

# Paper Review

이 한 준 교수

# Paper Critique

- 들어가며...
  - 세상에 완벽한 연구는 없음
    - 논문은 저자의 주장일 뿐 학계의 정설로 굳어지려면 많은 연구가 필요
    - 교과서조차도 시간이 지나 바뀌는 경우가 있음
    - 하물며 논리적 오류나 잘못된 주장도 얼마든지 있을 수 있음
    - 편집이 잘못된 부분도 있고, 오탈자도 있을 수 있음

# Paper Critique

- 들어가며...
  - AI가 이미 압도적으로 우수해진 시대, 인간이 개발해야할 것은?
  - AI가 할 수 없는 것은?



# Paper Critique

- 들어가며...
  - 질문 훈련
    - 아이디어를 생각하는 훈련
    - Research Question을 도출하는 훈련
    - 아이디어를 내는 것은 연구는 물론, 사업을 할 때도, 직장에서 일을 할 때도, 무슨 일에도 유용할 것
    - 어떻게 저런 아이디어를 냈을까를 생각해볼 것

# Paper Critique

- 들어가며...
  - 언제 아이디어가 나올까?

# Paper Critique

- 들어가며...

Research is fun!



# Paper Critique

- Paper 2. 기업 MBTI 분류
  - Research Question: 기업의 성격을 데이터분석으로 측정할 수 있을까?
  - Dataset: 잡플래닛
  - Methodology: Doc2Vec(Word embedding)

# Paper Critique

- Paper 2. 기업 MBTI 분류
  - Motivation
    - 어떻게 이미지를 판단할까?

그룹사를 떠올렸을 때 연상되는 이미지는? ①성별·나이(세) ②직업 ③이미지



삼성	현대차	LG	포스코	SK	롯데
남성·30~34	남성·30~34	남성·30~34	남성·40~44	여성·25~29	여성·25~29
연구개발직	전문기술직	연구개발직	전문기술직	판매서비스직	판매서비스직
지적, 권위적, 냉정	진취, 강인, 도시적	유행 민감, 대중적, 친근함	남성적인, 투박, 강인함	유행 민감, 세련됨, 대중적	대중적, 보수적, 여성스러움



## 기업이미지를 나타내는 특성용어들의 요인구조와 대기업간의 이미지 비교연구\*

유 태 용

광운대학교 산업심리학과

# Paper Critique

- Paper 2. 기업 MBTI 분류
  - Motivation

본 연구의 목적은 일상생활에서 사람이나 대상을 기술할 때 흔히 사용되는 특성용어(형용사)들 중에서 기업의 이미지를 나타내는 데 적절한 특성용어들을 찾아내고, 기업이미지를 나타내는 이러한 특성용어들에 내재되어 있는 요인구조를 밝히는 데 있다. 이를 위해 세가지 연구를 수행하였다. 연구 1에서는 총 926명의 대학생들을 사용하여 기업이미지를 나타내기 위한 770개의 특성용어들 각각이 얼마나 적절한지를 평정하였다. 연구 2에서는 총 755명의 대학생들이 연구 1에서 기업이미지를 나타내는 데 적절하다고 평정된 332개의 특성용어들이 얼마나 좋거나 또는 나쁜 이미지를 나타내는지를 평정하였다. 연구 3에서는 연구 1에서 기업이미지를 나타내는 데 적절하다고 밝혀진 332개의 특성용어들 중에서 적절도가 상대적으로 더큰 157개의 특성용어들을 사용하여, 총 764명의 대학생들이 5대기업(현대, 대우, 삼성, 럭키금성, 선경)의 이미지를 평정하였다. 이 자료를 요인분석하여 기업이미지를 나타내는 다섯가지 요인을 발견하였다. 다섯가지 요인들은 知性(31%), 生動性(27%), 非人間性(20%), 野性(12%), 堅實性(10%) 요인이었다. 이러한 요인에 기초하여 부가적으로 5대기업들 간의 이미지를 중다비교하였고, 다섯개 요인에 따른 전반적 기업이미지의 유사성을 알아보기 위해 군집분석을 실시하였다. 중다비교 결과, 삼성이 다섯개 차원에서 전반적으로 가장 좋은 이미지를 지니고 있었고 대우와 럭키금성은 다섯개 차원에서의 평정치가 전반적으로 낮았다. 현대는 다른 기업에 비해 상대적으로 가장 비인간적으로 평가되었다. 군집분석 결과, 다섯개 차원을 종합해 볼 때 대우와 럭키금성의 이미지와 삼성과 선경의 이미지가 전반적으로 유사하였다. 마지막으로 앞으로의 추가적인 연구과제가 논의되었다.

# Paper Critique

- Paper 2. 기업 MBTI 분류
  - Motivation

표 2. 7점 척도에서 좋거나 나쁜 기업이미지를 나타내는 대표적 특성용어들

호오도 (척도치)	특성용어	평균	표준편차
상당히 좋은 (6.00)	미래지향적이다	6.00	1.10
	전망이 밝다	5.91	1.14
	세계적이다	5.83	1.31
	창조적이다	5.81	1.36
	근면하다	5.72	1.14
	국제적이다	5.72	1.21
좋은 편 (5.00)	지구력이 있다	5.01	1.15
	매력적이다	5.01	1.24
	풍부하다	5.00	1.11
	튼튼하다	5.00	1.01
	꿇꿇하다	5.00	1.00
	산뜻하다	4.99	.98
중립적 (4.00)	서구적이다	4.06	1.05
	저돌적이다	4.05	1.29
	엄하다	4.03	1.04
	돌격적이다	3.97	1.24
	직선적이다	3.96	1.02
	고집스럽다	3.94	1.05

# Paper Critique

- Paper 2. 기업 MBTI 분류
  - Motivation

표 3. 기업이미지를 나타내는 특성용어의 요인구조

요인 I: 知性요인 ( $\alpha=.98$ )			요인 II: 生動性요인 ( $\alpha=.98$ )			요인 III: 非人間性요인 ( $\alpha=.95$ )			요인 IV: 野性요인 ( $\alpha=.91$ )			요인 V: 堅實性요인 ( $\alpha=.83$ )		
특성용어	요인부하	$R^2$	특성용어	요인부하	$R^2$	특성용어	요인부하	$R^2$	특성용어	요인부하	$R^2$	특성용어	요인부하	$R^2$
꼼꼼하다	.67	.53	발전적이다	.68	.64	몰인정하다	.75	.64	박력있다	.55	.53	튼튼하다	.54	.48
슬기롭다	.64	.58	도전적이다	.66	.54	위선적이다	.74	.67	남성적이다	.55	.40	탄탄하다	.52	.47
정확하다	.63	.57	창조적이다	.65	.63	돈밖에모른다	.74	.65	대담하다	.54	.58	견실하다	.44	.45
논리적이다	.63	.49	혁신적이다	.65	.58	비인간적이다	.74	.62	저돌적이다	.54	.44	강하다	.44	.43
틀림없다	.63	.49	진취적이다	.65	.57	이기적이다	.74	.59	화끈하다	.54	.56	크다	.43	.42
지성적이다	.60	.47	창의적이다	.64	.67	인간미없다	.71	.59	대범하다	.54	.55	조직적이다	.42	.34
산뜻하다	.60	.49	도약적이다	.63	.56	약삭빠르다	.70	.54	집념이강하다	.50	.54	경제적이다	.42	.38
모범적이다	.60	.64	동적이다	.62	.51	기회주의적이다	.69	.54	당당하다	.48	.55	빈틈없다	.41	.45
깔끔하다	.60	.48	독창적이다	.61	.56	차갑다	.68	.49	웅대하다	.45	.45	실질적이다	.37	.27
세련미있다	.59	.47	생동감있다	.61	.52	계산적이다	.67	.51	근성있다	.45	.46	생산적이다	.36	.39
야무지다	.59	.53	생동적이다	.60	.61	인색하다	.66	.48	강력하다	.41	.54	계획적이다	.34	.26
현명하다	.59	.54	활기있다	.58	.65	소비조장적이다	.66	.47				부실하다	-.45	.41
사리판단이 밝다	.59	.42	참신하다	.58	.57	이윤추구적이다	.64	.50						
도덕적이다	.59	.59	의욕적이다	.58	.50	기만적이다	.63	.49						
날카롭다	.58	.48	열의있다	.57	.54	물질적이다	.63	.47						
신빙성있다	.58	.57	실천적이다	.57	.50	냉정하다	.62	.42						
체계적이다	.57	.55	역동적이다	.56	.60	타산적이다	.57	.39						
매력적이다	.57	.52	패기있다	.56	.64	개인주의적이다	.57	.34						
학구적이다	.57	.43	과감하다	.56	.54	이해타산적이다	.56	.40						
양심적이다	.57	.58	적극적이다	.56	.63	관료적이다	.55	.38						
곧다	.56	.57	열성적이다	.56	.49	정치적이다	.54	.32						
근면하다	.55	.58	행동적이다	.54	.62	공격적이다	.51	.34						
완벽하다	.55	.45	활동적이다	.54	.60	실리적이다	.50	.41						

# Paper Critique

- Paper 2. 기업 MBTI 분류
  - Motivation

표 4. 5대기업에 대한 요인점수의 평균과 표준편차

	지성요인	생동성요인	비인간성요인	야성요인	견실성요인
현 대	4.32(.79)	4.55(.90)	4.47(.90)	4.64(.84)	4.81(.67)
대 우	4.14(.79)	4.45(.83)	4.17(.74)	4.32(.88)	4.40(.67)
삼 성	4.70(.79)	5.01(.85)	4.22(.80)	4.66(.85)	5.00(.67)
럭키금성	4.29(.73)	4.41(.81)	4.15(.70)	4.23(.81)	4.53(.66)
선 경	4.61(.67)	4.79(.70)	3.65(.69)	4.33(.77)	4.63(.56)

주. 괄호 안은 표준편차를 나타냄.

# 논문을 이해하기 위한 사전지식

- Doc2Vec(Document to Vector)
  - 문서(여러 문장으로 이루어진 텍스트)를 벡터(Vector)로 바꾸는 워드 임베딩(word embedding)이라는 자연어 처리 기법의 한 종류
- 자연어 처리(텍스트 분석)이란?
  - 인간의 언어를 컴퓨터가 이해할 수 있도록 하는 기술
  - 컴퓨터는 비정형데이터(텍스트, 이미지 등)를 숫자화해야 인식가능함
    - 컴퓨터가 색깔을 이해하는 방법? cf. RGB
    - 그렇다면, 컴퓨터가 글자를 이해하는 방법?
- 데이터사이언스프로그래밍(최한별 교수님) 과목에서 공부할 수 있음



# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - Bag of Words (BoW): 문맥에 대한 고려 없이 단어의 등장만을 분석

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - Bag of Words (BoW): 문맥에 대한 고려 없이 단어의 등장만을 분석
    - 활용 예시 - 워드 클라우드
      - 인터스텔라(2014) 영화 리뷰로 제작



# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - TF-IDF(Term Frequency-Inverse Document Frequency)
    - 단어의 횟수가 많이 나올수록, 그리고 그 단어가 덜 일반적으로 사용될수록 중요도를 부여하는 방식
    - 예시) 어떤 단어가 중요한 단어일까?
      - 리뷰1 – 이 영화를 보니 옛날 할머니와의 추억이 떠오르더군요. 어릴 적으로 돌아가는 듯한 기분이었습니다.
      - 리뷰2 – 배경음악이 참 좋았던 영화였습니다. 첫 음악도 무척 좋았구요. 음악 감독이 누군지 찾아봤는데 이번 작품이 첫 작품이더군요. 음악이 좋아서 OST도 구매했습니다.
      - 리뷰3 – 저는 영화가 그렇게까지 인상적이진 않았습니다. 주위에서 꼭 보라고 해서 보긴 봤는데, 그냥 그랬습니다.



# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - TF-IDF(Term Frequency-Inverse Document Frequency)
    - 단어의 횟수가 많이 나올수록, 그리고 그 단어가 덜 일반적으로 사용될수록 중요도를 부여하는 방식
    - TF - 특정 문서 d에서 특정단어 t의 등장 횟수
    - DF - 특정 단어 t가 등장한 문서의 수
  - TF와 IDF(DF의 역수)의 곱으로 계산
    - I(Inverse)는 역수의 의미
    - TF-IDF값은 TF에 비례하고 DF값에 반비례
- 빈번하게 등장하되 일반적이지 않고 특정 문서에 집중될수록 중요도 부여

# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - TF-IDF(Term Frequency-Inverse Document Frequency)
    - 연구 활용 예시 – DBPIA에 검색

The screenshot shows the DBPIA search results for the query 'tf-idf'. The search bar at the top contains 'tf-idf' and the results are displayed in a list format. The first result is titled 'TF-IDF를 활용한 초·중등학교 저작권 교육 자료의 내용 분석' (Content Analysis of Copyright Education Materials for Elementary and Middle Schools Utilizing TF-IDF). The second result is titled '웹셀 탐지를 위한 TF-IDF 및 응답 용량 변동 기반 사용자 행위 임베딩' (User Behavior Embedding Based on TF-IDF and Response Capacity Variation for Web Cell Detection). The left sidebar shows filters for document type (e.g., '외부링크', '원문이용') and subject (e.g., '사회과학', '교육학', '경영학').

DBpia

전체 tf-idf

상세검색

주제분류 BEST 매거진 저널 DBpia AI AI 검색 idea Citeasy 개인구독 지원사업

명지대학교 로그인 회원가입 고객센터 ENG

주제 추천결과 (5건) 검색 키워드 알림

idea tf-idf 연관 리포트/소논문 주제 리포트/소논문

검색결과 (1,341건)

☐ 내서재 답기 ☒ 정확도순 ☐ 이용순 ☐ 최신순 20개씩

☐ TF-IDF를 활용한 초·중등학교 저작권 교육 자료의 내용 분석

목적 달성을 위해 한국저작권위원회의 초·중등 표준 교재 및 강의자료에서 추출한 텍스트 데이터에 대하여 TF-IDF 및 동시출현 빈도를 분석하였다. 분석 결과, 현재 교재와 강의자료에서는 디지털 저작권 관련 단어가 높은 ...

#TF-IDF #저작권 #디지털 저작권

이경민, 김자미, 이원규 | 한국컴퓨터교육학회 | 컴퓨터교육학회 논문지 | 27(3) | 2024.5 | 65-74(10pages)

학술저널 | KCI등재 | 이용수 197

☐ 웹셀 탐지를 위한 TF-IDF 및 응답 용량 변동 기반 사용자 행위 임베딩

본 연구에서는 웹 서버의 access log에서 웹 서버에 접근한 사용자별 행동 특성을 벡터화하여 정상 상태를 정의 하고, 이를 벗어난 비정상 상태, 즉 웹셀 공격을 탐지하는 TF-IDF 기반의 임베딩 기법을 제안한다. 본 기법은 응...

#TF-IDF #Anomaly detection #threat detection

김강문, 이인섭 | 한국정보보호학회 | 정보보호학회논문지 | 34(6) | 2024.12 | 1,231-1,238(8pages)

학술저널 | KCI등재 | 이용수 9

# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - TF-IDF(Term Frequency-Inverse Document Frequency)
    - 연구 활용 예시 – DBPIA에 검색

The screenshot shows the DBpia search results page for the query 'tf-idf 주식'. The page is in Korean and displays search results for the topic '주제 추천결과 (5건)' and '검색결과 (7건)'. The search results are listed in a table with columns for '제목' (Title), '저자' (Author), '연도' (Year), and '페이지' (Pages). The first result is '증권신고서의 TF-IDF 텍스트 분석과 기계학습을 이용한 공모주의 상장 이후 주가 등락 예측' (Prediction of stock price fluctuations after public offering of companies using TF-IDF text analysis and machine learning). The second result is '딥러닝 기반의 주식시장 예측을 위한 동적 텍스트 변수 가중치에 관한 연구' (Research on dynamic text variable weights for stock market prediction based on deep learning). The page also includes a sidebar with filters for '형태' (Format), '분류' (Classification), and '연도' (Year). A tooltip message says '클릭! 내서재에 논문을 담아보세요.' (Click! Add the paper to your collection.).

DBpia

전체 tf-idf 주식

상세검색

idea

내서재

일

제분류 BEST 매거진 저널 + DBpia AI AI 검색 idea Citeasy 개인구독 지원사업

명지대학교 로그인 회원가입 고객센터

주제 추천결과 (5건)

idea tf-idf 주식 연관 리포트/소논문 주제

리포트/소논문

검색결과 (7건)

내서재 담기

정확도순 이용순 최신순 20개씩

증권신고서의 TF-IDF 텍스트 분석과 기계학습을 이용한 공모주의 상장 이후 주가 등락 예측

본 연구는 개인투자자들의 투자의사결정에 도움을 주고자, 증권신고서의 TF-IDF 텍스트 분석과 기계학습을 이용해 공모주의 상장 5거래일 이후 주가 등락 예측하는 모델을 제시한다. 연구 표본은 2009년 6월부터 202...

#TF-IDF #주가 예측 #기계학습

양수연, 이재록, 원종관 외 1명 | 한국지능정보시스템학회 | 지능정보연구 | 28(2) | 2022.6 | 237-262(26pages)

학술저널 | KCI등재 | 이용수 295

딥러닝 기반의 주식시장 예측을 위한 동적 텍스트 변수 가중치에 관한 연구

비선형적이고 급변하는 주식시장을 예측하기 위해 기존에 사용하지 않던 텍스트 데이터를 활용한 연구가 활발히 이루어졌다. 그러나 한국어 텍스트 데이터를 머신러닝 기반 예측모델의 입력으로 사용하는 것은 까다롭다. 선...

#주식시장 예측, 머신러닝, Local Interpretable Model-agnostic Explanations, 한국어 텍스트 데이터

서민석 | 연세대학교 대학원 | 2022

학위논문 | 이용수 37

클릭! 내서재에 논문을 담아보세요.

# 논문을 이해하기 위한 사전지식

- 토큰이란?

- 크롤링 등으로 얻은 코퍼스(말뭉치) 데이터를 분석에 적합한 단위로 나누는 것을 토큰화(tokenization)할 때 그 기본 단위를 말함

- 예시

- 입력: **Time is an illusion. Lunchtime double so!**
- 출력: "Time", "is", "an", "illusion", "Lunchtime", "double", "so"

# 논문을 이해하기 위한 사전지식

- 애매한 경우

- 예시

- **Don't** be fooled by the dark sounding name, Mr. **Jone's** Orphanage is as cheery as cheery goes for a pastry shop.

- 특수문자(')가 들어간 경우 가능한 토큰화 방법

- Don't

- Don t

- Dont

- Do n't

- Jone's

- Jone s

- Jone

- Jones

# 논문을 이해하기 위한 사전지식

- 애매한 경우
  - 특수 문자를 단순 제거할 수 없는 경우
    - Ex. m.p.h, ph.D, \$25.55, 01/04/12, 125,500, michael.lee@gmail.com
- 줄임말이 있는 경우
  - Ex. Nat'l, We're, rock 'n' roll

# 논문을 이해하기 위한 사전지식

- 한국어에서의 토큰화

- 한국어는 교착어(조사, 어미를 사용하는 언어)의 특성을 가짐
  - 띄어쓰기만으로 대략의 토큰화가 가능한 영어와 달리 한 어절 안에 조사나 어미를 포함
  - 예시
    - 문장 : 민수가 머신러닝책을 읽었다
    - 자립 형태소 : 예지, 머신러닝책
    - 의존 형태소 : 가, 을, 읽, 었, 다

- 한국어에서는 어절 단위의 토큰화가 아니라 형태소 토큰화가 필요

# 논문을 이해하기 위한 사전지식

- 한국어에서의 토큰화
  - 띄어쓰기가 영어보다 잘 지켜지지 않는다.
  - 예시
    - Tobeornottobethatisthequestion
    - 죽느냐사느냐그것이문제로다
    - 제가이렇게띄어쓰기를전혀하지않고글을썼다고하더라도글을이해할수있습니다.



# 논문을 이해하기 위한 사전지식

- 한국어 형태소 분석기
  - Kkma, Komoran, Hannanum, Twitter(Okt로 변경), Mecab 등 여러 형태소 분석기가 개발
  - 형태소 분석기를 통해 문장의 품사를 태깅하여 사용
  - 진짜 잘 될까???

# 논문을 이해하기 위한 사전지식

- 한국어 형태소 분석 예시
  - 직접 테스트해보자(Jupyter notebook이나 Colab 사용)

```
from konlpy.tag import Okt
```

```
# 형태소 분석기 객체 생성
```

```
okt = Okt()
```

```
# 분석할 문장
```

```
text = "자연어 처리는 정말 재미있는 분야입니다!"
```

```
# 형태소 분석
```

```
print("Okt 형태소 분석 결과:")
```

```
print(okt.morphs(text)) # 형태소 단위 분리
```

```
print(okt.pos(text)) # 품사 태깅
```

```
print(okt.nouns(text)) # 명사 추출
```

Okt 형태소 분석 결과:

```
['자연어', '처리', '는', '정말', '재미있는', '분야', '입니다', '!']
```

```
[('자연어', 'Noun'), ('처리', 'Noun'), ('는', 'Josa'), ('정말', 'Noun'), ('재미있는', 'Adjective'), ('분야', 'Noun'), ('입니다', 'Adjective'), ('!', 'Punctuation')]
```

```
['자연어', '처리', '정말', '분야']
```

# 논문을 이해하기 위한 사전지식

- 한국어 형태소 분석 예시
  - 직접 테스트해보자
  - 과연 이것도???

```
from konlpy.tag import Okt

# 형태소 분석기 객체 생성
okt = Okt()

# 분석할 문장
text = "아버지가방에들어가신다"

# 형태소 분석
print("형태소 분석 결과:")
print(okt.morphs(text)) # 형태소 단위 분리
print(okt.pos(text)) # 품사 태깅
print(okt.nouns(text)) # 명사 추출
```

# 논문을 이해하기 위한 사전지식

- 텍스트를 숫자로 표현하는 기법
  - Word2Vec
    - 2013년 구글에서 발표한 모델
    - 문맥을 반영할 수 있도록 주변 단어를 중심으로 중심 단어를 예측하거나 반대로 중심 단어로 주변 단어를 예측함
    - Doc2Vec도 Word2Vec에 기반을 두고 있음

# 논문을 이해하기 위한 사전지식

- Word2Vec 예시
  - 텍스트를 학습시키면 각 단어에 대한 벡터를 얻을 수 있음

```
from gensim.models import Word2Vec
sentences = [["I", "love", "machine", "learning"], ["deep", "learning", "is", "amazing"]]
model = Word2Vec(sentences, vector_size=100, window=5, min_count=1, workers=4)
print(model.wv["machine"]) # 특정 단어의 벡터 출력
```

```
[-0.00713902  0.00124103 -0.00717672 -0.00224462  0.0037193  0.00583312
 0.00119818  0.00210273 -0.00411039  0.00722533 -0.00630704  0.00464722
-0.00821997  0.00203647 -0.00497705 -0.00424769 -0.00310898  0.00565521
 0.0057984 -0.00497465  0.00077333 -0.00849578  0.00780981  0.00925729
-0.00274233  0.00080022  0.00074665  0.00547788 -0.00860608  0.00058446
 0.00686942  0.00223159  0.00112468 -0.00932216  0.00848237 -0.00626413
-0.00299237  0.00349379 -0.00077263  0.00141129  0.00178199 -0.0068289
-0.00972481  0.00904058  0.00619805 -0.00691293  0.00340348  0.00020606
 0.00475375 -0.00711994  0.00402695  0.00434743  0.00995737 -0.00447374
-0.00138926 -0.00731732 -0.00969783 -0.00908026 -0.00102275 -0.00650329
 0.00484973 -0.00616403  0.00251919  0.00073944 -0.00339215 -0.00097922
 0.00997913  0.00914589 -0.00446183  0.00908303 -0.00564176  0.00593092
-0.00309722  0.00343175  0.00301723  0.00690046 -0.00237388  0.00877504
 0.00758943 -0.00954765 -0.00800821 -0.0076379  0.00292326 -0.00279472
-0.00692952 -0.00812826  0.00830918  0.00199049 -0.00932802 -0.00479272
 0.00313674 -0.00471321  0.00528084 -0.00423344  0.0026418 -0.00804569
 0.00620989  0.00481889  0.00078719  0.00301345]
```

# 논문을 이해하기 위한 사전지식

- Word2Vec의 활용
  - N차원의 공간 상 특정 위치에 해당 단어가 위치하게 됨
  - 거리가 유사성을 의미하게 되며 의미상 유사성이 높은 단어일수록 거리가 가까움

한국 - 서울 + 도쿄 = ?

