대규모 언어 모델(LLM)을 활용한 AIP 데이터 구조화 절차 연구

김석호*, 이학태
인하대학교

Research on AIP Data Structuring Procedures Using Large Language Models (LLMs)

Seok-ho Kim*, Hak-Tae Lee

Key Words : Aeronautical Information Publication(항공정보간행물), Large Language Model(대규모 언어 모델)

서 론

항공종사자들이 항공업무를 수행하기 위해 참고하는 항공정보간행물(Aeronautical Information Publication, AIP)을 해석하고 이를 정형화된 구조로 정리하는 것은 대량의 운항데이터를 분석하기 위해 선제적으로 수행해야 할 조건 중 하나이다. 특히 AIP의 주요 구성 중 Standard Instrument Departure(SID), Standard Terminal Arrival Route(STAR), Instrument Approach Chart 정보는 항공기의 운항 경로 및 절차 분석에 직접적인 영향을 미친다.

그러나 국토교통부에서는 매월 수정된 AIP를 배포하고 있으며, 그때마다 변경된 내용을 수동으로 구조화하는 것은 과도한 업무 부하를 발생시킨다.

이에 본 연구에서는 대규모 언어 모델(Large Language Model, LLM)에 AIP 이미지와 정형화된 프롬프트를 입력하여 구조화된 CSV 파일로 자동 변환 이후 다수결, 우선순위 기반 알고리즘을 통해 복수 AI 모델의 결과를 비교·평가하여 가장 신뢰도 높은 데이터를 자동 추출하는 방안을 제시한다.

본 론

AIP는 여객기 운항시 위치별 경로, 속도, 고도 등 비행 정보와 공항별 접근, 착륙 절차 등을 포함한 규정들을 정리한 지침서이며, Table 1은 그 중 인천공항 AIP의 구성 목록이다.

Table 1. AIP의 구성 목록

AIP의 구성 목록
AD CHART
AIRCRAFT PARKING DOCKING CHART
AD GROUND MOVEMENT CHART
AD OBSTACLE CHART
PRECISION APP TERRAIN CHART
AREA CHART - ICAO(DEP)
SID
AREA CHART - ICAO(ARR)
STAR
ATC SURVEILLANCE MINIMUM ALTITUDE CHART
INSTR APCH CHART
VISUAL APCH CHART
BIRD CONCENTRATION CHART

RNAV BINIL 2C

Serial Number	Path Descriptor	Waypoint Identifier	Fly-over	Course/Track °M(°T)	Distance (NM)	Turn direction	Altitude (ft)	Speed (kt)	Coordinates	VPA/RDH	Navigation specification	Remarks
001	CF	CG050	-	15L-140(131.3) 15R-138(129.7)	7.7	R	-5 000	-	37°22'18.4"N 126°35'07.0"E	-	RNAV 1	-
002	TF	CG100	-	169 (160.9)	6.4	-	-10 000	-	37°16'16.3"N 126°37'43.8"E	-	RNAV 1	-
003	TF	CP100	-	224 (215.3)	9.9	-	-10 000 +9 000	-	37°08'11.8"N 126°30'35.8"E	-	RNAV 1	-
004	TF	HP100	-	278 (269.6)	7.6	-	-10 000	-	37°08'07.8"N 126°21'03.5"E	-	RNAV 1	-
005	TF	HP120	-	278 (269.5)	13.0	-	+12 000	-	37°07'59.2"N 126°04'48.1"E	-	RNAV 1	-
006	TF	BELTU	-	296 (287.9)	14.1	-	+FL 170	-	37°11'21'8.0"N 125°47'59.0"E	-	RNAV 1	-
007	TF	BINIL	-	302 (293.2)	29.5	-	-	-	37°23'49.2"N 125°13'59.3"E	-	RNAV 1	-

