

4.2 분석 결과 해석 및 활용

나는 너를 ~한다.

개념 1

• 데이터의 해석

↓
[5 , 4 , 3]

1. 임베딩 기법

기계가 이해할 수 있게 숫자 형태인 벡터로 바꾼 결과를 임베딩 벡터 그리고 이러한 과정을 임베딩이라고 한다.

2. 토폴로지 데이터 분석

토폴로지란 개체를 분리하지 않고 변형할 때 '보존된 기하학적 특징'을 연구하는 분야이다. 즉, 기하학적인 특징은 클러스터, 루트, 덩굴 형태 등을 이용해서 데이터를 분석하는 방법이다.

개념 2

• 비즈니스 기여도 평가의 이해

빅데이터의 비즈니스 기여도 평가를 단순히 빅데이터 분석을 통해 비즈니스 성과가 어느 정도 향상되었는지를 살펴보는 관점에서 바라봐서는 안 된다. 현재는 BSC를 비롯해 재무적 기여도 평가지표, IT 기여도 평가지표, 마케팅 평가지표 등을 활용해 다양한 방법이 현재 개발 및 실험되고 있다.

개념 3

● 시각화

시각화는 데이터를 분석해서 한눈에 알기 쉽게 이미지 형태로 정리 그리고 커뮤니케이션을 목적으로 한다.

1. 데이터 시각화

데이터의 시각적인 표현, 마인드맵, 의사결정 트리, 통계 그래픽, 명확하고 정확하게 커뮤니케이션 목적

2. 정보 시각화

트리맵 분기도, 수지도, 히트맵, 대규모의 비수량 정보를 시각적으로 표현

3. 정보 디자인

인포그래픽, 좀 더 명확하게 의미를 이해

개념 4

● 시각화 프로세스

1. 구조화 단계

강조하고 싶은 데이터 분석 결과가 무엇인지 파악한 뒤 데이터 표현 규칙과 패턴을 탐색해서 사용자에게 따른 시나리오를 작성하고 스토리를 구성하는 단계

2. 시각화 단계

구조화 단계에서 정의된 요건과 스토리를 기반으로 시각화를 구현하는 단계

3. 시각 표현 단계

시각화 단계를 통해 생성된 결과물을 검토하고 구조화 단계에서 정의한 초점을 흐리지 않았는지 검토하여 마무리하는 단계

개념 5

● 시간 시각화

데이터를 시간순으로 변화주는 보여주는 방법

사용되는 그래프는 막대그래프, 산점도, 선 그래프, 계단식 그래프 그리고 영역차트 가 있다.

개념 6

● 공간 시각화

기준.

사용되는 그래프들은 다음과 같다.

1. 등치 지역도

지도 위에 지리적 단위를 기준으로 하여 데이터의 의미를 색상으로 구분해 나타내는 시각화 방법이다.

2. 도트 플롯맵

지도상의 위도와 경도에 해당하는 지리적 좌표점에 산점도와 같이 점을 찍어서 시각화하는 방법

3. 버블 플롯맵

지리적 좌표 위에 정량적 데이터 값의 크기를 서로 다른 크기의 원형으로 표시하는 방법

4. 등치선도

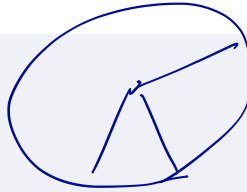
지리적으로 같은 데이터를 가진 곳을 선으로 이어서 시각화 하는 방법

5. 카토그램

선거인단 수와 같은 특정 성질을 가진 인구의 분포를 데이터 값의 변화에 따라 지도 위에 나타낸 것이다.

개념 7

● 분포 시각화



1. 파이 차트

원 형 모형으로 데이터가 차지하는 비율에 따라 여러 조각으로 나누어서 나타내는 시각화 방법

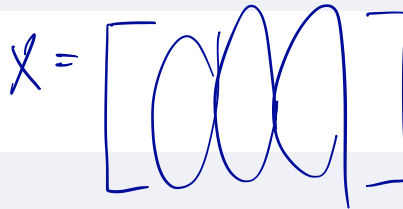
2. 도넛 차트

파이 차트와 같이 원형에 데이터를 표현하지만, 중심 부분이 비어 있는 도넛 모양으로 보이는 차이가 있다.

