

Lecture 09 - Contents

Overview of the lecture on few/zero-shot learning and medical applications.

Part 1

Zero-shot & Prompting

Part 2

Few-shot Learning

Part 3

Medical Applications

Hands-on

Prompt Engineering Hands-on

This outline is for guidance. Navigate the slides with the left/right arrow keys.

Lecture 9:

Few-Shot and Zero-Shot Medical Applications

Medical AI with Limited Data

Ho-min Park

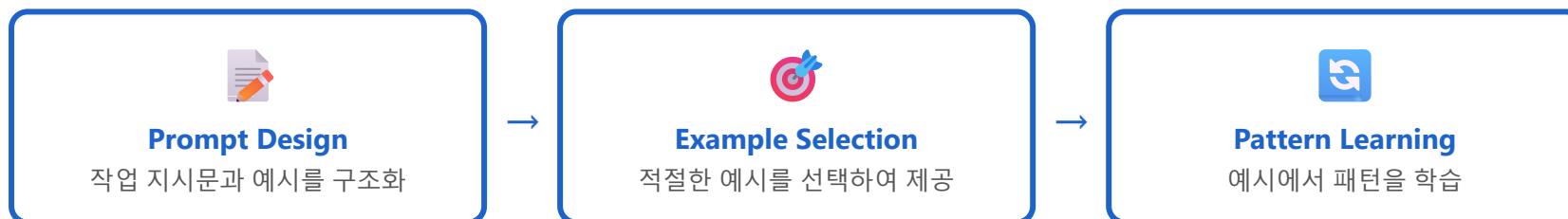
homin.park@ghent.ac.kr

powersimmani@gmail.com

In-Context Learning Overview

In-Context Learning (ICL)은 모델을 재학습하지 않고 **프롬프트에 예시를 포함**하여 작업을 수행하는 기법입니다.

ICL 메커니즘



의료 분야 적용



Part 1/3:

Zero-Shot Medical AI

- 1.** Zero-Shot Prompting Techniques
- 2.** Task Instruction Design
- 3.** Role Prompting for Medical Contexts
- 4.** Constraint Specification
- 5.** Output Format Control
- 6.** Cross-Lingual Transfer
- 7.** Domain Adaptation Strategies

Zero-Shot Prompting

Zero-Shot is a technique where the model performs new tasks with only task instructions, without any examples.

Design Principles



Clear Instructions

Describe tasks specifically and clearly



Structured Output

Specify the desired output format



Provide Context

Include background information needed for task execution



Constraints

Specify constraints for safety and quality



Medical Example: Symptom-Based Initial Assessment

You are a medical AI assistant. Analyze the following symptoms and provide a preliminary assessment with possible conditions (ranked by likelihood):

Symptoms: persistent cough, fever, fatigue

Output format: JSON with conditions array

✓ Clear role and task definition, structured output format specified

Task Instruction Design

Effective **task instructions** should include clarity, structure, and context.

Instruction Patterns



Direct

Direct imperative instructions



Step-by-Step

Sequential execution procedures



Conditional

Includes conditional logic



Role-Based

Role-based instructions



Template

Template-based generation



Contextual

Utilizes contextual information

Effectiveness Factors

✓ Effective Instructions

- ✓ Specific and measurable goals
- ✓ Clear input/output format definition
- ✓ Explicit necessary constraints
- ✓ Use of domain-specific terminology

✗ Ineffective Instructions

- ✗ Vague and abstract requests
- ✗ Unspecified output format
- ✗ Overly complex multi-tasks
- ✗ Generic and unprofessional expressions

Role Prompting for Medical Contexts

Role prompting assigns an expert persona to AI to elicit more accurate and contextually appropriate responses.



Clinical Specialist

Role of a specialist in a specific clinical field

"You are a cardiologist with 15 years of experience..."



Research Analyst

Expert in medical research and literature analysis

"You are a medical researcher analyzing clinical trials..."



Pharmacologist

Expert in drug actions and interactions

"You are a clinical pharmacologist evaluating drug interactions..."



Medical Coder

Expert in medical coding and classification

"You are a certified medical coder specializing in ICD-10..."

Performance Difference (Role-based vs. Generic Prompts)

Diagnostic Accuracy:

+18%

Response Relevance:

+25%

Technical Terminology:

+32%

Context Understanding:

+22%

Constraint Specification

Constraints are critical elements that ensure the safety, quality, and format of AI outputs.