RNAV BOPTA 2C

Serial Number	Path Descriptor	Waypoint Identifier	Fly-over	Course/Track °M(°T)	Distance (NM)	Turn direction	Altitude (ft)	Speed (kt)	Coordinates	VPA/RDH	Navigation specification	Remarks
001	CF	CG050	-	15L-140(131.3) 15R-138(129.7)	7.7	R	-5 000	-	37°22'18.4"N 126°35'07.0"E	-	RNAV 1	-
002	TF	CG100	-	169 (160.9)	6.4	-	-10 000	-	37°16'16.3"N 126°37'43.8"E	-	RNAV 1	-
003	TF	CD120	-	190 (181.1)	8.8	-	-13 000 +9 000	-	37°07'27.8"N 126°37'31.2"E	-	RNAV 1	-
004	TF	HD130	-	190 (181.1)	5.7	-	-13 000	-	37°01'44.4"N 126°37'23.0"E	-	RNAV 1	-
005	TF	HD140	-	188 (179.5)	8.6	-	+14 000	-	36°53'08.6"N 126°37'29.0"E	-	RNAV 1	-
006	TF	BOPTA	-	191 (182.6)	9.0	-	+FL 170	-	36°44'06.4"N 126°36'58.4"E	-	RNAV 1	-

Fig. 1. Standard Instrument Departure Procedure Coding Tables

Figure. 1은 인천국제공항의 SID 절차 중 RNAV BINIL 2C, RNAV BOPTA 2C 절차이다. AIP는 이처럼 PDF(이미지) 형태로 제공되며, 1페이지로 구성되지 않고 다수의 절차가 여러 페이지에 걸쳐 포함된다. 따라서 사람이 직접 내용을 해석하기에는 적절하지만, 알고리즘을 적용하거나 대량의 운항 데이터를 분석하기 위해서는 이를 기계가 읽을 수 있는 형식으로 구조화하는 작업이 필요하다.

대규모 언어 모델 기반 데이터 구조화

본 연구에서는 AIP 내 SID 차트(총 8개 절차, 4페이지)를 대상으로, 4개의 대규모 언어 모델(ChatGPT, Claude, Gemini, Deepseek)을 활용한 데이터 구조화를 수행하고 각 모델의 신뢰도를 비교 분석하였다.

첫 단계에서는 구조화 항목으로 Entry/Exit Node(이름, 위치 정보, 고도/속도 제한), RNAV 절차명을 정의하였다. 이후 동일한 프롬프트(SID 차트 이미지 포함)를 각 모델에 7회 반복 적용하여 총 28개의 CSV 결과물을 생성하였다. 프롬프트에는 열 구성, 데이터 형식, 표기 규칙 등을 명시하여 결과의 일관성 및 정합성을 확보하고자 하였다.

다음으로, 각 모델이 생성한 데이터의 신뢰도를 평가하기 위해 2단계 데이터 정제 및 평가 알고리즘을 구현하였다.

1단계는 다수결 기반의 데이터 통합 과정이다. 각 모델별로 생성된 7개 결과물을 Entry Node와 Exit Node 조합을 기준으로 전체 데이터를 경로별 그룹으로 나눈 후에 각 그룹 내에서 가장 빈도수가 높은 최빈값을 경로의 대표 값으로 선정하였다.

2단계는 규칙 기반의 신뢰도 평가 과정이다. 1단계에서 생성된 데이터를 대상으로, 사전에 정의된 데이터 유효성 규칙과의 부합 여부를 검증하였다. 첫째, Entry/Exit 노드명이 유효한 형식인지 (알파벳 3~5자 or 알파벳2자 + 숫자3자 조합) 판단하였고, 다음 고도 및 속도 형식 유효성을 판단하였다. 마지막으로 기타 오류 항목으로 셀 값이 '?', '-' 등 비정상적인 값이지 여부를 확인하였다.

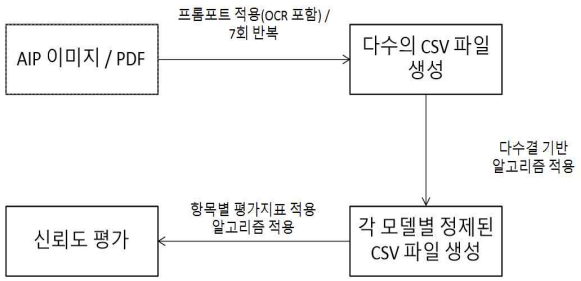

Fig. 2. 구조화 절차 흐름도

이상 데이터를 구조화하는 프로세스는 Fig. 2와 같다.

