информации наибольшее распространение получили универсальные нейросетевые алгоритмы обработки. В отличие от классической концепции, где класс обрабатываемых данных зафиксирован, нейросетевой подход предлагает адаптируемость к обрабатываемой информации [2]. Процесс настройки такого алгоритма под конкретный класс данных называется обучением. Эффективность полученного решения напрямую зависит от результата обучения [3].

Несмотря на достигнутые в данном направлении результаты, проблема обучения по-прежнему является актуальной.

Существуют два основных подхода к обучению нейронных сетей. Один базируется на методологии градиентного спуска, а другой – на случайном поиске. Наиболее популярными являются градиентные методы, которые на практике обучают нейронные сети за приемлемое время. Представители другого подхода – методы отжига, обеспечивают хорошее качество, но работают существенно медленнее.

В работе сравнивается эффективность метода отжига и градиентного спуска для обучения нейронной сети на примере решения задачи сжатия цветных изображений.

Анализ проблемы

Обучение нейронных сетей представляет из себя процесс настройки отдельных параметров сети для достижения наилучшего решения. Эффективность результата оценивается с помощью заданного для решаемой задачи функционала качества.

Таким образом, обучение нейронной сети является оптимизационной задачей, где решением являются конкретные значения параметров нейронной сети, а целевой функцией – функционал качества, описывающий решаемую задачу.

Для решения оптимизационных задач используются итерационные методы, которые, в свою очередь, подразделяются на два класса – направленные и ненаправленные.

При направленном поиске каждое последующее приближение вычисляется по строго определенному правилу на основе текущего приближенного решения. К направленным относятся все методы, основанные на вычислении градиентов: метод простого градиентного

спуска, метод секущих, метод Ньютона и т.д. Ненаправленные методы (методы случайного поиска) характеризуются случайностью выбора последующего приближения. К ним можно отнести метод дифференциальной эволюции, метод отжига и т.д.

Методы случайного поиска имеют практически неограниченное пространство для поиска решения, в то время как направленные методы ограничены строгими правилами выбора последующего решения, что существенно сужает пространство поиска решения. Этот факт дает интуитивное преимущество случайного поиска в качестве полученного решения над направленными методами. Таким образом, ненаправленные методы также имеют определенную перспективу.

Для сравнения этих двух принципиально разных подходов к решению оптимизационных задач рассмотрим два сильнейших представителя из каждого класса. Из ненаправленных методов – метод отжига, из направленных методов – метод адаптивного градиентного спуска.

Алгоритм обучения на основе метода отжига

Любая нейронная сеть представима в виде набора параметров – компонент. Например, многослойный перцептрон задается двумя наборами параметров: множеством весов W и набором смещений нейронов B . Такое представление сети необходимо для точного описания разработанного подхода к обучению. Алгоритм реализуется в два основных этапа.

Предварительный этап. На данном этапе производится инициализация (задание начальных значений) параметров сети и начальной температуры T_0 .

Основной этап. На данном этапе обучения реализуется процедура последовательного обновления значений параметров с использованием некоторого функционала качества.

Опишем более подробно процедуру обновления параметров.

Процедура обновления. Для простоты изложения рассмотрим ее на примере множества параметров W . Для других множеств эта процедура полностью совпадает.

Множеству параметров W поставим в соответствие отрезок L_w длины l . После этого каждый элемент множества W последовательно помещаем в центр заданного отрезка. Для определения направления изменения значений параметров генерируем случайное число от 0 до 1. Если полученная реализация числа больше 0,5, то значение параметра увеличивается, в противном случае – уменьшается.

Новое значение параметра определяется в результате реализации равномерно распределенной случайной величины на отрезке, концами которого являются текущее значение параметра, и конец отрезка, в сторону которого осуществляется изменение.

Аналогично действия последовательно выполняются для других параметров сети.

Для вновь полученных значений параметров вычисляется функционал качества.

Далее принимается решение о переходе в новое состояние:

– если значение функционала больше текущего, то переходим в новое состояние с вероятностью:

$$P(y|x) = \exp\left(-\frac{F(y) - F(x)}{T_i}\right), \quad (1)$$

где x – текущее состояние, y – выбранное для перехода состояние, F – минимизируемая целевая функция, T_i – температура i -ой итерации.

– в противном случае происходит безусловный переход в новое состояние.

– в случае смены состояния происходит охлаждение по правилу:

$$T_{k+1} = \frac{T_0}{\ln(k+1)}, \quad (2)$$

где k – количество совершенных итераций.

– в противном случае температура не изменяется.

После охлаждения производится проверка полученного решения на оптимальность:

решение является оптимальным, если в течение последних S итераций не проводился переход в новое состояние, либо время, отведенное на обучение, истекло.

Если полученное решение оптимально, то:

- останов алгоритма,
- в противном случае переход на следующую итерацию.

Конец алгоритма. Главным достоинством данного метода является гарантированная сходимость к точке глобального минимума, т.е. к оптимальному решению. К недостаткам можно отнести логарифмическую скорость сходимости метода.

Метод градиентного спуска

На каждой итерации градиентного спуска производится вычисление производной от целевой функции по каждому из настраиваемых параметров нейронной сети, с последующим их обновлением в противоположном градиенту направлении.

Данный подход имеет существенный недостаток: рассматриваемый метод, как правило, завершается в точках, где градиент равен нулю. Это приводит к получению решения равному локальному минимуму значения функционала качества. Как правило, значение локального минимума значительно больше значения глобального минимума.

Различные модификации метода градиентного спуска отличаются лишь правилом пересчета корректировок параметров на основе статистики изменений на предыдущих итерациях.

В проводимых нами экспериментах использовалась следующая модификация градиентного метода – метод адаптивного градиента [4]. В литературных источниках она считается наилучшей модификацией градиентного метода обучения, применяется для обучения глубоких нейронных сетей сложной архитектуры [4, 5].

Основная идея модификации заключается в «стабилизации» градиентов. Метод осуществляет вычисление корректировок с помощью специальных статистик корректировок из предыдущих итераций. Это обеспечивает снижение разности по модулю между корректировками соседних итераций. Данное свойство частично решает проблему локальных минимумов. А с другой стороны путем сглаживания предотвращаются резкие изменения параметров, что приводит к более «тонкой» настройке параметров.

К преимуществам данного подхода можно отнести линейную скорость сходимости.

Эксперименты

Для сравнения эффективности алгоритмов обучения нейронных сетей на основе методов градиентного спуска и отжига будем решать задачу сжатия изображений на выборках CIFAR-10 [6] и STL-10 [7].

Для всех степеней сжатия построена нейронная сеть из набора ограниченных машин Больцмана типа Гаусс-Бернулли.

При 32-кратном сжатии входной слой отдельной машины состоит из 48 нейронов, а выходной из 12. При этом каждая машина принимает в качестве входных данных фрагмент изображения размером 4 на 4 пикселя, но так у цветных изображений 3 цветовых канала то фрагмент изображения задается 48 значениями.

При 64-кратном сжатии входной слой отдельной машины состоит из 96 нейронов, а выходной из 12. При этом каждая машина принимает в качестве входных данных фрагмент изображения размером 4 на 8 пикселей, но так у цветных изображений 3 цветовых канала то фрагмент изображения задается 96 значениями.

При 128-кратном сжатии входной слой отдельной машины состоит из 192 нейронов, а выходной из 12. При этом каждая машина принимает в качестве входных данных фрагмент изображения размером 8 на 8 пикселей, но так у цветных изображений 3 цветовых канала то фрагмент изображения задается 192 значениями.

Следует отметить, что при более низких степенях сжатия классические алгоритмы сжатия, как правило, не уступают нейросетевому подходу, поэтому использовать более низкие степени сжатия в экспериментах для нейронных сетей нецелесообразно. С другой стороны, при более высоких степенях сжатия происходит практически полная потеря исходных изображений и сжатие в таком случае не имеет смысла.

Экспериментально было замечено, что меньшие размеры нейронных сетей приводят к более плохим результатам. Потому, что при высоких степенях сжатия (свыше 8 кратного) необходимо обобщать большие фрагменты входных данных, чтобы обеспечить качественное сжатие высокой степени. Архитектуры с большим количеством нейронов требуют гораздо большего времени на обучение, что усложняет проведение экспериментов, а также требуется большой объем входных данных для настройки большого количества параметров, что создает проблему нехватки данных для обучения.

Существует множество градиентных алгоритмов обучения ограниченных машин Больцмана: CD-k, PCD, N-PT_k. Они отличаются друг от друга способом оценки градиента функции правдоподобия.

PCD - быстрый и достаточно эффективный метод обучения. Однако, при большом количестве итераций обучения его эффективность снижается, и он уступает даже CD-1 [8]. В то же время алгоритм обучения CD-10 по эффективности примерно равен 10-PT5, однако требует гораздо меньшего объема вычислений [9]. При этом алгоритм CD-100 более эффективный, чем CD-10 [10].

Поэтому для сравнения эффективности обучения нейронных сетей в экспериментах использован описанный выше алгоритм на основе метода отжига и сравнительного расхождения (CD-100).

Для реализации процесса обучения использовалось 10000 изображений (3000 непосредственно для обучения и 7000 для валидации) для тестирования качества сжатия использовались остальные 50000 и 90000 изображений выборки соответственно.

Время, отводимое на обучение, определялось индивидуально для каждого алгоритма. Это было сделано для того, чтобы обеспечить сходимость алгоритма обучения и напрасно не расходовать вычислительные ресурсы.

Эксперименты проводились на системе Lubuntu 20.04 с процессором intel i7-4770k, видеокартой nvidia 1070 ti.

Как показали результаты экспериментов (см. табл. 1, 2) теоретическая сходимость метода отжига на практике обеспечила огромное превосходство в качестве сжатия по сравнению с градиентным методом (более чем в 2,7 раза).

Табл. 1. Результаты обучения методом градиентного спуска

Выборка	CIFAR-10			STL-10		
	0,75	0,375	0,1875	0,75	0,375	0,1875
Степень сжатия, бит/пиксель	0,75	0,375	0,1875	0,75	0,375	0,1875
MSE	1898	2385	3230	1728	2220	2949
PSNR	15,5	14,6	13,3	15,9	14,8	13,6
PSNR-HVS	15,6	14,7	13,4	16	14,9	13,8
SSIM	0,383	0,27	0,16	0,307	0,255	0,225
Время обучения, ч	0,142	0,08	0,064	1,27	0,714	0,554

Табл. 2. Результаты обучения методом отжига

Выборка	CIFAR-10			STL-10		
	0,75	0,375	0,1875	0,75	0,375	0,1875
Степень сжатия, бит/пиксель	0,75	0,375	0,1875	0,75	0,375	0,1875
MSE	649	855	1104	626	835	1061
PSNR	20,1	18,9	17,8	20,3	19	18
PSNR-HVS	20,3	19,1	18	20,5	19,2	18,1
SSIM	0,684	0,602	0,498	0,562	0,469	0,38
Время обучения, ч	1	1	1	8	8	8

В то же время теоретическая логарифмическая скорость сходимости отжига на практике оказалась существенно медленнее линейной градиента. Обучение методом отжига оказалось в среднем примерно в 12 раз медленнее градиентного метода. По количеству итераций

необходимых для сходимости алгоритма разница еще более заметная. Для сходимости отжига понадобилось порядка 10^6 итераций в то время как для градиентного метода порядка 10^3 .

Заключение

Экспериментально было показано, что метод градиентного спуска застревает в точках локального минимума и полученное решение существенно хуже оптимального. В тоже время разработанный, на основе метода отжига, алгоритм на практике показал высокое качество обучения, что подтверждает теоретические результаты о сходимости метода к оптимальному решению. Разработанный нами алгоритм был ограничен только временем, отведенным на обучение. Это означает, что при росте вычислительной мощности компьютера возрастает число совершенных итераций отжига за отведенное время и тем самым повышается качество обучения. В то время как у градиентного алгоритма число итераций с ростом мощности компьютера не увеличивается, а лишь уменьшается время обучения.

Из полученных результатов можно сделать вывод, что метод отжига имеет в перспективе определенный потенциал в решении задачи обучения нейронных сетей. Он может существенно повысить качество обучения нейронных сетей.

