Amazon SageMaker AI를 사용하여 화물 용량에 대한 수요 예측

AWS 권장 가이드

AWS 권장 가이드: Amazon SageMaker AI를 사용하여 화물 용량에 대한 수요 예측

Table of Contents

Amazon SageMaker AI를 사용하여 화물 용량에 대한 수요 예측

Tianxia Jia 및 Hengzhi Chen, Amazon Web Services(AWS)

2024년 5월(문서 기록)

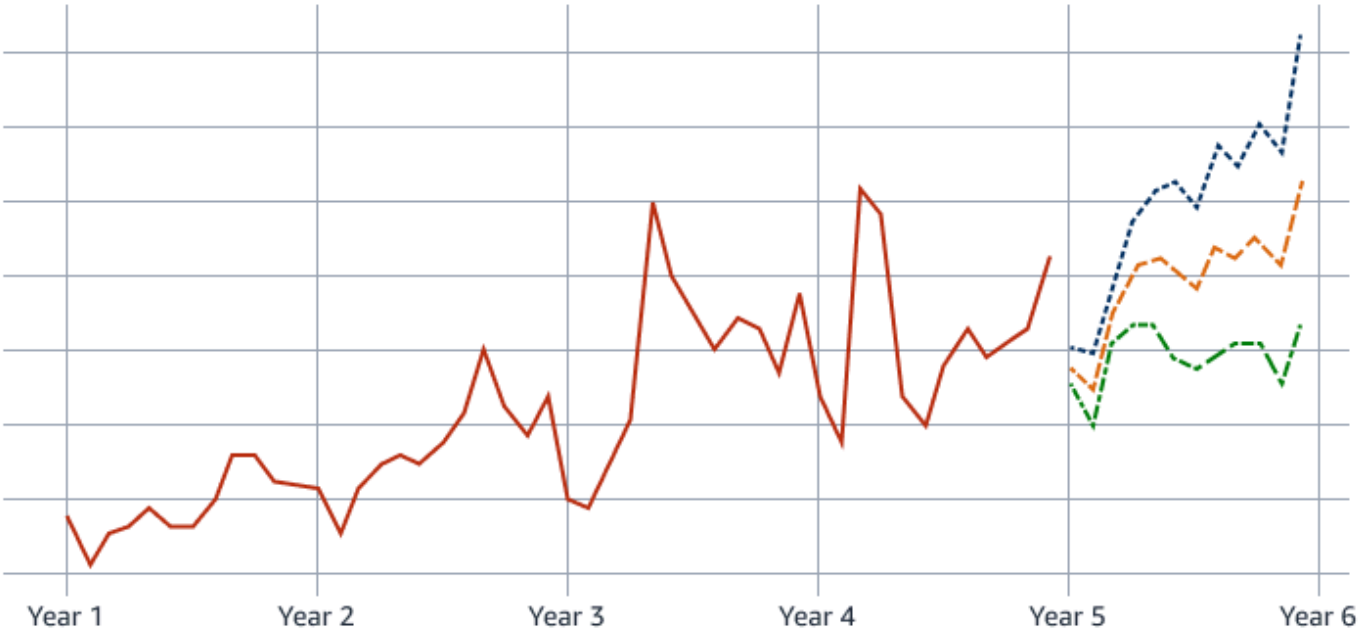
수요 예측은 운송 및 물류 산업, 특히 공급망 제약 기간에는 매우 중요합니다. 정확한 화물 수요 추정치는 컨테이너 및 항공 화물을 국경을 넘어 배송하는 기업과 같이 물류 및 공급망과 관련된 기업에 도움이 됩니다. 이를 통해 기업은 전송 네트워크 보안 비용을 효과적으로 관리할 수 있으므로 배송 비용을 관리하고 수익과 수익을 극대화할 수 있습니다.

정확한 예측을 할 수 있는 기계 학습(ML) 모델은 고품질 훈련 데이터에 따라 달라집니다. 수요 예측의 경우 훈련 데이터에는 과거 수요 볼륨과 가격, 재고, 영업 팀 인원 등 볼륨과 관련될 수 있는 기타 내부적으로 생성된 데이터가 포함될 수 있습니다. 또한 경쟁자, 시장 환경, 공휴일, 날씨 및 거시 경제와 같은 외부 데이터도 수요량에 영향을 미칠 수 있습니다. 이러한 내부 및 외부 데이터 요소를 ML 모델의 기능으로 사용할 수 있습니다.

모든 기능이 식별되면 비즈니스는 자신이 제어할 수 있는 일부 기능에 대한 입력을 제공하려고 할 수도 있습니다. 예를 들어 기업은 배송 가격을 미리 설정하거나 프로모션 및 할인 시기를 결정할 수 있습니다. 이러한 유형의 사용자 입력은 예측 시 모델에 통합할 수 있습니다.