Safety Constraints

Prevent dangerous recommendations



Length Constraints

Limit output length



Format Constraints

Adhere to specific structure



Scope Constraints

Answer within specific topics



Accuracy Constraints

Evidence-based responses



Ethical Constraints

Comply with medical ethics



Safety Rules Examples

- ▶ Specify that diagnosis requires physician confirmation
- ▶ Recommend immediate hospital visit in emergencies
- ▶ Prohibit self-treatment recommendations
- ▶ Restrict provision of prescription information



Output Format Examples

- ▶ Use JSON or XML structure
- ▶ Include confidence scores
- ▶ Cite supporting literature
- ▶ Present key information first

Output Format Control

출력 형식 제어

출력 형식 제어의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

지원 형식



JSON

구조화 데이터



XML

계층적 데이터



Markdown

문서 형식



CSV

표 형식

제어 방법

🎯 명시적 지정

- 형식 템플릿 제공
- 출력 스키마 정의
- 예시 구조 포함
- 필드 명세 명시

✓ 검증 규칙

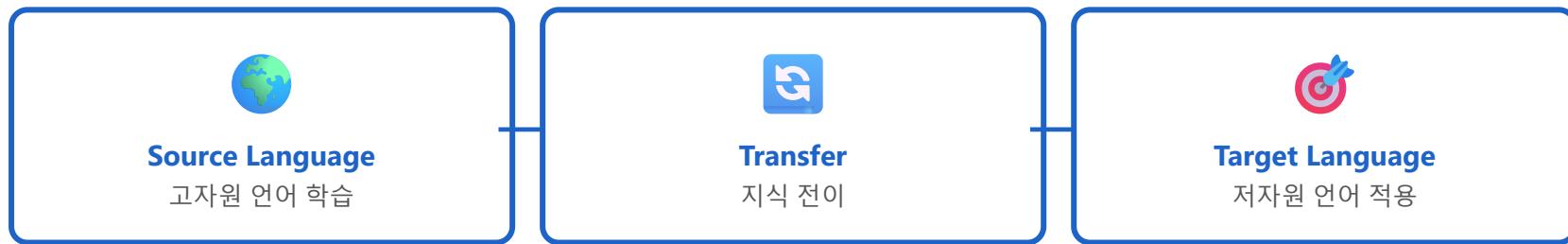
- 형식 일관성 체크
- 필수 필드 검증
- 데이터 타입 확인
- 구조적 무결성

Cross-Lingual Transfer

다언어 전이 학습

다언어 전이 학습의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

전이 메커니즘



전이 전략

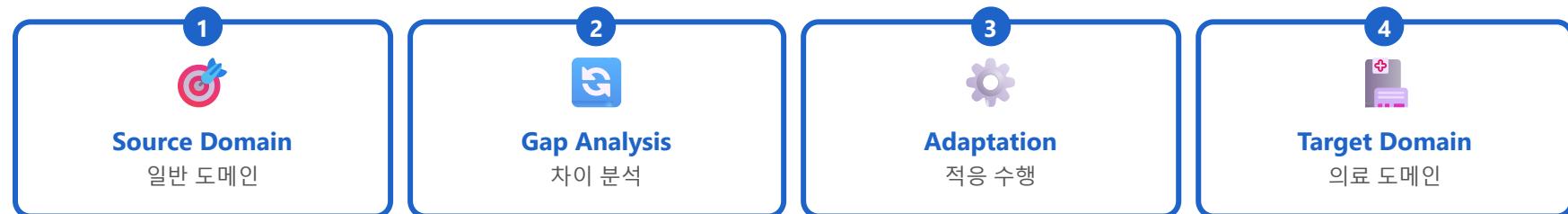


Domain Adaptation

도메인 적응

도메인 적응의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

적응 프로세스



적응 기법

Feature-Level

- 특성 공간 정렬
- 도메인 불변 표현
- 절대적 학습
- 최대 평균 불일치

Instance-Level

- 인스턴스 재가중치
- 샘플 선택
- 데이터 증강
- 합성 데이터 생성

Part 2/3:

Few-Shot Medical Learning

Example Selection Strategies

예시 선택 전략

예시 선택 전략의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.