고양이 + 애교 = ?

# 논문을 이해하기 위한 사전지식

- Doc2Vec의 개념과 활용
  - Word2Vec을 확장하여 문장(문서) 수준의 벡터를 학습하는 방법
  - 문서 간 유사도 분석, 주제 분류에 활용

위계질서가 강한 군대문화이다.



선후배 사이가 친근하다.



조직문화가 수평적이고 자유롭다.

# 논문을 이해하기 위한 사전지식

- Doc2Vec 예시

- 텍스트를 학습시키면 각 문장(문서)에 대한 벡터를 얻을 수 있음

```
from gensim.models.doc2vec import Doc2Vec, TaggedDocument
sentences = ["I love machine learning", "Deep learning is amazing"]
documents = [TaggedDocument(doc.split(), [i]) for i, doc in enumerate(sentences)]
model = Doc2Vec(documents, vector_size=100, window=5, min_count=1, workers=4, epochs=20)
print(model.dv[0]) # 첫 번째 문서의 벡터 출력
```

```
[-5.2420530e-03 -5.9914179e-03 -9.9011911e-03  8.5702864e-03
 3.5719262e-03  2.6319752e-04 -9.9002197e-03 -5.1755174e-03
-9.7358748e-03  2.0153220e-03  2.8358467e-03  4.6518533e-03
-4.3060547e-03 -3.1528224e-03 -3.0832796e-03 -8.7405331e-03
 2.1750084e-03  9.2435982e-03 -9.5198806e-03 -3.4634629e-03
-3.7757580e-03  2.6135899e-03 -5.7038167e-03  2.6273353e-03
 5.8122678e-03 -8.1226500e-03 -8.3472962e-03 -9.9743493e-03
 4.9431883e-03 -9.1415560e-03  5.8541209e-03  6.8138600e-03
-6.5192655e-03 -4.5276694e-03 -1.2580878e-03  1.6492606e-03
-1.4840655e-03 -8.5588032e-03 -3.6112086e-03  1.7341374e-03
-2.0595840e-03 -7.2440738e-03  4.1934596e-03 -8.5927974e-03
 2.7165171e-03 -4.6223975e-03  6.4714340e-04 -2.0598315e-03
 5.4239063e-03 -8.0190990e-03 -2.1245850e-03 -9.4326359e-05
-6.6505056e-03 -6.5405234e-03 -1.9373804e-03  8.8217044e-03
-1.2654938e-03  3.5444486e-03 -5.7626227e-03  8.8315783e-03
 2.9226553e-03  9.2988499e-03  4.3599159e-03 -4.2083622e-03
 2.2467098e-03 -4.4223112e-03  5.7872459e-03  1.8340871e-03
-2.2846858e-03 -5.8930484e-03 -8.0422228e-03 -8.5503148e-04
-8.9563886e-03 -9.2408545e-03 -7.9541961e-03  2.1737749e-03
-6.5123220e-03 -7.8043258e-03  2.1345178e-03  2.0564871e-03
 8.3657689e-03  4.6779057e-03 -9.4300210e-03 -3.3995794e-04
 7.8680422e-03  2.6782395e-03  2.6852784e-03 -4.8933132e-03
 6.4813667e-03  1.6523676e-03 -7.6183625e-03  6.8786242e-03
-9.7877542e-03 -8.1732254e-03 -4.8844120e-03  9.9579543e-03
 3.1200501e-03 -2.0166950e-03  8.9130057e-03  2.3567996e-03]
```



# 논문을 이해하기 위한 사전지식

- 최신 워드 임베딩 모델
  - BERT(Bidirectional Encoder Representations from Transformers)
    - 구글이 개발한 사전학습된 Transformer 기반 모델
    - 양방향 문맥을 고려하여 단어의 의미를 동적으로 결정
  - ELMo(Embeddings from Language Models)
    - 문장의 전체 문맥을 반영하는 임베딩 모델
    - 단어의 의미를 문맥에 따라 달리 표현
  - GPT(Generative Pre-trained Transformer)
    - OpenAI가 개발한 언어 생성 모델
    - 트랜스포머의 디코더 구조를 활용하여 텍스트를 생성하는데 강점

# Paper Critique

- 성격유형에 대한 선행연구
  - 개인 성격유형에 대한 연구는 심리학에서 오랜 연구주제
  - Big 5
    - 인간의 성격을 다섯 가지 차원(Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism)으로 설명
    - 가장 많이 연구된 성격 모델 중 하나이며, 다양한 문화권에서 검증됨
    - <https://together.kakao.com/big-five>

# Paper Critique

- 성격유형에 대한 선행연구
  - 개인 성격유형에 대한 연구는 심리학에서 오랜 연구주제
  - MBTI
    - Carl Jung의 심리 유형 이론을 바탕으로 만들어진 성격 유형 검사
    - 16가지 성격 유형(예: INTJ, ENFP 등)으로 사람들의 성향을 구분
    - 직업 상담, 조직 관리, 자기 이해 등에 활용되나, 신뢰성과 타당성에 대한 논란 있음

# Paper Critique

- 조직의 성격에 대한 선행연구
  - “구성원의 성격으로 조직의 성격을 판단할 수 있다.”
  - 조직의 성격에 대한 개념

Table 1. 성격요인별 기업의 특징

E (외향적 기업)	I (내향적 기업)
조직 경계가 개방적이다. 의사결정에 다수가 참여한다. 결정하기 전에 의견일치를 본다. 회의로 의사소통한다. 회의가 많다. 부서간 협업이 많다. 외부전문가에 의존한다. 다른 기업과의 교류가 많다.	조직 경계가 폐쇄적이다. 의사결정에 소수만 참여한다. 결정하고 나서 합의를 도출한다. 상황파악에 많은 시간을 보낸다. 문서로 의사소통한다. 서면보고가 많다. 부서간 협업이 적다. 내부적으로 문제를 해결한다. 각자 경쟁한다.
S (감각적 기업)	N (직관적 기업)
현실과 세부적인 것에 관심을 둔다. 반복적인 규칙을 강조한다. 점진적이고 단계적인 변화를 추구한다. 목표와 계획을 강조한다. 경험과 권위가 중요하다. 상세한 정보에 의해 의사결정이 이루어진다. 안정을 추구한다. 전통을 중시한다. 표준화된 매뉴얼이 중요하다.	미래와 가능성에 관심을 둔다. 직관을 강조한다. 혁신적인 변화를 추구한다. 목적과 비전을 강조한다. 통찰력과 창의성이 중요하다. 전반적인 추세에 의해 의사결정이 이루어진다. 성장을 추구한다. 새로운 것을 만들어낸다. 매뉴얼이 중요하지 않다.
T (사고적 기업)	F (감정적 기업)
원리 원칙에 따라 결정한다. 규칙을 중시한다. 예외를 인정하지 않는다. 논리적 사고를 강조한다. 목표 중심적이다. 비판적이다. 목표를 달성하도록 돕는다. 효율을 강조한다.	인간적 가치관에 의해 결정한다. 인간적 상황을 인정한다. 감정적 배려를 중시한다. 사람 중심적이다. 최선을 다하도록 돕는다. 인화를 강조한다. 화합을 강조한다.
J (판단적 기업)	P (인식적 기업)
결정을 내리도록 강요한다. 성급하게 결정한다. 구체적으로 정의한다. 만장일치를 선호한다. 결정을 바꾸지 않는다. 의사결정의 신속성을 강조한다. 절차가 상세하다.	결정을 미룬다. 신중하게 결정한다. 정의되지 않은 채로 둔다. 의사결정에서 토의에 개방적이다. 더 나은 선택의 여지를 남겨둔다. 의사결정이 느리다. 절차가 불명확하다.

