

Г.А. Султанова¹, Б.-Б.С. Есмагамбетов², А.А. Мусабеков¹, Г.Е. Тилеуов*, Р.А. Советбаев³

¹М. Әуезов атындағы Оңтүстік Қазақстан зерттеу университеті,
160012, Қазақстан Республикасы, Шымкент қ., Тауке хан көшесі 5

²Ж.А.Тәшенев университеті, 160011, Қазақстан Республикасы, Шымкент қ., Қонаев көшесі 21

³Шәкәрім университеті, 071404, Қазақстан Республикасы, Семей қ., Глинки көшесі 20а

*e-mail: gamidulla.tileuov@mail.ru

АҚПАРАТТЫҚ-ӨЛШЕУ ЖҮЙЕЛЕРІНДЕ ДЕРЕКТЕРДІ ҚЫСУ ӘДІСТЕРІН ЗЕРТТЕУ

Аңдатпа: Ақпараттық техниканың дамуындағы заманауи тенденцияның бірі ақпараттық-өлшеу жүйелерінің (АӨЖ) техникалық мүмкіндіктері барлық салаларда кеңінен қолданылады. Заманауи технологиялардың сапасын арттыру, сондай-ақ ақпараттың көлемі байланыс желілері арқылы берілгендіктен АӨЖ деректерді қысу қажеттілігі туындайды. Байланыс арналарында ақпараттың үлкен массивтерін сақтауда, өңдеуде және жіберуде ақпараттың көлемін қысудың әртүрлі әдістері қарастырылады. Ақпараттылығы жоғары деректерді қысу, телеметриялық өлшеу жүйелерінде тиімді жүргізіледі. Мәселен, ғарыш аппараттары мен спутниктік жүйелерде деректерді жинауда және оларды басқару станциясына беруде әртүрлі қысу әдістерін қолдану арқылы байланыс арнасының өткізу қабілеттілігі азайтылады. Бұл әртүрлі параметрлерді жинауға, өлшеуге және қашықтағы құрылғыларға беруге мүмкіндік беретін инновациялық ғылыми бағыт. Қысу әдістері әртүрлі дерек көздерде деректерді бөлек өңдеуге бағытталған. Ашық немесе жасырын ақпараттық байланыстар деректерді қысу мүмкіндіктерін арттыруы ықтимал тәсілдермен өзара байланысты болып келеді. АӨЖ жобалау кезеңінде күрделі өлшеу параметрлерінен алынған деректердің қысу әдістерінің жүргізілуін нығайтады. Осылайша, деректерді өлшеу, өңдеу және қысу әдістерін қарастыру өзекті міндет болып табылады. Мақалада деректерді қысу әдістері бойынша зерттеу нәтижелері көрсетіледі.

Түйін сөздер: ақпараттық – өлшеу жүйелері, деректерді өлшеу, деректерді қысу алгоритмі, деректерді түрлендіру, телеөлшеу.

Кіріспе

Қазіргі таңда өлшеу жүйелерінде деректерді түрлендірудің қарқынды дамуының нәтижесінде өлшеу ақпаратының үлкен көлемін алуды, өңдеуді және сақтауды қажет ететін ғылыми зерттеулердің нақты міндеттері шешілуде. Өлшеу құралдары қол жетімсіз мәселелерді шешуге мүмкіндік береді, олардың заманауи дамуы өлшеу ақпараттарының көбеюіне әкеледі. АӨЖ ғылыми эксперименттерді технологиялық үрдіспен адаптивті деңгейде модельдейтін инновациялық өлшеу құралдары қолданысқа енуде. Өлшеу үрдісін автоматтандырудың деңгейі және сенімділігі, нәтиженің жоғары ақпараттық көрсетілуін қамтамасыз етеді. Осылайша, қысу әдістерін қолдану тасымалданатын және сақталатын деректер көлемін азайту бойынша сұранысқа ие. Деректерді қысу әдістері АӨЖ байланыс арналарының өткізу қабілеттілігіне қойылатын талаптарды төмендетуге және қажетсіз деректерді жоғалту арқылы ақпаратты өңдеу жүйелерінің тиімділігін арттыруға мүмкіндік береді. Деректер көлемін азайту қысу әдістерін белсенді қолдану арқылы жүзеге асырылады. Бұл мәселені көптеген жетекші ғалымдар зерттеген, соның ішінде М.П. Цапенко, П.В. Новицкий, В.Л. Волков, В.А. Виттах, И.С. Гоноровский, И.С. Еремеев, А.Н. Колмогоров, Р.Е. Кричевский, Е.А. Ломтев, А.И. Лоскутов, Д. Хаффман (D. Huffman), К.Шеннон (C.E. Shannon), Дж. Зив (J. Ziv), А. Лемпел (A. Lempel), К. Вуттон (C. Wotton), Д. Саломан (D. Salomon).

Материалдармен тәсілдер

Дерек көздерде деректерді қысудың әртүрлі әдістері қарастырылған. Зерттеу тақырыбын ашу және түсіну үшін деректерді толық қалпына келтіру мүмкіндігі тұрғысынан қысу әдістері екі топқа бөлінеді:

- Қайтымды қысу;
- Қайтымсыз қысу.