핵심 포인트

의료 데이터의 특수성을 고려한 접근

제한된 데이터 환경에서의 효율적 학습

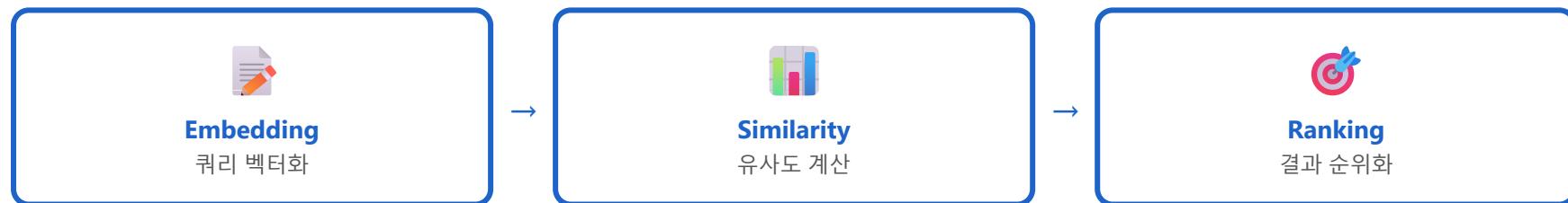
안전성과 신뢰성을 보장하는 설계

Similarity-Based Retrieval

유사도 기반 검색

유사도 기반 검색의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

검색 프로세스



주요 기법

1 2 3 4 유사도 측정

- Cosine Similarity
- Euclidean Distance
- Semantic Similarity
- Domain-specific Metrics

의료 응용

- 유사 증례 검색
- 관련 문헌 추출
- 진단 패턴 매칭
- 치료법 추천

Diversity Sampling

다양성 샘플링

다양성 샘플링의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

샘플링 전략



Random Sampling

무작위 다양성 확보



Stratified

계층적 균형 샘플링



Cluster-Based

클러스터 기반 선택



Maximal Diversity

최대 다양성 추구



Balanced

클래스 균형 유지



Feature-Based

특성 기반 다양화

핵심 이점



다양성의 장점

일반화 성능 향상

편향 감소

포괄적 학습



의료 적용

다양한 증상 패턴 커버

희귀 질환 포함

인구 통계학적 균형

견고성 증가

치료 반응 다양성

Dynamic Example Selection

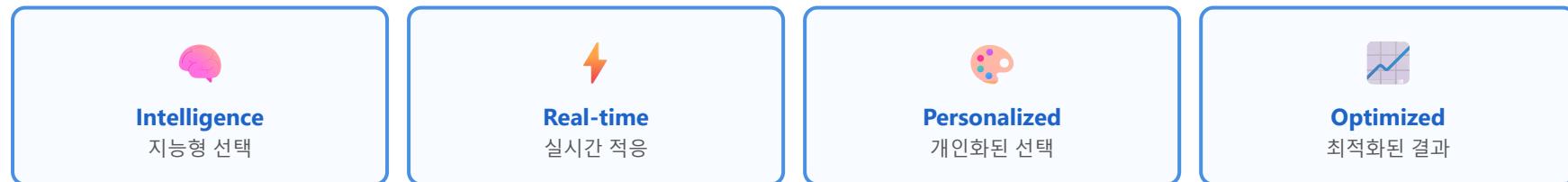
동적 예시 선택

동적 예시 선택의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

적응형 선택 프로세스



주요 특징



Template Engineering

템플릿 엔지니어링

템플릿 엔지니어링의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.



주요 개념

템플릿 엔지니어링의 이론적 배경과 핵심 원리



의료 적용

실제 의료 환경에서의 활용 사례



성능 평가

효과성 측정 및 개선 방법론



실습 예제

실제 데이터를 활용한 구현

핵심 포인트

의료 데이터의 특수성을 고려한 접근

제한된 데이터 환경에서의 효율적 학습

안전성과 신뢰성을 보장하는 설계

Chain Prompting

체인 프롬프팅

체인 프롬프팅의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.



주요 개념

체인 프롬프팅의 이론적 배경과 핵심 원리



의료 적용

실제 의료 환경에서의 활용 사례



성능 평가

효과성 측정 및 개선 방법론



실습 예제

실제 데이터를 활용한 구현

핵심 포인트

의료 데이터의 특수성을 고려한 접근

제한된 데이터 환경에서의 효율적 학습

안전성과 신뢰성을 보장하는 설계

Part 3/3:

Clinical Applications

Rare Disease Diagnosis

Diagnosis of Rare Diseases

Learn about the core concepts of **Rare Disease Diagnosis** and the application of medical data science.