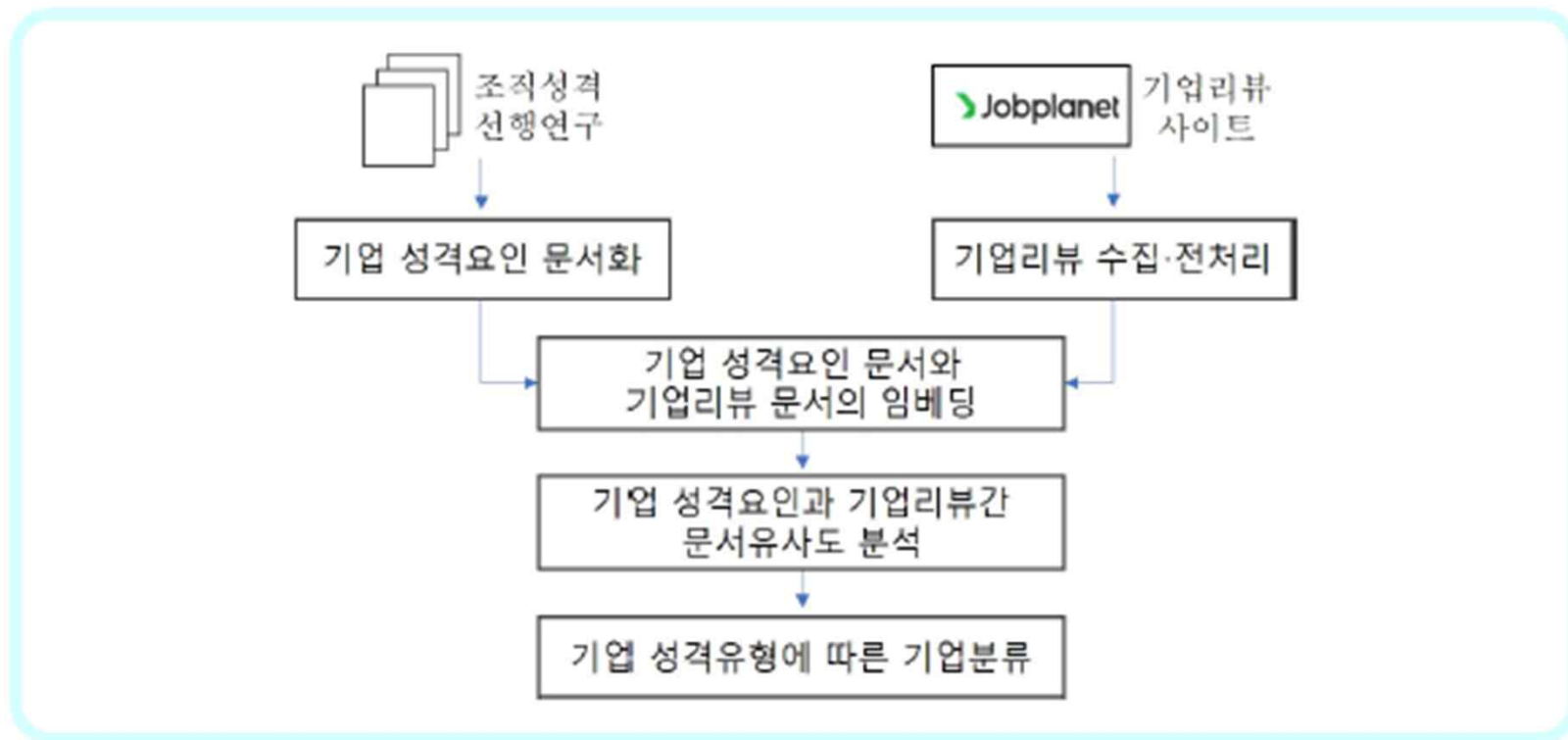
# Paper Critique

## • Paper 2. 기업 MBTI 분류

Table 3. 잡플래닛에 게시된 기업리뷰 예시		
게시일	장점	단점
2020.08.02	"수평 조직문화이고 복지제도도 잘 갖춰져 있어서 좋습니다. 복장에 대한 자율성 보장, 구내식당 등 직원복지도 괜찮습니다."	"성과주의 원칙 때문에 직원간의 내부 경쟁이 조금 있습니다. 그리고 업무 강도가 부서별로 다르긴 한데 강할수도 있습니다."
2020.07.29	"일하는 데 필요한 시스템이 잘 갖춰져 있음. 조직문화도 수평적이고 괜찮은 편."	"치열한 경쟁으로 끊임없는 자기계발을 하지 않으면 치고 올라오는 후배들에게 압박을 받을 수 있음"
2019.06.30	"아무래도 자유로운 분위기를 시도하려는 노력이 있고 우리나라에서 가장 큰 기업이니 안정적으로 다닐 수 있고 복지는 좋아요"	"보고를 위한 보고가 너무 많고 총총이 누적된 임원들에게 같은 보고를 여러번 해야하는 비효율적인 면이 많습니다. 그리고 조직이 커서 융통성이 별로 없습니다"
2017.02.19	"인재를 소비하기 보다는 키우는 개념을 가짐. 핵심인력이 되면 좋은 기회가 많아짐."	"조직이 비대해져 의사결정 체계가 복잡. 조직간 지나친 경쟁으로 협력이 어려워 큰 혁신을 이뤄내기 어렵다."
2016.02.26	"모든 것이 시스템으로 구성되어 있으며, 예외없이 원칙적으로 업무가 진행되고 있음."	"업무와 삶의 밸런스가 맞지 않으며, 잦은 인사발령과 이에 따른 사람을 상대하기 위한 조직생활이 고단합니다."

# Paper Critique

- Paper 2. 기업 MBTI 분류



# Paper Critique

## • Paper 2. 기업 MBTI 분류

개시일	장점	단점
2020.08.02	"수준 조직문화이고 복지제도도 잘 갖춰져 있어서 좋습니다. 복지에 대한 관심이 높고, 구내식당 등 직원복지도 괜찮습니다."	"성과주의 원칙 때문에 직원간의 내분 경쟁이 조금 있습니다. 그리고 업무 조도가 분석결과 다르긴 한데 강압수도 있습니다."
2020.07.29	"일하는 데 필요한 시스템이 잘 갖춰져 있음. 조직문화도 수평적이고 괜찮은 편."	"지열한 방향으로 끊임없는 자기개발을 하지 않으면 지고 올라오는 후배들에게 압박을 받을 수 있음"
2019.06.30	"아무래도 자유로운 분위기를 시도하려는 노력이 있고 우리나라에서 가장 큰 기업이나 안정적으로 다닐 수 있고 복지는 좋아요"	"보고를 위한 보고가 너무 많고 총총이 부족한 임원들에게 같은 보고를 여러번 해야하는 비효율적인 면이 있습니다. 그리고 조직이 커서 융통성이 별로 없습니다"
2017.02.19	"인재를 소비하기 보다는 키우는 개념을 가짐. 핵심인력이 되면 좋은 기회가 많아짐."	"조직이 비대해져 의사결정 체계가 복잡. 조직간 자나인 경쟁으로 협력이 어려워 큰 혁신을 이뤄내기 어렵다."
2016.02.26	"모든 것이 시스템으로 구성되어 있으며, 예외일이 원칙적으로 업무가 진행되고 있음."	"업무와 삶의 밸런스가 맞지 않으며, 많은 인사발령과 이에 따른 사별을 상대하기 위한 조직생활이 고전합니다."

Table 1. 성격요인별 기업의 특징

E (외향적 기업)		I (내향적 기업)	
조직 경계가 개방적이다.	조직 경계가 폐쇄적이다.	조직 경계가 소수만 참여한다.	조직 경계가 소수만 참여한다.
의사결정에 다수가 참여한다.	결정하고 나서 합의를 도출한다.	결정하고 나서 합의를 도출한다.	결정하고 나서 합의를 도출한다.
회의로 의사소통한다.	회의로 의사소통한다.	상황적이며 많은 시간을 보낸다.	상황적이며 많은 시간을 보낸다.
회의가 많다.	회의가 많다.	문자로 의사소통한다.	문자로 의사소통한다.
부서간 협력이 많다.	부서간 협력이 적다.	서면보고가 많다.	서면보고가 적다.
외부전문가에 의존한다.	내부적으로 문제를 해결한다.	부서간 협력이 적다.	내부적으로 문제를 해결한다.
다른 기업과의 교류가 많다.	각자 경쟁한다.	내부적으로 문제를 해결한다.	내부적으로 문제를 해결한다.
S (감정적 기업)		N (리스크적 기업)	
원리와 세분적인 것에 관심을 둔다.	미래와 가능성에 관심을 둔다.	미래와 가능성에 관심을 둔다.	미래와 가능성에 관심을 둔다.
반복적인 규칙을 강조한다.	직관을 강조한다.	직관을 강조한다.	직관을 강조한다.
감정적이고 단계적인 변화를 추구한다.	혁신적인 변화를 추구한다.	혁신적인 변화를 추구한다.	혁신적인 변화를 추구한다.
목표와 계획을 강조한다.	목표와 비전을 강조한다.	목표와 비전을 강조한다.	목표와 비전을 강조한다.
강령과 권위가 중요하다.	통찰력과 창의성이 중요하다.	통찰력과 창의성이 중요하다.	통찰력과 창의성이 중요하다.
상세한 정보에 의해 의사결정이 이루어진다.	전반적인 추세에 의해 의사결정이 이루어진다.	전반적인 추세에 의해 의사결정이 이루어진다.	전반적인 추세에 의해 의사결정이 이루어진다.
안정을 추구한다.	성장을 추구한다.	성장을 추구한다.	성장을 추구한다.
전통을 중시한다.	새로운 것을 만들어낸다.	새로운 것을 만들어낸다.	새로운 것을 만들어낸다.
표준화된 매뉴얼이 중요하다.	매뉴얼이 중요하지 않다.	매뉴얼이 중요하지 않다.	매뉴얼이 중요하지 않다.
T (사고적 기업)		F (감정적 기업)	
원리 원칙에 따라 결정한다.	인간의 가치관에 의해 결정한다.	인간의 가치관에 의해 결정한다.	인간의 가치관에 의해 결정한다.
규칙을 중시한다.	인간의 상황을 인정한다.	인간의 상황을 인정한다.	인간의 상황을 인정한다.
예외를 인정하지 않는다.	강령과 비전을 중시한다.	강령과 비전을 중시한다.	강령과 비전을 중시한다.
논리적 사고를 강조한다.	사람 중심적이다.	사람 중심적이다.	사람 중심적이다.
목표 중심적이다.	최선을 다하도록 돕는다.	최선을 다하도록 돕는다.	최선을 다하도록 돕는다.
비판적이다.	화해를 강조한다.	화해를 강조한다.	화해를 강조한다.
효율을 강조한다.	화해를 강조한다.	화해를 강조한다.	화해를 강조한다.
J (판단적 기업)		P (인식적 기업)	
결정을 내리도록 강요한다.	결정을 미룬다.	결정을 미룬다.	결정을 미룬다.
상급하게 결정한다.	신중하게 결정한다.	신중하게 결정한다.	신중하게 결정한다.
구체적으로 정의한다.	강의되지 않은 정보도 다룬다.	강의되지 않은 정보도 다룬다.	강의되지 않은 정보도 다룬다.
안정성을 선호한다.	의사결정에서 표의에 개방적이다.	의사결정에서 표의에 개방적이다.	의사결정에서 표의에 개방적이다.
결정을 바꾸지 않는다.	더 나은 선택의 여지를 남겨둔다.	더 나은 선택의 여지를 남겨둔다.	더 나은 선택의 여지를 남겨둔다.
의사결정의 신속성을 강조한다.	의사결정이 느리다.	의사결정이 느리다.	의사결정이 느리다.
결정이 상세하다.	결정이 불명확하다.	결정이 불명확하다.	결정이 불명확하다.