АӨЖ деректерді қайтымды қысу әдістерін зерттеуде төмен көрсеткіштегі шығынмен қысу әдістерін талдауды қамтиды, бірақ шығынсыз қысу сияқты қайтымды болмайды. Бұл әдістер деректердің деградациясына жол берілетін жүйелерде маңызды, бірақ жоғары қысу коэффициенті мен өңдеу жылдамдығы басымдыққа ие. Нақты мәндерді шамалық мәндермен ауыстыру ақпарат көлемін азайтады, сондай-ақ бұрмалануды тудырады. Болжамды және нақты мән арасындағы айырмашылықтың кодталуы деректер көлемін азайтады. Бастапқы деректерді абсолютті дәлдікпен қалпына келтіруге болатын шығынсыз қысудан айырмашылығы, қайтымды әдістер деректерді белгілі бір дәрежеде бұрмалануымен қалпына келтіреді. Қысу алгоритмі мен параметріне байланысты «Қайтымдылық» дәрежесі үлкен шығындар әкеледі. Үлкен көлемдегі деректерді өңдеуді қажет ететін жүйелер үшін дәлдіктің шамалы төмендеуі, қайтымды қысу үшін тиімді болып келеді. Мәселен, әр өлшемнің абсолютті дәлдігі емес, бақылау жүйелерінде ресурстарды үнемдеудің орнына аз мөлшерде деректердің жоғалуы қолайлы. Мысалы, шектеулі жағдайларда ғарыш аппаратының бортық компьютерлерінде қайтымды қысу, байланыс арналарының жадымен өткізу қабілеттілігінің белгіленген шектеулеріне сәйкес келеді. Жоғары қысу коэффициентінің арқасында құрылғылар көбірек деректерді жинап, сақтай алады. Берілетін деректер көлемі байланыс арналарының өткізу қабілеттілігіне қойылатын талаптарды азайтады және деректерді беру құнын төмендетеді. Қайтымды әдістерінде қысудың жоғары деңгейіне қол жеткізіп жоғалмайтын маңызды деректер сақталады. Қайтымды қысу көбінесе толығымен қайтымды немесе шығынсыз қысу деп аталады. Бұл әдіс ақпаратты жоғалтпай деректер көлемін азайтады, яғни кодталған (қысылған) деректерді толықтай өзгеріссіз қалпына келтіреді, мәтіндік ақпаратты кодтау кезінде мәліметтер қорында қолданылады. Қайтымды қысу тобында келесі әдістер кең таралған: статистикалық кодтау және жалпыланған дискреттеу. Статистикалық кодтау мәндерін әртүрлі ұзындықтағы кодтармен кодтауға негізделген, хабарламалар ең қысқа код комбинацияларына сәйкес нәтиженің қайтымдылығы осы әдістердің артықшылығында. Хабарламаның толық статистикасын алдын-ала білу қажеттілігімен байланысты болуы біршама кемшіліктерге әкеледі. Әйтседе, телеөлшеу деректерін беру кезінде мұндай статистика белгісіз. Статистикалық әдістер екі топқа бөлінеді:

1. Адаптивті немесе ағынды әдістер. Өңделген деректер бойынша статистика негізінде есептелген жаңа деректердің ықтималдығын қолдана отырып жасалады.
2. Блоктық әдістер. Блок статистикасы бөлек кодталып, қысылған блокқа қосылады. Блоктық әдістерде Хаффман, Шеннон-Фано әдістерінің статикалық нұсқалары, элементтердің ағындарына арналған арифметикалық кодтау мысалдары жатады [2, 5].

Код ұзындығының символ энтропиясын қолданып, элементтің пайда болу ықтималдығы арқылы анықтауға болады. S -символ энтропиясы, P_s – пайда болу ықтималдығы бар, N_s – ақпарат саны.

$$S, N_s = P_s * \log_2 P_s \quad (1)$$

Шеннонның дерек көзіне сәйкес кодтауды шифрлау теоремасында (Shannon's sourcecoding theorem) S элементі пайда болу ықтималдығымен P_s теңеседі, кодтың ұзындығы битпен көрсетіледі.

$$L_s = \log_2 P_s \quad (2)$$

Тиімділікті бағалау үшін қолданылатын негізгі сипаттамалар қысу коэффициенті K және R қысу жылдамдығы жатады.

$$\text{Қысу коэффициенті } K = \frac{\text{Шығыс файлының өлшемі}}{\text{Кіріс файлының өлшемі}}$$

$0 < K < 1$ – кезінде қысу жақсы жүргізіледі

Егер бұл шама 1 – ден аз болса, қысу жүреді, 1 – ден көп болса, онда кеңейтілу жүреді. Қысу коэффициентіне кері шаманы қысу факторы деп атайды. Қысу факторы кіріс файлының өлшемін шығыс файлының өлшеміне бөлу арқылы анықталады.

$$R = \frac{k}{n} \quad (3)$$

Мұндағы, R - қысу жылдамдығы

k – қысу коэффициенті

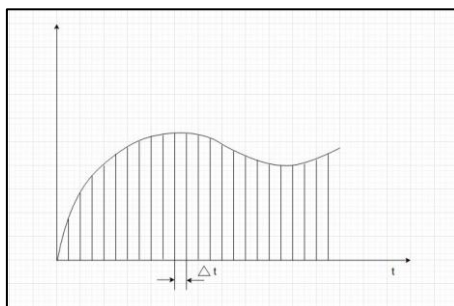
n – биттегі дерек көздеректерінің өлшемі

$$\text{Қысу сапасы } 100 * (1 - k) \quad (4)$$

Мұндағы, k - қысу коэффициенті, қысу сапасын көрсетеді.

Нәтижелер

Қайтымды қысу тобына жататын жалпыланған дискреттеу әдістері анағұрлым жақсы дамыған. Параметрді сұрау жиілігі тұрақты, 1-суретте көрсетілгендей хабарламаның кейбір максималды сипаттамаларына (параметрдің өзгеру жылдамдығына) сүйене отырып таңдалады.