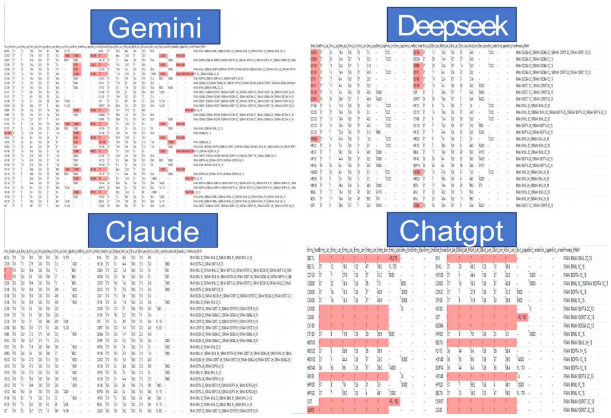

Fig. 3. LLM별 정제된 CSV 파일

Figure. 3.는 각 모델의 정제 결과를 시각화한 예시로, 신뢰할 수 없는 데이터는 빨간색으로 표시되며, 각 모델의 총점도 함께 표기되도록 알고리즘을 구성하였다. 평가 결과 모든 AI 모델에서 입력 데이터에 있는 Entry Exit 노드의 조합을 동일하게 검출하지 못했으며 모델에 따라 누락 또는 없는 조합이 추가되는 현상이 발생했다.

Claude가 전반적으로 가장 높은 신뢰도를 보였고, Deepseek는 노드 이름 인식 및 고도(FL) 입력 부분에서 약간의 오류를, Gemini는 노드 이름 인식 및 고도(-) 입력 부분에서 많은 오류를, ChatGPT는 좌표 미생성, 열 병합 오류, OCR 정확도 등의 많은 오류로 인해 가장 낮은 신뢰도를 보였다.

Table 2. LLM별 세부 오류 통계

모델명	정답 셀 개수	생성 셀 개수	오류 셀 개수
Claude 3.7 Sonnet	529	575	2
Deepseek-R1		552	16
Gemini 2.5 pro		805	65
Chatgpt 4o		437	100
4개 모델 통합		1610	182

세부 결과는 Table 2와 같다.

결과 및 제약점

AIP 이미지의 품질과 텍스트 배치에 따른 OCR 정확도 저하와 중요 정보(좌표, 고도)의 분리·단위 혼동 오류가 발생하였다.

또한, 현재 적용한 알고리즘 평가 방식상 생성된 셀의 오류만 탐지하므로, 모델이 셀 생성을 누락한 경우는 집계하지 못하는 한계가 있다.

프롬프트 기반 구조화는 일관성이 부족하며, 형식상 문제가 없어도 실제 데이터와 불일치하는 경우가 존재해 신뢰성 확보에는 과제가 남아있다.

결 론

본 연구에서는 대규모 언어 모델을 활용한 AIP 내 SID 차트의 구조화 가능성을 검토하고, 각 모델의 데이터 신뢰도를 정량적으로 비교하였다. Claude 모델이 상대적으로 우수한 결과를 보였으며, ChatGPT는 가장 낮은 결과를 보였다. 그러나 가장 우수한 단일모델에만 의존하는 것은 여전히 편향 및 불안전성의 한계를 가질 수 있다. 다만 여러 모델의 결과를 단순히 취합할 경우, 본 연구에서 채택한 다수결 기반 정제 알고리즘이 신뢰도가 낮은 ChatGPT, Gemini가 생성한 다수의 오류 데이터를 오히려 정답으로 오인하였고, 이로 인해 Claude의 우수한 결과가 희석되어 전체 데이터의 정확도를 저해하는 문제가 발생하므로, 향후 각 모델의 결과를 선별적으로 통합·보정하는 고도화된 시스템 및 프롬프트 자동 생성, 데이터 구조화 유형 확장(SID → STAR, APP) 연구를 통해 사람의 개입 없는 자동화 시스템을 구축할 예정이다.

후 기

본 연구는 국토교통부의 '데이터 기반 항공교통관리 기술 개발 (RS-2021-KA163373)'연구의 지원에 의하여 이루어진 연구로, 관계 부처에 감사드립니다.

참고문헌

1) Hwang, H. S., Lee, H. T., "Introduction to AIP and Proposed Data Structure for Utilizing AIP," Proceeding of the 2022 KSAS Spring Conference, 2022, pp. 203~204.

2) Oh, S. Y., Moon, S. H., The Role of Large Language Models in Large-scale Aerospace Data Processing and the Need for MLOps. Aerospace Magazine, 17(3), 2023, 30-37.

3) Han, D. G., Choi, S. B., Kim, B. S., Bae, S. M., Song, S. M., Shin, B. C., Park, S. J., "Comparison of programming performance by type of Large Language Model," Proceeding of the 2024 Summer Conference of the Institute of Electronics Engineers of Korea, pp. 2553~2557.