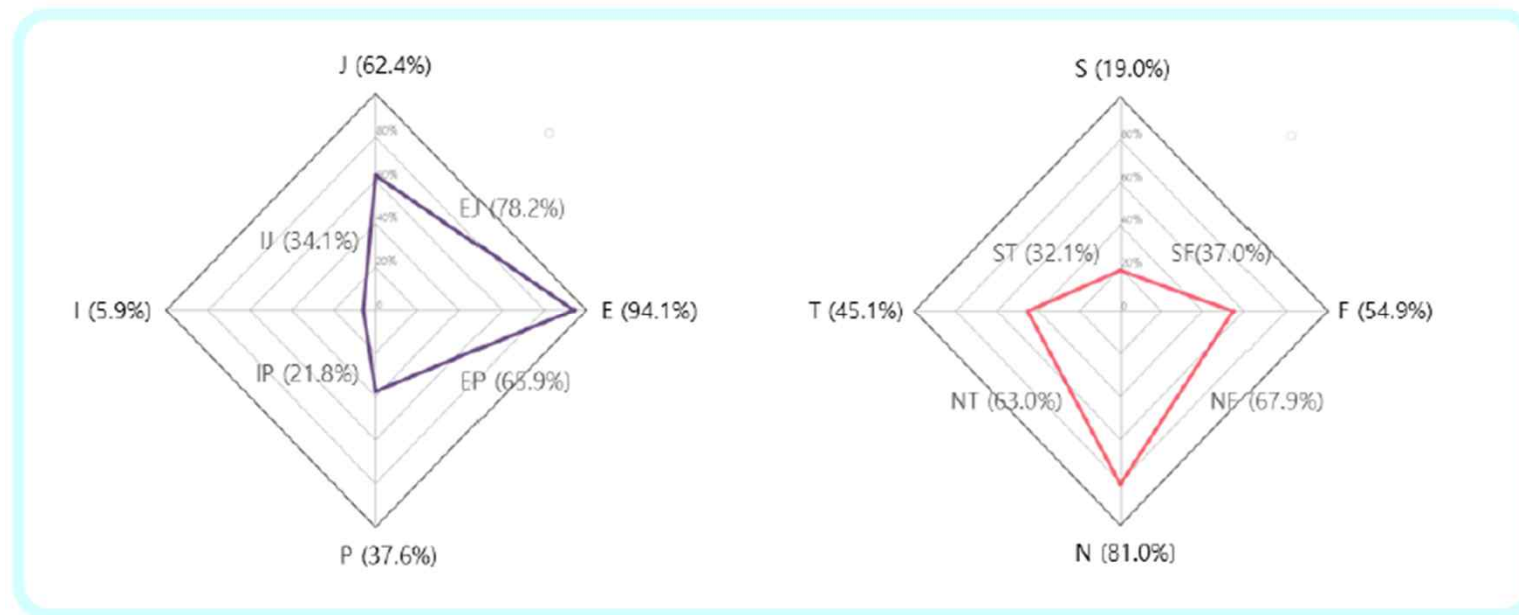


Table 4. 리뷰별 8개 기업 성격요인 Fit 측정 예시

리뷰	E	I	S	N	T	F	J	P
"대기업의 수직구조 문화, 인화의 엘지이지만 결국 술자리들 빠지지 않고 참석해야 하는 문화, 오래 앉아 있어야만 일 열심히 하는 것으로 인정받는 문화, 리프레쉬 기간이 많아 ..."	1.20572	1.06547	1.18936	1.23911	1.10819	1.24801	1.15760	1.13728
"부서장끼리 파워게임을 해서 일이 밀려오기도 한다. 임원 계약이 달려있어 단기간의 성과를 중시해서 끊임없이 뭔가 내려오기 때문에 개인이 뭔가 개선을 하기는 어렵다. 많은 사람들이 번아웃을 겪고있고 심리적으로 지쳐있는 경우가 많다. ...."	1.12319	1.06801	1.13526	1.13854	1.17278	1.22798	1.12993	1.19573

# Paper Critique

- Paper 2. 기업 MBTI 분류





# Paper Critique

- Paper 2. 기업 MBTI 분류

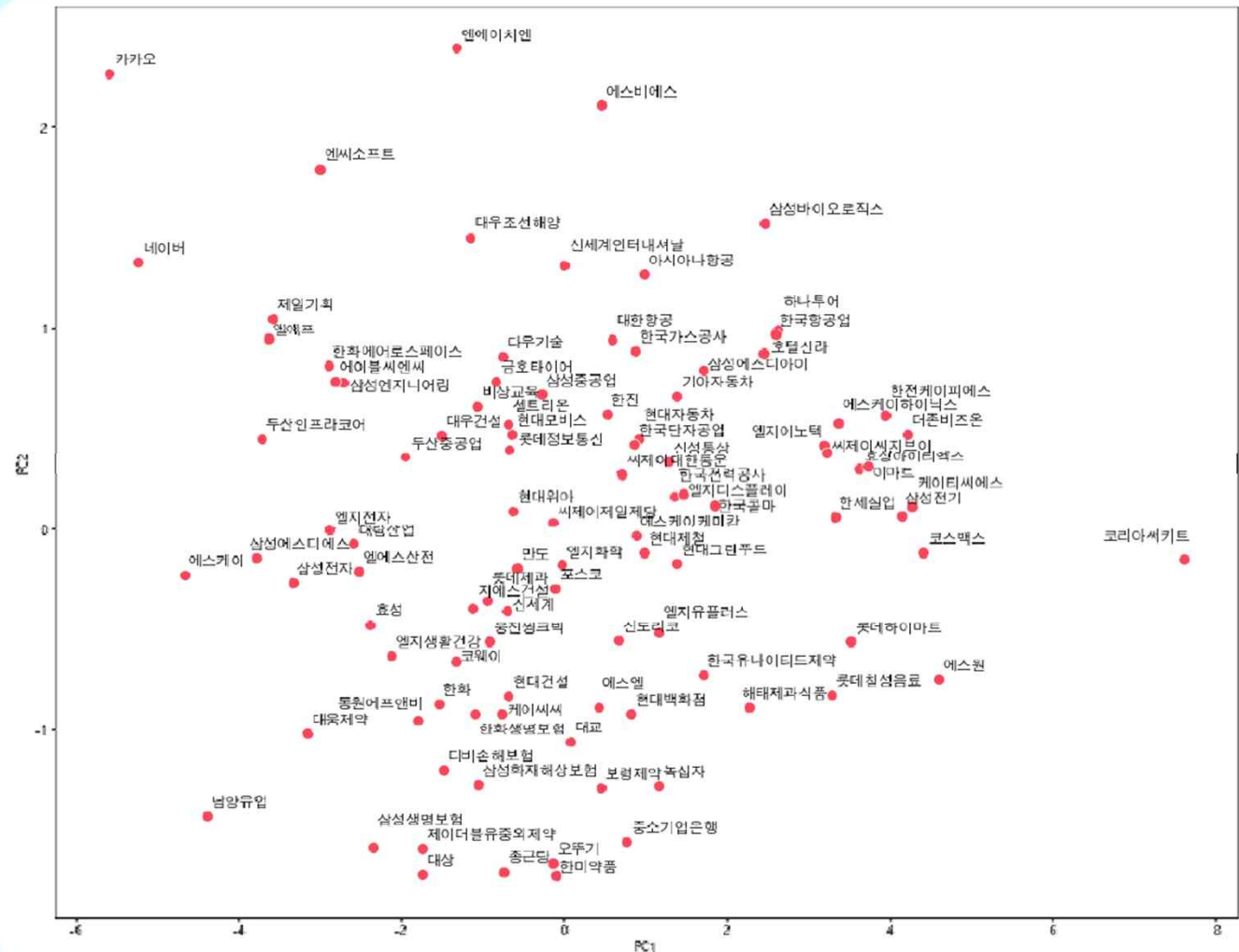
Fig. 6. 산업군별 기업 성격유형 분포



# Paper Critique

- Paper 2. 기업 MBTI 분류

Fig. 7. 기업규모 기준 100대 기업에 대한 성격유형 분포



# Paper Review

이 한 준 교수

# Introduction

- Research Background

## 취업시장의 새로운 스펙, MBTI?

‘ENFP, ESFJ 우대합니다.’



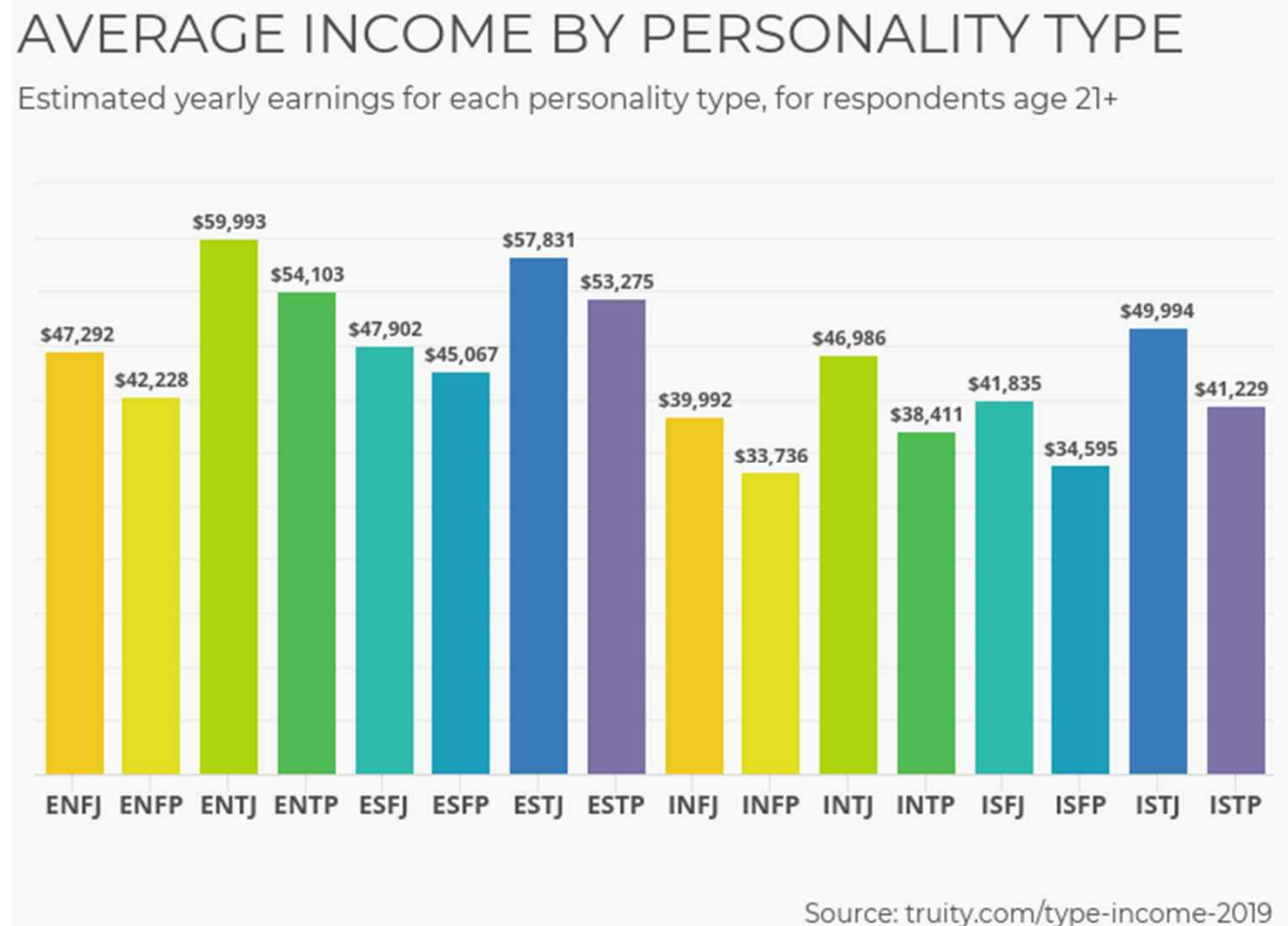
# Paper Critique

- Research Background



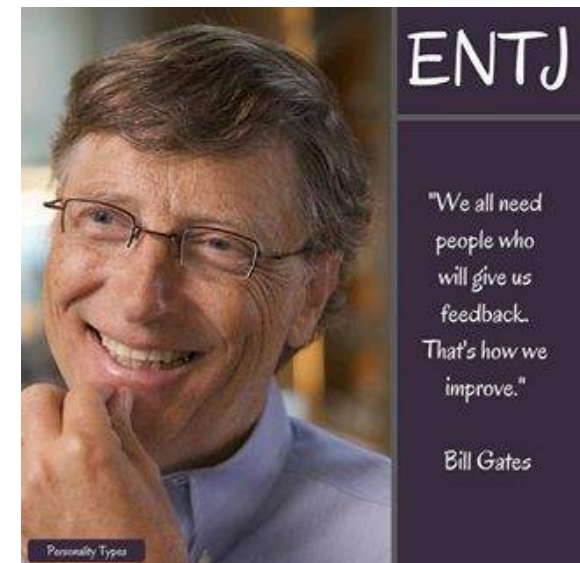
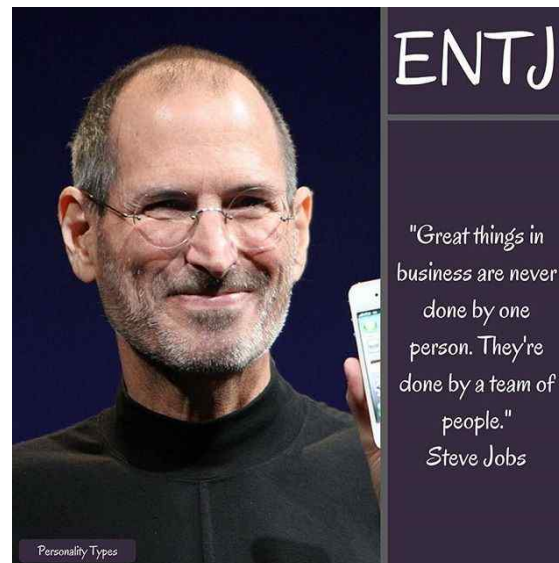
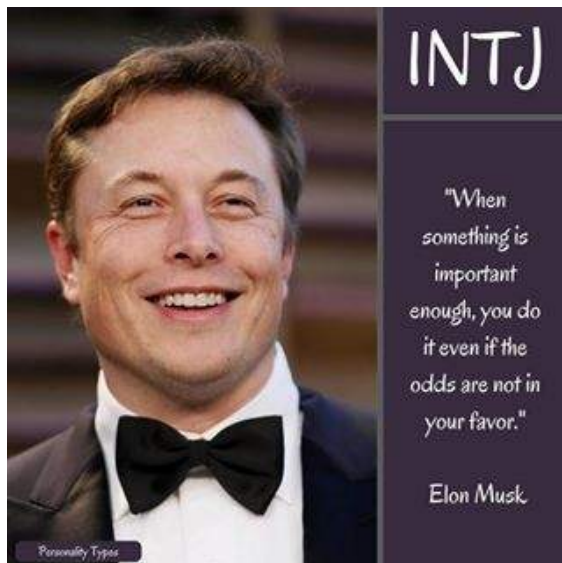
# Introduction