ANNEALING METHOD OPPORTUNITIES IN NEURAL NETWORKS TRAINING PROBLEM

V.V. MATSKEVICH

Abstract. The paper deals with a neural network training problem. Training algorithm based on annealing method is proposed. It is shown that constructed algorithm has significantly higher efficiency that existed gradient decay methods.

Keywords: annealing method, gradient decay method, training, neural network.

Список литературы

1. Краснопрошин В.В., Мацкевич В.В. // Вестник Брестского государственного технического университета. Серия - физика математика, информатика. 2018. № 5 (113). С. 15–18.
2. Zhengxue C., Heming S., Masaru T., Jiro K. // IEEE Transactions on multimedia 2019. Vol. 14. No. 8.
3. Krasnoproshin V.V., Matskevich V.V. // Research Papers Collection “Open Semantic Technologies for Intelligent Systems” 2019. Iss. 3. P. 265–268.
4. Kingma D.P., Ba J.L. // Proc. of the 3rd International Conference on Learning Representations. 2015. P. 1–15.
5. Hamis S., Zaharia T., Rousseau O. // IEEE 23rd International Symposium on Consumer Technologies (ISCT), P. 128–133.
6. Выборка CIFAR-10 [Электронный ресурс]: – Режим доступа: <https://www.cs.toronto.edu/~kriz/cifar.html> – Дата доступа: 04.03.2020.
7. Выборка STL-10 [Электронный ресурс]: – Режим доступа: <https://web.archive.org/web/20110803194852/http://www.stanford.edu/~acoates/stl10/> – Дата доступа: 24.04.2019.
8. Oswin K., Fischer A., Igel Ch. // Pattern Recognition Letters. 2018. Vol. 102. P. 1–7.
9. Brugge K., Fischer A., Igel Ch. // Machine Learning. 2013. Vol. 93. Iss. 1. P. 53–69.
10. Li X., Gao X., Wang Ch. // IEEE ACCESS. 2021. Vol. 9. P. 21939–21950.

OBJECT RECOGNITION BASED ON SKELETON

Y.P.Y. QIU, J. MA, Z.M. LIAO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 01, 2022

Abstract. In computer vision, representing graphics with important information can reduce the workload of recognition. As an important research content of computer graphics representation, skeleton features are simple in structure, so images can be matched effectively and quickly. Skeleton provides an effective and compact representation of an image object by reducing its dimensionality to a skeleton while preserving the topologic and geometric properties of the object, so using the skeleton to identify the shape has certain advantages. In this paper, we present the key processes of the object recognition based on the skeleton.

Keywords: skeleton, object recognition.

Introduction

With the development of modern technology, the image resources that can be obtained and utilized are increasing rapidly. Using the theory and method of image analysis to quickly recognize the object in the image has an urgent demand in various fields applications. Because the skeleton of object has the characteristics of hierarchy, multi-scale, consistent with the original target topology and adapting to large changes, people begin to pay attention to the object recognition based on skeleton [1].

Assuming that there is a grayscale image and a library containing different skeletons, object recognition based on skeleton is to extract the skeleton from the input image through a series of processes, and then compare the skeleton with the library to realize object recognition.

In this paper, we present the key processes of the object recognition based on skeleton. The whole process can be summarized as follows: image segmentation, skeletonization, skeleton pruning, model transformation and skeleton matching.

Image segmentation

The function of image segmentation is to separate the object from the background. The most commonly used method of gray image binarization is the threshold method. It uses the difference between the object and the background in the image to set the image to two different levels, and selects an appropriate threshold to determine whether a pixel is the object or the background, Mark all pixels in the sub image whose gray value is greater than or equal to the optimal threshold as the object image, and mark all pixels in the sub image whose gray value is less than the optimal threshold as the background image, so as to obtain the binarized image.

Skeletonization

The next step is to conduct skeletonization algorithm. The purpose of skeletonization is to reduce redundant information and leave enough useful information for topology analysis and shape analysis. The foundation of skeletonization was originally introduced by Blum [2] through an analogy with grassfires. Skeletonization methods can be divided into three major approaches, geometric, curve propagation and digital approaches [3]. The quality of skeleton is affected by many factors, including the noise of the image, the thickness of lines, the determination of endpoints and the selection of line

intersections. An excellent skeletonization algorithm should have the following characteristics: ensure the connectivity of the skeleton after skeletonization, maintain the detailed features of the original image well, the result of skeletonization is the center line of the original line, the endpoints of the line are intact, the intersection of the line can't be distorted and the skeletonization time can't be too long. An example of skeletonization is shown in Figure 1.



Figure 1. Skeletonization example of a letter of the alphabet

Skeleton pruning

Skeleton pruning is an essential part of the processing and analysis of skeletons. The purpose of this step is to filter out the unnecessary branches caused by boundary noise, which is the preprocessing before skeleton recognition. It is still quite a challenging problem because of the lack of standard measurements for the importance or significance of a branch. The relative significance of the same branches will be different if we see them from different perspectives with different objectives [4]. In the past decade, there are many approaches based on different perspectives have emerged for trying to tackle this problem. Pruning methods based on DCE [5] and pruning based with bending potential ratio [6] are recommended for use. A pruned example is present in Figure 2.

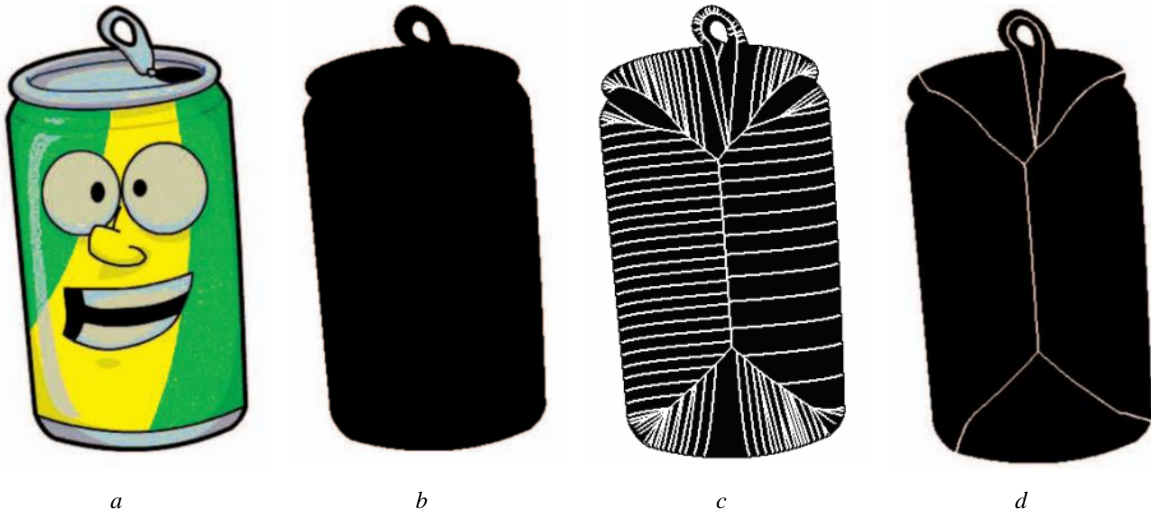


Figure 2. Example to illustrate the function of the skeleton pruning: *a* – the input object; *b* – binary object mask; *c* – the initial skeleton; *d* – pruned skeleton

Skeleton recognition

Skeleton recognition is the last step of object recognition. Comparing the processed skeleton with the skeleton in the library, it can retrieve the graphics corresponding to the graphics library by the matching degree of object skeleton. Therefore, the key step of shape retrieval is to perform effective bone matching [7]. At present, there are several common skeleton matching methods, such as skeleton graph matching based on path similarity, skeleton matching based on clustering and affine invariant of image feature matching algorithm.

Conclusion

This paper briefly introduces each key step of object recognition based on skeleton. Skeleton provides an effective and compact representation of an image object by reducing its dimensionality to a

skeleton while preserving the topologic and geometric properties of the object. Therefore, the skeleton based object recognition algorithm has important application significance in image processing.

References

1. Chen X. Research on target representation and recognition technology based on skeleton. University of Defense Science and technology, 2004.
2. Blum H. // Biological prototypes and synthetic systems. 1962. P. 244–260.
3. Saha P.K., Borgefors G., Baja G.S. // Pattern recognition letters. 2016. P. 3–12.
4. Liu H., Wu Z., Zhang X., Hsu D. // Pattern Recognition Letters. 2013. P. 1138–1145.
5. Bai X., Latecki L.J., Liu W.Y. // IEEE transactions on pattern analysis and machine intelligence. 2007. P. 449–462.
6. Shen W., Bai X., Hu R., Wang H., Latecki L.J. // Pattern Recognition. 2011. P. 196–209.
7. Gao B. Skeleton matching based on clustering. Dalian University of technology, 2012.

UDC 004.391.

HUMAN ACTIVITY RECOGNITION BASED ON RANDOM FOREST

Y. XUEYING

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received February 26, 2022*

Abstract. With the improvement of pattern recognition algorithm, human activity recognition (HAR) based on smart phone sensor data has become a highly concerned and rapidly developing research field. According to the basic principle of HAR, this paper proposes an activity recognition system based on random forest classifier. Data collection, data preprocessing and feature extraction are specified. The random forest classifier is summarized. The accuracy rate, TPR, FNR, PPV and FNR were used to evaluate HAR system proposed in this paper. Experimental results show that this study can accurately distinguish six basic activities, and the average accuracy of the random forest algorithm is 94,6 %.

Keywords: human activity recognition, classifier, random forest algorithm.

Introduction

Human Activity Recognition (HAR) has become a very active research topic in the field of machine learning [1-2]. The purpose of this work is to use sensor data and machine learning methods to detect human movement. Advanced mobile devices, such as smart phones, are usually integrated by multiple sensors. Accelerometers, gyroscopes, and magnetometers are commonly embedded. Mobile devices can obtain a large amount of user-related information by monitoring and tracking the user's movement. Classic pattern recognition steps include data preprocessing, feature extraction/selection, classifier training, and sample recognition. The research is mainly as follows:

1 Bagging algorithm [3] was proposed in 1996. The new training dataset generated by random sampling contains the same number of data instances as the original training dataset D_i , and the data is repeated due to uneven selection. In literature [4], the D_i ($i=1,2,\dots,T$) the model building method of decision tree CART [5] algorithm is used for improvement. There are T decision trees in the decision forest to provide information for their decisions.

2 AdaBoost algorithm is a classical algorithm among Boost algorithms [6]. In this kind of algorithm, the base classifier is derived from the training data set D with weight factors added.

3 Random Subspace algorithm [7] re-selects data attribute subset and randomly selects attribute subset f from all feature space m in the process of each iteration to form m . The decision tree is trained from the new training data set. The accuracy of Random Subspace algorithm is between Bagging algorithm and AdaBoost algorithm, but the diversity is not as high as the above two algorithms.

This paper puts forward a kind of human activity recognition based on random forest algorithm, the target is to adopt mobile phone embedded sensor data through human activity recognition technology, achieved the identification of common 6 class activities. The experimental results show that the human activity recognition based on random forest algorithm has very high accuracy.

The processing procedures of accelerometer data

According to the classical pattern recognition process, we introduce in detail the data processing process as the pre-part of activity recognition method, including: data acquisition, data pretreatment, data segmentation, feature extraction and selection.

The user activity data is collected to analyze and verify the activity identification method proposed in this paper. The quality of data will directly affect the final result of activity identification. The difficulty of data collection can be greatly reduced by using sensors embedded in portable intelligent devices to collect data. Simple activities are characterized by high monotonicity and repeatability and obvious feature distribution.

Therefore, based on the experience of previous studies and the basic characteristics of simple human activities, we selected the sensors built into smart phones to obtain the original signals, Acceleration sensor, Gravity sensor, Gyroscope, Magnetic sensor

During the data acquisition process, due to interference from factors such as external environment and human error, the original sensor data usually contains noise (missing values, incorrect values, or abnormal values, etc.). Therefore, the original data needs to be preprocessed. The commonly used data preprocessing methods are mainly data filtering and windowing segmentation.

In data processing, the function of a filter is to remove unwanted parts of the signal, such as random noise, or to extract useful parts of the signal, such as the components lying within a certain frequency range. In this paper, Chebyshev Type II sensor is used to remove the influence of gravity. Chebyshev Type II filters have flat passbands (no ripple), making them a good choice for DC and low frequency measurement applications.