Сурет 1 – Сауалнама жиілігі

Практикада жалпыланған дискреттеу әдістері қажетсіз үлгілерді алып тастау үшін қолданылады. Жуықтау әдістеріне негізделе сауалнаманың тұрақты жиілігімен алдын ала іріктеу жүргізіледі. Деректерді қысу мәселесін C_k коэффициенттері бар алгебралық көпмүшемен өлшеу нәтижелерінің сызықтық жуықтауы ретінде ұсынуға болады C_k :

$$y_i \sim = \sum_{k=0}^m C_k t_i^k \quad (5)$$

Бұл жағдайда жуықтау қатесі рұқсат етілгеннен артық болмауы керек:

$$du_i = \max |y_i - y_i \sim| \leq D \quad (6)$$

Жуықтау алгоритмдерінің келесі нұсқаларын синтездеуге болады:

1. Жуықтау аралығы бекітіліп, көпмүшенің дәрежесі ізделеді;
2. Жуықтайтын көпмүшенің дәрежесі бекітіліп, жуықтау аралығы таңдалады;
3. Жуықтау аралығы және көпмүшенің дәрежесі анықталады.

Алғашқы екі нұсқа бір параметрлі адаптивті нұсқалар болып табылады. Бір параметрлі адаптивті көпмүшенің бекітілген дәрежесі, жуықтау интервалында адаптивті алгоритм нақты уақыт қарқынымен өлшеуге мүмкіндік береді.

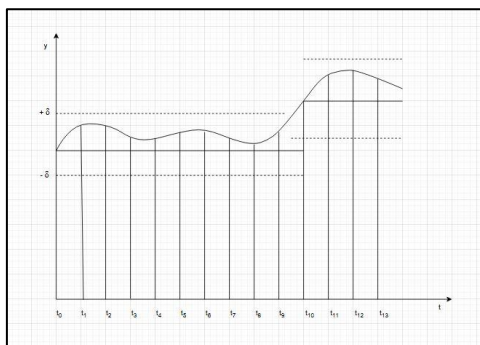
Үшінші нұсқа екі параметрлі адаптивті деректерді қысу алгоритмдерін қамтиды. Ең алдымен көпмүшенің белгіленген дәрежесі мен жуықтау аралығы таңдалады немесе ауыстырылады. Екі параметрлі адаптивті алгоритмдер аппараттық және бағдарламалық қамтамасыз етуді іске асыру күрделілігін ескере отырып, жоғары тиімділікке қарамастан сирек қолданылады.

Маңызды ординаттарды іріктеу кезеңіне байланысты экстраполяциялық және интерполяциялық кезеңдері ажыратылады. Жуықтау көпмүшесінің дәрежесі бойынша алгоритмдер нөлдік, бірінші, екінші және одан жоғары ретті болуы мүмкін [1].

Экстраполяцияға негізделген деректерді қысудың қарапайым алгоритмдерін қарастыратын болсақ, көпмүшенің таңдалған дәрежеде өлшенетін параметрі мүмкін болатын аралықтарда рұқсат етілген қателік қисығына жуықтайды, яғни N санауларының мүмкін болатын үлкен саны шамамен ауыстырылады. Жуықтау аралығы бойынша адаптивті негізделген бұл тәсіл ең ұтымды болып келеді, өйткені көпмүшенің дәрежесі адаптивті кезінде жуықтаудың белгіленген аралығындағы коэффициенттері қателіктерге сезімтал болады да, айырмашылық формулаларын (Ньютон, Лагранж) қолдана отырып есептеледі.

Нөлдік ретті экстраполяторды зерттей келе, u_0 соңғы маңызды санақты білдіреді, оған қатысты $\pm \delta$ кіру өрісі орнатылады. Егер кейінгі алынған санақтар u_0 өсі кіру өрісінің ішінде болса, олар артық болып саналады және жойылады.

Қол жеткізу өрісінің шекарасынан шыққан кез келген u_k өлшемдері маңызды болып саналады, оған қатысты $\pm \delta$ жаңа төзімділік өрісі орнатылады және ережесі қайталанады (2-сурет). $u(t)$ параметрінің функциясын қалпына келтіру келесі маңызды санақ пайда болғанға дейін маңызды санақтар арқылы көлденең сызықтарды жүргізу арқылы жүзеге асырылады.



Сурет 2 – Нөлдік ретті экстраполятор

Бірінші ретті экстраполяторды қолдануда телеөлшенетін функцияның сызықтық өзгеру заңының болу ықтималдығы болжанады. Маңызды ординаттарды іріктеу кезеңі жоғарыда сипатталғанға ұқсас, бұл жағдайда жуықтайтын көпмүшелік сызықтық және үздіксіз функция болып табылады. Алгоритмнің негізгі кемшілігі оның шуға сезімталдығы. Интерполяцияға негізделген қысу әдістерінің басты айырмашылығы $y(t)$ параметрінің әрекетін бірқатар нүктелерде сәйкес келетін белгілі бір уақыт тәуелділіктерімен шамамен ауыстыру болып табылады. Бұл жағдайда интерполяциялық түзудің орны белгілі бір қадамдардан немесе әр қадамда жуықтау интервалының әр нүктесі үшін интерполяциялық мәндерді тексеру арқылы нақтыланады. Бұл үлкен қысу коэффициенттеріне қол жеткізуге мүмкіндік береді, бірақ есептеу уақытының айтарлықтай шығындары болады [4]. Нөлдік ретті интерполяторды зерделей келе, бастапқы функцияны жуықтау артық емес санақтар арқылы өтетін көлденең түзулер жүзеге асырады. Артық емес ординат мәндері, артық ординат мәндерінің орташа мәні ретінде есептеледі.

Зерттеу нәтижесінде алгоритм бірінші және кейінгі есептеулер арасындағы айырмашылықты есептей отырып жұмыс істейді.