Diagnosis Pipeline



Key Elements



Challenges

- Data Scarcity (Rarity)
- Complex Symptom Patterns
- Long Diagnosis Period
- Lack of Specialists



AI Solutions

- Few-Shot Learning Application
- Similar Case Retrieval
- Rapid Early Diagnosis
- Knowledge-Based Augmentation

Rare diseases are very difficult to diagnose due to **small patient populations and diverse symptoms**. On average, it takes **5-7 years** to reach an accurate diagnosis, during which patients cannot receive appropriate treatment.

Recent **advances in AI technology** have made it possible to address these challenges. Few-Shot Learning enables learning with limited data, and similar case retrieval allows utilization of past diagnostic cases. Additionally, integration with medical knowledge graphs has

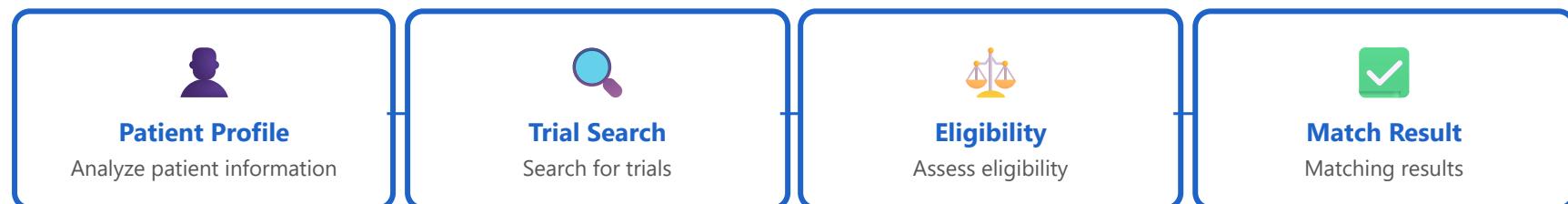
enabled more accurate diagnoses.

Clinical Trial Matching

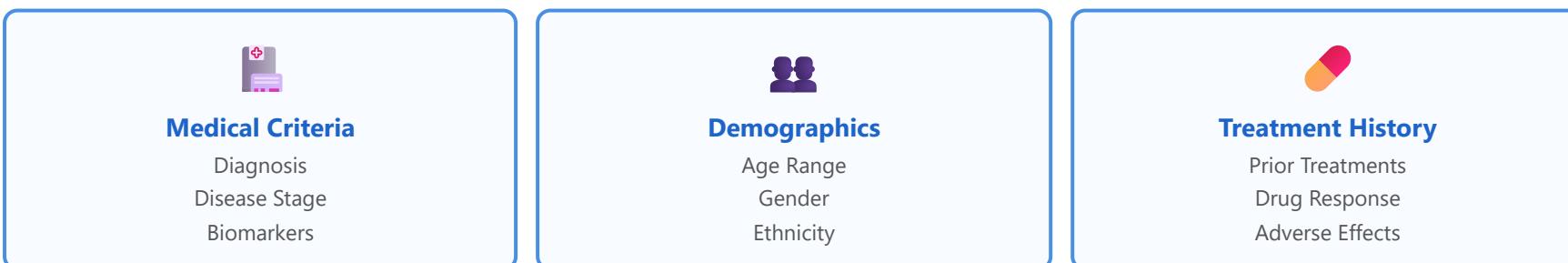
Connecting Patients with Appropriate Clinical Trials

Learn about the core concepts of **Clinical Trial Matching** and its application in medical data science.

Matching Process



Matching Criteria



💡 What is Clinical Trial Matching?

Clinical trial matching is a system that automatically identifies eligible clinical trials for patients by analyzing their medical information. It extracts data such as diagnosis, disease stage, and biomarkers from patient profiles, then compares this information against the inclusion and exclusion criteria of clinical trials to recommend suitable trials. Medical data science technologies enable efficient and accurate matching throughout this process.

Through a data science approach, the most appropriate options for patients can be quickly identified from thousands of clinical trials, which helps expand treatment opportunities for patients and increase participation rates in clinical trials.

Adverse Event Detection

부작용 탐지

부작용 탐지의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.