The Chebyshev high-pass filter shown in Figure 1 is used to separate the components of bulk acceleration (BA) and gravitational acceleration (GA).

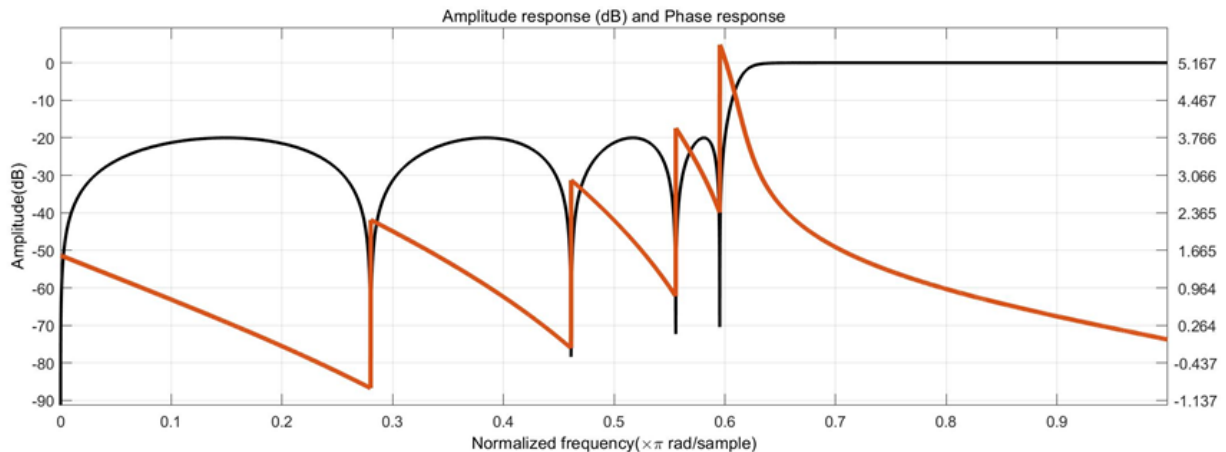


Figure 1. The amplitude and phase response of 9th-order high-pass Chebyshev Type II filter

Due to the data from the sensor is time domain data points in a row, high dimension, not suitable for directly used for feature extraction, otherwise it will affect the accuracy of the system and efficiency. In this paper, a 50 % overlap slider Window is selected to windowize the data.

Feature extracting. Extracting key features to represent human activities is the key to activity analysis and behavior analysis. Feature selection has a very important impact on the recognition effect. Features with significant discriminative power can greatly improve recognition accuracy or reduce model complexity. Active features usually include time domain features, frequency domain features and some other domain features. In this paper, we extracted 66 temporal features from the ACC data of 6 sports activities, as shown in Figure 2.

It can be seen from Figure 2, *a* that walking and lying down can be easily distinguished by using the feature vector consisting of histogram values of ACC signal. It can be seen that walking and standing can be easily distinguished by using the standard deviation, as shown in Figure 2, *b* shown.

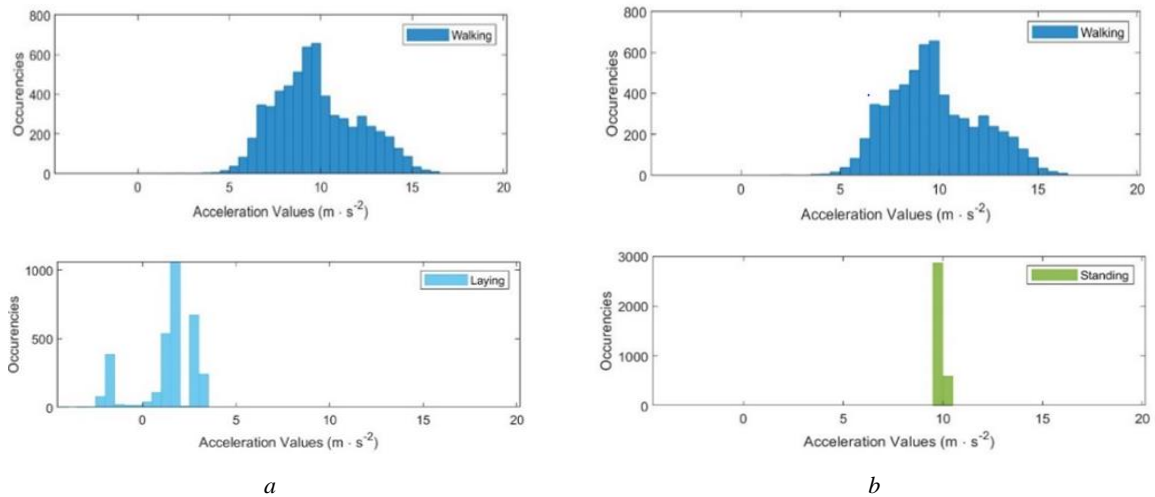


Figure 2. Histograms of ACC data for 3 classes of human physical activities: *a* – Walking and Lying; *b* – Walking and Standing

Random forest (RF) is a supervised machine learning algorithm that has become one of the most commonly used algorithms due to its accuracy, simplicity and flexibility. It has nonlinear properties that enable it to be used for classification and regression tasks and is highly adaptable to a range of data and situations. RF use Bagging (picking one observation sample instead of all) and random subspace methods (picking one feature sample instead of all features, in other words – attribute bagging) to grow a tree. RF has simplicity, diversity, robustness and reliability, and can greatly improve model accuracy by adjusting hyperparameters and selecting key features.

RF learning algorithm is composed of three parts like any other machine learning algorithm. Training, validation and testing. RF learning provides us with several Spaces (optimal combination of features) in which we can perform optimal domain partitioning and achieve high classification accuracy.

The base classifier of RF is decision tree, which belongs to a single classifier. A decision tree can provide a set of rules to distinguish between different characteristics of data. The basic idea of decision tree algorithm is to use the training set of known categories to train classifiers and generate rules. These rules are used to classify and mine unknown data sets.

The decision tree model is a tree structure, including three types of nodes (Figure 3).

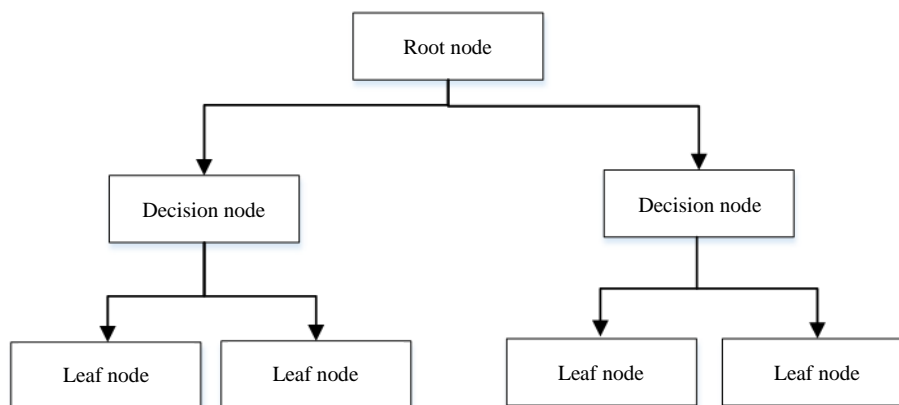


Figure 3. Decision tree structure used in RF classifier

The nodes include root decision node, intermediate decision node and leaf node. Among them, the decision child node represents the determination of feature attributes, and the leaf node represents the determination of feature attributes. The child node represents a category, and the path of the connected node represents the output of the characteristic attribute within a certain range of values. The decision-making process of decision tree is the process of testing the characteristic attributes of the data to be classified, selecting the output branches, and arriving from the root node to the leaf node.

The most critical step in constructing decision tree is to determine split attribute. Split attribute refers to dividing and constructing different branches according to a certain characteristic attribute at

non-leaf nodes. The key lies in attribute selection metric, which is a criterion for selecting split attribute. There are many node splitting algorithms, which generally adopt the top-down recursive divide and conquer method and the greedy strategy without backtracking.

This paper adopts CART algorithm, which adopts the splitting rule *Gini* index minimum principle when nodes are split. *Gini* coefficient calculation formula is as follows:

$$Gini(P) = 1 - \sum_{i=1}^C (p_i)^2, \quad (2)$$

where $P = (p_1, p_2, \dots, p_n)$; p_i is the probability of an object being classified to a particular class; C is the number of physical activities and i is an index of the class.

The *Gini* index is applied to categorize the target variables as "success" or "failure". Since it only performs binary splitting, a higher *Gini* index means greater inequality and heterogeneity.

The construction process of random forest mainly includes three steps. First, sample the data set and generate a training set for each decision tree; then use each training set to build a decision tree. The process of generating a decision tree does not require pruning processing; and finally generate a random forest and perform a classification algorithm.

The growth of each decision tree is as follows:

1. Random record selection: each tree training accounts for roughly two-thirds of the total training data was (63,2 %). Cases were randomly selected and replaced from the original data. The sample will be used as a training set for growing trees.

2. Random variable selection: randomly selected from all the predictor variable m , for example, some predict variables, and use these m the best segmentation to segment nodes.

The combination of multiple decision trees established through the above two steps will become a random forest. The decision tree uses voting to participate in decision-making, and the category with the most votes will be the final output of the random forest algorithm. The random forest pseudocode used for classification application.

Activity recognition performance evaluation

The random forest algorithm is used to verify the existing model using a verified subset of the research data set. The 10-fold cross-verification method is used for the training subsets to verify the cross-verification model. The experiment used different random seeds for multiple experiments. Use the obfuscation matrix to verify the performance of the model. The results are shown in Table 1.

Table 1. Confusion matrix of recognized activities

Human Activity Recognition	Walking	Walking upstairs	Walking downstairs	Sitting	Standing	Laying
Walking	1668	15	39	0	0	0
Walking upstairs	21	1502	21	0	0	0
Walking downstairs	44	54	1308	0	0	0
Sitting	0	0	0	1624	152	1
Standing	0	0	0	206	1699	1
Laying	0	0	0	0	0	1944

The confusion matrix represents the dispositions of the set of instances, on which a classifier is tested. The example is given in Table 2.

Table 2. The estimations of activity recognition performance

Human Activity Recognition	TPR (%)	FNR (%)	PPV (%)	FDR (%)	Precision (%)
Walking	96,9	3,1	96,2	3,8	96,9
Walking upstairs	97,3	2,7	95,6	4,4	97,3
Walking downstairs	93,0	7,0	95,6	4,4	93,0
Sitting	91,4	8,6	88,7	11,3	91,4
Standing	89,1	10,9	91,8	8,2	89,1
Laying	100	0	99,9	0,1	100

The highest values in the result percentage of the confusion matrix indicate a higher success in recognition performance i.e. The error percentage is minimal. It can be seen from the confusion matrix that the recognition of human activities based on random forests can accurately identify basic human activities: walking, walking upstairs, walking downstairs, Sitting, Standing, and Laying.

TPR (true positive rate), FNR (false negative rate), PPV (positive predictive value) and FDR (false discovery rate) were used to demonstrate the identification performance of using random forest, the experimental results show that this study can accurately distinguish the six basic activities, and clearly shows that the random forest algorithm in the activity recognition experiment of higher accuracy, the average accuracy is 94,6 %.

Conclusion

With the development of wearable devices, human activity recognition through sensor data has become an important research field. All kinds of classifiers in machine learning show excellent performance in their own fields. This paper reviews and summarizes human activity recognition based on random forest classifier. Firstly, the set of embedded sensors for data acquisition and their functions, Chebyshev Type II high-pass filter for data processing and feature extraction using time domain features are introduced. Secondly, the random forest classifier, its base classifier, and the decision tree node splitting algorithm are described and analyzed. Finally, with TPR, FNR, PPV, FDR and accuracy as evaluation indexes, the human activity recognition system is evaluated through experiments. Research shows that the recognition of human activities based on random forest has high average accuracy (94,6 %) for 6 classes of physical activities using time features.