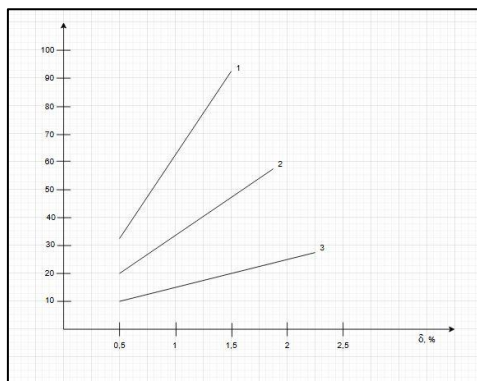
Егер бұл айырмашылық рұқсат етілген δ мәнінен аз болса, өлшемдер артық деп саналады. Қандай да бір қадамда мәндер айырмашылығы δ деңгейінен асып кетсе, бұл мән артық емес болып саналады және барлық алдыңғы мәндердің орташа мәні есептеледі. Бұл орташа мән маңызды санақ болып табылады. Бұл алгоритм нөлдік ретті көпмүшелер арасында қысу тиімділігі бойынша оңтайлы. Бірінші ретті интерполятордың жұмыс принципі келесі зерттеу мысалында түсіндіріледі. u_0 өлшенетін параметрдің артық мәні, ал u_1 , u_2 – кейінгі өлшемдер болсын. Әрбір санаққа қатысты $\pm\delta$ төзімділік өрістері белгіленеді. u_0 нүктесінен $u_1 + \delta$ және $u_1 - \delta$ нүктелері арқылы өтетін түзулер сызылады. Егер u_2 мәні осы түзулермен шектелген өрістің ішінде болса, u_2 ординатасы артық деп саналады. Әрі қарай, екі шекара сызығы қайтадан қалыптасады, бірақ $u_2 + \delta$ және $u_2 - \delta$ нүктелерінен өтіп, үшінші ординаттың артықтығы тексеріледі. Ереже бойынша кез келген ординат толеранттылық шегінен шыққанға дейін қайталанады. Мұндай ординат маңызды болып саналады, ал шекаралық сызықтар пайда болатын координаттар жаңа бастама ретінде қабылданады. Жуықтау сызығы маңызды ординаттар арқылы жүзеге асырылады. Қайтымды қысу алгоритмдерінің тиімділігін бағалау оларды келесі көрсеткіштер бойынша салыстыру кезінде жүргізіледі:

- 1) Қысу коэффициенті K ;
- 2) η тиімділік коэффициенті;
- 3) M_n ұсынудың шуға төзімділігі;
- 4) Алгоритмнің күрделілік коэффициенті.

Негізгі көрсеткіш – теледидарлық өлшемдердің артықтығын жою дәрежесін сипаттайтын қысу коэффициенті. Ол қатынаспен анықталады:

$$K = \frac{n_{\text{э}}}{n}, \quad (7)$$

Мұндағы, n – қысу нәтижесінде өлшеулердің орташа саны, $n_{\text{э}}$ – хабарламадағы өлшеулер саны. 3-суретте қысу коэффициентінің төзімділік мәніне тәуелділігін көрсететін эксперименттік деректер келтірілген. Берілген тәуелділіктерді талдау интерполяциялық үрдістерде ең үлкен қысу әсері болатынын көрсетеді, алайда олардың кемшілігі компьютерде іске асыру қиындықтары кездеседі [8].



Сурет 3 – Қысу коэффициентінің төзімділік шамасына тәуелділігі

Бірөлшемді өңдеу барысында нөлдік ретті болжаушыны пайдаланған кезде шамамен 10 элементарлы операция қажет, ал нөлдік ретті интерполяторды пайдаланған кезде шамамен 25 операция қажет етіледі. Сонымен қатар, интерполяциялық қысу алгоритмдерінің маңызды кемшілігі олардың уақыт бойынша деректерді беру кезінде уақыттық кешіктруді қажет етеді. Егер нақты уақыт режимінде деректерді өңдеу қажет болса, әсіресе төтенше жағдайларды анықтаған кезде, бұл алгоритмдерді қолдану мүмкін емес. Экстраполяциялық алгоритмдердің ішінде нөлдік ретті экстраполятор жоғары қысу коэффициентіне ие, өйткені экстраполяторлар шуға жоғары сезімтал, сондықтан олардың тиімділігі төмен. Осылайша, практикалық іске асыру тұрғысынан экстраполятор ең оңтайлы [3].

Жылдам өзгеретін үрдістерді (ЖӨҮ) өңдеу үшін нөлдік ретті болжаушыны (НРБ) қолдану зерттеледі [6, 7]. Мысалы, сигнал спектрінің тиімді ені 16 кГц болсын. Котельников теоремасына сәйкес бастапқы циклдік іріктеудің жиілігі:

$$f_g = (5+10) \Delta F_{эф.} = (5+10) \times 16 = (80-160) \text{ кГц} \quad (8)$$

Бір санақты өңдеуге арналған уақыт, бұл жағдайда

$$t_g = \frac{1}{f_g} = \frac{1}{(80-160)} = (12,5-6,2) \text{ мкс} \quad (9)$$

1. Бірінші ретті интерполятор
2. Нөлдік ретті интерполятор
3. Нөлдік ретті экстраполятор

Бір санақты нөлдік ретті болжаушымен өңдеу үшін шамамен 20 мкс қажет, сондықтан нақты қарқынмен өңдеу мүмкін еместігін көруге болады. Деректерді өңдеумен берудің кешігуі қосымша жадты қажет етеді. Бір өлшеу арнасын 30 минут (1800 с) өңдеу кезінде қажетті жад көлемі шамамен осы көлемге жетеді.

$$N_n = \frac{20}{12,5} \cdot 1800 \approx 3000 \text{ бит} \quad (10)$$

Осылайша, ЖӨҮ-ні өңдеу үшін қайтымды қысудың қарапайым алгоритмдерін қолдануда бірқатар кемшіліктер орын алды:

- бастапқы дискреттеу жиілігіне қойылатын жоғары талаптар;
- нақты уақыт масштабында өңдеу мүмкіндігінің болмауы;
- қосымша аппараттық шығындар.

