연관규칙&순차패턴분석

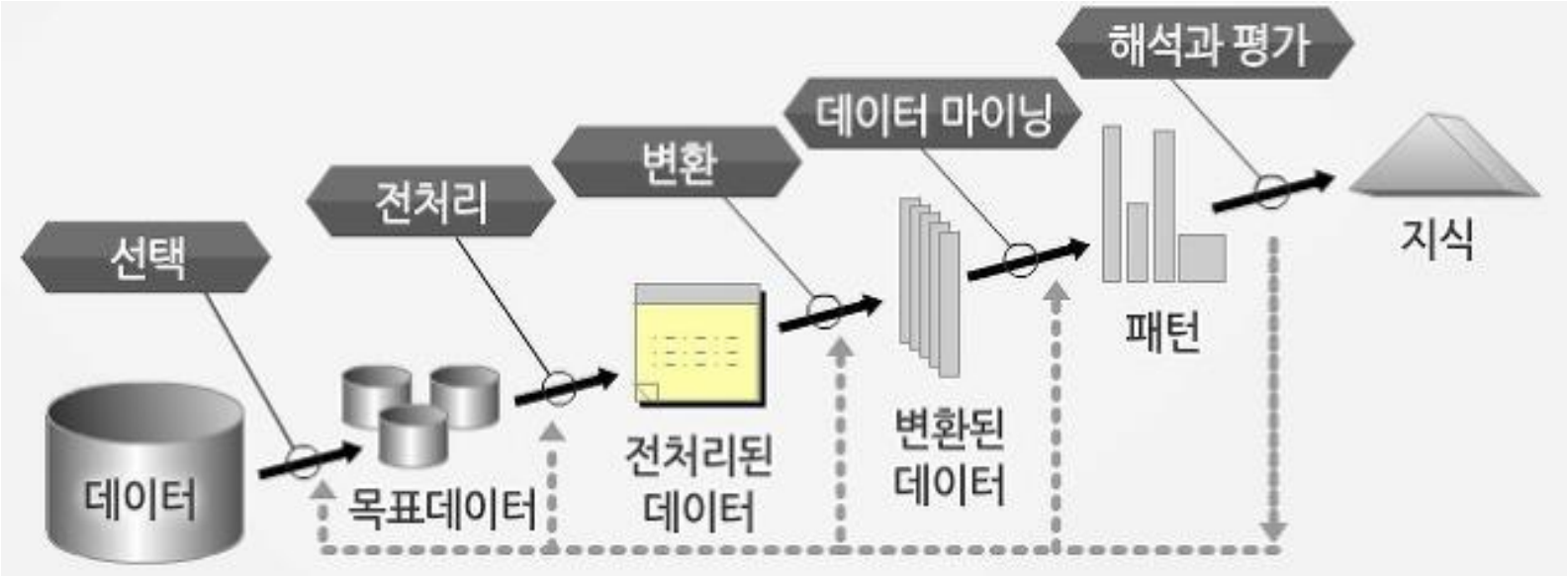
- 권수태 교수

JEONJU UNIVERSITY
The Place for Superstars

1. 연관규칙

❖ 데이터마이닝

➢ 복잡한 통계적인 분석이나 모형구축 기법을 사용하여 대용량의 데이터 내에 패턴이나 규칙 등을 탐색하고 모형화함으로써 유용한 지식을 추출하는 일련의 과정