References

1. Thakur D., Biswas S. // Smartphone based human activity monitoring and recognition using ML and DL: a comprehensive survey. 2020. P. 433–444.
2. Beddiar D. R., Nini B. // Vision-based human activity recognition: a survey. 2020. P. 509–555.
3. Breiman L. // Machine Learning. 1982. Vol. 1. P. 123–140.
4. Han Jiawei. Data mining: concepts and techniques. Morgan Kaufmann Publishers Inc, 2005.
5. Loh W. Y. // Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery. 2011. Vol. 1. P. 14–23.
6. Freund Y., Schapire R. E. // Experiments with a new boosting algorithm. 2017. Vol. 1. P. 12–15.
7. Fawagrehk G. // Systems Science and Control Engineering. 2014. P. 602–609.

DIFFERENT APPLICATIONS AND IMPLEMENTS OF THE HAMMING PRODUCT CODES

Y.M. CHEN, X.H. REN

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 28, 2022

Abstract. Hamming code are a family of linear error-correcting codes, which can detect one-bit and two-bit errors, or correct one-bit errors without detection of uncorrected errors. Hamming product codes are an improved variety of Hamming codes. This paper briefly describes Hamming product codes and compares the different implementations.

Keywords: Hamming codes, product codes, Multi bit error, Extended hamming product code.

Introduction

Hamming Product Code, or Product Code in brief, is a serially concatenated linear block code with two or more component codes. It is a "long" code made by a simple combination of "short" codes, and it can be decoded by a simple component-wise decoding [1].

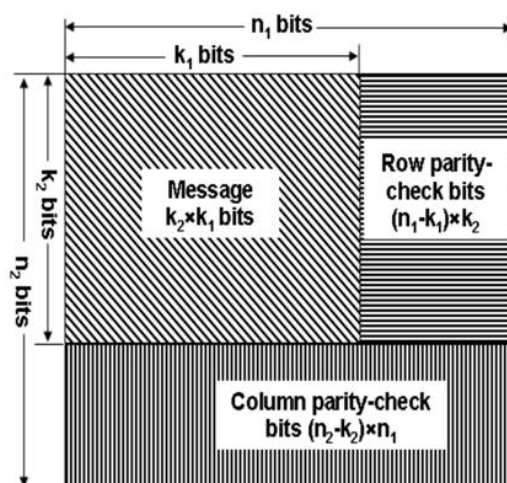


Figure 1. Construction of two-dimensional product codes [2]

As shown in Figure 1, a block of product codes contains a data matrix of size $k_1 \times k_2$, row parity bits of size $(n_1 - k_1) \times k_2$, and column parity bits of size $(n_2 - k_2) \times n_1$. The row parity bit consists of the parity value of the data bits in the row, and the column parity bit consists of the parity value of the column (either the data bits or the row parity bit). If the rows and columns and their corresponding parity checks have minimum Hamming distances d_1 and d_2 , respectively, then in theory, the minimum Hamming distance of this two-dimensional product code is $d_1 \times d_2$.

A product codes constructed in this way can achieve a better effect, also in the meantime, can have an acceptable performance overhead. Years of research have led to a variety of implementations, including hardware and software.

Among the many implementations, the hardware-based implementations appear to be more useful and attractive. Because the hardware-based implementation has a stable and superior performance.

These hardware-based implementations go beyond simply implementing the algorithm and go on to incorporate other techniques to improve overall performance.

Applications

The simplicity of Hamming code makes it suitable for applications with a higher sensitivity to performance overhead and latency. The ability to correct multiple random and burst errors makes the product code a good choice for a lightweight on-chip interconnect error correction solution.

Mixing multiple technologies can lead to more gains than just improved algorithms. The product code is often used in conjunction with other error control techniques to improve its overall performance and reliability. Common technology portfolios include automatic repeat requests (ARQ), forward error correction (FEC), and hybrid ARQ (HARQ).

At the same time, the product code can also be applied to scenarios requiring high-speed coding and decoding, such as flash memory error correction. The experiments show that the coding efficiency is 11 % and the actual coding efficiency is 12,6 % when the effective data two-dimensional structure in flash memory system is 1024×128 with two-dimensional Hamming code structure. The encoding speed and decoding speed reached 102 Gbps, which is suitable for high-speed storage system [3].

Codec Design

Common application scenarios where encoders and decoders use different designs. In addition, the codec also integrates other error correction mechanism. Most designs will use semaphores to communicate between the encoder and decoder. Due to the design of the product code, the coding and decoding process includes repetitive operations, so it is also a common design to add cache and logic control modules while reusing a single codec unit.

A design with multiplexed row or column codecs applying types-II of HARQ is illustrated in Figure 2. In this design, rows and columns are processed separately and the results are finally merged by a MUX and sent. The design also builds the communication structure between the encoder and the decoder. With this communication structure, the decoder can control the encoder to perform actions in case of unrecoverable errors. Because the communication between the codecs of this design includes at least two signal quantities (ACK/NACK), the design uses a triple redundant communication line design in order to avoid communication failures between the encoder and decoder.

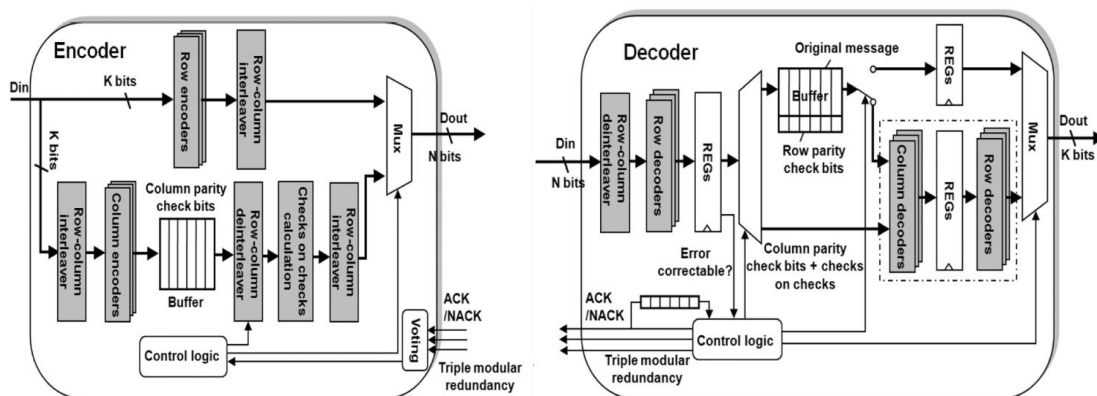


Figure 1. Encoder and decoder by Bo Fu [4]

Conclusion

The Hamming product codes are based on the Hamming code and uses the idea of cross-checking to improve the existing algorithm, which theoretically has a superior performance and effect than the existing algorithm. On-chip interconnect error correction mechanisms using such algorithms can effectively reduce link power consumption and improve stability. This algorithm is also effective in improving compiler efficiency when applied to flash compilers.

References

1. Nam Y., Young K. and Pil J. // Proceedings ISCC 2000 Fifth IEEE Symposium on Computers and Communications. 2000. P. 732–737.
2. Bo F., Ampadu P. // 2008 IEEE International SOC Conference. 2008. P. 59–62.
3. 余辉龙,张健,李清,覃翠,赵静,花涛. // 微电子学与计算机. 2018. P. 1–3.
4. Bo F., Ampadu P. // In IEEE Transactions on Circuits and Systems I: Regular Papers. 2009. Vol. 56. P. 2042–2054.
5. Chlaab A., Flayyih W., Rokhani, F. // Bulletin of Electrical Engineering and Informatics. 2020.

MULTI-CLASS CLASSIFICATION SVM METHODS

ZHAO DI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 9, 2022

Abstract. Support Vector Machine (SVM) was first proposed by Cortes and Vapnik in 1995. It is developed from the linear separable classification problem. The basic principle of SVM is to use a hyperplane to divide data into two categories, but in many classification problems, the sample is not linear, and the classification problem is multi-classification. We need to solve the multi-classification problem on a dichotomy basis.

Keywords: Support Vector Machine (SVM), multi-classification.

Introduction

Multi-class classification SVM has been applied in many fields, and there are two main categories of SVM ideas. One is to calculate all the classification decision functions simultaneously to solve the multi-classification problem at a time. However, the optimization problem solving process of this method is very complex, very computational and difficult to achieve, so it has not been widely used. The other is the decomposition of multi-classification problems into multiple binary classification problems, which are also divided into three categories: one-vs-one (OVO), one-vs-rest (OVR) and so on [1].

OVO multi-class classification SVM algorithm

The OVO multi-class classification SVM algorithm designs an SVM between any two classes of samples. When the unknown sample is classified, and classified by the voting method, the category with the most votes is the category of the unknown sample. Suppose samples have k categories, then we need to design binary SVM. When each sample was tested, each classifier judged it and output the category, and recorded the number of times judged as each category, and the category with the largest number of discriminants is the final category of the sample, the flowchart of this classification method shown as Figure 1.

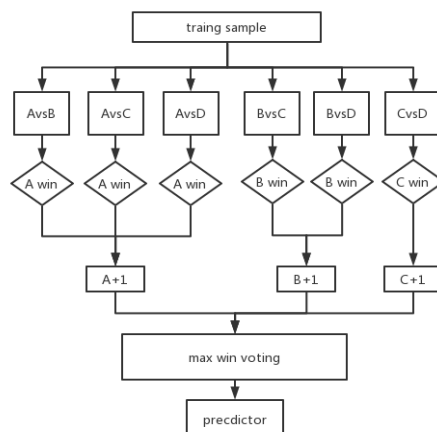


Figure 1. The flowchart of One-vs-One (OVO) multi-class classification SVM method

Pros of One-vs-One (OVO) multi-class classification SVM method:

- relative speed is faster when training individual models;
- when new samples are added without retraining all SVM, only retraining and increasing the sample-related classifier.

Cons of One-vs-One (OVO) multi-class classification SVM method:

- the number of binary classifiers required to construct and test regarding k into quadratic functions grows, and the total training time and test time are relatively slow.

Directed Acyclic Graph(DAG) multi-class classification SVM algorithm

From the OVO method, the Directed Acyclic Graph multi-class classification appears [2]. The training process is similar as OVO method, but the specific test process is shown in Figure 2.

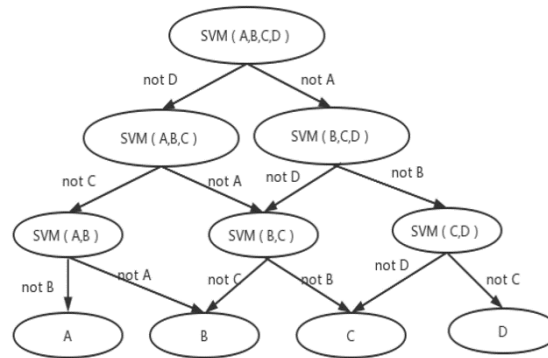


Figure 2. The flowchart of DAG multi-class classification SVM method

Pros of DAG multi-class classification SVM method:

- without having to traverse through all the dichotomized SVM, the decision speed is faster than the OVO voting method;
- it optimizes the training and decision time, is easier to calculate, learning effect is better;
- there is no inseparable region, and the sample number asymmetry problem is solved.

Cons of DAG multi-class classification SVM method:

- this method needs to construct more dichotomy SVM and slower training for classification problems with more samples.

OVR multi-class classification SVM algorithm

OVR multi-class classification is a strategy to compare the original and simple combined multi-classified SVM. Assuming a k classification problem, when training the classifier, the sample space is divided into categories and non, and the sample space is constructed into k SVM classifiers. In the test stage, the test samples were input into k SVM classifiers to compare the classification results of each classifier, and the classifier corresponding to the maximum classification function value was selected to determine the category of the sample, the flowchart of this classification method shown as Figure 3.

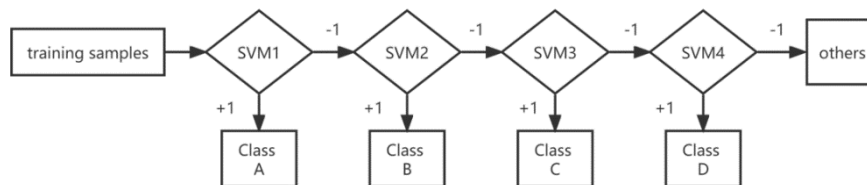


Figure 3. The flowchart of One-vs-Rest (OVR) multi-class classification SVM method

Pros and cons of OVR multi-class classification SVM method:

- this method just need k classifiers and its classification speed is relatively fast.