Талқылау

Қайтымсыз қысу ақпаратты жояды, деректерді толықтай бастапқы қалпына келтіру мүмкін болмайтын қысу әдісі. Деректердің бір бөлігі жоғалған кезде бастапқы деректердің максималды қысу коэффициенті қамтамасыз етіледі. Бұл қысу әдісі әдетте графикалық, дыбыстық және бейнелі ақпараттық кодтау кезінде қолданылады. Өртүрлі типтегі ақпараттық қысудың екі түрі бар: арнайы; әмбебап.

Арнайы әдістер белгілі бір түрдегі ақпаратты қысуды жүзеге асырады. Мысалы, суретті, дыбысты немесе бейнені қабылдау ерекшеліктерін ескереді. Қажетсіз ақпаратты жоғалтудың нәтижесінде ақпаратты қабылдау жоғары қысу деңгейіне жетеді, сондықтан мұндай әдістер қайтымсыз болып келеді [9].

Әмбебап әдістер ақпаратты биттер тізбегі ретінде береді, әсіресе сандық ақпаратты қысу үшін қолданылады. Мұндай әдістер қайтымды, өйткені қысылған деректерді толық қалпына келтіруге болады.

Хаффман қысу алгоритмі канондық және адаптивті нұсқалары бойынша жоғары тиімділікке ие. Таңбаның кейбір қайталанатын элементі, баспа белгісі, кез келген бит тізбегі, код немесе код сөзі (ASCII) арқылы да түсіндіріледі. Бұл алгоритм негізінен қысу екі кезеңнен тұрады:

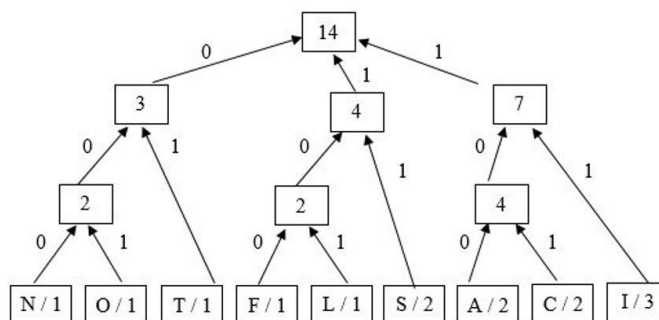
- бірінші кезеңде таңбалардың пайда болу жиілігі немесе ықтималдығы есептеледі,
- таңбалар массиві есептелген жиілік мәндеріне сәйкес сұрыпталады.
- екінші кезеңде осы мәліметтер бойынша Хаффман ағашы салынады, оған сәйкес таңбалардың кодтары есептеледі. Бұл деректерді энтропияға дейін қысуға мүмкіндік береді, сондықтан Хаффман алгоритмі шығынсыз қысудың оңтайлы әдісі болып табылады. Алгоритм код ағашын төменнен жоғары қарай салу арқылы жұмыс істейді [11].

Хаффман әдісін қолдана отырып зерттеу жүргізілді. Бастапқы жолды CLASSIFICATION деп алынып, оның жиілігі анықталды (1 кесте).

Кесте 1 – Жиілік кестесі

I	C	A	S	L	F	T	O	N
3	2	2	2	1	1	1	1	1

Жиілікті қолдана отырып Хаффман ағашын құру жүргізілді (4 – сурет). Әрбір таңба 0 және 1 тізбегі ретінде сақталады.



Сурет 4 – Хаффман ағашы

Түзілген Хаффман ағашына байланысты код құрылды (2 кесте).

Кесте 2 – Хаффман коды

I	C	A	S	L	F	T	O	N
11	101	100	11	101	100	01	001	000

Жиілікті 8 битке көбейтіп бастапқы мәні көрсетілді, сонымен қатар жиілікті кодға көбейту арқылы файл көлемі қысылды (3 кесте).

Кесте 3 – Таңбаларды қысуды есептеу

Жиілігі	Бастапқы	Қысу	Азайтылды
I 3	$3 \cdot 8 = 24$	$3 \cdot 2 = 6$	18
C 2	$2 \cdot 8 = 16$	$2 \cdot 3 = 6$	10
A 2	$2 \cdot 8 = 16$	$2 \cdot 3 = 6$	10
S 2	$2 \cdot 8 = 16$	$2 \cdot 2 = 4$	12
L 1	$1 \cdot 8 = 8$	$1 \cdot 3 = 3$	5
F 1	$1 \cdot 8 = 8$	$1 \cdot 3 = 3$	5
T 1	$1 \cdot 8 = 8$	$1 \cdot 2 = 2$	6
O 1	$1 \cdot 8 = 8$	$1 \cdot 3 = 3$	5
N 1	$1 \cdot 8 = 8$	$1 \cdot 3 = 3$	5
Барлығы:	112 бит	36 бит	76 бит

Зерттеу барысында бастапқы файлдың көлемі 112 бит болған, Хаффман әдісі арқылы қысылған файл 36 бит болды. Нәтижесінде файлдың көлемі 76 битке азайтылды, қысылған

файлдың көлемі 67,8% пайызға қысқарды. Сондай-ақ жұмыс барысында бірқатар кемшіліктер кездесті, олар:

- код ағашының алгоритмі код ағашының сақталуына байланысты еңбек сыйымдылығын арттырады;
- хабарламаның мазмұнын қалпына келтіру үшін ықтималдық кестесін білуі;
- кодтау үшін жиілік кестесін құру және ағаш салу, нақты кодтауды анықтау.