1. 연관규칙

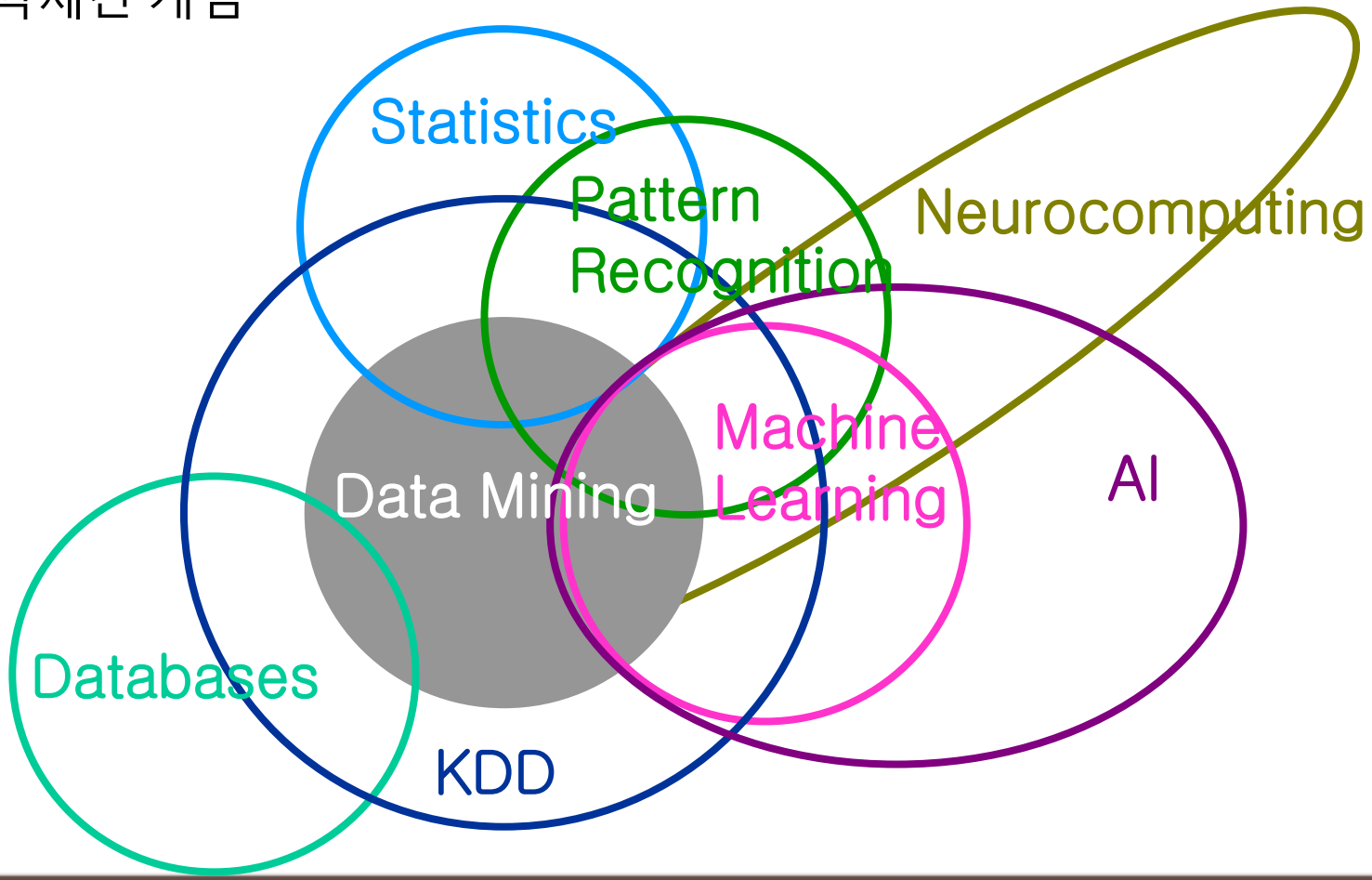
❖ 데이터마이닝

- ➢ 지식발견
  - ✓ 데이터로 부터 유용한 정보를 추출하는 프로세스의 전 과정
- ➢ 지식발견프로세스
  - ✓ 데이터의 선택 (selection)
  - ✓ 데이터의 정제 (cleaning)
  - ✓ 데이터의 보완 (enrichment)
  - ✓ 데이터의 변환 (transformation)
  - ✓ 데이터마이닝 기법 선택 및 적용 (selection & application)
  - ✓ 모형의 평가 (model evaluation)

1. 연관규칙

❖ 데이터마이닝

➢ 다학제간 개념



JEONJU UNIVERSITY

1. 연관규칙

❖ 데이터마이닝

➢ 기본방법론

- ✓ 회귀분석(regression analysis): 선형회귀모형, 변수선택, 로지스틱 회귀
- ✓ 모형평가(model assessment): 모형들의 예측성능을 평가하는 방법
- ✓ 의사결정나무(decision trees)
- ✓ 신경망(neural networks)
- ✓ 기타 지도학습기법: 단순 베이즈 분류(naive Bayes classifier), k -근접 분류, SVM
- ✓ 차원축소기법: 주성분분석(principal component analysis), 인자분석(factor analysis), 다차원 척도법(multidimensional scaling)
- ✓ 연관규칙분석(association rule analysis)
- ✓ 군집분석(cluster analysis): k -평균군집법(k -means clustering), 계층적 군집법 (hierarchical clustering)

1. 연관규칙

❖ 데이터마이닝

➢ 적용사례

✓ 소매/유통업

- Wall Mart에서 매장내의 상품들과 고객들의 구매패턴의 연관성을 발견

✓ 신용카드회사

- 카드의 부정사용 방지를 통하여 고객의 자산 보호 및 회사의 손해액 감소

✓ 의료분야

- 과거의 환자들에 대해서 종양검사의 결과를 근거로(즉, 종양의 크기, 모양, 색깔 등) 종양의 악성/양성 여부를 구별하는 분류모형을 만든 후, 새로운 환자에서 얻은 입력변수를 이용하여 암을 진단