Cons of OVR multi-class classification SVM method:

- the training of each classifier takes all the samples as the training samples, so that when solving the quadratic planning problem, the training speed will slow down sharply as the number of training samples increases;
- because the data of negative sample is much greater than that of positive sample, sample asymmetry will appear, and this situation tends to be serious with the increase of training data;
- retraining of all models is required when new samples are added.

Decision Tree (DT) multi-class classification SVM algorithm

Following the OVR method, we derived decision tree multi-class classification SVM algorithm [2]. This method first divides all categories into two categories and further divides the subclasses into two secondary subclasses so that all nodes contain only a single category, which are also leaves of the binary tree species. This classification also decomposes the original classification problem into a series of two-class classification problems, where the classification function between the two subclasses adopts SVM. As shown in Figure 4.

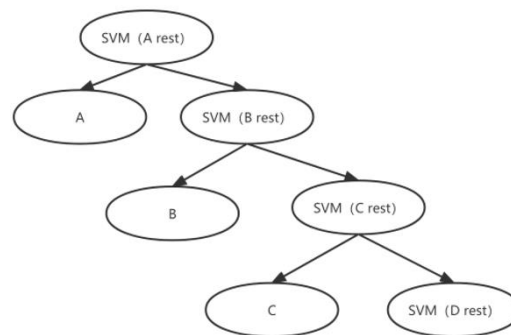


Figure 4. The flowchart of DT multi-class classification SVM method

Pros and cons of DT multi-class classification SVM method:

- the decision tree is easy to understand and interpret, visually analyze and extract rules;
- both nominal and numerical data;
- more suitable for samples with missing attributes;
- it is able to handle irrelevant features;
- faster operation for testing data sets;
- feasible and effective results for large data sources in a relatively short time.

Cons of OVR multi-class classification SVM method:

- easy to overfitting;
- easy to ignore the correlation of attributes in the dataset;
- different judgment criteria will bring different attribute selection tendencies in the decision tree.

Retraining of all models is required when new samples are added.

Conclusion

For multi-classification problems, we can combine multiple binary classifiers into multi-classifiers to handle the multi-classification problem. The basic methods include OVO method and OVR method, which have different advantages. In our application, we should choose our own methods according to the characteristics of the data.

References

1. Fleury A., Vacher M., Nury N. // IEEE Trans. On ITB. 2010. Vol. 14. P. 274–283.
2. Sun Z., Fox G. // Proc. of PDPTA. 2012. P. 16–19.

UDC 004.71

INFORMATIONAL LINEAR CAPACITY OF FIBER-OPTIC BACKBONES OF THE COUNTRIES OF THE WEST COAST OF AFRICA

D.V. MARCHENKO, MBUYI MICHELLE MULUMBA,
KURUMA MOHAMMED, KOUA LEONEL IGNAS

Educational institution "Francisk Skorina Gomel State University", Republic of Belarus

Received February 18, 2022

Abstract. This article describes an assessment of the ways in which the international digital data backbone is serviced, which is determining the development of the economy. The focus of the article is on the west coast of Africa and direct interregional data delivery channels. The concept of informational linear capacity was introduced and substantiated.

Keywords: optical fiber, digital cable, international Internet backbone, Tbit/sec, linear information capacity.

Introduction

The organization of information flows and access to them for the modern economy is as relevant as access to waterways and energy resources in the 19th and 20th centuries, respectively. In modern conditions, fiber-optic backbones are used as large elements of the infrastructure for organizing and supporting information flows. Laying fiber-optic lines on land is costly, but a fairly straightforward task. The advantage of such a solution is the ability to organize data logistics and start a return on investment in stages until the completion of the entire project, since the communication and intermediate routing strongholds through which the fiber-optic backbone passes provide access to the user audience that is geographically located around it.

The essential complexity of this solution lies in the need to lay the bandwidth of the backbone, taking into account the potential rapid growth of information growth in the load on the network both at each of the endpoints of the backbone and at its intermediate nodes.

Relevance of laying fiber-optic lines along the sea and ocean bottom

It can be argued that laying fiber-optic backbones over land solves local regional problems and cannot always provide targeted traffic to regions that are geographically distant from routing centers. An additional limitation on the choice of such a solution may be unfavorable geological conditions, environmental restrictions, natural disasters and the risk of military conflicts, which can damage the infrastructure of the backbone networks and complicate its operation or maintenance.

Subsea fiber-optic trunklines have similar risks: extraterritorial deployment, fixed cable structure configuration (from laying to decommissioning), potential damage from vessel traffic (commercial fishing, use of anchors), potential intentional damage (Figure 1) [1] and even harm from the local fauna (for example, aggressive behavior of sharks).

Based on these risks, all types of fiber-optic trunk cables must be protected, and the degree of protection depends on the topography of the ocean or seabed and the depth of their placement. The decision to place fiber-optic lines at the bottom in the coastal zone in some cases turns out to be the only alternative and the most economical way.

For example, in 1989, a Scotland-Northern Ireland 1 cable connection between Donaghadee and Portpatrick, UK, across the North Strait Between the Atlantic Ocean and the Irish Sea, was commissioned. The length of the highway is about 35 km. The data transfer rate at the time of

commissioning was 3,36 Gbps (0,00336 Tbps). It was supposed to complete the operation of this highway by 2014, but its use continues in 2021.

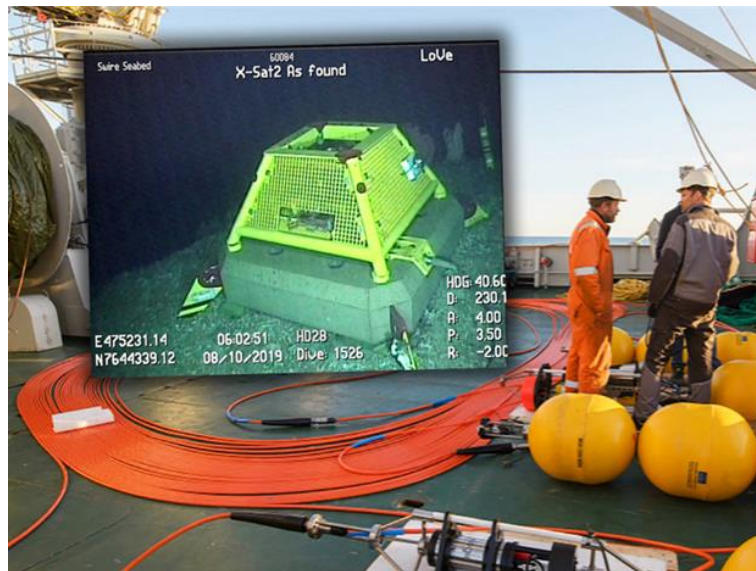


Figure 1. Precedent for damage to submarine cables at the Lofoten-Vesterålen Marine Observatory of the coast of Northern Norway

In the same 1989, three more fiber-optic trunk lines were created between Belgium and Great Britain, between Denmark and Sweden and two Danish islands in the Baltic Sea. The main task of the sea lanes of that period was the organization of traffic transit between separated networks.

The disadvantage of organizing work of this type of highways is the technical complexity of maintenance and the need to introduce redundancy in the number of optical fibers to ensure reliability, maintain the bandwidth and the possibility of its expansion with increasing load in the future. The confirmed possibility of extending the service life of highways in relation to copper solutions, as well as the potential to increase the speed characteristics of the fiber used by upgrading the terminal interfaces, has led to a rapid increase in investments in these technologies.

Another convenience that was quickly mastered on the market was the guaranteed quality of communication over long distances. Since the number of bends between the point of immersion of the mainline and its withdrawal to the shore was strictly regulated, therefore, the "intermediate analysis of traffic" characteristic of fiber-optic mains on land did not interfere with the solution of the target problem.

Already in 1998, the Atlantic Crossing-1 (AC-1) transatlantic project was completed. He connected communication nodes Sylt (Germany), Beverwijk (Netherlands), Whitesands Bay (Great Britain), Brookhaven, NY (USA) into a ring structure. The length of the highway is about 13,168 km. The data transfer rate at the time of commissioning is 40 Gbps (0,04 Tbps). For 2014, the data transfer rate was increased to 2,35 Tbps (2.35 Tbps). Its operation is expected to be completed by 2023. At the moment, about 20 fiber-optic lines cross the Atlantic Ocean, most of which were put into operation after 2008. Their resource for modernization, that is, increasing information capacity, is very large.

From this point on, it makes sense to introduce a classification of submarine fiber-optic highways based on the principle of geographic coverage and nationality of communication nodes that are connected to the backbone directly, or as a "tap-off point from the backbone".

1. National fiber-optic backbone – connects communication nodes on the territory of one state.
2. Regional fiber-optic backbone – connects communication nodes on the territory of neighboring states of the same region.
3. Interregional fiber-optic backbone – connects communication nodes on the territory of different states located far from each other.
4. Intercontinental fiber-optic backbone – connects communication centers on the territory of states located on different continents.

Atlantic Crossing-1 (AC-1) is an example of an intercontinental fiber backbone, Scotland-Northern Ireland 1 is an example of a national fiber backbone.

Backbone of the west coast of Africa

The most developed communication center is the port of Lagos (Nigeria) in the Gulf of Guinea (Table 1). Of the six operating fiber-optic highways with access to Lagos (Nigeria), five can be classified as intercontinental. In 2022 and 2023, two more intercontinental fiber-optic highways are expected to be commissioned with access to Lagos (Nigeria). Traffic processing is done on the coast. At the moment, technical sites of six data centers are involved in this work. Most often, cities in the geographical environment of Lisbon (Portugal) are the extreme European points of contact for this direction.

Table 1. List of underwater fiber-optic communication lines with access to Lagos [2]

Name	Trunk type	Length, km	Commissioning year	Baud rate, Tbps
South Atlantic 3/West Africa Subm. Cable (SAT-3/WASC)	intercontinental	14400	2002	0,8
Glo-1	intercontinental	9800	2010	2,5
MainOne	intercontinental	7000	2010	10
Africa Coast to Europe (ACE)	intercontinental	17000	2012	12,8
West Africa Cable System (WACS)	intercontinental	14500	2012	14,5
Nigeria Cameroon Subm. Cable System (NCSCS)	regional	1100	2015	12,8
Equiano	intercontinental	12000	2022	200
2Africa	intercontinental	45000	2023	180

The cost of designing and installing communication systems of this scale is many times higher than the economic capabilities of the countries of the region and their construction is carried out at the expense of foreign investments. In particular, the owners of the investment packages of the 2Africa highway are China Mobile, Facebook, MTN Group, Orange, Saudi Telecom, Telecom Egypt, Vodafone, WIOCC.

Countries neighboring Nigeria have more modest opportunities for technical connection to fiber-optic backbones.

In the Republic of the Congo, for operational use, there are only two inputs of submarine fiber-optic lines:

– West Africa Cable System (WACS) intercontinental fiber backbone (Pointe-Noire, Congo, Rep.) launched in 2012. The total length of the highway is 14,530 km. Terminal nodes Portugal-South Africa. The data transfer rate at the time of commissioning was 14,5 Tbit/s (14,5 Tbps).

– Intercontinental fiber optic backbone 2Africa (hub at Pointe-Noire, Congo, Rep.) with an official commissioning date of 2023. The total length of the highway is 45,000 km. End nodes UK-India, UK-Djibouti-Spain (Figure 2). The estimated speed of information transfer at the time of commissioning is 180 Tbps (180 Tbps).

It is extremely difficult to forecast the analysis of the information traffic of this highway by country. In information flows, tense areas are likely to be highlighted. Probably in the future this highway will be divided into independent projects.



Figure 2. 2Africa cable map (Source: 2Africa Consortium)

The local terrestrial optical operator in the Republic of Congo Congo Telecom has been operating since 1999 and has at its disposal a section of local fiber-optic backbones of 1,529 km. Processing of information traffic between terrestrial and submarine fiber-optic lines is carried out by means of the Congo Telecom Carrier Facility PNR1 data center.

In Guinea, a solution was applied to distribute network traffic between north and south. Two information highways have created independent technical sites and infrastructure in coastal cities that are quite remote from each other.