Қорытынды

Жүргізілген зерттеуде АӨЖ деректерді қысу үрдістерін талдаудың нәтижесінде жад ресурсында, байланыс арналарында өткізу қабілеттілігімен есептеу қуатының ұтымдылығы көрсетілді. Статистикалық кодтау арқылы код ұзындықтарының символ энтропиясы қолданылып, элементтің шығу ықтималдық формуласын және қысу коэффициентін, жылдамдығы мен сапасын анықтау қажеттілігі талданды. Жалпыланған дискреттеу әдістерінде параметрді сұрау жиілігінің сипаттамасына негізделі сауалнама жиілігін жүргізудің нәтижесінде практикада қажетсіз үлгілерді алып тастау оңтайлығы ұсынылды. Жуықтау алгоритмдері үш нұсқа бойынша синтездеу жұмыстарын ұйымдастыру қажеттілігі сипатталды. Экстраполяция және интерполяцияға негізделі деректерді қысу алгоритмдерін қарастырудың маңыздылығы көрсетілді. Нөлдік ретті экстраполятор маңызды санақтар арқылы көлденең сызықтарды жүргізу арқылы телеөлшенетін функцияның сызықтық өзгеру заңының болу ықтималдығы болжанды.

Қысу коэффициентінің төзімділік шамасына тәуелділік алгоритмінің шешу жолдарын талдау зерделенді. Әйтседе, жылдам өзгеретін үрдістерді өңдеуде қысу алгоритмдерін қолдануда біршама кемшіліктер айқындалды. Хаффман әдісі арқылы деректерді қысу алгоритмінің нәтижесінде файлдың көлемі 76 битке азайтылды, қысылған файлдың көлемі 67,8% пайызға қысқартылды, дегенмен бұл әдісті зерттеу барысында кемшіліктер де анықталды. Осылайша, өлшеу нәтижесін алудың жылдамдығы, цифрлық ресурстарда ақаулар мен кідірістерді азайту және алмасудың сенімділігін күшейту арқылы АӨЖ жұмысы нығайды. Деректер көлемінің өсуіне байланысты қысу әдісі өзекті, сондай – ақ мультимедиялық, бұлттық және әлеуметтік желі орталарында қысу әдістерін қолдану тиімді болып табылады. АӨЖ өнімділігі дәлдік пен жылдамдықты, экономикалық ұтымдылықты арттыру үшін деректерді қысу әдістерін пайдалану қажет етіледі.

Әдебиеттер тізімі

1. Адаптивные телеизмерительные системы / Б.Я. Авдеев и др. – Л.: Энергия, 1981. – 248 с.
2. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин и др. – М.: ДИАЛОГ-МИФИ, 2003. – 384 с.
3. Бакулина М.П. Эффективное сжатие изображений на основе кодирования низкоэнтропийных источников / М.П. Бакулина // Автометрия. – 2011. – Т. 47, № 1. – С. 59-66.
4. Карпенко Д.С. Метод сжатия телеизмерений с использованием бинарного арифметического кодирования / Д.С. Карпенко, И.А. Максимов, В.В. Иванов // Труды МАИ. – 2013. – № 65. – С. 26-36.
5. Петрянин Д.Л. Сжатие текстовых данных / Д.Л. Петрянин, Е.А. Сидорова, А.В. Зорькин // Новые информационные технологии в автоматизированных системах: Материалы семнадцатого научно-практического семинара. – Москва: ИПМ им. М.В. Келдыша, 2014. – С. 505-507.
6. Левенец А.В. Принципы разработки перспективных способов сжатия телеметрических данных / А.В. Левенец // Вестник ТОГУ. – 2017. – № 2. – С. 31-38.
7. An enhanced run-length encoding compression method for telemetry data / Yanhu Shan et al // Metrology and measurement systems. – 2017. – № 4. – P. 551-562.
8. Обзор способов снижения избыточности телеметрической информации / Ф.Н. Байбекова и др. // Радиопромышленность. – 2019. – Т. 29, № 2. – С. 8-16.
9. A hierarchical adaptive spatio-temporal data compression scheme for wireless sensor networks / S. Chen et al // Wireless Networks. – 2019. – № 25. – P. 429-438.
10. Chen C. A new lossy compression algorithm for wireless sensor networks using Bayesian predictive coding / C. Chen, L. Zhang, R.L.K. Tiong // Wireless Networks. – 2020. – Vol. 26, Issue 8. – P. 5981-5995.
11. Алгоритмы сжатия данных без потерь: учебное пособие для вузов / Е.Р. Пантелеев, А.Л. Алыкова. – Санкт-Петербург: Лань, 2021. – 172 с.

References

1. Adaptivnye teleizmeritel'nye sistemy / B.Ya. Avdeev i dr. – L.: Ehnergiya, 1981. – 248 s. (In Russian).
2. Metody szhatiya dannykh. Ustroistvo arkhivatorov, szhatie izobrazhenii i video / D. Vatolin i dr. – M.: DIALOG-MIFI, 2003. – 384 s. (In Russian).
3. Bakulina M.P. Ehffektivnoe szhatie izobrazhenii na osnove kodirovaniya nizkoehtropiinykh istochnikov / M.P. Bakulina // Avtometriya. – 2011. – T. 47, № 1. – S. 59-66. (In Russian).
4. Karpenko D.S. Metod szhatiya teleizmerenii s ispol'zovaniem binarnogo arifmeticheskogo kodirovaniya / D.S. Karpenko, I.A. Maksimov, V.V. Ivanov // Trudy MAI. – 2013. – № 65. – S. 26-36. (In Russian).
5. Petryanin D.L. Szhatie tekstovykh dannykh / D.L. Petryanin, E.A. Sidorova, A.V. Zor'kin // Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh: Materialy semnadsatogo nauchno-prakticheskogo seminar. – Moskva: IPM im. M.V. Keldysha, 2014. – S. 505-507. (In Russian).
6. Levenets A.V. Printsipy razrabotki perspektivnykh sposobov szhatiya telemetricheskikh dannykh / A.V. Levenets // Vestnik TOGU. – 2017. – № 2. – S. 31-38. (In Russian).
7. An enhanced run-length encoding compression method for telemetry data / Yanhu Shan et al // Metrology and measurement systems. – 2017. – № 4. – P. 551-562. (In English).
8. Obzor sposobov snizheniya izbytochnosti telemetricheskoi informatsii / F.N. Baibekova i dr. // Radiopromyshlennost'. – 2019. – T. 29, № 2. – S. 8-16. (In Russian).
9. A hierarchical adaptive spatio-temporal data compression scheme for wireless sensor networks / S. Chen et al // Wireless Networks. – 2019. – № 25. – P. 429-438. (In English).
10. Chen C. A new lossy compression algorithm for wire-less sensor networks using Bayesian predictive coding / C. Chen, L. Zhang, R.L.K. Tiong // Wireless Networks. – 2020. – Vol. 26, Issue 8. – P. 5981-5995. (In English).
11. Algoritmy szhatiya dannykh bez poter': uchebnoe posobie dlya vuzov / E.R. Panteleev, A.L. Alykova. – Sankt: Lan', 2021. – 172 s. (In Russian).

