

Тақырыбы: Ассоциативті ережелерді өндіру: Apriori, FP-growth және Бизнеске Қолданылуы

Ассоциативті ережелерді өндіру (Association Rule Mining) — бұл деректерді өндірудің (Data Mining) маңызды бағыттарының бірі. Оның мақсаты — үлкен деректер жиынындағы айнымалылар арасындағы күшті байланыстарды, корреляцияларды немесе қатынастарды табу. Ең жиі қолданылатын мысал — нарықтық себетті талдау (Market Basket Analysis), мұнда клиенттердің бір транзакцияда қандай тауарларды бірге сатып алатыны анықталады.

Ассоциативті ереже келесідей түрде көрсетіледі:

$$A \Rightarrow B$$

Бұл "Егер A болса, онда B болуы мүмкін" дегенді білдіреді.

- A (Антецедент - Antecedent): Ереженің сол жағындағы элемент немесе элементтер жиынтығы (мысалы, нан).
- B (Консеквент - Consequent): Ереженің оң жағындағы элемент немесе элементтер жиынтығы (мысалы, сүт).
- A және B — бір-біріне тәуелсіз, бірақ бір транзакцияға жататын элементтер жиынтығы.

Ассоциативті ереженің негізгі өлшемдері

Табылған ережелердің қаншалықты маңызды және пайдалы екенін бағалау үшін үш негізгі метрика қолданылады:

1. Қолдау (Support - S): Барлық транзакциялардың ішінде ереженің сол және оң жағындағы элементтердің $(A \cup B)$ бірге кездесу жиілігін көрсетеді. Бұл ереженің жалпы танымалдылығын білдіреді.

$$S(A \Rightarrow B) = \frac{\text{A және B бірге кездесетін транзакциялар саны}}{\text{Барлық транзакциялар саны}}$$

2. Сенімділік (Confidence - C): A элементін қамтитын транзакциялардың ішінде B элементін де қамтитын транзакциялардың пайыздық үлесін көрсетеді. Бұл ереженің дұрыстығын өлшейді.

$$C(A \Rightarrow B) = \frac{\text{A және B бірге кездесетін транзакциялар саны}}{\text{A кездесетін транзакциялар саны}} = \frac{S(A \cup B)}{S(A)}$$

3. Лифт (Lift - L): Ереженің кездейсоқ жағдайдан қаншалықты артық немесе кем екенін көрсетеді. Егер A мен B тәуелсіз болса, онда $Lift = 1$ болады.

$$L(A \Rightarrow B) = \frac{C(A \Rightarrow B)}{S(B)} = \frac{S(A \cup B)}{S(A) \cdot S(B)}$$

- $Lift > 1$: A мен B арасында оң корреляция бар (бірін-бірі ынталандырады).

- $Lift < 1$: А мен В арасында теріс корреляция бар (бірін-бірі болдырмайды).

Ассоциативті ережелерді өндіру алгоритмдері

Ассоциативті ережелерді өндіру екі кезеңнен тұрады:

1. Жиі кездесетін элементтер жиынтығын (Frequent Itemsets) табу: Минималды қолдау табылмайды (Min_Support) жоғары жиілікпен кездесетін элементтер жиынтықтарын анықтау.
2. Ережелерді құру: Табылған жиі кездесетін элементтер жиынтығынан минималды сенімділік табылмайды (Min_Confidence) жоғары ассоциативті ережелерді құру.

Apriori Алгоритмі

Apriori — жиі кездесетін элементтер жиынтығын табуға арналған классикалық және ең көп қолданылатын алгоритм. Ол элементтер жиынтығын біртіндеп, төменгі деңгейден жоғарғы деңгейге қарай кеңейту арқылы жұмыс істейді.

Apriori принципі (Apriori қасиеті):

Егер элементтер жиынтығы жиі кездесетін болса, онда оның кез келген ішкі жиынтығы да (subset) жиі кездесетін болады.

Немесе контрапозитивті түрде:

Егер элементтер жиынтығы сирек кездесетін болса, онда оны қамтитын кез келген үлкен жиынтық та сирек кездесетін болады.

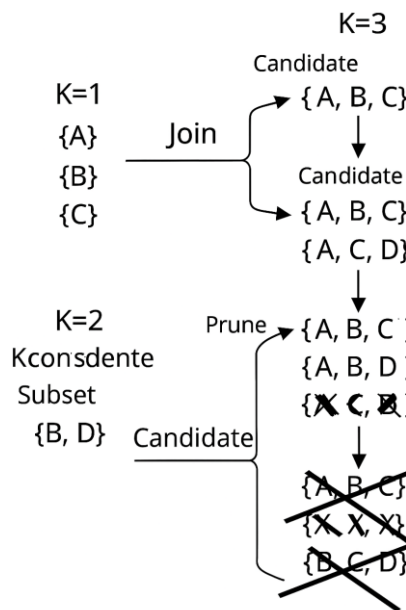
Бұл қасиет алгоритмнің тиімділігін арттырады, себебі сирек кездесетін жиынтықтарды тексеру қажеттілігін жояды.

Apriori қадамдары:

1. $K=1$ жиынтықтарды сканерлеу: Әрбір жеке элементтің қолдауын есептеп, минималды қолдаудан төмендерін алып тастау.
2. Генерациялау: ($K-1$)-ші деңгейдегі жиі кездесетін элементтерді біріктіру арқылы K -шы деңгейдегі ықтимал кандидат жиынтықтарды генерациялау.
3. Қасиетті қолдану (Pruning): Егер K -шы деңгейдегі кандидат жиынтықтың кез келген ішкі жиынтығы ($K-1$ деңгейінен) жиі кездесетін элементтер жиынтығына кірмесе, онда бұл кандидатты алып тастау.
4. Сканерлеу: Қалған кандидаттардың қолдауын есептеп, олардың ішіндегі жиі кездесетіндерін анықтау.
5. Қайталау: Бұл процесс жаңа жиі кездесетін элементтер табылмайынша жалғасады.

