

Дәріс 5. Әдістің теориялық негіздері. Ақпаратты жіберуді оңтайландыру үшін кодтар құру.

Дәріс жоспары.

1. Әдістің теориялық негіздері.
2. Ақпаратты жіберуді оңтайландыру үшін кодтар құру.
3. Хаффман және Шеннон–Фано кодтау әдістері.

Әдістің теориялық негіздері

Әдістің теориялық негіздері ақпарат теориясының негізгі қағидаларына сүйенеді, атап айтқанда К. Шеннон ұсынған энтропия және ақпарат мөлшерін өлшеу ұғымдарына. Ақпаратты кодтау мен жіберу процесі — бұл ақпарат көзінен шыққан хабарларды белгілі бір ережелер бойынша символдар тізбегіне айналдыру. Бұл кодтар ақпаратты жоғалтпай, сенімді және тиімді түрде жеткізуге бағытталған. Кодтау әдістерінің теориялық негізінде ықтималдық үлестірімі, энтропия, ақпараттың артықшылығы және кодтың ұзындығы сияқты түсініктер жатыр.

Ақпарат теориясы бойынша тиімді код дегеніміз — ақпарат көзінен алынған хабарлардың ықтималдығын ескере отырып, олардың орташа ұзындығын минимизациялайтын код. Шеннонның негізгі теоремасына сәйкес, орташа код ұзындығы энтропиядан кіші бола алмайды, яғни $H(X)L_{\min} \geq H(X)$. Демек, хабардың ықтималдығы жоғары болған сайын, оған қысқа код, ал ықтималдығы төмен хабарға ұзын код беру тиімді. Бұл тәсіл ақпаратты жіберу кезінде артық бит санын азайтып, арна жүктемесін төмендетеді.

Ақпаратты жіберуді оңтайландырудың басты мақсаты — кодтау тиімділігін арттыру. Тиімді кодтау әдістері деректерді қысу (компрессия) мен арналық өткізу қабілетін үнемдеуге мүмкіндік береді. Бұл үшін кодтардың бірімәнді шешілуі (декодталуы) және қателерге төзімділігі қамтамасыз етілуі қажет. Егер код бірімәнді болмаса, қабылдаушы құрылғы хабарды дұрыс қалпына келтіре алмайды. Сондықтан, кодтау кезінде префикстік кодтар (мысалы, Хаффман кодтары) жиі қолданылады.

Префикссіз (яғни, бір кодтық комбинация екінші кодтың басы болмайтын) кодтар декодтау кезінде шатасуды болдырмайды. Мұндай кодтар Фано және Хаффман әдістерімен жасалады. Бұл әдістердің теориялық негізі — ықтималдығы жоғары хабарларға қысқа код, ал сирек кездесетін хабарларға ұзын код беру принципі. Осылайша, ақпаратты жіберудің орташа ұзындығы азаяды, яғни байланыс арнасы тиімдірек пайдаланылады.

Кодтаудың теориялық тұрғыдан маңызды сипаттамаларының бірі — артықтылық (redundancy). Артықтылық хабар құрамындағы қосымша биттерді білдіреді, олар ақпараттың мағынасын өзгертпейді, бірақ байланыс қателерін анықтауға немесе түзетуге көмектеседі. Мысалы, циклдік кодтар мен Хэмминг кодтары деректерді қорғау үшін қолданылады. Бұл кодтар арқылы ақпаратты тек тиімді емес, сонымен бірге сенімді түрде жеткізуге болады.

Ақпаратты жіберуді оңтайландыру тек код ұзындығын азайтумен шектелмейді. Ол сондай-ақ арна өткізу қабілеттілігін (channel capacity) және

шу деңгейін ескеруді талап етеді. Егер арнада шу көп болса, онда қателік ықтималдығын азайту үшін қосымша түзету кодтары енгізіледі. Бұл кезде Шеннонның арна сыйымдылығы туралы теоремасы пайдаланылады: $C = B \log_2(1 + \frac{S}{N})$, мұнда C - арна сыйымдылығы, B — өткізу жолағы, S/N — сигнал мен шудың қатынасы.

Ақпаратты кодтау кезінде ықтималдық үлестірімі де шешуші рөл атқарады. Егер хабарлар тең ықтималды болса, онда бірдей ұзындықтағы кодтар пайдаланылады (мысалы, ASCII). Ал хабарлар тең емес ықтималды болған жағдайда, тиімді кодтау әдістері (мысалы, Хаффман немесе арифметикалық кодтау) қолданылады. Бұл әдістер хабарларды ықтималдығына қарай бейімдейді, соның нәтижесінде кодтың орташа ұзындығы азаяды және арна ресурсы үнемделеді.

