

Дәріс 6. Оптималды кодтар және Хаффман әдісі

Дәріскер: PhD Темирбекова Ж.Е.

Оптималды код деген не?

- ▶ • Орташа код ұзындығын минимизациялайтын код
- ▶ • Энтропияға жақын ұзындық
- ▶ • Бірмәнді декодталатын болуы тиіс
- ▶ • Префикстік код талап етіледі

Энтропия және теориялық шек

- ▶ • Энтропия — ақпараттағы белгісіздік өлшемі
- ▶ • $H(X)$ — минималды орташа код ұзындығы
- ▶ • Оптималды кодтар $L \approx H(X)$ болуға тырысады
- ▶ • Шеннон теоремасы негіз

Префикстік код қасиеті

- ▶ • Ешбір код басқа кодтың басы болмайды
- ▶ • Бірмәнді және қателіксіз декодтау
- ▶ • Хаффман, Шеннон-Фано әдістерінің негізі

Деректерді қысу түрлері

- ▶ • Жоғалтусыз қысу (lossless)
- ▶ • Жоғалтумен қысу (lossy)
- ▶ • Жоғалтусызда дерек толық қалпына келеді
- ▶ • Жоғалтумен — адам байқамайтын ақпарат жойылады

Жоғалтусыз қысу әдістері

- ▶ • Хаффман кодтау
- ▶ • Шеннон-Фано
- ▶ • Lempel-Ziv (LZ77, LZ78)
- ▶ • RLE (қайталануды есептеу)

Жоғалтумен қысу әдістері

- ▶ • JPEG (сурет)
- ▶ • MP3 (дыбыс)
- ▶ • MPEG (бейне)
- ▶ • Адам қабылдамайтын ақпаратты алып тастау

Хаффман әдісі

- ▶ • 1952 жылы Д. Хаффман ұсынған
- ▶ • Жиі кездесетін символ — қысқа код
- ▶ • Сирек кездесетін символ — ұзын код
- ▶ • Жоғалтусыз қысу алгоритмі

Хаффман ағашын құру

- ▶ 1. Символ жиіліктерін санау
- ▶ 2. Ең кіші екі символды біріктіру
- ▶ 3. Жаңа түйінді тізімге қосу
- ▶ 4. Бір түбір қалғанға дейін қайталау

Хаффман кодын тағайындау

- ▶ • Әр тармаққа 0 немесе 1 жазылады
- ▶ • Түбірден жапыраққа дейінгі жол — код
- ▶ • Жиі символдар ағашқа жақын орналасады
- ▶ • Префикстік код автоматты түрде пайда болады

Хаффман мысалы

- ▶ Мысал хабар: $A=2, B=2, C=1, D=1$
- ▶ • Ең аздары: $C + D \rightarrow$ біріктіру
- ▶ • А мен В кейінгі деңгейде орналасады
- ▶ • Мүмкін кодтар: $A=0, B=10, C=110, D=111$

Хаффман әдісінің артықшылықтары

- ▶ • Оптималды жоғалтусыз қысу
- ▶ • Префикстік — қате болдырмайды
- ▶ • Жылдам кодтау/декодтау
- ▶ • ZIP, JPEG, MP3 форматтарында қолданылады

Қолдану салалары

- ▶ • Файлдарды қысу (ZIP, GZIP)
- ▶ • Сурет, дыбыс, бейне кодтау
- ▶ • Архивтеу жүйелері
- ▶ • Телекоммуникация және деректер беру

Қорытынды

- Хаффман әдісі — деректерді жоғалтусыз қысудың ең тиімді және математикалық тұрғыдан дәлелденген тәсілдерінің бірі. Ол ықтималдыққа негізделген құрылымы арқасында хабардың артық ақпаратын азайтып, ақпаратты сақтау мен тасымалдауды оңтайландырады. Хаффман кодтау — қазіргі заманғы деректерді өңдеу мен компрессия жүйелерінің іргетасын қалаған классикалық әрі әмбебап әдіс болып табылады.