

Қазақстан Республикасы Тәуелсіздігінің 30 жылдығына арналған «Сейфуллин оқулары – 17: «Қазіргі аграрлық ғылым: цифровизация трансформация» атты халықаралық ғылыми – тәжірибелік конференцияға материалдар = Материалы международной научно – теоретической конференции «Сейфуллинские чтения – 17: «Современная аграрная наука: цифровая трансформация», посвященной 30 – летию Независимости Республики Казахстан.- 2021.- Т.1, Ч.4 - С.250-253

АҚПАРАТТЫҚ ЭНТРОПИЯ ТЕНДЕУІН ПРАКТИКАЛЫҚ МАҚСАТТА ҚОЛДАНУ

Заманбекова Жұлдыз Серікболқызы

Қазіргі қоғамның барлық сферасы ақпараттанып жатқан кезде, ақпаратттық-коммуникациялық технологиялар адамның барлық іс-әрекетінде кең түрде қолданып жатқанда және ақпараттың тасқынды өсу жағдайында математика ғылымындағы сүбелі бір теория *ақпарат теориясы* да табысты дамып отыр. Атап айтқанда, «ақпарат», «ақпарат мәлшері», «ақпарат алмасу», «оқиғалар ықтималдылығы», «энтропия», «ақпараттың орташа мәні», «ақпарттың бағалылығы және артық болуы», «ақпаратты кодтау» және т.б. мәселелердің шешімі табылмаған тұстары жетерлік.

Ақпараттың қолдану аясының ауқымдылығы оны сипаттайтын параметрлерді өлшеу процедурасын жетілдіруді талап етіп отыр. Ақпаратты өлшеу ғылыми танымның әдіснамасының басты мәселесіне айналды. Ақпаратты өлшеу ұғымы мына аксиомамен тікілей байланысты: белгілі бір нысан туралы ақпарат келіп түскен, яғни ақпарат өскен сайын, осы нысан туралы білмеушілік азая береді.

Мысалы, А оқушы С қаласында тұрады. Ал оқушының осы қаланың Ертіс көшесінде тұратыны туралы хабарлама анықталмаушылықты азайтты. Сөйтіп, бір ғана ақпарат дозасын алған соң, біз бұрынғыға қарағанда көп билетін болдық. Бірақ ақпараттық анықталмаушылық (информационная неопределенность) азайғанымен жоғалып кеткен жок. Ақпараттық анықталмаушылықтың кері шамасының логарифмі ықтималдық деп аталады.

Ал анықталмаушылықты екі есе азайтатын хабарламада 2 бит информация болады. Бит – информация мәлшерінің бірлігі.

1 мысал. Шарик 4 жәшіктің біреуінде орналасқан.

Бұл мәліметтегі ақпараттық анықталмаушылық сегізге тең, ал шарикті тауып алу ықтималдылығы $\frac{1}{4}$ ге тең.

2-мысал. Кітап екі полканың (жоғарғы және төменгі) бірінде жатыр.

Кітаптың жоғары полкада жатқаны туралы хабарламада бір бит информация бар.

3-мысал. Кітап үш полканың (жоғарғы, ортаңғы және төменгі) бірінде жатыр. Кітап ортаңғы полкада жатыр деген хабарламада бір биттен үлкен информация бар.

1928 жылы американдық инженер Р.Хартли информацияны бағалайтын өзінің формуласын ұсынды. Бұл формула – Хартли формуласы деп аталады.

$$I = \log_2 N \quad (1)$$

Мұндағы N – тең ықтималдылықты оқиғалар саны, I хабарламадағы бит саны.

4-мысал. Кітап үш полканың бірінде жатыр деген информацияның мөлшері

$$I = \log_2 3 = 1,585 \text{ бит ақпарат.}$$

Хартли формуласын (1) басқаша да жазуға болады. N оқиғаның әрқайсысының ықтималдықтары бірдей болғандақтан былайша жазамыз: $P=1/N$, то $N=1/P$, демек

$$I = \log_2 N = \log_2 (1/P) = -\log_2 P \quad (2)$$

Енді N әр түрлі ықтималдықтағы оқиғалардың біреуінің ақпарат мөлшерін есептеп шығарайық. Әлдебір генератор өз экранында 1, 2, 3, ..., i , ... K сандарының кез келгенін шығарсын. Осы сандардың экранда пайда болу ықтималдықтары $P_1, P_2, P_3, \dots, P_i, \dots, P_k$ болсын. Көрнекілік үшін 1-кестені толтырамыз.