✓ 제조업

- 반도체회사에서 불량품 자동검색장치 개발

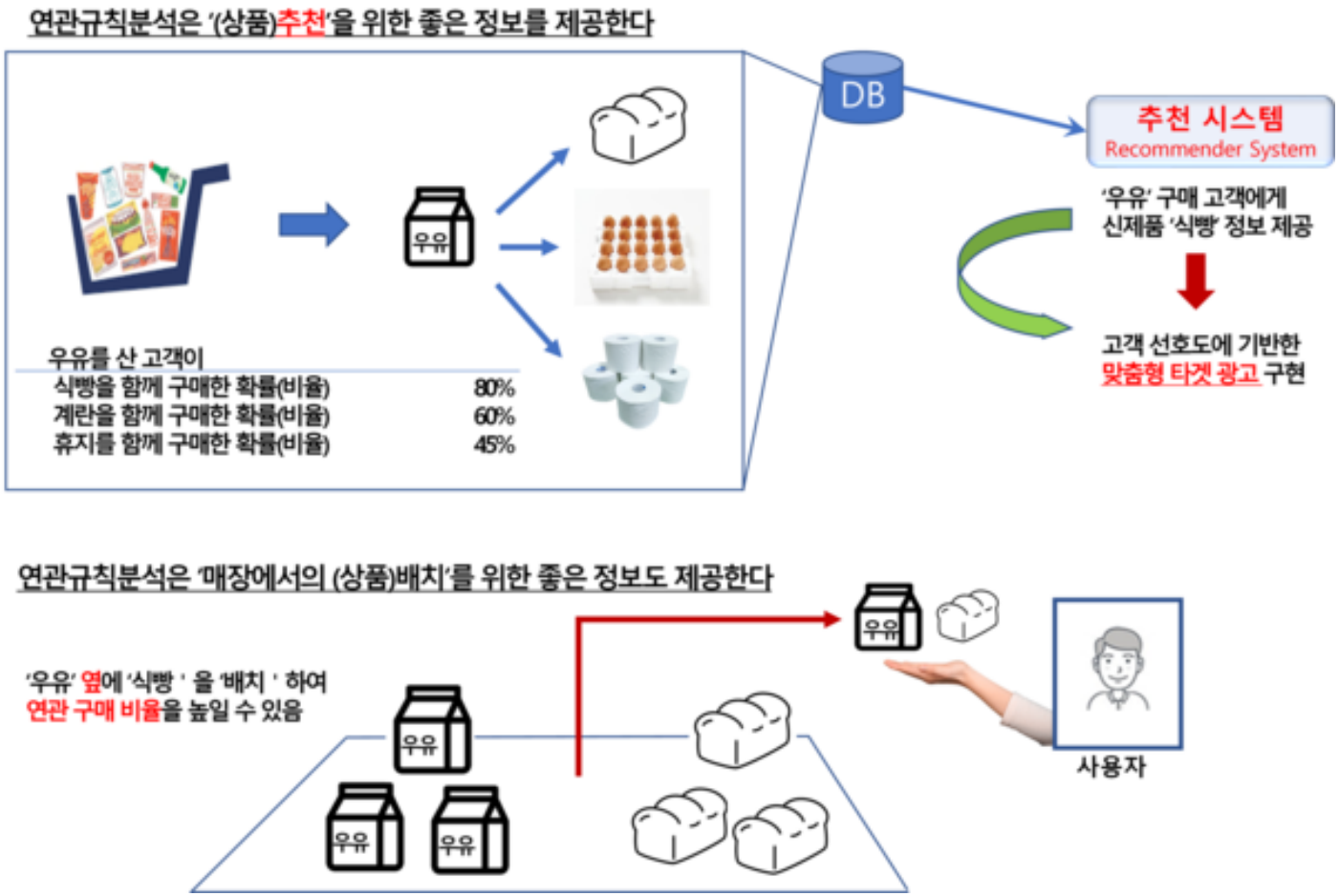
1. 연관규칙

❖연관 분석(Association Rule)

- 연관규칙이란 어떤 사건이 일어나면 다른 사건이 일어나는 관련성을 의미
- 대규모 데이터세트에서 항목간 관련성을 파악하는 탐색적 데이터분석기법
- 대형 데이터베이스에서 변수 간의 흥미로운 관계를 발견하기 위한 규칙-기반 기계 학습 방법(장바구니 분석)
- 대량의 트랜잭션 정보(예: 고객의 쇼핑 이력)로부터 개별 데이터(변수) 사이에서 연관규칙(x면 y가 발생)을 찾는 것

1. 연관규칙

❖연관 분석



JEONJU UNIVERSITY

1. 연관규칙

❖연관 분석

- 1990년도 초반에 IBM의 라케시 아그라왈(Rakesh Agrawal)이 영국의 Marks & Spencer 마켓에서 CRM(고객 관계 관리, Customer Relationship Management) 시스템을 바탕으로 소비자 분석을 실시
  - ✓ '기저귀'와 '맥주'의 구매 여부에 대한 상관관계를 분석
  - ✓ '기저귀를 산 사람 중 40%'가 맥주를 구매했다는 사실에 주목하고 기저귀를 산 사람에게 맥주를 '추천'하면?
  - ✓ 매출증가

1. 연관규칙

❖연관 분석

- 좋은 규칙을 판단하는 세가지 지표
- 신뢰도(Confidence)
  - ✓ 항목 A를 포함한 거래 중에서 항목 A와 항목 B가 같이 포함될 확률

$$\frac{P(A \cap B)}{P(A)} = \frac{A\text{와 } B\text{가 동시에 포함된 거래 수}}{A\text{를 포함하는 거래 수}}$$

  - ✓ 규칙에 대한 확실성을 나타내는 지표
- 지지도(Support)
  - ✓ 전체 거래 중 항목 A와 B를 동시에 포함하는 거래의 비율

$$P(A \cap B) = \frac{A\text{와 } B\text{가 동시에 포함된 거래 수}}{\text{전체 거래 수}}$$

  - ✓ 규칙에 대한 유용성을 나타내는 지표

1. 연관규칙

❖연관 분석

➢향상도(Lift)

- ✓A가 주어지지 않은 상태에서 B의 확률에 대하여 A가 주어졌을 때 B의 확률 증가비율
- ✓A에 대해 B가 등장 가능성이 높은지 C가 등장 가능성이 더 높은지를 확인하는 지표

$$\frac{P(B|A)}{P(B)} = \frac{A\text{를 포함하는 거래 중 } B\text{를 포함하는 거래의 수(지지도)}}{\text{전체 거래 중 } B\text{를 포함하는 거래 수(신뢰도)}}$$