Кодтау теориясында Шеннон-Фано әдісі — ақпаратты ықтималдыққа негіздеп кодтаудың алғашқы тәсілдерінің бірі. Бұл әдісте хабарлар ықтималдығына қарай реттеліп, жиын екіге бөлінеді: әр топтың ықтималдығы шамамен тең болуы керек. Осылайша, кодтық ағаш құрылады, әрбір хабарға 0 немесе 1 таңбасы тағайындалады. Кейінірек Хаффман алгоритмі бұл әдісті жетілдіріп, код ұзындығы бойынша ең тиімді нұсқаны ұсынды.

Ақпаратты жіберуді оңтайландырудың заманауи бағыттары — адаптивті кодтау және ықтималдықты болжауға негізделген кодтар. Мұндай әдістер хабар ағынының құрылымын динамикалық түрде талдап, кодтарды автоматты түрде өзгертіп отырады. Бұл әсіресе бейне, аудио және мәтіндік деректерді қысу кезінде тиімді. Мысалы, MPEG және JPEG стандарттарында осындай бейімделгіш кодтау әдістері қолданылады.

Қорытындылай келе, ақпаратты жіберуді оңтайландыру үшін кодтар құру — теория мен тәжірибені біріктіретін маңызды сала. Ол ақпаратты жоғалтпай, жылдам әрі сенімді жеткізуге бағытталған. Теориялық негіздері ықтималдық заңдылықтары мен энтропия ұғымына сүйенеді, ал практикалық іске асыруы Хаффман, Фано, және Грей сияқты кодтау әдістерінде көрініс табады. Осының арқасында ақпараттық жүйелер мен байланыс арналарында жоғары өнімділік пен сенімділік қамтамасыз етіледі.

Ақпаратты жіберуді оңтайландыру үшін кодтар құру.

Ақпаратты жіберуді оңтайландыру үшін кодтар құру - бұл ақпарат теориясындағы маңызды бағыттардың бірі, оның негізгі мақсаты - ақпаратты мүмкіндігінше қысқа, дәл және сенімді түрде жеткізу. Ақпаратты кодтау кезінде хабарлар символдар жиыны арқылы беріледі, ал оңтайлы кодтар - осы хабарларды артық биттерсіз және қателіксіз жеткізуге мүмкіндік беретін кодтар жүйесі. Мұндай кодтарды құру кезінде хабарлардың ықтималдығы, энтропиясы және арнаның өткізу қабілеті есепке алынады.

Ақпаратты жіберуді оңтайландырудың негізі - К. Шеннонның энтропия ұғымы. Энтропия хабар көзі шығаратын орташа ақпарат мөлшерін сипаттайды және ол кодтың ең төменгі мүмкін орташа ұзындығын анықтайды. Яғни, кодтың тиімділігі энтропиямен шектеледі: орташа код ұзындығы энтропиядан

кем бола алмайды. Бұл принцип ақпаратты артықшылықсыз, бірақ толық жеткізудің теориялық шегі болып саналады.

Кодтарды оңтайландыру кезінде хабарлардың ықтималдық үлестірімін ескеру қажет. Егер хабарлардың пайда болу ықтималдығы бірдей болмаса, тиімді кодтау әдісі әр хабарға әртүрлі ұзындықтағы код береді: жиі кездесетін хабарларға қысқа, ал сирек кездесетін хабарларға ұзын код. Мұндай тәсіл орташа код ұзындығын азайтып, ақпаратты тиімді жеткізуге мүмкіндік береді. Бұл принципке негізделген ең танымал әдістердің бірі - Хаффман кодтау әдісі.

Хаффман коды - ықтималдыққа негізделген префикстік код, онда бірде-бір кодтық комбинация басқа комбинацияның басы болмайды. Осылайша, хабарды бізмәнді түрде декодтауға болады. Хаффман әдісі ақпаратты қысу саласында, мысалы, ZIP, JPEG, және MP3 форматтарында кеңінен қолданылады. Ал Шеннон–Фано әдісі де ұқсас қағидаларға сүйенеді, бірақ Хаффман кодтауымен салыстырғанда тиімділігі аздау.