Г.А. Султанова¹, Б.-Б.С. Есмагамбетов², А.А. Мусабеков¹, Г.Е. Тилеуов*, Р.А. Советбаев³

¹Южно-Казахстанский исследовательский университет имени М. Ауэзова

160012, Республики Казахстан, г. Шымкент, улица Тауке хана 5

²Университет имени Ж.А. Ташенева

160011, Республики Казахстан, г. Шымкент, улица Кунаева 21

³Университет Шакарима

071404, Республики Казахстан, г. Семей, улица Глинки 20а

*E-mail: gamidulla.tileuov@mail.ru

ИССЛЕДОВАНИЕ МЕТОДОВ СЖАТИЯ ДАННЫХ В ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫХ СИСТЕМАХ

Одной из современных тенденций в развитии информационной техники является то, что технические возможности информационно-измерительных систем (ИИС) широко используются во всех отраслях. В зависимости от возможностей современных технологий возникает необходимость в совершенствовании сетей связи и повышении качества их пропускной способности, а также в сжатии данных ИИС, поскольку объем информации передается по сетям связи. Рассматриваются различные способы сжатия объема информации при хранении, обработке и передаче больших массивов информации в каналах связи. Сжатие высокоинформативных данных эффективно проводится в системах телеметрических измерений. Например, в космических кораблях и спутниковых системах пропускная способность канала связи уменьшается за счет использования различных методов сжатия при сборе данных и их передаче на станцию управления. Это инновационное научное направление, которая позволяет собирать, измерять и передавать различные параметры удаленным устройствам. Методы сжатия направлены на раздельную обработку данных в разных источниках. Открытые или скрытые информационные связи становятся все более взаимосвязанными способами, что потенциально увеличивает возможности сжатия данных. На этапе проектирования ИИС усиливают проведение методов сжатия данных, полученных из сложных параметров измерений. Таким образом, рассмотрение методов измерения, обработки и сжатия данных является актуальной задачей. В статье отражаются результаты исследований по методам сжатия данных.

Ключевые слова: информационно-измерительные системы, измерение данных, алгоритм сжатия данных, преобразование данных, телеизмерение.

STUDY OF DATA COMPRESSION METHODS IN INFORMATION-MEASURING SYSTEMS

One of the modern trends in the development of information technology is that the technical capabilities of information and measurement systems (IMS) are widely used in all areas. Due to the capabilities of modern technologies, there is a need to improve communication networks and improve the quality of their bandwidth, as well as to compress data in the IMS due to the fact that the amount of information is transmitted through communication networks. Various methods of compressing the amount of information in the storage, processing and transmission of large arrays of information in communication channels are considered. Highly informative data compression is effectively performed in telemetry measurement systems. For example, in spacecraft and satellite systems, the bandwidth of the communication channel is minimized by using various compression methods in collecting data and transmitting them to the control station. This is an innovative scientific direction that allows you to collect, measure and transfer various parameters to remote devices. Compression methods focus on separate data processing in different data sources. Open or hidden information connections are increasingly interconnected in ways that can increase data compression capabilities. Strengthens the implementation of compression methods for data obtained from complex measurement parameters at the design stage of IMS. Thus, the consideration of methods for measuring, processing and compressing data is an urgent task. The article reflects the results of research on data compression methods.

Key words: information and measurement systems, data measurement, data compression algorithm, data conversion, tele-measurement.

Авторлар туралы ақпарат

Гулбану Султанова – докторант, М. Әуезов атындағы Оңтүстік Қазақстан зерттеу университеті, Шымкент; e-mail: gasultanova@mail.ru. ORCID: <https://orcid.org/0000-0002-5168-7805>.

Бұлат-Батыр Есмағамбетов – д.т.н., профессор, Ж.А. Тәшенев атындағы университеті, Шымкент; e-mail: Bulatbatyr@mail.ru. ORCID: <https://orcid.org/0000-0003-2825-5200>.

Ахметбек Мусабеков – к.т.н., доцент, М. Әуезов атындағы Оңтүстік Қазақстан зерттеу университеті, Шымкент; e-mail: Musabekov_A@rambler.ru. ORCID: <https://orcid.org/0000-0003-0499-472X>.

Гамидулла Тилеуов* – магистр, аға оқытушы, М. Әуезов атындағы Оңтүстік Қазақстан зерттеу университеті, Шымкент; e-mail: gamidulla.tileuov@mail.ru. ORCID: <https://orcid.org/0000-0002-6049-6066>.

Раил Советбаев – PhD доктор, Шәкәрім университеті, Семей; e-mail: rsovetbayev93@gmail.com. ORCID: <https://orcid.org/0009-0007-3605-515X>.