- ✓향상도(lift) 값이 1이면 서로 독립적인 관계이며 1보다 크면 두 품목이 서로 양의 상관관계, 1보다 작으면 두 품목이 서로 음의 상관관계

1. 연관규칙

❖연관 분석

➢ 장점

- ✓ 연관 분석은 조건 결과의 빈도수를 기반으로 표현되기 때문에 비교적 결과를 쉽게 이해할 수 있음
- ✓ 구매내역의 자료 구조를 가지기 때문에 특별한 전처리 과정을 필요로 하지 않음

➢ 단점

- ✓ 품목의 개수가 늘어남에 따라 분석에 필요한 계산의 수가 기하급수적으로 증가

1. 연관규칙

❖연관 분석

➢ 어떤 규칙의 지지도가 10%라면 그 의미는 전체트랜잭션 중에서 그 규칙을 따르고 있는 트랜잭션이 10%를 차지한다는 것을 의미

구매번호	구매항목
1	{라면, 오렌지쥬스, 커피}
2	{라면, 소시지}
3	{라면, 커피}
4	{오렌지쥬스, 비누, 샴푸}

'{라면} → {커피}'라는 연관 규칙은 '라면을 산 사람은 커피도 같이 산다'는 의미
네 가지 트랜잭션 중 1번과 3번 소비자가 구매한 물건들에 들어 있는 규칙이므로
지지도는 50%
신뢰도는 규칙의 왼쪽에 있는것을 산 사람들 중에서 오른쪽에 있는 물건들을 모두 산 사람들의 퍼센트
라면을 산 사람들은 세 사람인데 그 중에서 커피를 산 사람은 두사람이므로 이 규칙의 신뢰도는 66.7%

1. 연관규칙

❖연관 분석

➢ Apriori 알고리즘(1994)

- ✓ 후보항목집합을 구성한 후 사전지식(priori knowledge)을 이용하여 빈발항목집합을 생성하는 방법
- ✓ Step 1: k개의 아이템을 가지고 단일항목집단을 생성
- ✓ Step 2: 단일항목집단에서 지지도 계산 후 최소 지지도 값(minimum support) 이상의 항목만 선택
- ✓ Step 3: Step 2에서 선택된 항목 만을 대상으로 2개항목집단을 생성
- ✓ Step 4: 2개항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택(가지치기)
- ✓ Step 5: 위의 과정을 k개의 k-item frequent set을 생성할 때까지 반복

1. 연관규칙

❖연관 분석

➢ Apriori 알고리즘

데이터베이스 D

구매 번호	구매 항목
1	{1, 3, 4}
2	{2, 3, 5}
3	{1, 2, 3, 5}
4	{2, 5}

D를 스캔함 →

C1

항목 집합	지지도(%)
{1}	50%
{2}	75%
{3}	75%
{4}	25%
{5}	75%

F1

항목 집합	지지도(%)
{2}	75
{3}	75
{5}	75

C2

항목 집합
{2, 3}
{2, 5}
{3, 5}

D를 스캔함 →

C2

항목 집합	지지도(%)
{2, 3}	50
{2, 5}	75
{3, 5}	50

F2

항목 집합	지지도(%)
{2, 5}	75

1. 연관규칙

❖연관 분석

➢ Apriori 알고리즘

1단계

1.{coke,bread,hamburger}
2.{coke,hamburger,juice}
3.{milk,sandwich,juice}
4.{sandwich,milk,juice,bread}
5.{hamburger,juice,coke}
6.{coke,bread,hamburger}
7.{coke,hamburger,juice}
8.{hamburger,juice}
9.{milk,hamburger,sweater}
10.{coke,milk,juice}
11.{coke,juice}
12.{coke,sweater}

단일항목
{coke} : 8
~~{bread} : 3~~
{hamburger} : 7
{juice} : 8
~~{milk} : 4~~
~~{sandwich} : 2~~
~~{sweater} : 2~~

다수항목(2)
{coke,hamburger} : 5
{coke,juice} : 5
~~{hamburger,juice} : 4~~

다수항목(3)
~~{coke,hamburger,juice} : 2~~

최소 지지도 :40%
0.4 * 12 = 4.8

~~{coke} → {hamburger} : 62.5%~~
{hamburger} → {coke} : 71%
~~{coke} → {juice} : 62.5%~~
~~{juice} → {coke} : 62.5%~~

최소 신뢰도 : 70%

신뢰도(%)
= P(결론부)/P(조건부) *100

1. 연관규칙

❖연관 분석

conda install -c conda-forge mlxtend

➢ Apriori 알고리즘

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori

dataset=[[ ' 라면', ' 오렌지 쥬스 ' , ' 커피’], [ ' 라면 ' , ' 소시지'],
[ ' 라면 ' , ' 커피’], [ ' 오렌지 쥬스 ' , ' 비누 ' , ' 샴푸’]]

te = TransactionEncoder()
te_array = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_array, columns=te.columns_) #데이터프레임으로 변경

#가나다 순으로 Column값을 생성 데이터가 있으면 True로 없으면 False로 표시
frequent_itemsets = apriori(df. min_support=0.5, use_colnames=True)
frequent_itemsets
```

	support	itemsets
0	0.75	(라면)
1	0.50	(오렌지 쥬스)
2	0.50	(커피)
3	0.50	(라면, 커피)

1. 연관규칙

❖연관 분석

➢ Apriori 알고리즘