Сурет 7.1: Apriori алгоритмінің логикасы (Pruning)

Frequent Itemset Generation



K=3 кандидаттардың қалай генерацияланып, K-2-дегі сирек жиынтықтың ішкі жиынтығы болғандықтан қалай жойылатынын көрсететін мысал.

FP-growth (Frequent Pattern Growth) Алгоритмі

FP-growth (Жиі кездесетін үлгіні өсіру) — бұл Apriori-ге қарағанда едәуір тиімді алгоритм. Ол Apriori сияқты кандидаттарды генерациялау және қайталанған сканерлеуді қажет етпейді.

FP-growth принципі:

1. FP-Tree құру: Алгоритм бүкіл транзакциялық деректер базасын FP-Tree деп аталатын арнайы ағаш құрылымына сығады (compress). Элементтердің жиілігі бойынша ағаштағы түйіндердің орналасу реті анықталады.
2. Рекурсивті өндіру: FP-Tree-ні соңынан (сирек элементтерден) бастап, шартты (conditional) базаларды құру арқылы рекурсивті түрде бөлшектейді. Әрбір шартты базадан кішігірім шартты FP-Tree құрылады, және бұл процесс бір түйін қалғанша немесе минималды қолдау табалдырығына жеткенше жалғасады.

Артықшылығы:

- Деректер базасын тек екі рет сканерлейді (біріншісі жиілікті есептеу үшін, екіншісі FP-Tree құру үшін).

- Кандидат жиынтықтарды генерациялау қажет емес, бұл оны Apriori-ге қарағанда әсіресе тығыз (dense) деректер жиынтықтарында әлдеқайда жылдам етеді.

Ассоциативті ережелер бизнес саласында

Ассоциативті ережелерді өндірудің практикалық құндылығы оның бизнестің әртүрлі аспектілерінде қолданылуында.

Нарықтық себетті талдау (Market Basket Analysis - MBA)

MBA — ассоциативті ережелердің ең негізгі қолданылуы.

- Мысал ереже: $\{\text{Нан, Жұмыртқа}\} \Rightarrow \{\text{Сүт}\} (C = 80\%, L = 1.5)$
- Бизнес-шешім: Егер клиент нан мен жұмыртқаны алса, оның сүт сатып алу ықтималдығы 80%. Бұл ереже супермаркетке осы үш тауарды бір-біріне жақын орналастыруға немесе екі тауарды алған клиентке сүтке жеңілдік ұсынуға мүмкіндік береді.

Қолданудың басқа салалары

Сала	Қолданылуы	Шешімдердің түрі
Ритейл және E-commerce	Өнімді орналастыру, Кросс-сату (Cross-selling)	Онлайн дүкенде "Бұл тауарды сатып алған клиенттер..." ұсыныстарын қалыптастыру.
Денсаулық сақтау	Медициналық диагностика	Белгілі бір симптомдар жиынтығының (A) белгілі бір ауруға (B) әкелу ықтималдығын табу.
Телекоммуникация	Қызметтерді топтастыру	Клиенттердің A қызметін пайдаланғанда B қызметін де қосу ықтималдығын анықтау.
Веб-аналитика	Веб-сайттағы навигациялық жолдар	Пайдаланушы A бетінен B бетіне өткеннен кейін C бетіне өту ықтималдығын табу, сайт құрылымын оңтайландыру.

Шешімдер қабылдаудағы рөлі

Ассоциативті ережелер менеджерлерге келесі стратегияларды әзірлеуге көмектеседі:

- Инвентаризацияны басқару: Бірге жиі сатылатын тауарлардың қорларын толықтыруды қамтамасыз ету.

- Баға саясаты: Бірге сатылатын тауарлардың бірін арзан бағамен ұсынып, екіншісінің бағасын көтеру арқылы жалпы пайданы арттыру.
- Акциялар мен ұсыныстар: Клиенттің себетіндегі тауарларға сәйкес келетін қосымша ұсыныстарды (up-selling, cross-selling) жекелендіру.

Қорытынды

Ассоциативті ережелерді өндіру — деректерді өндірудегі ең тиімді және практикалық қолданысы бар әдістердің бірі. Apriori және FP-growth алгоритмдері үлкен транзакциялық деректер базаларынан құнды заңдылықтарды тиімді түрде тауып, бизнестің шынайы әлемдегі стратегиялық шешімдерін қабылдауға көмектеседі. Қолдау, Сенімділік және Лифт сияқты метрикалар табылған ережелердің нақты іскери құндылығын бағалауға мүмкіндік береді.

Қолданылған әдебиеттер тізімі

1. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. Elsevier. (Apriori және FP-growth алгоритмдерінің егжей-тегжейлі сипаттамасы)

2. Agrawal, R., Imieliński, T., & Swami, A. (1993). *Mining Association Rules between Sets of Items in Large Databases*. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. (Apriori алгоритмінің бастапқы ұсынысы)
3. Han, J., Pei, J., & Yin, Y. (2000). *Mining Frequent Patterns without Candidate Generation*. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. (FP-growth алгоритмін ұсынған мақала)
4. Tan, P. N., Steinbach, M., & Kumar, V. (2023). *Introduction to Data Mining*. Pearson. (Ассоциативті ережелердің метрикалары мен интерпретациясы)
5. Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer. (Әртүрлі ассоциативті ереже алгоритмдерін салыстыру)