주요 개념

부작용 탐지의 이론적 배경과 핵심 원리



의료 적용

실제 의료 환경에서의 활용 사례



성능 평가

효과성 측정 및 개선 방법론



실습 예제

실제 데이터를 활용한 구현

핵심 포인트

- 의료 데이터의 특수성을 고려한 접근
- 제한된 데이터 환경에서의 효율적 학습
- 안전성과 신뢰성을 보장하는 설계

부작용 탐지의 중요성

부작용 탐지(Adverse Event Detection)는 의료 데이터 과학에서 환자 안전을 보장하는 핵심 기술입니다. 전자건강기록(EHR)과 같은 대규모 의료 데이터에서 약물이나 치료와 관련된 예상치 못한 부작용을 조기에 발견함으로써, 의료진은 신속한 대응과 치료 계획 수정을 통해 환자의 건강을 보호할 수 있습니다.

머신러닝과 자연어 처리 기술을 활용하여 임상 노트, 검사 결과, 투약 기록 등 다양한 형태의 의료 데이터를 분석하고, 패턴 인식을 통해 잠재적 위험 신호를 자동으로 식별합니다. 이는 기존의 수동적 보고 시스템의 한계를 극복하고, 보다 포괄적이고 즉각적인 약물 감시 체계를 구축하는 데 기여합니다.

Medical Coding Automation

의료 코딩 자동화

의료 코딩 자동화의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.



주요 개념

의료 코딩 자동화의 이론적 배경과 핵심 원리



의료 적용

실제 의료 환경에서의 활용 사례



성능 평가

효과성 측정 및 개선 방법론



실습 예제

실제 데이터를 활용한 구현

핵심 포인트

의료 데이터의 특수성을 고려한 접근

제한된 데이터 환경에서의 효율적 학습

안전성과 신뢰성을 보장하는 설계

Literature Summarization

문헌 요약

문헌 요약의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.



주요 개념

문헌 요약의 이론적 배경과 핵심 원리



의료 적용

실제 의료 환경에서의 활용 사례



성능 평가

효과성 측정 및 개선 방법론



실습 예제

실제 데이터를 활용한 구현

핵심 포인트

의료 데이터의 특수성을 고려한 접근

제한된 데이터 환경에서의 효율적 학습

안전성과 신뢰성을 보장하는 설계

Performance Analysis

성능 분석

성능 분석의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

핵심 평가 메트릭



Accuracy

정확도 측정



Precision

정밀도 평가



Recall

재현율 분석



F1-Score

종합 성능

분석 프레임워크



정량적 분석

- 통계적 유의성 검정
- Cross-validation 결과
- 벤치마크 비교
- 성능 추세 분석



정성적 분석

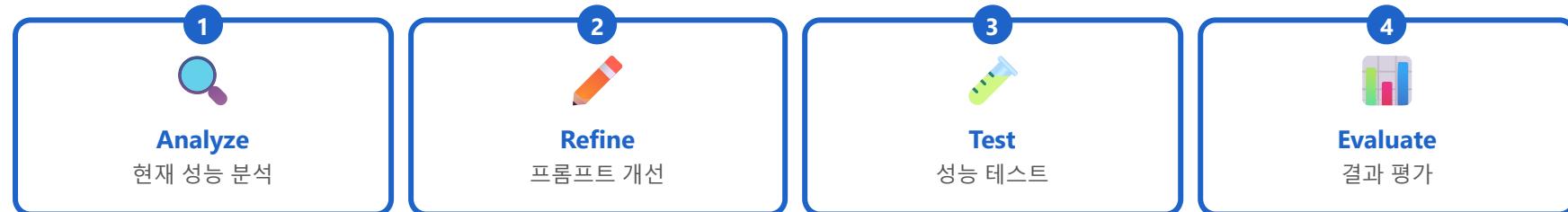
- 오류 사례 분석
- 임상적 타당성 검증
- 전문가 평가
- 사용자 피드백

Prompt Optimization

프롬프트 최적화

프롬프트 최적화의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

최적화 사이클



최적화 기법

🎯 구조적 최적화

- 명확한 지시문 작성
- 예시 품질 개선
- 출력 형식 표준화
- 맥락 정보 최적화

⚡ 성능 최적화

- 토큰 효율성 향상
- 응답 시간 단축
- 배치 처리 최적화
- 캐싱 전략 적용

Active Learning Integration

능동 학습 통합

능동 학습 통합의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

피드백 루프



Query Selection

불확실성 높은 샘플 선택



Expert Labeling

전문가 라벨링



Model Update

모델 성능 향상

주요 이점



효율성

레이블링 비용 감소



정확도

모델 성능 개선



적응성

지속적 학습

Case Study: Rare Genetics

Medical Data Science Application

Learn about the core concepts of **case studies** and their application in medical data science.