- Research Background
  - MBTI type and success



# Introduction

- Research Background
  - MBTI type and success

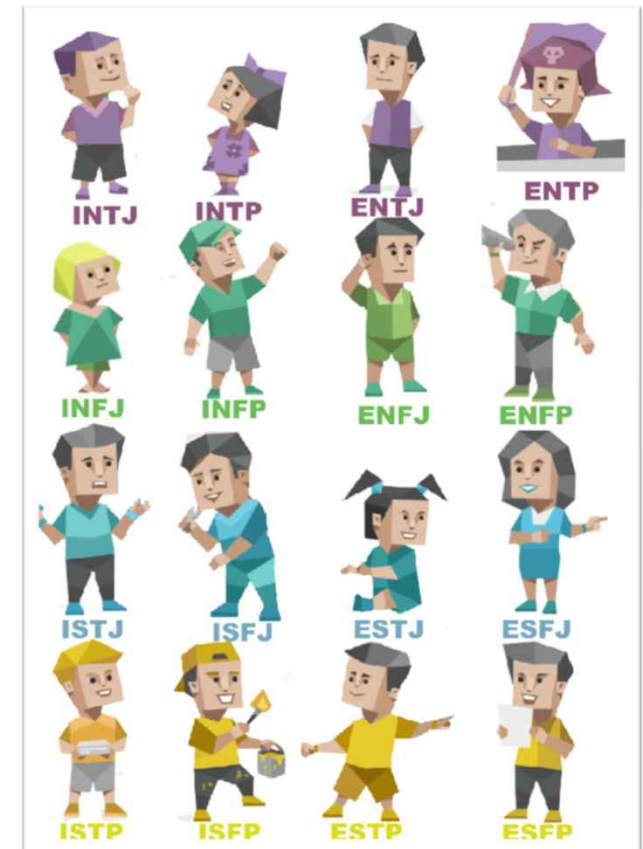




# Introduction

- MBTI (Myers-Briggs Type Indicator)

<b>E</b>	<b>Extroverts</b> are energized by people, enjoy a variety of tasks, a quick pace, and are good at multitasking.	<b>I</b>	<b>Introverts</b> often like working alone or in small groups, prefer a more deliberate pace, and like to focus on one task at a time.
<b>S</b>	<b>Sensors</b> are realistic people who like to focus on the facts and details, and apply common sense and past experience to come up with practical solutions to problems.	<b>N</b>	<b>Intuitives</b> prefer to focus on possibilities and the big picture, easily see patterns, value innovation, and seek creative solutions to problems.
<b>T</b>	<b>Thinkers</b> tend to make decisions using logical analysis, objectively weigh pros and cons, and value honesty, consistency, and fairness.	<b>F</b>	<b>Feelers</b> tend to be sensitive and cooperative, and decide based on their own personal values and how others will be affected by their actions.
<b>J</b>	<b>Judgers</b> tend to be organized and prepared, like to make and stick to plans, and are comfortable following most rules.	<b>P</b>	<b>Perceivers</b> prefer to keep their options open, like to be able to act spontaneously, and like to be flexible with making plans.





# Introduction

- Personality type literature
  - Using the MBTI in management and leadership (Walck, 1997)
  - The effect of personality type on team performance (Bradley & Hebert, 1997)
  - What makes a successful leader (Richmond et. al., 2004)
  - MBTI personality types of project managers and their success (Cohen et. al., 2013)
  - The myers-briggs type indicator (MBTI) and promotion at work (Furnham & Crump, 2015)
  - Top-Down Influence? Predicting CEO Personality and Risk Impact from Speech Transcripts (Theil, et al., 2022)

# Introduction

- Organizational character (Bridges, 1992)
  - "Organizations differ in their size, structure and purpose, and in character as well."
  - "A manufacturing company has a very different character from a new start-up IT company."
  - "As with personality type, organizational character can be established."

# Introduction

- Organizational Character and four opposing tendencies (Bridges, 1992)
  - An **Extroverted** organization looks outward to its markets, the environment in which it operates, its competition. A primarily **Introverted** organization is inwardly focused on its own culture, hierarchy, style, technology
  - An **iNtuitive** organization focuses on complexity, the bigger picture and unearthing new possibilities and opportunities; whereas **Sensing** organizations focus on current reality, events, required responses
  - The **Thinking** organization makes logical decisions grounded on established processes and systems. **Feeling** organizations operate from a basis of feelings
  - **Judging** organizations prefer clear 'black or white' answers, while **Perceiving** organizations seek context, interconnections, open-ended, and non-dualistic approaches

# Introduction

- Organizational Character Index (Bridges, 1992)
  - Survey-based company character measuring tool

방법론적 한계점

QUESTION	RATING			
1. Does the organization pay more attention to the demands of its customers or to what it knows how to do best?	Customers 1	2	3	What it does best 4
2. Is the organization better at producing and delivering established products/services or at planning or creating new ones?	Producing/ delivering 1	2	3	Planning/ creating 4
3. Which is more important to the organization: its efficient systems or its dedicated people?	Systems 1	2	3	People 4

1	2	3	4
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
TOTAL <input type="text"/>	TOTAL <input type="text"/>	TOTAL <input type="text"/>	TOTAL <input type="text"/>
[22 = E; 23 = I]	[22 = S; 23 = N]	[22 = T; 23 = F]	[22 = J; 23 = P]
According to the Organizational Character Index, your organization's character is			
<input type="text"/>			

# Introduction

- Possible approach
  - Company reviews contain clues to infer company character

유동/무역 | 전직원 | 서울 | 2021. 06

★★★★★

승진 기회 및 가능성

복지 및 급여

업무와 삶의 균형

사내문화

경영진

"맨땅에 헤딩하는데 머리만 깨지는 중. 매일 사람을 뽑지만 매일 사람이 나가며, 다들 자기 회사가 체계 없다고 하지만 이렇게 체계 없는 곳은 또 처음 봄."

**장점**

대외적 이미지? 남들 보기엔 유니콘 기업처럼 보일 수 있으나 그건 일부 직무 한정이고 그나마도 우형과는 다름.

**단점**

비효율의 끝판왕. 주먹구구식으로 문제를 해결하며 관리라고는 도대체 되지 않음. 거의 모든 직 기계약적으로 뽑으며 정규직 전환 케이스 본 적 없음. 일하는 사람만 일이 계속 생기고 노는 사람 속 놀며 일루. 소속감 제로에 급여는 거의 최저시급이며 근무환경 매우 열악함.

**경영진에 바라는 점**

소통을 중시하는 척, 급여와 복지 좋은 척 좀 그만 했으면.

이 기업을 추천하지 않습니다.

"개발자가 하고싶은 말을 할 수 있으며 협업의 의미가 있는 회사"

**장점**

업무에 필요한 모든건 지원해줌. 1인 1법카를 주며 자유로운 회식과 워라밸

**단점**

많은 의견이 존재하기 때문에 이를 취합하는 과정이 필요함. 수평적인 문화의 단점으로 작용하는 점은 여러 의견이 부딪히고 있을 때 누군가는 의견을 정해주어야 하는데 그러기 쉽지 않다는 점. 그 때문에 일정이 늦어지거나 마찰이 생기는 경우도 생김

"실무보단 기획을 해야하는 곳  
업계 중에선 가장 일다운 일을 하는 곳"

**장점**

눈치를 안보는 분위기  
파주라는 위치가 최대단점이었으나 청계사옥 이전후 위치 단점 사라짐

**단점**

보고를 위해 문서 작성하는 시간이 많이 걸림  
특히 회장님 보고는 디자이너까지 붙는 경우도 있음

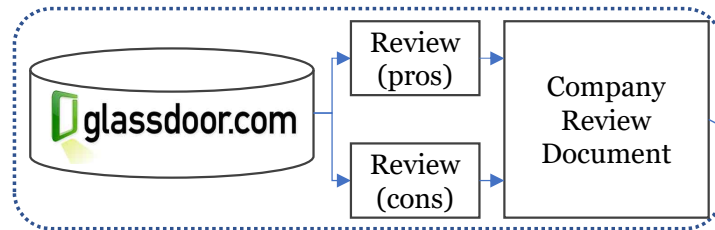
# Introduction

- Possible approach
  - Document Similarity Analysis
    - Very popular research topic in NLP
    - Distance calculation between two vectors (documents)
    - Similarity analysis requires vectorization
    - Various embedding models were developed: Doc2Vec(Word2Vec), BERT, GPT