```
from mlxtend.frequent_patterns import association_rules
association_rules(frequent_itemsets, metric="confidence", min_threshold=0.3)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(라면)	(커피)	0.75	0.50	0.5	0.666667	1.333333	0.125	1.5
1	(커피)	(라면)	0.50	0.75	0.5	1.000000	1.333333	0.125	inf

1. 연관규칙

❖연관 분석

- FP-Growth 알고리즘(Frequent Pattern 알고리즘)
  - ✓ FP-Tree를 생성한 후에 최소 지지도 이상의 패턴만을 추출
  - ✓ Step 1: 모든 거래를 확인해 각 아이템마다의 지지도를 계산하고 최소 지지도 이상의 아이템만 선택
  - ✓ Step 2: 모든 거래에서 빈도가 높은 아이템 순서대로 순서를 정렬
  - ✓ Step 3: 부모 노드를 중심으로 거래를 자식 노드로 추가해주면서 tree를 생성
  - ✓ Step 4: 새로운 아이템이 나올 경우에는 부모 노드부터 시작하고, 그렇지 않으면 기존의 노드에서 확장
  - ✓ Step 5: 위의 과정을 모든 거래에 대해 반복하여 FP-Tree를 만들고 최소 지지도 이상의 패턴만을 추출

1. 연관규칙

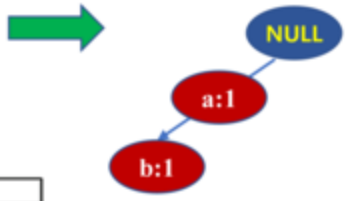
❖연관 분석

➢ FP-Growth 알고리즘

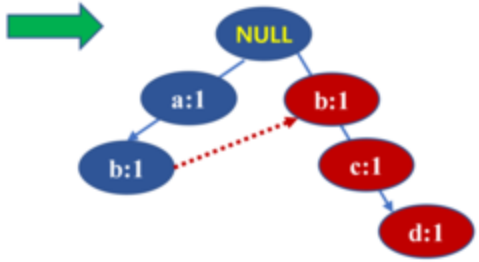
*TID: Transaction ID

TID	상품 품목들
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}

------>	노드 링크 포인터
———	부모 링크 포인터
상품:k	품목명:Support 값



[1] TID=1에서의 정보 조회 및 FP-Tree 생성



[2] TID=2에서의 정보 조회 및 FP-Tree 생성

[3] TID=3에서의 정보 조회 및 FP-Tree 생성

1. 연관규칙

❖연관 분석

➢ FP-Growth 알고리즘

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
dataset=[[ ' 라면', ' 오렌지 쥬스 ' , ' 커피’], [ ' 라면 ' , ' 소시지'],
[ ' 라면 ' , ' 커피’], [ ' 오렌지 쥬스 ' , ' 비누 ' , ' 샴푸’]]
te = TransactionEncoder()
te_array = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_array, columns=te.columns_) #데이터프레임으로 변경
#가나다 순으로 Column값을 생성 데이터가 있으면 True로 없으면 False로 표시
from mlxtend.frequent_patterns import fpgrowth
fp_g=fpgrowth(df, min_support=0.5, use_colnames=True)