In the north, in the Kamsar region, and in the south, in the Conakry region, the connection is provided by the Africa Coast to Europe (ACE) intercontinental fiber backbone (Pointe-Noire, Congo, Rep.), Which was commissioned in 2012. The total length of the highway is 17,184 km. Terminals France-South Africa. The data transfer rate at the time of commissioning was 12,8 Tbit/s (12,8 Tbps).

Two independent inputs allow simultaneously and without delays in the implementation of local provider projects to develop the country's economy much faster.

Informational linear capacity of the fiber-optic backbone and the possibility of its expansion

Within the framework of the study, the term informational linear capacity of a fiber-optic backbone is used. This is a numerical characteristic equal to the ratio of the data transmission rate in terms of the total number of light guides involved in the physical backbone to the maximum distance of the information signal delivery.

The error in assessing intercontinental fiber-optic backbones, which is incorporated during integration into the topology of optical fibers for organizing national and regional communication sections, will have to be neglected. In conditions when standard cable types with unified requirements for suppliers are used within the framework of one project, the total number of optical fibers (used and unused, that is, in reserve) along the entire length of the trunk is considered the same.

For example, according to the official technical specifications, the 2Africa backbone uses 16 fiber pairs for the entire operating distance (45,000 km), while the WACS backbone uses only 4 fiber pairs (14,530 km).

The information capacity of the assessment of these highways is:

$$2\text{Africa} = 80 \text{ Tbps} / 45,000 \text{ km} = 4 \text{ Gbps} / \text{km},$$

$$\text{WACS} = 14,5 \text{ Tbps} / 14,530 \text{ km} = 1 \text{ Gbps} / \text{km}.$$

The practice of long-term operation under modernization conditions has shown the following values of changes in the information capacity of the Atlantic Crossing-1 (AC-1) optical highway:

$$1999 = 0,04 \text{ Tbps} / 14,301 \text{ km} = 0,0003 \text{ Gbps} / \text{km},$$

$$2002 = 0,12 \text{ Tbps} / 14,301 \text{ km} = 0,008 \text{ Gbps} / \text{km},$$

$$2007 = 0,635 \text{ Tbps} / 14,301 \text{ km} = 0,044 \text{ Gbps} / \text{km},$$

$$2008 = 0,815 \text{ Tbps} / 14,301 \text{ km} = 0,057 \text{ Gbps} / \text{km},$$

$$2009 = 1,165 \text{ Tbps} / 14,301 \text{ km} = 0,081 \text{ Gbps} / \text{km},$$

$$2010 = 1,455 \text{ Tbps} / 14,301 \text{ km} = 0,1 \text{ Gbps} / \text{km},$$

$$2012 = 1,76 \text{ Tbps} / 14,301 \text{ km} = 0,12 \text{ Gbps} / \text{km},$$

$$2013 = 2,1 \text{ Tbps} / 14,301 \text{ km} = 0,14 \text{ Gbps} / \text{km},$$

$$2014 = 2,35 \text{ Tbps} / 14,301 \text{ km} = 0,16 \text{ Gbps} / \text{km}.$$

Later speed indicators for AC-1 are not given in open sources.

The backbone is built according to the specifications of the Synchronous Digital Hierarchy (SDH) standard and includes 8 DWDM (Dense Wavelength Division Multiplexing) channels in two fiber-optic pairs [3].

The distance between the carrier frequencies in DWDM systems can be from 25 to 200 GHz; in modern networks, a grid of channels with a step of 50 GHz is used. Depending on the type of cable used,

the spectral ranges C (1530...1565 nm), S (1460...1530 nm) and L (1565...1625 nm) can be used for transmission. An example of the channel filling density is shown in Figure 3.

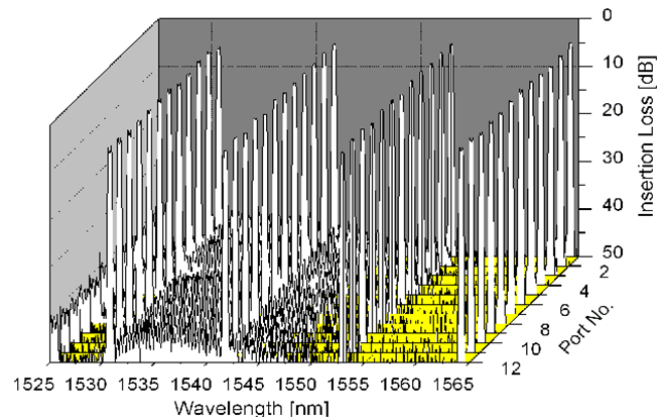


Figure 3. Channel allocation with a step of 50 GHz for a 12-port WSS fabric

To assess the growth of information linear capacity on the west coast of Africa, let us consider the most promising of the region's intercontinental fiber-optic backbones. Africa will use PROADM Wavelength Selective Switching (WSS) technology for the first time in Africa, allowing for more flexible bandwidth management.

The WSS consists of one common optical port on one side and X split multi-wavelength ports on the other side, where each DWDM wavelength input from the common port can be switched (routed) to any of the X multi-wavelength ports regardless of how all the remaining X wavelength channels are routed.

The first industrial WSS systems were based on movable mirrors using microelectromechanical systems (MEMS), now technology development using liquid crystal elements based on silicon (LCoS) is relevant. The existing samples propose the implementation of 1 WSS×4 DWDM channel multiplexing, the transition to 1 WSS×20 DWDM solutions is predicted. The technical implementation of this forecast will give a fivefold increase in the informational linear capacity of optical lines.

Unfortunately, a method for producing submillisecond SLM LCoS crystals is currently not commercially available.

Conclusion

Intercontinental fiber-optic backbones made it possible to organize information flows from end-use networks or the generation of targeted traffic in African countries to large routing nodes and data processing centers in Europe and America. This provides the conditions for the future leveling of commercial activity, the development of the education system, modern production and living standards in these countries.

Planning the growth of the information capacity of the involved fiber-optic backbones should take into account the technical limitations of the cable systems and communication devices used that provide signal retransmission.

References

1. Norwegian Undersea Surveillance Network Had Its Cables Mysteriously Cut. [Electronic resource]. URL: <https://www.thedrive.com/the-war-zone/43094/norwegian-undersea-surveillance-network-had-its-cables-mysteriously-cut>.
2. Global Submarine Cable Systems. [Electronic resource]. URL: <https://www.submarinenetworks.com>.
3. Synchronous Digital Hierarchy. [Electronic resource]. URL: https://www.ee.columbia.edu/~bbathula/courses/HPCN/chap04_part-2.pdf.

RESEARCH ON SKELETON MATCHING CLASSIFICATION

Z.M. LIAO, J. MA, Y.P.Y. QIU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 01, 2022

Abstract. Skeleton matching plays a very important role in image retrieval, which can retrieve the graphics corresponding to the graphics library by the matching degree of object skeleton. Therefore, the key step of shape retrieval is to perform effective bone matching. The skeleton can also be used for surface segmentation, especially when the two deformation models are divided into the same number of surface slices, the difference between the surface slices of the two models needs to be used to build a continuous map, which is the case. In this paper, the skeleton matching methods are summarized, and the advantages and disadvantages of each algorithm are described, which provides theoretical basis for target recognition, shape retrieval, medical image analysis and processing, augmented reality and other fields.

Keywords: skeleton, skeleton matching, target recognition.

Introduction

The rapid development of information technology has made it easier to obtain image resources, and the amount of data has also increased dramatically. Using image analysis and processing methods to process a large number of image data can greatly reduce the workload and improve the efficiency of image data processing. For example, the use of image processing technology can quickly detect image features, this work can be further used in the field of image recognition, greatly improving the work efficiency of industry, medical industry, aerospace industry, and achieve multiplier effect [1].

In computer vision, representing graphics with important information can reduce the workload of recognition [2]. As an important research content of computer graphics representation, skeleton features are simple in structure, so images can be matched effectively and quickly. The skeleton, also known as the central axis, is a simple shape representation, which contains the distance information to the boundary, according to which the boundary of the original shape can be reconstructed, so using the central axis to identify the shape has certain advantages. But the skeleton is greatly affected by the boundary disturbance, so the skeleton matching belongs to a kind of rough matching.

This paper proposes several skeleton matching algorithms, and expounds the theory of each algorithm to achieve skeleton matching from different perspectives. These algorithms have laid a theoretical foundation in areas such as image retrieval, robot navigation, medical image analysis, path planning, and text recognition.

Skeleton graph matching based on path similarity

The main idea is to compare the paths of the endpoints between different skeleton diagrams [3], and do not consider the different stonewall structures between different diagrams, and match the endpoints by passing the path between the comparison endpoints.

The main method is to first extract the skeleton of the calibration type, and then prune and branch to get the skeleton path, and dust the path distance dissimilarity matrix to complete the matching of the end point values.

The biggest advantage of this algorithm is that it avoids the structural changes caused by the instability of the skeleton joints because only the shortest path between the endpoints is considered.

This method only matches endpoints, not intersections. What we want is that both the endpoint and the intersection node participate in the matching, and we hope that more points participate in the matching, not limited to the case where one point corresponds to one point. In the two skeleton diagrams, one point may correspond to multiple points, or multiple points may correspond to multiple points. The matching example is shown in Figure 1.

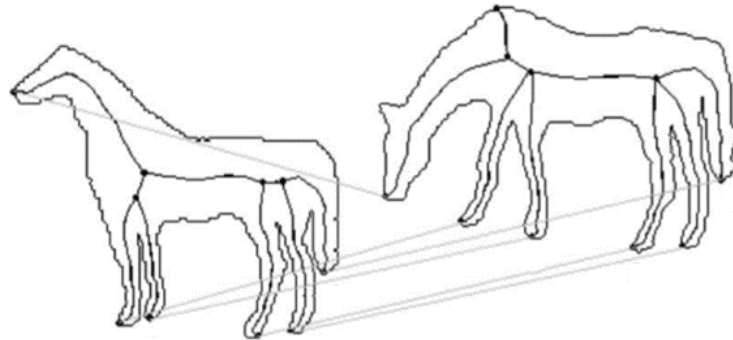


Figure 1. Matching example by using method based on path similarity

Skeleton matching based on clustering

For continuous sampling of moving objects [4], our goal is to divide the skeleton of the object in two adjacent frames into corresponding small parts, and the number of divisions should be the same. The specific method is to extract the key points of the skeleton, and these key points are divided into the intersection and the end of the skeleton, and the matching of the intersection and the end of the skeleton is carried out separately. The end points of the skeleton are first clustered, and the clusters of the end points are matched according to the skeleton paths between the clusters [5], and then a similar method is used to determine the matching relationship of the intersections.

Affine invariant of image feature matching algorithm

The improved method extracts feature points with affine invariance and handle images with repetitive regions well under noise and lighting conditions [6]. We have improved it from the following three angles. First, based on the principle of eight affine simulations, find the most similar affine transformed image to the image to be matched and match the best affine simulated image with the original image to ensure a high matching rate. Secondly, the idea is matched in two directions, and the coincident matching pair of the two-way matching is taken as the stable matching pair. Through bidirectional matching, the stable feature points are guaranteed to be matched first and used as the basis for the matching of other feature points. Then use the stable matching pair to calculate the affine invariant (triangle area ratio invariant) and then improve the accuracy of the remaining point matching. The feature point matching pair has affine invariance.

Conclusion

Shape matching has gone through many years of research. Although there are many mechanisms by which the human brain recognizes shapes is still unclear, the analysis of shapes has become increasingly mature and widely used in many aspects. In this paper, several problems that need to be solved in shape matching are studied, namely, extracting multi-scale skeleton representation that satisfies human visual laws, improving the accuracy and speed of skeleton map matching, improving the similarity measure after shape matching, and using natural images in natural images. Improve the efficiency of learning and modeling in the matching and detection of target shapes. This paper review classification of skeleton matching algorithms skeleton graph matching based on path similarity, skeleton matching based on clustering, affine invariant of image feature matching algorithm, and expounds the theory of each algorithm to achieve skeleton matching from different perspectives. These algorithms have laid a theoretical foundation in areas such as image retrieval, robot navigation, medical image analysis, path planning, and text recognition.