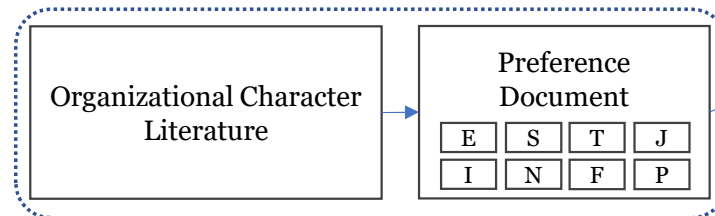
# Approach

- Overall Process

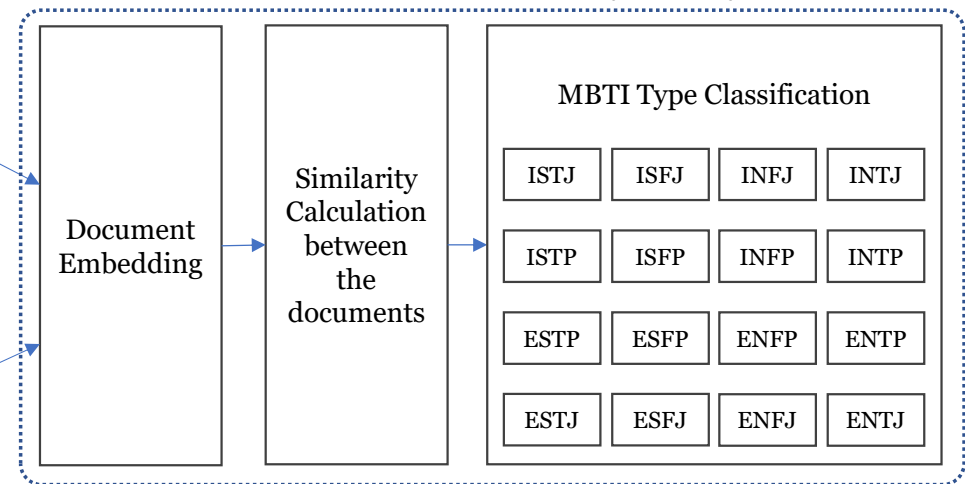
## ① Company Review Documentation



## ② Preference Documentation

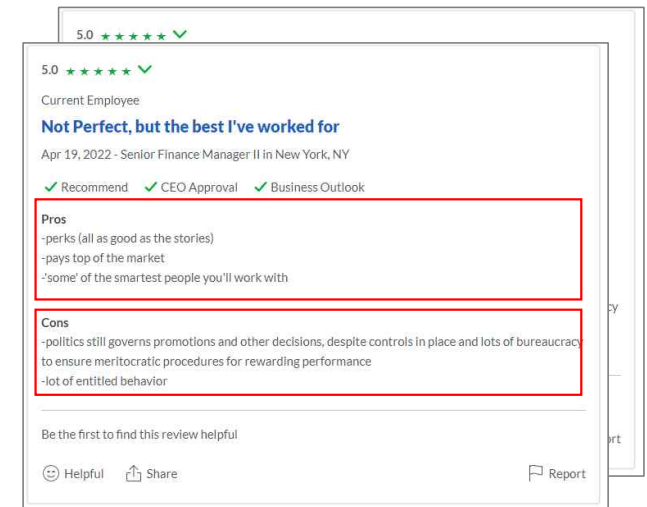


## ③ Document Similarity Analysis



# Approach

- Company Review Documentation
  - Glassdoor(<http://glassdoor.com>)
    - The most popular company review site
    - launched in 2008 and contains 50M+ reviews
  - Data Collection
    - S&P 1500 companies only were targeted
    - 548,619 reviews from 963 companies were collected
  - Preprocessing
    - Each review text were merged and tokenized
    - 7,624 Bi-gram and 439 tri-gram were extracted using Genism library





# Approach

- Documentation of Preferences (1/4)
  - Description for each preference were generated with reference to Bridges(1992) and related works

## **Extraversion (E)**

- Organizational boundaries are open.
- The majority participates in decision-making.
- See a consensus before making a decision.
- Communicate in the meeting.
- A lot of meetings
- There is a lot of collaboration.
- Rely on external experts.

## **Introversion (I)**

- Organizational boundaries are closed.
- Only a minority participate in decision-making.
- Make a decision and then reach an agreement.
- Communicate in writing.
- Many written reports.
- There is little collaboration.
- Troubleshoot problems internally.

# Approach

- Documentation of Preferences (2/4)

- Description for each preference were generated with reference to Bridges(1992) and related works

## **Sensing (S)**

- Pays attention to reality and details.
- Emphasize repetitive rules.
- Pursuing gradual and step-by-step changes.
- Emphasize goals and plans.
- Experience and authority are important.
- Decisions are made by detailed information.
- Seeking stability
- Tradition is valued.
- A standardized manual is important.

## **Intuition (N)**

- Pays attention to the future and possibilities.
- Emphasize intuition.
- Pursuing innovative changes.
- Emphasize purpose and vision.
- Insight and creativity are important.
- Decisions are made by the overall trend.
- Pursuing growth
- Create new things.
- Manual is not important.

# Approach

- Documentation of Preferences (3/4)
  - Description for each preference were generated with reference to Bridges(1992) and related works

## **Thinking (T)**

- Make decisions based on principles.
- Rules are valued.
- No exceptions are allowed.
- Emphasize logical thinking.
- Goal-oriented
- Help you achieve your goals.
- Emphasize efficiency.

## **Feeling (F)**

- Make decisions based on human values.
- Situation is considered.
- Admit exceptions.
- Emotional consideration is valued.
- People-oriented
- Help you do your best.
- Emphasize harmony.

# Approach

- Documentation of Preferences (4/4)
  - Description for each preference were generated with reference to Bridges(1992) and related works

## **Judging (J)**

- Forcing you to make decisions.
- Make hasty decisions.
- Define in detail.
- Prefer unanimity.
- Do not change decisions.
- Emphasize the speed of decision-making.
- The procedure is detailed.

## **Perceiving (P)**

- Postpone decision.
- Decide carefully.
- Leave undefined.
- Open to discussion in decision-making.
- Leaves room for better choices.
- Decision-making is slow.
- The procedure is unclear.

# Approach

- Document Similarity Analysis
  - Document embedding
    - Preference and review documents are embedded
    - GENSIM python library was used
  - Similarity calculation between two vectors

# Approach

- Document Similarity Analysis
  - Similarity calculation between two vectors
    - Cosine similarity
      - *distance* ranges from -1 to 1

$$distance_{cos}(v, s) = \frac{v \cdot s}{|v| \cdot |s|} = \frac{\sum_{i=1}^n v_i \cdot s_i}{\sqrt{\sum_{i=1}^n v_i^2} \cdot \sqrt{\sum_{i=1}^n s_i^2}}$$

# Approach

- Document Similarity Analysis
  - Similarity calculation between review document and preference document
    - *Similarity* equation was proposed to give weight for similar documents
      - *Similarity* ranges from 0.5 to  $\infty$

$$Similarity(d_1, d_2) = \frac{1}{1 - distance_{\cos}(v_{d1}, v_{d2})}$$

# Results

- Examples of reviews and similarity calculation results

Review		Similarity			
Pros	Cons				
“Project based, communicate with a lot of stakeholders from various companies, internal alignment is also strong.”	“Time management is crucial as workload is project based.”	E	<b>0.4474</b>	I	0.3039
		S	<b>0.2522</b>	N	0.2114
		T	<b>0.2753</b>	F	0.1690
		J	0.1780	P	<b>0.3085</b>
“Logical and rationale people. Always doing the right thing”	“Process driven. Slow to react/lead the market”	E	0.1910	I	<b>0.2479</b>
		S	0.2868	N	<b>0.3104</b>
		T	<b>0.4462</b>	F	0.2736
		J	<b>0.3623</b>	P	0.3632
“Smart people, flexible, working arrangements, value based decisions.”	“Difficult to advance at certain levels”	E	<b>0.2918</b>	I	0.2337
		S	<b>0.2562</b>	N	0.2151
		T	0.2703	F	<b>0.4610</b>
		J	0.2923	P	<b>0.3179</b>



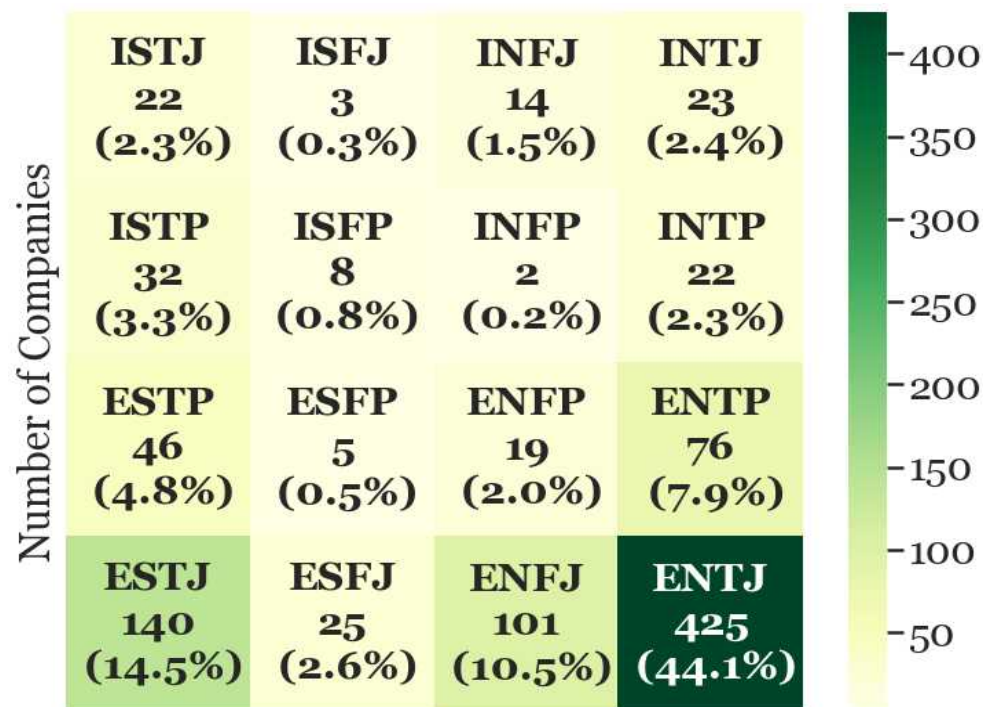
# Results

- Descriptive Statistics of Similarity Differences (N = 963)

	EI			SN			TF			JP		
	E	I	E-I	S	N	S-N	T	F	T-F	J	P	J-P
Mean	1.192	1.176	<b>0.016</b>	1.177	1.189	<b>-0.011</b>	1.210	1.200	<b>0.010</b>	1.230	1.216	<b>0.013</b>
Std.	0.023	0.021	<b>0.018</b>	0.022	0.024	<b>0.024</b>	0.022	0.025	<b>0.023</b>	0.022	0.028	<b>0.026</b>
Min	1.070	1.025	<b>-0.077</b>	1.034	1.037	<b>-0.197</b>	1.087	1.052	<b>-0.216</b>	1.035	1.035	<b>-0.132</b>
25%	1.180	1.166	<b>0.007</b>	1.168	1.178	<b>-0.024</b>	1.200	1.189	<b>0.002</b>	1.220	1.203	<b>0.001</b>
50%	1.192	1.175	<b>0.016</b>	1.179	1.189	<b>-0.010</b>	1.211	1.199	<b>0.011</b>	1.231	1.219	<b>0.012</b>
75%	1.203	1.186	<b>0.025</b>	1.189	1.200	<b>0.001</b>	1.222	1.210	<b>0.018</b>	1.241	1.231	<b>0.024</b>
Max	1.424	1.445	<b>0.149</b>	1.344	1.461	<b>0.134</b>	1.350	1.462	<b>0.138</b>	1.358	1.402	<b>0.249</b>