Key Concepts

Theoretical background and core principles of case studies



Medical Application

Real-world use cases in medical environments



Performance Evaluation

Effectiveness measurement and improvement methodology



Practical Examples

Implementation using real-world data

Key Points

- Approach considering the specificity of medical data
- Efficient learning in limited data environments

- Design ensuring safety and reliability

Rare Genetics Case Study Overview

Approximately 7,000 rare genetic diseases are known worldwide, and while individual disease incidence rates are low, they collectively affect about 6-8% of the population, making them an important public health issue. This case study covers data science approaches for diagnosing and treating rare genetic diseases, and teaches methods for building effective analysis and predictive models in limited data environments.

Real-World Application Examples

For certain rare genetic diseases, there may be only a few hundred patients worldwide. In such situations, traditional large-scale data-driven machine learning approaches are difficult to apply. Therefore, innovative approaches using transfer learning, meta-learning, and few-shot learning techniques are necessary.

Research Methodology

1. Data Collection & Integration

Collect and integrate information from various sources (genomic data, clinical records, imaging data) to construct comprehensive patient profiles.

2. Feature Extraction & Selection

Identify key biomarkers and genetic variants related to diseases from high-dimensional genomic data and apply dimensionality reduction techniques.

3. Model Development & Validation

Use ensemble techniques and regularization methods suitable for small datasets to prevent overfitting and

4. Clinical Application & Evaluation

Validate developed models in actual clinical settings and evaluate performance using metrics such as sensitivity,

improve generalization performance.

specificity, and positive predictive value.

Learning Objectives & Expected Outcomes

Through this case study, learners will understand the unique challenges of rare disease research and develop the ability to perform effective analysis even in limited data situations. In particular, the following practical skills can be cultivated:

- ▶ **Data Augmentation Techniques:** Methods to effectively expand training data from limited samples
- ▶ **Domain Knowledge Integration:** Strategies to incorporate biological and clinical knowledge into models
- ▶ **Ensuring Interpretability:** Building explainable AI models that clinicians can trust
- ▶ **Ethical Considerations:** Compliance with patient privacy protection and data security

Future Outlook

The use of AI and data science in the field of rare genetic diseases continues to expand. Through data sharing via multi-institutional collaborative networks, advances in synthetic data generation technology, and the introduction of privacy-preserving technologies such as federated learning, accurate diagnosis and personalized treatment will become available to more patients.

Hands-On: Prompt Engineering

Practical Workshop

Learn about core concepts of **Prompt Engineering** and its application in medical data science.



Key Concepts

Theoretical background and core principles of the workshop



Medical Applications

Real-world use cases in healthcare environments



Performance Evaluation

Effectiveness measurement and improvement methodologies



Practical Examples

Implementation using real-world data



What is Prompt Engineering?

Prompt engineering is the technique of systematically designing and optimizing input instructions to effectively communicate with AI language models. In the medical field, where accuracy and safety are directly linked to patient lives, designing clear, specific, and context-aware prompts is essential.

Applications in Medical Data Science

When applying AI to various medical tasks such as patient diagnosis assistance, medical image analysis, clinical note summarization, and drug interaction checking, proper prompt design significantly enhances model performance. It is also important to write prompts that consider medical ethics, privacy protection, and clinical validity, while clearly defining the role as a tool to assist healthcare professionals' final judgment.

Key Points

- Approach considering the specificity of medical data
- Efficient learning in limited data environments
- Design ensuring safety and reliability

Evaluation Frameworks

평가 프레임워크

평가 프레임워크의 핵심 개념과 의료 데이터 과학 적용에 대해 학습합니다.

평가 레이어



Task-Level Evaluation

작업별 성능 평가: 정확도, 완성도, 관련성



Clinical Validation

임상적 타당성: 의학적 정확성, 안전성, 실용성



Efficiency Metrics

효율성 평가: 응답 시간, 리소스 사용, 비용

평가 차원



정량적 평가

F1-Score
AUROC
Specificity



정성적 평가

전문가 검토
사용성 평가
오류 분석



윤리적 평가

공정성
투명성
책임성

Thank you

Few-Shot and Zero-Shot Medical AI

Ho-min Park

homin.park@ghent.ac.kr

powersimmani@gmail.com