Информация об авторах

Гулбану Султанова – докторант, Южно-Казахстанский исследовательский университет имени М. Ауэзова, Шымкент; e-mail: gasultanova@mail.ru. ORCID: <https://orcid.org/0000-0002-5168-7805>.

Бұлат-Батыр Есмағамбетов – д.т.н., профессор, университет имени Ж.А. Ташенева, Шымкент; e-mail: Bulatbatyr@mail.ru. ORCID: <https://orcid.org/0000-0003-2825-5200>.

Ахметбек Мусабеков – к.т.н., доцент, Южно-Казахстанский исследовательский университет имени М. Ауэзова, Шымкент; e-mail: Musabekov_A@rambler.ru. ORCID: <https://orcid.org/0000-0003-0499-472X>.

Гамидулла Тилеуов – магистр, старший преподаватель, Южно-Казахстанский исследовательский университет имени М. Ауэзова, Шымкент; e-mail: gamidulla.tileuov@mail.ru. ORCID: <https://orcid.org/0000-0002-6049-6066>.

Раил Советбаев – Phd доктор, университет Шәкәрім, Семей; e-mail: rsovetbayev93@gmail.com. ORCID: <https://orcid.org/0009-0007-3605-515X>.

Information about the authors

Gulbanu Sultanova – doctoral student, M.Auezov South Kazakhstan Research University, Shymkent; e-mail: gasultanova@mail.ru. ORCID: <https://orcid.org/0000-0002-5168-7805>.

Bulat-Batyr Yesmagambetov – doctor of technical sciences, professor, J.A. Tashenev University, Shymkent; e-mail: Bulatbatyr@mail.ru. ORCID: <https://orcid.org/0000-0003-2825-5200>.

Akhmetbek Musabekov – candidate of technical sciences, docent, M. Auezov South Kazakhstan Research University, Shymkent; e-mail: Musabekov_a@rambler.ru. <https://orcid.org/0000-0003-0499-472X>.

Gamidulla Tileuov – master, senior teacher, M. Auezov South Kazakhstan Research University, Shymkent; e-mail: gamidulla.tileuov@mail.ru. ORCID: <https://orcid.org/0000-0002-6049-6066>.

Rail Sovetbayev – doctor PhD, Shakarim University, Semey; e-mail: k-mamadieva@mail.ru. ORCID: <https://orcid.org/0009-0007-3605-515X>.

Редакцияға енуі 05.01.2026
Өңдеуден кейін түсуі 05.03.2026
Жариялауға қабылданды 17.03.2026

[https://doi.org/10.53360/2788-7995-2026-1\(21\)-9](https://doi.org/10.53360/2788-7995-2026-1(21)-9)



МРНТИ: 81.93.29

Д.Т. Тюлемисова¹, А.К. Шайханова*, В.П. Марценюк², Г.Б. Бекешова¹, Б.Т. Смаилова³

¹Евразийский национальный университет имени Л.Н. Гумилева,
10000, Республика Казахстан, г.Астана, ул. Сатпаева, 2

²Университет Бельско-Бяла,
ul.Willowa 2, 43-300, Бельско-Бяла, Польша

³Шәкәрім университет,
071412, Республика Казахстан, г. Семей, ул. Глинки 20 А
*e-mail: shaikhanova_ak@enu.kz

ОПТИМИЗАЦИЯ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ДЕТЕКЦИИ ФЕЙКОВЫХ НОВОСТЕЙ: ПОДБОР ГИПЕРПАРАМЕТРОВ И АНАЛИЗ ROC-КРИВЫХ

Аннотация: В условиях активного и неконтролируемого распространения информации в цифровой среде, особенно в социальных медиа и новостных агрегаторах, особую актуальность приобретает задача автоматического выявления фейковых новостей. Рост объемов пользовательского контента и высокая скорость его распространения существенно усложняют ручную верификацию информации, что обуславливает необходимость применения методов машинного обучения. Целью настоящего исследования является анализ, сравнительная оценка и оптимизация базовых моделей машинного обучения для детекции фейковых новостей с акцентом на подбор гиперпараметров, вычислительную эффективность и интерпретацию результатов классификации. В работе использован набор текстовых данных ISOT Fake News Dataset, содержащий новостные сообщения на английском языке с бинарными метками «правда» и «ложь», прошедшие этапы очистки и векторизации с применением метода TF-IDF. В рамках исследования реализованы и проанализированы модели логистической регрессии, дерева решений, случайного леса и градиентного бустинга. Оценка качества классификации проводилась на основе метрик Accuracy, Precision, Recall, F1-score, а также ROC-анализа и значения площади под кривой (AUC). Показано, что модель градиентного бустинга обеспечивает наивысшую точность классификации, тогда как логистическая регрессия демонстрирует сопоставимое качество при значительно меньших вычислительных затратах и времени обучения. Дополнительно в работе выполнен контроль утечки данных и анализ причин завышенных метрик качества, что позволило обеспечить корректную интерпретацию результатов. Полученные выводы подтверждают, что оптимизированные классические модели машинного обучения способны обеспечивать высокое качество детекции фейковых новостей и могут рассматриваться как ресурсоэффективная альтернатива более сложным нейросетевым подходам в прикладных задачах кибербезопасности и мониторинга информационных потоков.

Ключевые слова: фейковые новости; машинное обучение; оптимизация гиперпараметров; классификация текстов; ROC-анализ; AUC; логистическая регрессия.

1. Введение

В условиях стремительного роста объемов информации в сети Интернет проблема распространения фейковых новостей приобретает особую актуальность. Социальные медиа и онлайн-платформы стали основным источником новостных сообщений, однако наряду с достоверной информацией всё чаще распространяются ложные или искажённые данные, способные оказывать значительное влияние на общественное мнение, политические

©Д.Т. Тюлемисова, А.К. Шайханова, В.П. Марценюк, Г.Б. Бекешова, Б.Т. Смаилова, 2026