Ақпаратты жіберу процесін оңтайландыру кезінде артықтылық (redundancy) ұғымы да маңызды рөл атқарады. Артық биттер кейде ақпаратты қорғау мақсатында әдейі енгізіледі. Мысалы, Хэмминг кодтары мен циклдік кодтар (CRC) қателерді табу және түзету үшін қолданылады. Бұл кодтар арнада шу болған жағдайда да ақпараттың бұрмаланбауын қамтамасыз етеді.

Арифметикалық кодтау — заманауи тиімді кодтау әдістерінің бірі. Бұл әдіс әр хабардың ықтималдығын интервал түрінде бейнелейді және бүкіл хабар тізбегін бір нақты санмен кодтайды. Мұндай тәсіл жоғары ықтималдық дәлдігін қамтамасыз етеді және бейне мен мәтінді қысу алгоритмдерінде (мысалы, MPEG, H.264) кеңінен пайдаланылады.

Ақпаратты жіберуді оңтайландыру тек кодты қысқарту ғана емес, сонымен қатар арна сыйымдылығын (channel capacity) тиімді пайдалану болып табылады. Егер код тым ұзақ болса, арнаға түсетін жүктеме артады және өткізу қабілеті төмендейді. Сондықтан кодтарды жобалау кезінде арнадағы шу деңгейі, қателік ықтималдығы және сигнал/шу қатынасы (S/N) сияқты факторлар ескеріледі.

Оңтайлы кодтар құрудың тағы бір маңызды аспектісі — бізмәнді декодталу. Яғни, қабылдаушы жақ кодталған хабарды еш шатаспай қайта қалпына келтіре алуы тиіс. Мұны қамтамасыз ету үшін Фано теңсіздігі, Крафт теңсіздігі сияқты математикалық шарттар қолданылады. Бұл шарттар кодтық жүйенің тиімді әрі дұрыс жұмыс істеуін қамтамасыз етеді.

Ақпаратты тиімді жіберу үшін адаптивті кодтау әдістері де қолданылады. Бұл тәсіл хабар ағынының құрылымын талдап, кодтарды динамикалық түрде өзгертеді. Мысалы, мәтіндік ақпаратта кейбір әріптер жиі қайталанса, оларға қысқа кодтар тағайындалады. Бұл әдіс әсіресе интернет арқылы деректерді жіберу кезінде өткізу қабілетін үнемдейді.

Қорытындылай келе, ақпаратты жіберуді оңтайландыру үшін кодтар құру — қазіргі заманғы байланыс және ақпараттық технологиялардың негізін құрайды. Бұл салада математикалық теория мен инженерлік тәжірибе үйлесім тапқан: ықтималдық, энтропия және кодтық құрылымдар арқылы деректерді жылдам, тиімді әрі сенімді түрде жеткізу қамтамасыз етіледі. Оңтайлы кодтар

энергияны үнемдейді, арна жүктемесін азайтады және ақпараттың бүтіндігін сақтайды - бұл цифрлық дәуірдің басты талабы.

Хаффман және Шеннон–Фано кодтау әдістері.

Хаффман және Шеннон–Фано кодтау әдістері — ақпарат теориясындағы ең танымал және кең қолданылатын деректерді қысу тәсілдері. Бұл әдістердің басты мақсаты — хабарды кодтау кезінде оның ұзындығын азайту, яғни ақпаратты мүмкіндігінше аз битпен жеткізу. Мұндай кодтау тәсілдері ықтималдық теориясына негізделеді: жиі кездесетін символдарға қысқа код, сирек кездесетіндерге ұзын код тағайындалады. Осылайша, орташа код ұзындығы минимизацияланады, ал деректерді сақтау немесе жіберу тиімдірек болады.

Шеннон–Фано әдісі — алғашқы ықтималдыққа негізделген кодтау алгоритмдерінің бірі. Бұл әдісті 1940 жылдары Клод Шеннон мен Роберт Фано ұсынған. Оның негізгі идеясы — хабардағы символдардың ықтималдығына қарай кодтарды теңестіріп бөлу. Яғни, ең жиі кездесетін символдарға қысқа, ал сирек кездесетін символдарға ұзын код беріледі. Кодтарды құру процесі келесідей жүреді: алдымен барлық символдар олардың пайда болу ықтималдығы бойынша кему ретімен реттеледі, содан кейін жиын екі бөлікке бөлінеді — әр бөліктегі ықтималдықтардың қосындысы бір-біріне шамалас болатындай етіп. Әр бөлікке 0 және 1 тағайындалады, содан кейін процесс рекурсивті түрде қайталанады.