References

1. Bo J. Research on image description and matching algorithm based on graph theory. Anhui University, 2012.
2. Guo M. Research on human skeleton point detection based on deep learning. Xi'an University of Science and Technology, 2019.
3. Bai X. Research on some problems in shape matching based on skeleton. Hua Zhong University of Science and Technology, 2009.
4. Gao B. Skeleton matching based on clustering. Dalian University of Technology, 2012.
5. Zhu B. Graphic image retrieval and automatic clustering based on skeleton graph. Xi Dian University, 2014.
6. Zang Y. Research on Image Feature Matching Algorithm Based on Affine Invariance. North China University of Technology, 2015.

OPTIMIZING SMART HOME NETWORK PROTOCOL SELECTION BY HIERARCHY ANALYSIS METHOD

U.A. VISHNIAKOU, D. ZONGQI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 28, 2022

Abstract. The article analyzes four protocols of the Internet of Things smart home network with corresponding characteristics. The process of choosing the best protocol depending on the expert requirements for the IoT SH network is considered. An example of choosing a protocol based on four indicators based on the hierarchy analysis method is given.

Keywords: IoT network, protocols, smart home, hierarchy analysis method.

Introduction

The Internet of things (IoT) is a set of embedded systems, networks of wireless sensors, control systems and automation tools for processing information received from sensors. An important point in the development of IoT is the analysis of the types of protocols for collecting information from sensors used to transmit data between components of IoT network on short distance [1]. However, finding the best match between the protocol and the IoT network is an optimization task, since it is difficult to understand the influence of parameters when choosing a protocol. In this paper, we consider the choice of the appropriate protocol of the IoT smart home (SH) network using the hierarchy analysis method (HAM).

Algorithm of hierarchy analysis method

The hierarchy analysis method was developed in the early 1970s by T.L. Saati, an operations researcher at the University of Pittsburgh, USA [2]. It is based on a deep analysis of the nature, influencing factors and internal relationships of complex decision-making problems, using less quantitative information to algorithmize the decision-making process. The HAM algorithm includes steps [3].

Step 1. Identification of the problem and formulation of the goal.

Step 2. Definition of the main criteria and alternatives.

Step 3. Building a hierarchy: from goals through criteria to alternatives.

Step 4. Construction of a matrix of pairwise comparisons of criteria in order to select alternatives by criteria.

Step 5. Application of the method of analysis of the obtained matrices.

Step 6. Determination of alternative weights according to the hierarchy system.

The importance of different system options and different quality indicators are compared in pairs. The results of paired comparisons of elements are reduced to a matrix form:

$$\mathbf{A} = \|a_{ij}\|, \quad (1)$$

where $a_{ij} = w_i/w_j$ are estimates of paired comparisons of w_i and w_j selection elements. The diagonal of this matrix is filled with single values, and the matrix elements lying below the diagonal are filled with inverse values, for example, for the value 2 will be $1/2$, for the value 3 – $1/3$, etc.

The expert matrix formed by the pairwise comparison results is called the judgment matrix.

The scaling method of the judgment matrix element a_{ij} is based on comparison from Table 1. Experts form matrices of paired comparisons in relation to all quality indicators.

Table 1. **Element scale table**

Factor i is better than factor j	Value
Two elements of comparison are equally important	1
Moderate advantage of one element over another	3
Significant advantages of one element over another	5
The greater advantage of one element over another	7
The very strong superiority of one element over another	9
Intermediate decision between two decisions	2, 4, 6, 8

Next, the generated matrices of pairwise comparisons of the elements of quality indicators are processed. From a mathematical point of view, this processing problem is reduced to the calculation of the main eigenvector corresponding to the maximum eigenvalue of the matrix. As a result of processing the obtained matrixes, the components of the eigenvector V_j of the global priority vector P_j are obtained

$$V_j = \sqrt[n]{\sum_{i=1}^n a_{ij}}, j \in (1, n), \quad (2)$$

$$P_j = \frac{V_j}{S}, j \in (1, n), S = \sum_{j=1}^n V_j, \quad (3)$$

where n is the number of variants of self-organization algorithms, is compared.

Using the obtained data, the values of the components of the vector of global priorities C_j are calculated according to formula

$$C_j = \sum_{j=1}^n P_j Q_{ij}, i \in (1, n), Q_{ij} = V_i / \sum V_i. \quad (4)$$

Data transmission standards

Currently, there are several standards for data transmission in the IoT networks. Let's consider some of them [4].

1. The Wireless USB standard is a wireless data transmission standard developed by the Wireless USB Promoter Group. During the development, much attention was paid to improving energy efficiency. Devices manufactured in accordance with Specification 1.1 consume less power in idle mode. Wireless USB 1.1 supports NFC technology, which simplifies the setup and operation of wireless USB devices.

2. The Narrowband Internet of Things (NB-IoT) standard is a mobile communication standard for telemetry devices with low data exchange volumes. It was developed by the 3GPP consortium as part of the work on new generation mobile network standards. It was intended to connect a wide range of autonomous devices to the digital communication network, for example, medical sensors, resource consumption meters, smart home devices, etc. NB-IoT is one of three IoT standards developed by 3GPP for mobile communications: eMTC (enhanced machine-type communication), NB-IoT and EC-GSM-IoT [2]. The standard eMTC has the highest bandwidth and is built on the basis of LTE standard equipment. The EC-GSM-IoT standard provides the lowest bandwidth and goes beyond the GSM network.

There are many different communication protocols in the IoT SH network [4]. Taking a smart home as an example, a suitable communication protocol can optimize the network environment and reduce the energy consumption of the entire system.

Optimization the selection of the IoT network protocol

We optimize the smart home network by using the HAM to choose data transmission protocols. The first is the choice of data transmission standards with parameters [4] (Table 2).

Table 2. Comparative analysis of the data transmission standard LTE Cat 0, eMTC, NB-IoT, EC-GSM-IoTs

	LTE Cat 0	eMTC	NB-IoT	EC-GSM-IoT
Downlink speed	1 Mbit/s	1 Mbit/s	250 kbit/s	474 kbit/s or 2 Mbit/s
Delay	5 ms	10 ms–15 ms	1,6 s–10 s	700 ms–2 s
Device bandwidth	1,4–20 MHz	1,4 MHz	180 kHz	200 kHz
Device transmission power	23 dBm	20/23 dBm	20/23 dBm	23/33 dBm

We build the relevant hierarchy according to the Table 3. We create the structural model based on the four characteristics of download speed, delay, device bandwidth, and device transmission power consumption for the four protocols of LTE Cat 0, eMTC, NB-IoT, EC-GSM-IoT. The first level in this structure is the target of the preferred choice, the second level is the quality criteria, and the third level is the alternatives (Figure 1).

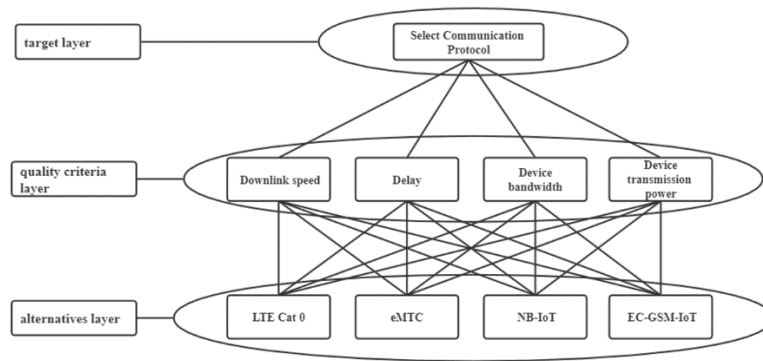


Figure 1. Decomposition of the choice problem

A pairwise comparison matrix was established by experts for this group of quality indicators according with HAM. In Table 3 experts, a pairwise comparison of the importance of selected quality metrics, in particular downlink speed, delay, device bandwidth and device transmission power builds. The diagonal of this matrix is filled with the value "1", and the matrix elements below the diagonal are filled with the inverse value.

Table 3. Computational estimation of matrix and vector components for pairwise comparison of communication protocols

	Downlink speed	Delay	Device bandwidth	Device transmission power	Eigenvector components V_j	Components of the priority vector P_j
Downlink speed	1	1/5	1/4	1	0,47	0,09
Delay	5	1	2	5	2,65	0,52
Device bandwidth	4	1/2	1	2	1,41	0,27
Device transmission power	1	1/5	1/2	1	0,56	0,11

Next, pairwise comparisons are made at step 3 in the form of the relative complexity of the alternatives in relation to each quality indicator. As a result of processing the obtained matrices, according to formula (2) and (3), the eigenvectors (V_{ij}) and priority vectors (Q_{ij}) are calculated, which are given in Tables 4–7.

Table 4. Pairwise comparison matrix of download speed

	LTE Cat 0	eMTC	NB-IoT	EC-GSM-IoT	V_{i1}	Q_{i1}
LTE Cat 0	1	1	4	2	1,68	0,36
eMTC	1	1	4	2	1,68	0,36
NB-IoT	1/4	1/4	1	1/2	0,42	0,09
EC-GSM-IoT	1/2	1/2	2	1	0,84	0,18

Table 5. Pairwise comparison matrix of delay

	LTE Cat 0	eMTC	NB-IoT	EC-GSM-IoT	V_{i2}	Q_{i2}
LTE Cat 0	1	3	9	7	3,70	0,58
eMTC	1/3	1	7	5	1,84	0,29
NB-IoT	1/9	1/7	1	1/2	0,29	0,04
EC-GSM-IoT	1/7	1/5	2	1	0,48	0,07

Table 6. Pairwise comparison matrix of device bandwidth

	LTE Cat 0	eMTC	NB-IoT	EC-GSM-IoT	V_{i3}	Q_{i3}
LTE Cat 0	1	3	9	7	3,70	0,59
eMTC	1/3	1	6	5	1,77	0,28
NB-IoT	1/9	1/6	1	2	0,43	0,06
EC-GSM-IoT	1/7	1/5	1/2	1	0,34	0,05

Table 7. Pairwise comparison matrix of device transmission power

	LTE Cat 0	eMTC	NB-IoT	EC-GSM-IoT	V_{i4}	Q_{i4}
LTE Cat 0	1	2	2	3	1,86	0,42
eMTC	1/2	1	1	2	1	0,22
NB-IoT	1/2	1	1	2	1	0,22
EC-GSM-IoT	1/3	1/2	1/2	1	0,53	0,12

Table 8 summarizes the component estimates of download speed, latency, bandwidth, and communication power consumption construct-related quality indicator priority vectors. Using these priority vectors, the values of each sub-vector of the global priority vector are calculated according to the formula 4 and are given in the last column of Table 8.

Table 8. The results of calculating the values of the components of the global vector of priorities

No.	Standard	Q_{i1}	Q_{i2}	Q_{i3}	Q_{i4}	C_i
1	LTE Cat 0	0,36	0,58	0,59	0,42	0,5395
2	eMTC	0,36	0,29	0,28	0,22	0,283
3	NB-IoT	0,09	0,04	0,06	0,22	0,0693
4	EC-GSM-IoT	0,18	0,07	0,05	0,12	0,0793
	P_j	0,09	0,52	0,27	0,11	–

According to the maximum value of the vector components of the global priority C_i , considering the introduction of the quality index, the better information transmission standards selected should be LTE Cat 0.

Conclusion

The algorithm of the hierarchy analysis method related to expert methods is given. Four protocol standards for creating and modeling networks and in a smart home are described. An example of choosing the best IV protocol using MAI is given. According to the maximum value of the vector components of the global priority the better information transmission standards selected should be LTE Cat 0.

References

1. IoT Data Transfer Protocols. [Electronic resource]. URL: <https://iot.ru/wiki/protokoly-peredachi-dannykh-iot>.
2. Saati T. Decision-making. Method of hierarchy analysis. M., 2003.
3. Nikulshin B.V. [et al.]. System analysis and decision-making in project and management activities. Minsk, 2021.
4. Jamalipur A. Wireless mobile Internet: architecture, protocols and services. M., 2009.
6. Bezruk V.M., Chebotareva D.V. Multi-criteria analysis and choice of means of telecommunications. Kharkov, 2017. P. 268.

STRUCTURAL DESIGN OF THE DATA BASE ON THE INTERNET OF THINGS FOR PRODUCT QUALITY CONTROL

LIU ZHENGHUA

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 5, 2022

Abstract. Database management system structure and function, providing data definition language DDL and data manipulation language DML, and some other user functions. Product quality control database structure design, including logical structure design and physical structure design.