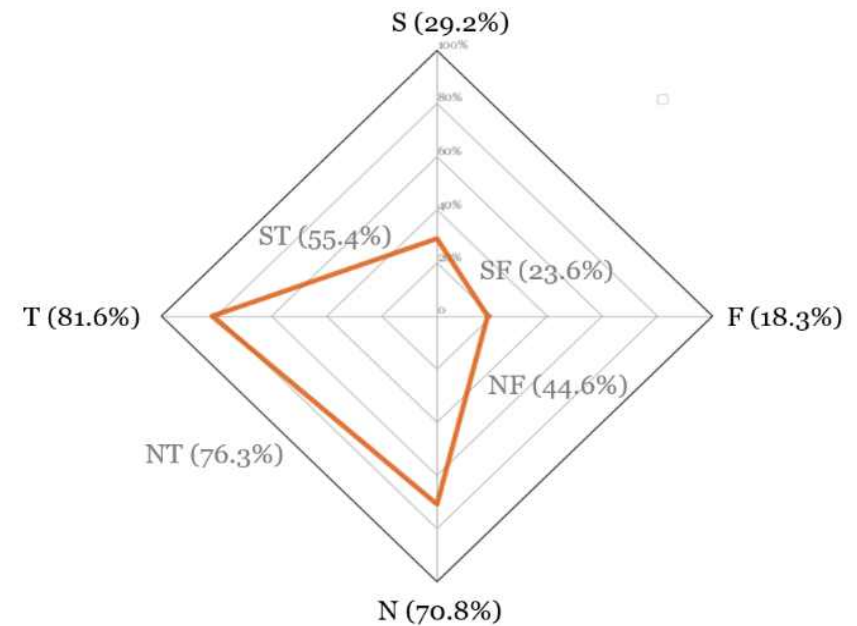
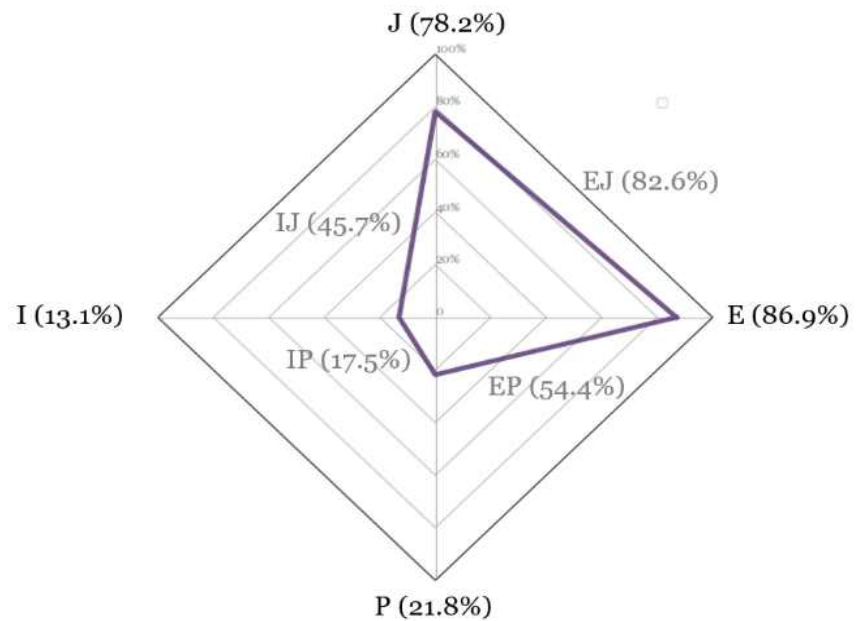
# Results

- Distribution of Company Character Type



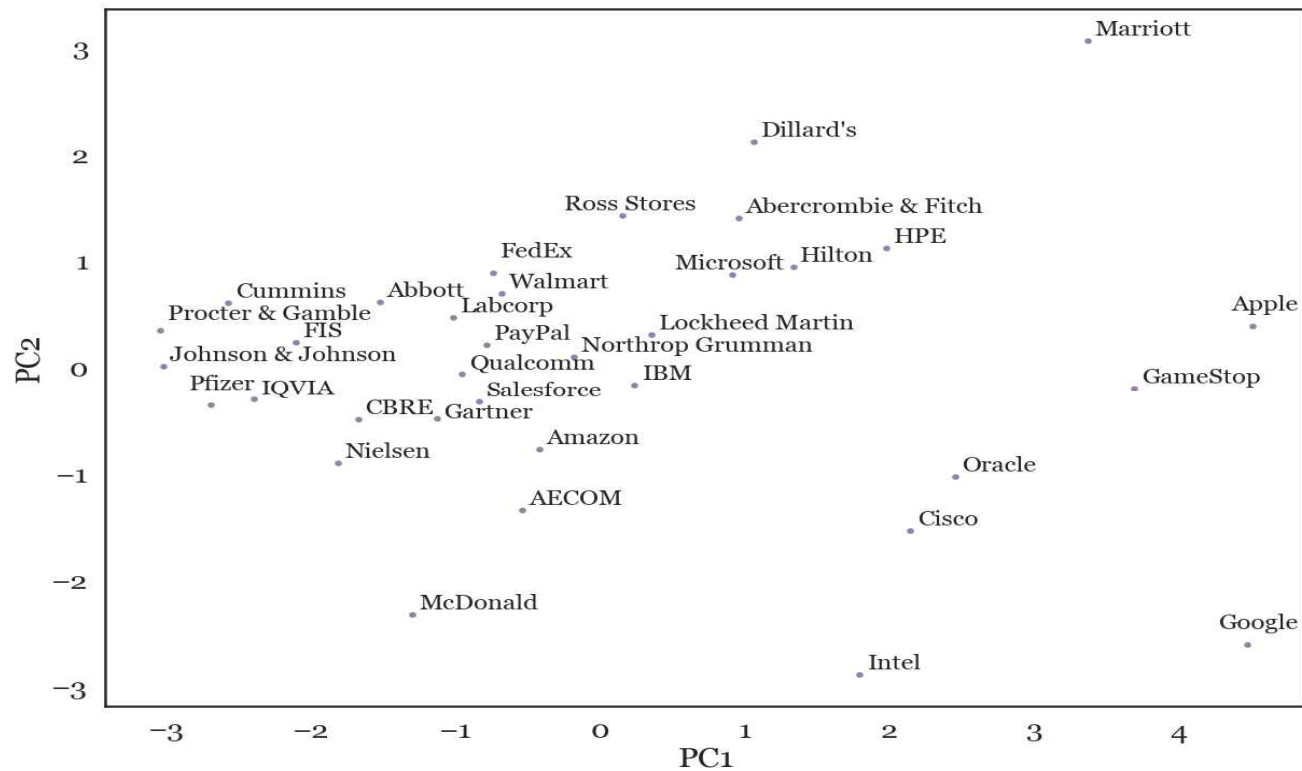
# Results

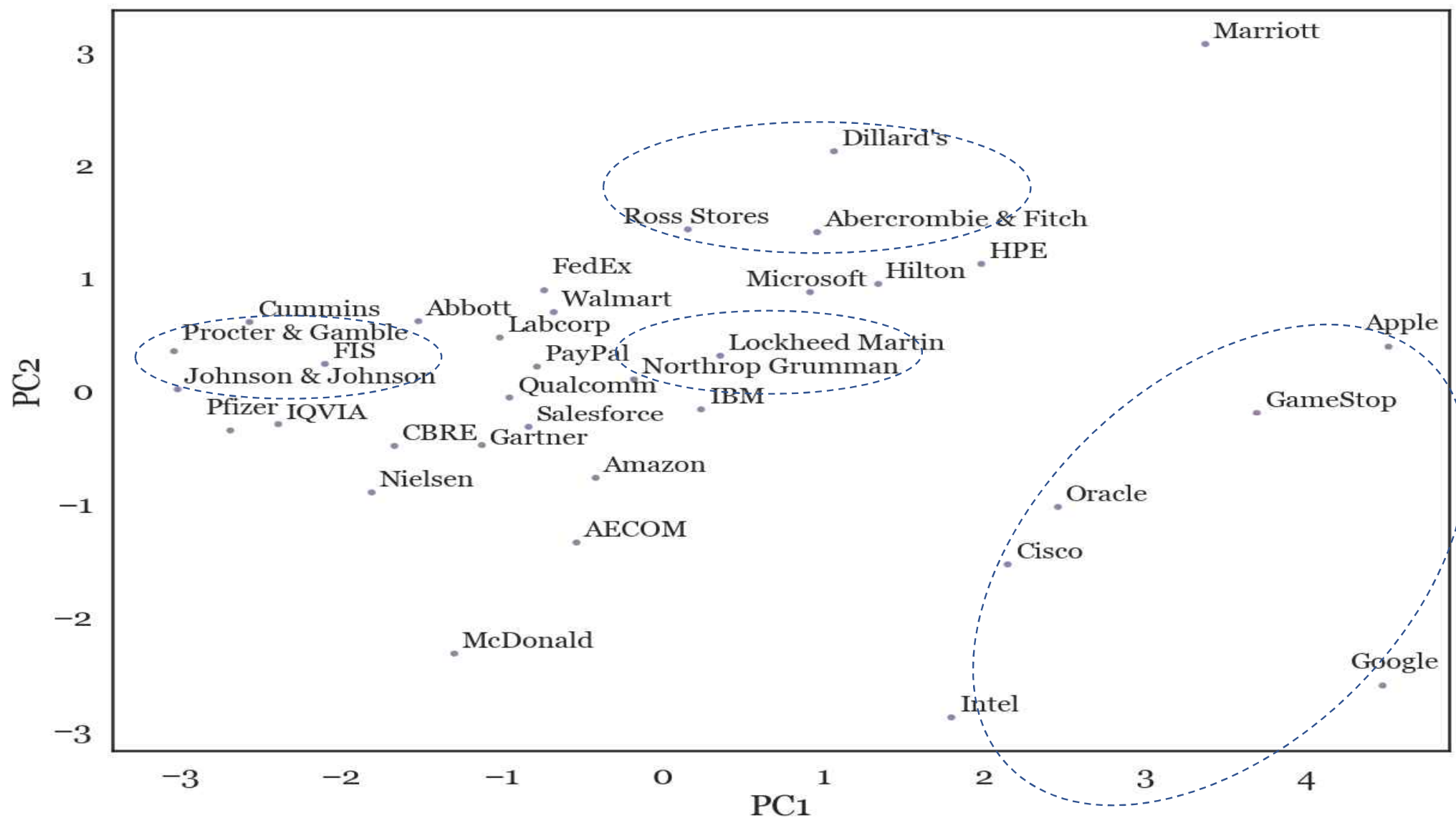
- Distribution of behavioral and decision-making types



# Results

- Positioning Map of Companies (N = 36)





# Discussion

- 'ENTJ' type found to be the most common organizational character among the target companies
  - Technical entrepreneurs are more extroverted (E), intuitive (N), thinking (T) and judging (J) (Roberts, 1989)
  - 'ENTJ' type is the most common character of chief executive officers by analyzing speech transcriptions of the officers (Theil, et. al., 2022)
- 'ENTJ', 'ESTJ' and 'ENTP' types are the top-3 prevalent types
  - A variety of officers, such as entrepreneurs, small business owners and managers, commonly have the three types of character (Reynierse, 1997)

# Discussion

- About 87% of companies are *Extroverted* and about 82% of ones are *Thinking*-oriented
  - Small company entrepreneurs, small bus owners and managers tend to be more *Extroverted* (Carland & Carland, 1992)
  - 87% of Inc. 500 founders are *Thinking*-oriented (Ginn & Sexton, 1990)
  - 'T' is a dominant character among various officers (Carland & Carland, 1992; Reynierse, 1997)

# Discussion

- 'EJ' type and 'NT' type occupy 82.6% and 76.3% respectively
  - Lower-level managers prefer the 'EJ' against the 'IJ' , 'IP' and 'EP' (Reynierse, 1997)
  - 51% of CEOs have the 'NT' with a future-oriented tendency in their decision-making (Ginn & Sexton, 1990)



# Discussion

- Types of industry seems to be associated with organizational character
  - The character of company differs according to the industry (Bridges, 1992)
- IT companies have relatively different character comparing to traditional companies
  - IT-related workers, such as data processing employees, system analysts, programmer analysts, and programmers, may have different personality profiles than the general population (Woodruff, 1980)

# Implication

- Academic implication
  - A novel approach to measure company's character relying on text analysis
    - Can enable empirical studies that have been difficult in organizational character area
  - Can be a good case study utilizing document similarity analysis
    - Document similarity analysis is a useful methodology, but has rarely been used in business field

# Implication

- Practical implication
  - For managers,
    - Decision makers and HR managers can obtain an insight for their management and establishing strategies
    - The proposed method can be utilized as an organization diagnosis
  - For job-seekers,
    - Organizational character can be an important factor for their job application decision

# Conclusion

- Limitation
  - Possible bias due to the single data source
  - Assumption that company character rarely changes
- Future research topics
  - Relationship between company character and financial performance, turnover rate, and employee satisfaction
  - Impact of business environmental change on company character
  - Company culture measuring method development