3. 트리맵

트리맵은 사각형을 이용하여 데이터의 값을 나타내는 것이다. 서로 다른 크기를 이용해 비율을 나타내며 사각형을 겹쳐 놓음으로써 대분류와 하위분류를 나타낸다.

성별 나이 지역



개념 8

● 관계 시각화

→ 기호

1. 산점도 행렬

다변량 변수를 갖는 데이터에서 가능한 모든 변수 쌍에 대한 산점도를 행렬의 형태로 나타낸 그래프이다.

2. 버블 차트

산점도에서는 데이터의 값을 점으로 나타내는데, 이를 확장하여 값의 비율에 따라 원형 버블의 크기를 다르게 하거나 버블의 모양을 데이터를 패턴에 따라 나타내는 시각화 방법이다.

3. 히스토그램

가로축(x축)은 정량적인 값으로 특정 간격을 가지고 있고 각 구간에 대응하는 값의 빈도를 높이로 세로축(y축)에 표현한다.



개념 9

● 비교 시각화

1. 플로팅 바 차트/ 간트 차트

가장 낮은 수치부터 가장 높은 수치까지 걸쳐 있는 막대로 표현하며 범주 내의 다양성과 범주 간 중복, 이상치 파악이 가능하다.

2. 히트맵

여러 변수를 한 번에 비교하기 위해 흔히 사용되는 그래프로, 표와 같이 행 방향은 관측 개체를 나타내고 열 방향은 각각의 변수를 나타낸다. 사각형의 각 칸에 색상을 이용해 값을 표현한다. 이 색은 채도의 차이로 큰 값과 작은 값을 구분해 나타낼 수 있다.

3. 평행 좌표계

여러 개의 축을 병렬적으로 배열하여 비교하는 방법이다. y축에서 윗 부분은 변수값 범위의 최대값, 아래는 최소값이다. 연결선 하나는 하나의 데이터를 의미하며 데이터 수만큼 연결선이 생성된다.

4. 스타 차트

거미줄 차트 또는 방사형 차트라고도 한다. 중앙에서 외부 링까지 이어지는 몇 개의 축을 그려서 별 모양의 다각형을 그린다. 변수마다 축 위에서 중앙으로부터 거리로 수치를 나타낸다.

개념 10

• 인포그래픽

인포그래픽이란 다양하고 복잡한 데이터를 이해하기 쉽고 한눈에 볼 수 있게 표현한 것을 말한다. 그래프, 다이어그램 등 다양한 시각화 도구를 이용해 데이터를 간소화한 후 데이터의 종류와 특징에 따라 분류함으로써 데이터의 의미 및 가치를 직관적으로 전달할 수 있게 한다.

중요정보를 하나의 그래픽으로 표현하는 방법.

정보를 SNS 상에 쉽고 빠르게 전달할 수 있다.

유형 = 지도형, 도표형, 스토리텔링형, 타임라인형,
비교분석형, 만화형.

분석 결과 활용의 이해

개념 11

- 빅데이터 큐레이션

빅데이터 큐레이션은 다양한 부서들끼리의 업무 조율 및 협조가 이루어지게 하는 관리자를 말한다.

* 분석 결과 활용.

① 분석 모형 전개.

(i) 빅데이터 모형 개발 및 운영 단계

(ii) 빅데이터 모형의 운영 시스템 적용 ~~연구~~ 방안.

(iii) 빅데이터 모형의 운영 시스템 적용 단계

② 분석 결과 활용 시나리오 개발.

(i) 분석모형 결과를 활용할 수 있는 분야 파악.

(ii) 분석결과 " 분야 분류.

(iii) ' 서비스 영역 도출

* 기를 정리

트리 맵 : 크기로 비율을 표현, 분포시각화.

히트 맵 : 색상으로 Data 표현, 비교시각화.

스타 차트 : 설명변수 $\uparrow \Rightarrow$ 축 \uparrow

평행좌표 : 다변량 Data.를 2차원에 표현.

여러 축을 평행 배치

4각선에 변수를 배치

비교시각화.