Keywords: DBMS, database for product quality control, database model diagram.

Introduction

A database management system is a software for managing databases. It provides functions such as database definition, data manipulation and database operation management. This article introduces the organization and functions of the database management system. The structure design of the product quality control database, taking the milk quality control database as an example, we introduced its logical structure through the database model diagram, and introduced its physical structure, fields, data types, whether it is empty by taking the milk information table as an example.

Database Management System Structure and Function

DBMS is a large-scale software for manipulating and managing databases, used to establish, use and maintain databases. It manages and controls the database uniformly to ensure the security and integrity of the database. Users access the data in the database through the DBMS, and the database administrator also maintains the database through the DBMS. It can support multiple applications and users in different ways to create, modify and query the database at the same time or at different times.

Most DBMSs provide data definition language DDL (Data Definition Language) and data manipulation language DML (Data Manipulation Language) for users to define the schema structure and permission constraints of the database and implement operations such as adding and deleting data. The database management system (DBMS) structure is shown in Figure 1 [1].

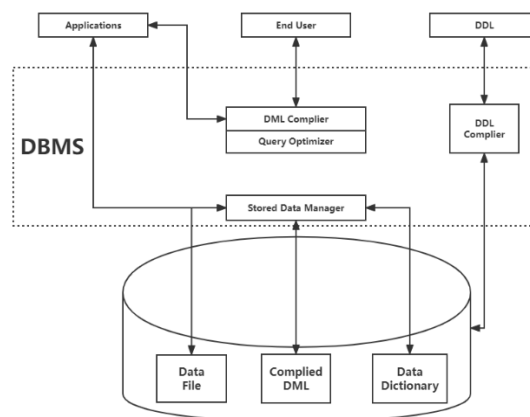


Figure 1. DBMS structure

Structure Design of Product Quality Control Database

Taking the milk quality control database as an example (Figure 2). Model diagram of milk quality control database shows the milk quality control database model, which introduces the associations between the various tables. For example, in the milk information table, MID (Milk ID) is used as the primary key in the table, and CID (Cow ID) is used as a foreign key, and CID is the primary key in the milk information table, so that the cow can be queried in the milk information table through CID, and MID is also used as the foreign key of the fat content table and other tables, so that it can be know which cow the fat content record came from, so as to manage and control milk quality [2].

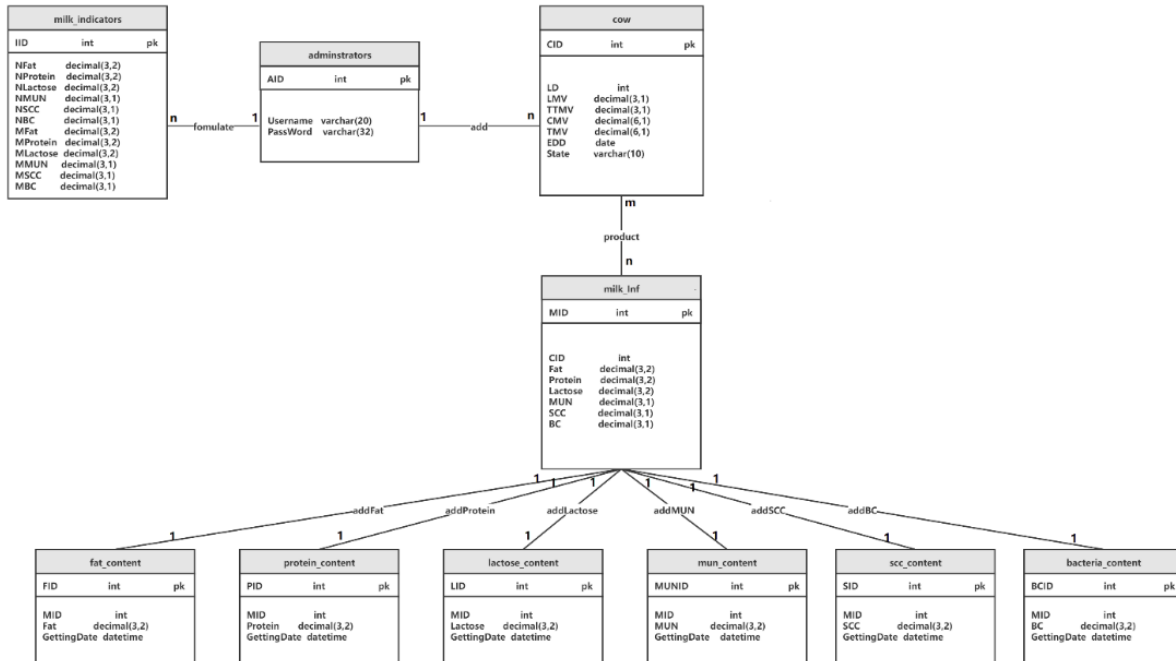


Figure 2. Model diagram of milk quality control database

The MID (Milk ID) field is an int data type. This field is used as the primary key of the table to uniquely identify each record. Whenever a new record is inserted into the table, the field will increment by 1, so the field is Unique.

The CID (Cow ID) field is of type int. It is the primary key in the cow information table. It uniquely identifies each cow. It is used as a foreign key in the milk information table. Through this field, you can know which cow the milk record comes from.

The Fat field is a decimal (3,2) type, similar to floating-point numbers in the C language. The first parameter 3 means that the total number of digits of the data is 3, and the second parameter is 2, indicating that the data is after the decimal point. There are two bits, indicating that the accuracy is 0,01.

The Protein field is of decimal (3,2) type, similar to floating-point numbers in C language. The first parameter 3 indicates that the total number of digits of the data is 3, and the second parameter is 2, indicating that the data is after the decimal point. There are two bits, indicating that the accuracy is 0,01.

The Lactose field is a decimal (3,2) type, which represents a floating-point number. The first parameter 3 indicates that the total number of digits of the data is 3 digits, and the second parameter is 2, indicating that there are two digits after the decimal point of the data, indicating the accuracy is 0,01.

The field of urea nitrogen in milk (MUN) is a decimal (3,1) type, which represents a floating-point number. The first parameter 3 means that the total number of digits of the data is 3 digits, and the second parameter is 1, indicating that the data has one digit after the decimal point. Indicating that the accuracy is 0,1.

The somatic cell (SCC) number field is of decimal (3,1) type, which represents a floating-point number. The first parameter 3 means that the total number of digits of the data is 3 digits, and the second parameter is 1, indicating that the data has one digit after the decimal point. Indicates that the accuracy is 0,1.

Bacteria content (BC) field decimal (3,1) type, which means floating point number, the first parameter 3 means that the total number of digits of the data is 3, the second parameter is 1, indicating that there is one digit after the decimal point of the data, indicating the precision is 0,1.

Table 1. Milk information table structure

Milk Information			
Field name	Data type	Is it empty?	Is it a primary key?
MID	int	No	Yes
CID	int	No	No
Fat	decimal (3,2)	No	No
Protein	decimal (3,2)	No	No
Lactose	decimal (3,2)	No	No
MUN	decimal (3,1)	No	No
SCC	decimal (3,1)	No	No
BC	decimal (3,1)	No	No

Conclusion

We first introduced the structure and function of the database management system. It provides the data definition language DDL (Data Definition Language) and the data manipulation language DML (Data Manipulation Language) for users to define the schema structure and permission constraints of the database to realize the addition of data, delete, etc.

Next, we introduce the logical structure of the milk quality control database through the database model diagram, and introduce the physical structure of the single table, field names, attributes, whether it is the primary key, whether it is the primary key, whether it is null or not by taking the milk information table as an example.

References

1. DBMS (Database Management System) function introduction. [Electronic resource]. URL: https://blog.csdn.net/weixin_45537987/.
2. Wang Lei. Design of dairy farm information management system. D., 2018.
3. Lu Shijie., Wu Hongxiang., Xi Xueyong // Journal of Effects of cow mastitis on main milk components. 2019. Vol. 5, Iss. 7. P. 27–32.

СВЕДЕНИЯ ОБ АВТОРАХ

1. Аксенов Вячеслав Анатольевич – старший преподаватель кафедры инфокоммуникационных технологий БГУИР
2. Алисеенко Маргарита Александровна – аспирант кафедры инфокоммуникационных технологий БГУИР
3. Борискевич Илья Анатольевич – к.т.н., доцент, доцент кафедры инфокоммуникационных технологий БГУИР
4. Вишняков Владимир Анатольевич – д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
5. Вороной Сергей Михайлович – к.т.н., доцент кафедры информационных технологий ГУОП
6. Вычужанин Владимир Викторович – д.т.н., заведующий кафедрой информационных технологий ГУОП
7. Гнедько Елена Михайловна – инженер кафедры инфокоммуникационных технологий БГУИР, магистрант БГУИР
8. Гу Фэншоу – д.т.н., Профессор Центра повышения эффективности и производительности, Университет Хаддерсфилда
9. Давыдов Максим Викторович – к.т.н., первый проректор БГУИР
10. Ди Чжао – магистрант кафедры инфокоммуникационных технологий БГУИР
11. Ду Цзунци – магистрант кафедры инфокоммуникационных технологий БГУИР
12. Егошина Анна Анатольевна – к.т.н., доцент кафедры информационных технологий ГУОП
13. Журавлев Вадим Игоревич – к.т.н., заведующий кафедрой теоретических основ электротехники БГУИР
14. Жэнь Сюнь Хуань – аспирант кафедры инфокоммуникационных технологий БГУИР

15. Кийко Вадим Николаевич – аспирант кафедры инфокоммуникационных технологий БГУИР
16. Конопелько Валерий Константинович – д.т.н., почетный профессор БГУИР
17. Коуа Леонель Игнас – магистрант кафедры автоматизированных систем обработки информации ГГУ имени Ф.Скорины
18. Курума Мохаммед – магистрант кафедры автоматизированных систем обработки информации ГГУ имени Ф.Скорины
19. Ляо Чжунминь – магистрант кафедры инфокоммуникационных технологий БГУИР
20. Марченко Дмитрий Владимирович – магистрант кафедры автоматизированных систем обработки информации ГГУ имени Ф.Скорины
21. Мацкевич Вадим Владимирович – аспирант кафедры информационных систем управления ФПМИ БГУ
22. Мбуйи Мишель Мулумба – магистрант кафедры автоматизированных систем обработки информации ГГУ имени Ф.Скорины
23. Наумович Андрей Игнатович – магистрант кафедры ИСУ, ФПМИ, БГУ
24. Наумович Николай Михайлович – к.т.н., заведующий НИЛ 1.6 БГУИР
25. Рабцевич Виолетта Викторовна – ассистент кафедры инфокоммуникационных технологий БГУИР
26. Рудиков Станислав Игоревич – магистр технических наук, заместитель директора по информационным технологиям Унитарного предприятия «НТЦ «ЛЭМТ» БелОМО»
27. Рудниченко Николай Дмитриевич – к.т.н., доцент кафедры информационных технологий ГУОП
28. Саломатин Сергей Борисович – к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
29. Турлай Андрей Петрович. – соискатель кафедры инфокоммуникационных технологий БГУИР
30. Фам Хю Ань – магистрант кафедры инфокоммуникационных технологий БГУИР

31. Ху Чжифэн – магистрант кафедры
инфокоммуникационных технологий
БГУИР
32. Цветков Виктор Юрьевич – д.т.н., профессор, заведующий кафедрой
инфокоммуникационных технологий
БГУИР
33. Цзюнь Ма – аспирант кафедры
инфокоммуникационных технологий
БГУИР
34. Цю Юэпэйянь – магистрант кафедры
инфокоммуникационных технологий
БГУИР
35. Чжэнхуа Лю – магистрант кафедры
инфокоммуникационных технологий
БГУИР
36. Чэнь Имин – магистрант кафедры
инфокоммуникационных технологий
БГУИР
37. Шibaева Наталья Олеговна – к.т.н., доцент кафедры информационных
технологий ГУОП
38. Шкадаревич Алексей Петрович – д.ф-м.н., профессор, академик НАН РБ,
директор Унитарного предприятия «НТЦ
«ЛЭМТ» БелОМО»
39. Ян Сюэин – магистрантка кафедры
инфокоммуникационных технологий
БГУИР