from mlxtend.frequent_patterns import association_rules
association_rules(fp_g, metric="confidence", min_threshold=0.3)
```


2. 순차패턴분석

❖순차패턴 분석

- 데이터에 공통으로 나타나는 순차적인 패턴을 찾아내는 것
- 순서가 있는 시퀀스 데이터 베이스에서 순서를 고려하여 빈번하게 나타나는 패턴(pattern)을 찾아주는 기술
- 시퀀스 데이터(categorical sequences)
  - ✓ 변수가 순서 의미를 갖는 데이터
  - ✓ 관측치 형태: event, item, element
  - ✓ 어떤 패턴이 빈번하면 그 패턴의 일부분도 모두 빈번

2. 순차패턴분석

❖순차패턴 분석

- ➢ 적용사례
  - ✓ 소비자 구매형태
    - 가전제품: 냉장고, TV, 세탁기, 컴퓨터, 김치냉장고, 청소기, 건조기
    - 가구: 침대, 옷장, 식탁, 화장대, 쇼파, 가구
  - ✓ 의료 치료
  - ✓ 자연재해 : 지진, 이상기후
  - ✓ DNA 시퀀스와 유전자 구조
  - ✓ 웹콘텐츠 탐색 :
  - ✓ 여행 : 국가, 여행지

2. 순차패턴분석

❖순차패턴 분석

- ➢ 트랜잭션
  - ✓ 같은 시간대에 일어나는 사건들이나 한번에 구매한 물품 : {i1, i2, i3,…}
- ➢ 시퀀스
  - ✓ 트랜잭션의 시간적 순서: ordered list 로 표현 (t1, t2, t3, …)
- ➢ 예제
  - ✓ 고객 A는 처음에는 담배와 술을, 다음날에는 담배와 신문을, 그 다음날에는 음료수와 과자를 구매
  - ✓ 트랜잭션 : {담배, 술}, {담배, 신문}, {음료수, 과자}
  - ✓ 시퀀스: ({담배, 술}, {담배, 신문}, {음료수, 과자})

✓ 모든 사용자 시퀀스 중 몇 % 이상 공통으로 나타내는 시퀀스를 찾는 것

2. 순차패턴분석

➢ 관련연구

- ✓ Concept introduction and an initial Apriori-like algorithm(Agrawal & Srikant. Mining sequential patterns, 1995)
- ✓ Apriori-based method: GSP (Generalized Sequential Patterns: Srikant & Agrawal 1996)
- ✓ Pattern-growth methods: FreeSpan & PrefixSpan (Han et al. 2000; Pei, et al. 2001)
- ✓ Vertical format-based mining: SPADE (Zaki 2000)
- ✓ Constraint-based sequential pattern mining (SPIRIT: Garofalakis, Rastogi, Shim 1999]; Pei, Han, Wang 2002)
- ✓ Mining closed sequential patterns: CloSpan (Yan, Han & Afshar 2003)

2. 순차패턴분석

- IBM 연구소에서 개발한 GSP(Generalized Sequential Pattern) 알고리즘과 캐나다의 Simon Fraser 대학에서 개발한 PrefixSpan 알고리즘
- GSP(Generalized Sequential Pattern) 알고리즘
  - ✓ 사용자가정해준 지지도보다 많은 사용자 시퀀스를 포함하고 있는 패턴을 빈번한 패턴(frequent pattern)
  - ✓ 크기가 1인 후보집합부터 시작해 빈번한 집합을 찾아가는 방식
  - ✓ 지지도를 기반으로 길이가 가장 긴 빈번한 순차 패턴 탐색
  - ✓ F_k: 빈번한패턴들의 집합
  - ✓ C_k: 후보 패턴, 빈번한패턴이 될 가능성이 있는 패턴
  - ✓ 후보패턴들의 실제 등장빈도를 데이터베이스에서 카운트해 빈번한 패턴을 찾는 것
  - ✓ 즉, C1→F1→C2→F2→C3→F3→…의 순서로 점점 길이가 긴 빈번한 패턴을 찾아가는 것

2. 순차패턴분석

➢ GSP(Generalized Sequential Pattern) 알고리즘