Кесте 1

1	2	3	...	i	...	$K-1$	K
P_1	P_2	P_3	\dots	P_i	\dots	P_{K-1}	P_K

Әрбір сан осы санның көріну ықтималдығына сәйкес экранға шығады. Экранды бақылаймыз: экранда N сан ($N >> k$) пайда болсын. Егер бізді і саны қызықтырса, онда ол экранда ($N P_i$) рет пайда болады. Әрбір і – інші санның экранда көрінуі ($-\log_2 P_i$) бит ақпарат береді. Ал барлық N сан көрінгеннен кейінгі жалпы ақпарат мөлшерін есептеу үшін әрбір санның көрінуіне сәйкес ақпарат мөлшерін есептеп алып, олардың қосындысын аламыз. Жалпы ақпарат

$$I = -N \sum_{i=1}^n P_i \log_2 (P_i) \quad (3)$$

Ал бір санға ғана сәйкес келетін *ортаса ақпарат мөлшері*:

$$I_{op} = - \sum_{i=1}^n P_i \log_2 (P_i) \quad (4)$$

Бұл теңдеуді 1948 жылы бірінші рет американдық математик К.Шеннон ұсынды. Сол себепті оны Шеннон теңдеуі деп те атайды.

Ақпарат теориясының элементтерін оқушының немесе студенттің білім алу параметрлерін өлшеуде қолдануға болады. Ақпараттық энтропия немесе ораша ақпарат теңдеуі көмегімен оқушылардың тест тапсырмасын орындау іс-әрекетінің нәтижелерін дәл есептеуге болады. Тестілеу процесі білімді бағалаудың статистикалық түріне жатады. Тестілеу көмегімен біз оқушының білім мөлшері туралы информацияның сандық сипаттамасын аламыз. Бұл ақпарат мынандай формулалармен анықталады:

$$I_i = - \ln P_i, \quad P_i = L_{+i} / L \quad (6)$$

Мұндағы P_i – i нөмірлі сұраққа берілген жауаптың ықтималдылығы, L_{+i} – дұрыс жауаптардың саны, L – тестке қатысқан оқушы саны. Іштималдық нөл және бір саны аралығында жатады, демек ақпарат нөл және шексіздік арасындағы он сандар бола алады.

Іштималдығы аз ($P_i \approx 0$) оқиғалардың (қыын сұраққа жауап беру) ақпараты үлкен болады. Егер $P_i \approx 1$ болса, яғни i нөмірлі сұраққа барлық оқушылар дұрыс жауап берсе, онда бұл сұрақ үшін ақпарат нөлге жуық. Демек $\ln 1 = 0$. Осы айтылғандардан мынандай қорытынды жасауға болады: ақпарат сұрақтардың салыстырмалы қындығының немесе сұрақтың құндылығының сандық сипаттамасы бола алады.

Осындай әдіспен анықталған ақпарат оқушы білімінің анықталуының элементар актісі болып табылады және тестің белгілі бір сұрағының құндылық коэффициентін анықтайды. Тест сұрақтарына берілген дұрыс жауаптың санына қарап оқушы білімінің нақты деңгейі туралы дәл баға беру мүмкін емес. Оқушының салыстырмалы білім деңгейі туралы дәл бағалау жүргізу

үшін ақпараттың орташа мәні немесе ақпараттық энтропия формуласын пайдалану керек.

$$S = - \sum_{i=1}^n P_i \ln P_i, \quad (7)$$

мұндағы i тестің дұрыс жауап берілген сұрақтарының нөміріне сәйкес келеді.

Энтропия ұғымы жалпы ғылымдық ұғым, ол күрделі жүйенің тепе-тендігін, жетілгендігін, анықталмағандық өлшемін сипаттайты. Тестен өткен оқушының басқа оқушымен салыстырғандағы білім деңгейі жоғары болуы үшін оның тест сұрақтары бойынша ақпараттық энтропиясы екінші оқушынің ізінен жоғары болуы керек. Яғни дұрыс жауап берілген қын сұрақтарының анықталмағандық өлшемдері (қындығы) жоғары болу керек.

Тест көмегімен конкурсның орындарды анықтаған кезде (мысалы, жоғары оқу орнына түсу және мемлекеттік қызметтің бос орындарына өту кезінде) тестен өтушіге қойылған барлық сұрақтардың құндылық коэффициенті есептеледі. Тестен өтушінің салыстырмалы білім деңгейін орташа ақпаратының мәні немесе ақпараттық энтропиясы аныктайды.

Мысал. Конкурстық тестілеудің нәтижелерін ақпараттық энтропия көмегімен өндөу керек.