Шеннон–Фано әдісінің артықшылығы — оның салыстырмалы түрде қарапайым және тез орындалатын алгоритм екендігі. Дегенмен, бұл әдіс әрқашан ең қысқа код бермеуі мүмкін. Кейбір жағдайларда алынған кодтың орташа ұзындығы Хаффман әдісіне қарағанда аздап ұзын болуы мүмкін. Сондықтан практикалық жүйелерде, әсіресе деректерді қысу стандарттарында, көбіне Хаффман кодтау қолданылады.

Хаффман кодтау әдісі 1952 жылы Дэвид Хаффман тарапынан ұсынылған. Бұл әдіс те символдардың ықтималдығына сүйенеді, бірақ кодты құру процесі біршама тиімді және дәл. Хаффман алгоритмінің негізгі принципі — ең аз ықтималдыққа ие екі символды біріктіріп, оларды бір түйін ретінде қарастыру. Бұл процесс барлық символдар бірыңғай бинарлық ағашқа біріктірілгенге дейін қайталанады. Нәтижесінде жиі кездесетін символдар ағаштың түбіне жақын орналасады, яғни қысқа код алады, ал сирек кездесетіндер ұзын код алады.

Хаффман әдісі арқылы алынған код әрқашан оптималды префикстік код болып табылады. Бұл дегеніміз — бір код ешқашан басқа кодтың басы бола алмайды. Мұндай қасиет кодтарды қате түсіндіру мүмкіндігін болдырмайды және декодтау процесін оңайлатады. Сонымен қатар, Хаффман коды деректерді жоғалтпай қысуға мүмкіндік береді, сондықтан ол lossless compression санатына жатады.

Шеннон–Фано және Хаффман әдістерінің айырмашылығы олардың құрылу логикасында. Шеннон–Фано алгоритмі ықтималдықтар тізімін екіге бөлуден басталса, Хаффман әдісі керісінше, ең аз ықтималдықты элементтерді біріктіруден бастайды. Осы себепті Хаффман әдісі әрқашан оңтайлы нәтиже

береді, ал Шеннон–Фано кейбір жағдайларда тек жуық шамада тиімді нәтиже көрсетеді.

Бұл екі әдіс те қазіргі заманғы ақпаратты қысу алгоритмдерінің негізін қалайды. Мысалы, Хаффман кодтау ZIP, JPEG, MP3, PDF және басқа да танымал форматтарда қолданылады. Ал Шеннон–Фано коды тарихи тұрғыда маңызды, себебі ол ықтималдыққа негізделген кодтау идеясының алғашқы іске асырылуы болды.

Теориялық тұрғыдан, бұл әдістер Клод Шеннонның ақпарат энтропиясы ұғымына сүйенеді. Энтропия хабар құрамындағы орташа ақпарат мөлшерін сипаттайды және кодтау кезінде қол жеткізуге болатын минималды орташа ұзындықты анықтайды. Хаффман коды энтропиямен тығыз байланысты, себебі ол орташа код ұзындығын энтропияға барынша жақындатады, яғни мүмкін болған ең тиімді кодты береді.

Сонымен қатар, Хаффман және Шеннон–Фано кодтары тек деректерді қысуға ғана емес, сонымен бірге байланыс арналарындағы өткізу қабілеттілігін арттыру, берілетін ақпараттың сенімділігін қамтамасыз ету, және артықшылықсыз кодтар құру сияқты міндеттерді шешуде де қолданылады. Бұл әдістер телекоммуникация, мультимедиа, деректер сақтау және желілік хаттамалар саласында кеңінен таралған.

Қорытындылай келе, Хаффман мен Шеннон–Фано кодтау әдістері ақпарат теориясының іргетасы болып табылады. Олар ықтималдық негізінде ақпаратты тиімді кодтауға мүмкіндік береді және қазіргі цифрлық технологияларда деректерді қысу мен жіберуді оңтайландырудың негізгі құралдарына айналған. Хаффман әдісі өзінің дәлдігі мен оптималды нәтижелерімен ерекшеленсе, Шеннон–Фано әдісі теориялық түсініктер мен оқу мақсаттары үшін маңызды болып қала береді.

Пайдаланылған әдебиеттер тізімі

1. Асанов С. *Физикалық энциклопедия*. – Алматы: Қазақ энциклопедиясы, 2003.
2. Шеннон К. *Математическая теория связи*. – М.: Наука, 1963.
3. Клаузиус Р. *О движущей силе тепла и законах, пригодных для этого движения*. – Берлин, 1865.
4. Больцман Л. *Избранные труды по статистической физике*. – М.: Наука, 1984.