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})

2. 순차패턴분석

➢ GSP(Generalized Sequential Pattern) 알고리즘

최소 지지도 = 100%

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})

C_1	지지도
맥주	4
땅콩	3
오징어	4
신문	4
소주	1
양주	1

JEONJU UNIVERSITY

2. 순차패턴분석

➢ GSP(Generalized Sequential Pattern) 알고리즘

최소 지지도 = 100%

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})

C_2	
다른시점	한시점
맥주 -> 맥주	(맥주, 오징어)
맥주 -> 오징어	(맥주,신문)
맥주 -> 신문	(신문, 오징어)
오징어 -> 맥주	
오징어 -> 오징어	
오징어 -> 신문	
신문 -> 맥주	
신문 -> 오징어	
신문 -> 신문	

F_1
맥주
오징어
신문

F_1	맥주	오징어	신문
맥주	*	*	*
오징어	*	*	*
신문	*	*	*

2. 순차패턴분석

➢ GSP(Generalized Sequential Pattern) 알고리즘

최소 지지도 = 100%

C_2	
다른시점	한시점
맥주 -> 맥주	(맥주, 오징어)
맥주 -> 오징어	(맥주,신문)
맥주 -> 신문	(오징어,신문)
오징어 -> 맥주	
오징어 -> 오징어	
오징어 -> 신문	
신문 -> 맥주	
신문 -> 오징어	
신문 -> 신문	

F_2	지지도
맥주 -> 맥주	1
맥주 -> 오징어	3
맥주 -> 신문	4
오징어 -> 맥주	0
오징어 -> 오징어	2
오징어 -> 신문	4
신문 -> 맥주	2
신문 -> 오징어	3
신문 -> 신문	3
(맥주, 오징어)	2
(맥주,신문)	1
(신문, 오징어)	1

F_2	지지도
맥주 -> 신문	4
오징어 -> 신문	4

2. 순차패턴분석

➢ GSP(Generalized Sequential Pattern) 알고리즘 최소 지지도 = 100%

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})

패턴길이	후보패턴	빈번한패턴
1	({맥주}), ({신문}), ({오징어}), ({땅콩}), ({소주}), ({양주})	({맥주}), ({신문}), ({오징어})
2	({맥주}, {맥주}), ({맥주, 신문}), ({맥주}, {신문}), ({맥주, 오징어}), ({맥주}, {오징어}), ({신문}, {맥주}), ({신문}, {신문}), ({신문, 오징어}), ({신문}, {오징어}), ({오징어}, {맥주}), ({오징어}, {신문}), ({오징어}, {오징어})	({맥주}, {신문}), ({오징어}, {신문})
3	없음	없음

2. 순차패턴분석

➢ PrefixSpan 알고리즘

✓ GSP 방법이 후보 패턴을 만들고, 그 후보 패턴이 데이터베이스에 몇 번 나오는가 세느라 시간이 걸리는 단점을 없애기 위해, 후보 패턴을 만들지 않으면서 빈번한 패턴을 찾는 방법

✓ PrefixSpan 트리를 만들어 가면서 빈번한 패턴을 찾게 됨

- 루트 노드에서 시작해 깊이 우선 검색($Depth\ First\ Search$) 순서로노드를 확장해 가면서 트리를 만듬
- 노드를 확장할 때에는 노드확장이외에 $projected$ DB도 만들게 됨
- projected DB란 전체데이터베이스의 사용자 시퀀스 중에서 그 노드가 나타내는 빈번한 시퀀스를 포함하고 있는 사용자 시퀀스만을 모은 후 노드가 나타내는 시퀀스 이후 부분만을 저장해 놓은 데이터베이스

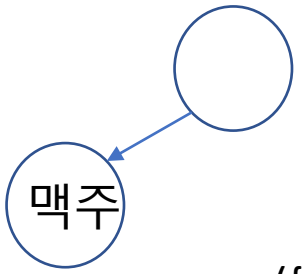
2. 순차패턴분석

C_1	지지도
맥주	4
땅콩	3
오징어	4
신문	4
소주	1
양주	1

➢ PrefixSpan 알고리즘

✓ Prefix를 먼저 할 대상을 찾는 기준: 지지도

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})



({맥주})-projected DB