2 – кестеде көрсетілген жеті тестен өтуші ($L = 7$) бес сұраққа жауап берген ($i = 5$). Кестенің он жағында Σ бағанында әрбір тестен өтуші үшін арифметикалық жолмен қосылған дұрыс жауапқа алған балдарының жиынтығы орналасқан. Бір дұрыс жауап – бір балл береді. Кестеден көргеніміздей балдардың қосындысына қарап қай тестен өтушінің білім деңгейі жоғары екенін дәл айта алмаймыз. Өйткені, бірнеше тестен өтуші әр түрлі сұраққа жауап берсе де, бірдей балл алып отыр ($L = 1, 3, 4, 7$ жағдайлар).

Осы жерде үлкен сұрақ туындаиды: қай тестен өтушінің білім деңгейі жоғары? Кестеден байқағанымыздай $i = 1, 4$ нөмірлі сұрақтар басқа сұрақтарға қарағанда салыстырмалы түрде қын болған. Яғни осы сұрақтарда ақпарат мөлшері жоғары, демек басқа сұрақтарға қарағанда құнды. Керісінше екінші сұрақтың тестен өтушілер үшін құндылығы нөлге тең. Қын сұраққа дұрыс жауап бере отырып, көп уақытын жұмсаған конкурсқа қатысушыларға қандай артықшылықтар жасалуы керек?

L_+ қатарында әрбір сұрақ бойынша дұрыс жауаптардың қосындысы орналасқан. P_i қатарында әрбір i нөмірлі сұрақтың ықтималдығы жазылған. Осы ықтималдықты табу үшін дұрыс жауаптар қосындысын (L_+) тестке

қатысушылар санына бөледі ($L = 7$), сөйтіп әрбір i – нөмірлі сұрақтың ықтималдығын (P_i) табады.

Осыдан кейін i – нөмірлі сұрақтың информациясын $I_i = -\ln P_i$ формуласы көмегімен есептейді. Әрбір i – нөмірлі сұрақ үшін P_i ді I_i ге көбейте отырып, алынған мәндерді келесі қатарға жазады. Осыдан кейін әрбір тестен өтуші үшін оның білімінің информациялық энтропиясын (S) табамыз да оны S бағанына жазамыз. Мысалы. Бірінші тестен өтушінің білімінің ақпараттық энтропиясы $S_1=0,3579 + 0 + 0,2403 = 0,5983$. Кімнің білімінің ақпараттық энтропиясы S жоғары болса, соның салыстырмалы білім деңгейі жоғары болады.

3 және 7 нөмірлі тестен өтушілердің информациялық энтропиялары бірдей болғандықтан олар 4 және 5 орындарды өзара бөлісу керек. Конкурстық орындардың қатысушылар арсындағы бөлінісі соңғы бағанда берілген.

Кесте 2 – Тестілеу нәтижесінде конкурстық орындарды анықтау

L/i	1	2	3	4	5	Сумма	S	
1	1	0	1	0	0	2	0,372466,5	
2	1	1	1	0	1	4	1,098721,2	
3	0	1	1	0	1	3	0,858384,3	
4	0	0	1	0	0	1	0,132129,7	
5	1	0	0	0	0	1	0,240337,6	
6	1	0	1	1	0	3	0,730398,4	
7	1	1	1	1	1	5	1,456654,1	
L+i	5	3	6	2	3			
Pi	0,714285	0,428571	0,857142	0,285714	0,428571			
Ii	0,336472	0,847297	0,154150	1,252762	0,847297			
Pi*Ii	0,240337	0,363127	0,132129	0,357932	0,363127			

Әдебиеттер

1. Кадомцев Б. Б. Динамика и информация. М.: Наука, 1997.

2. Мукушев Б.А., Мукушев С.Б., Турдина А.Б. Изучение основ теории информации в процессе обучения информатике // Информатика и образование. 2008. №5.
3. Райхерт Т.Н., Хеннер Е.К. Место теории информации в подготовке учителя информатики // Информатика и образование. 1999. №2.
4. Севастьянов Б. А. Курс теории вероятностей и математической статистики. – М.: - Наука, 1982.
5. Shaping Scientific Worldview of Schoolchildren by Including Synergetics into the Content of Education //Integration of Education. 2018; 22(4):632-647. DOI: 10.15507/1991-9468.093.022.201804.632-647 Scopus, Q3, процентиль 36, SJR 2018 0,27; CiteScore 2018 0,64 <http://edumag.mrsu.ru/index.php/ru/articles/82-18-4/697-10-15507-1991-9468-092-022-201804-3>