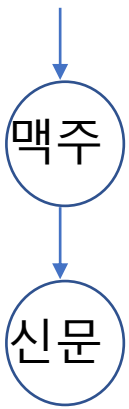
고객번호	구매기록
1	({_땅콩}, {맥주, 오징어}, {신문})
2	({신문, 땅콩})
3	({_신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({_오징어}, {신문, 양주}, {신문, 오징어})

2. 순차패턴분석

- PrefixSpan 알고리즘
  - ✓Prefix를 먼저 할 대상을 찾는 기준: 지지도
  - ✓각 노드의 projected-DB에서 Fk를 탐색

C_1	지지도
맥주	4
땅콩	3
오징어	4
신문	4
소주	1
양주	1

고객번호	구매기록
1	({_땅콩}, {맥주, 오징어}, {신문})
2	({신문, 땅콩})
3	({_신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({_오징어}, {신문, 양주}, {신문, 오징어})



({맥주},{{신문})-projected DB

고객번호	구매기록
1	Null
2	({_, 땅콩})
3	Null
4	({_, 양주}, {신문, 오징어})

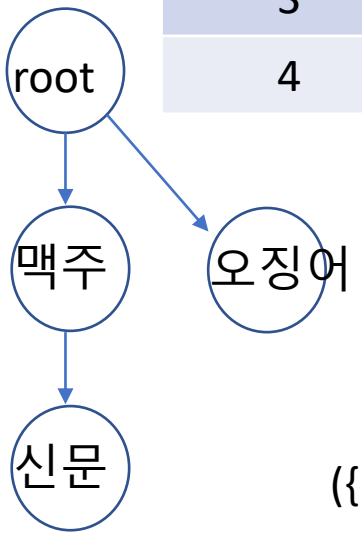
2. 순차패턴분석

➢ PrefixSpan 알고리즘

✓ Prefix를 먼저 할 대상을 찾는 기준: 지지도

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})

C_1	지지도
맥주	4
땅콩	3
오징어	4
신문	4
소주	1
양주	1



({오징어})-projected DB

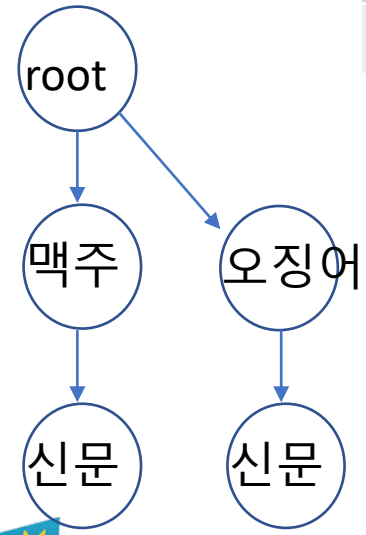
고객번호	구매기록
1	({신문})
2	({소주}, {맥주}, {신문, 땅콩})
3	({오징어, 땅콩}, {신문})
4	({신문, 양주}, {신문, 오징어})

2. 순차패턴분석

- PrefixSpan 알고리즘
  - ✓Prefix를 먼저 할 대상을 찾는 기준: 지지도

C_1	지지도
맥주	4
땅콩	3
오징어	4
신문	4
소주	1
양주	1

고객번호	구매기록
1	({신문})
2	({소주}, {맥주}, {신문, 땅콩})
3	({오징어, 땅콩}, {신문})
4	({신문, 양주}, {신문, 오징어})



({오징어},{신문})-projected DB

고객번호	구매기록
1	Null
2	({_땅콩})
3	Null
4	({_양주}, {신문, 오징어})

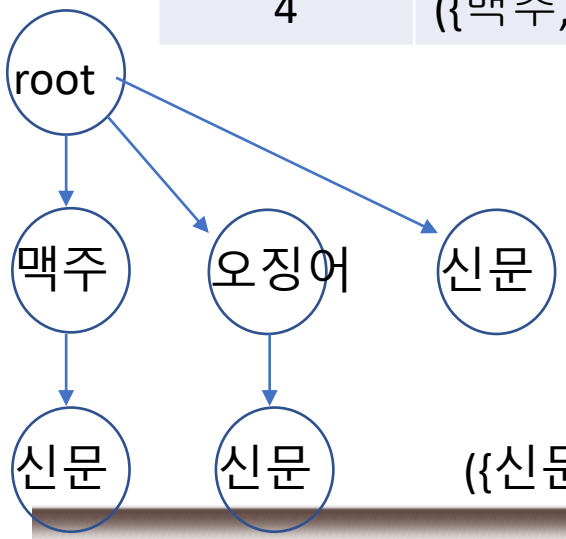
JEONJU UNIVERSITY

2. 순차패턴분석

C_1	지지도
맥주	4
땅콩	3
오징어	4
신문	4
소주	1
양주	1

- PrefixSpan 알고리즘
  - ✓Prefix를 먼저 할 대상을 찾는 기준: 지지도

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})



고객번호	구매기록
1	Null
2	({오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({오징어}, {오징어, 땅콩}, {신문})
4	({_양주}, {신문, 오징어})

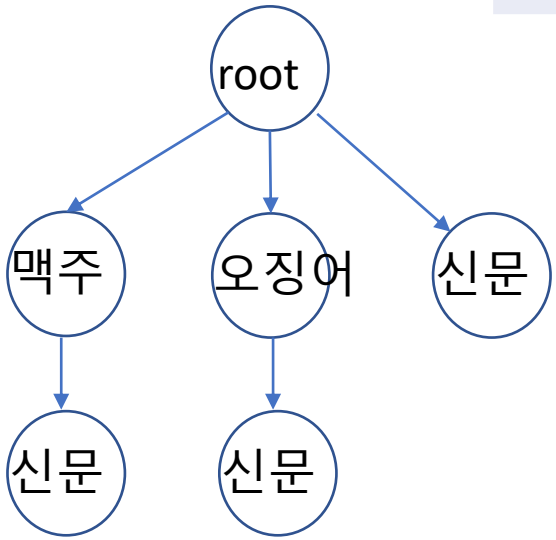
({신문})-projected DB

JEONJU UNIVERSITY

2. 순차패턴분석

➢ PrefixSpan 알고리즘

고객번호	구매기록
1	({맥주, 땅콩}, {맥주, 오징어}, {신문})
2	({신문}, {오징어}, {소주}, {맥주}, {신문, 땅콩})
3	({맥주, 신문}, {오징어}, {오징어, 땅콩}, {신문})
4	({맥주, 오징어}, {신문, 양주}, {신문, 오징어})

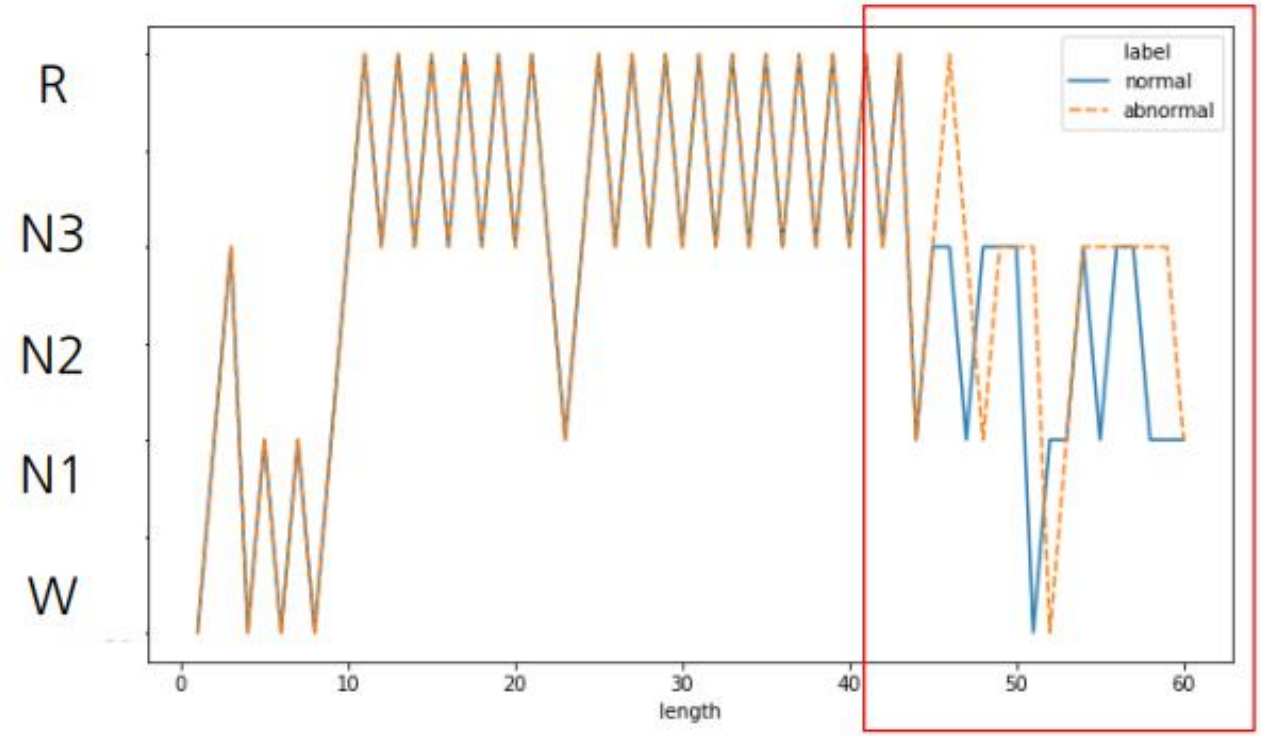


2. 순차패턴분석

➢ PrefixSpan 알고리즘

✓ 수면상태: W(얕은 수면), N1, N2, N3, R(깊은 수면) 순으로 수면 정도를 의미

✓ 정상인과 수면질환 환자의 수면 패턴 차이 분석



Thank You !

JEONJU UNIVERSITY
The Place for Superstars