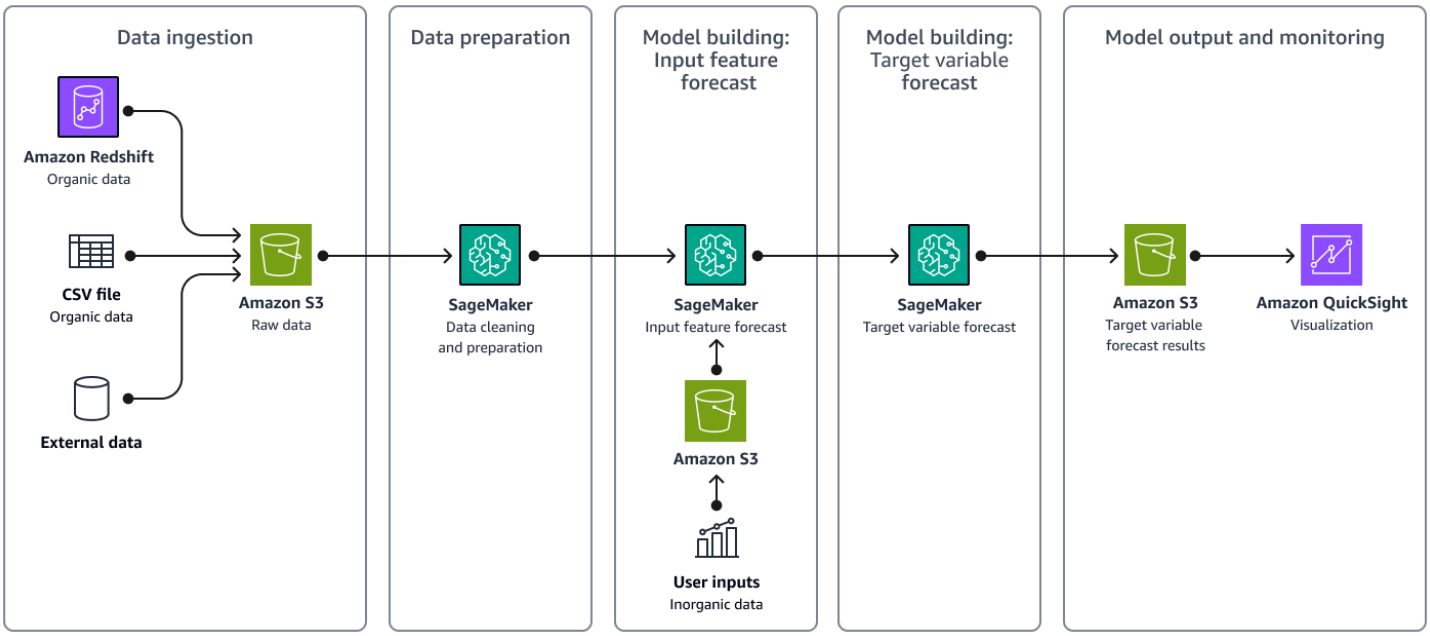
이 가이드에서는 ML 모델을 사용하여 정확한 로지스틱 수요를 예측 AWS 하는 솔루션을 구축하는 전략을 설명합니다. 수요 볼륨 및 수요와 관련된 기능이 포함된 기록 데이터 세트를 기반으로 모델을 훈련합니다. 이러한 기능과 지표에는 내부 유기 데이터와 외부 데이터가 모두 포함됩니다. 또한 이 솔루션은 사용자와 비즈니스 분석가가 입력을 제공할 수 있는 유연성을 제공하며, 이를 예측 모델에 통합할 수 있습니다.

다음 이미지는 과거 시계열 및 12개월 예측 범위의 예를 보여줍니다. 이 가이드의 권장 사항을 사용하여 이러한 유형의 예측을 생성하는 ML 모델을 생성할 수 있습니다.


Year 1
Year 2
Year 3
Year 4
Year 5
Year 6

화물 수요 예측을 위한 아키텍처

다음 이미지는 데이터 수집, 데이터 준비, 모델 구축, 최종 출력 및 모니터링을 포함한 솔루션의 워크플로를 보여줍니다.



솔루션 아키텍처에는 다음과 같은 주요 구성 요소가 포함되어 있습니다.

1. 데이터 수집 - 유기 데이터와 외부 데이터를 모두 Amazon Simple Storage Service(Amazon S3)에 저장합니다.
2. 데이터 준비 - Amazon SageMaker AI는 데이터를 정리하고 ML 모델 훈련을 위해 준비합니다. 자세한 내용은 SageMaker AI 설명서의 데이터 준비를 참조하세요.
3. 모델 구축: 입력 기능 예측 - SageMaker AI는 Prophet를 사용하여 각 입력 기능에 대한 시계열 예측을 생성합니다. 예측 결과를 검사합니다. 필요한 경우 사용자 입력을 제공하여 기능의 예측을 덮어씁니다.
4. 모델 구축: 대상 변수 예측 - SageMaker AI는 수정된 입력 기능을 사용하여 추론을 위한 회귀 모델을 생성합니다.
5. 모델 출력 및 모니터링 - 회귀 모델은 예측 결과를 Amazon S3에 출력합니다. Amazon Quick에서 예측을 시각화할 수 있습니다. 분석가는 예측 결과를 모니터링하고 예측을 실제 수요량과 비교하여 정확도를 평가할 수 있습니다.

데이터 수집부터 최종 모델 출력까지 전체 처리 파이프라인을 오케스트레이션하여 자동으로 실행할 수 있습니다. 예를 들어 월별 수요 예측을 위해 매월 자동으로 실행되도록 설정할 수 있습니다. 둘 이상의 제품에 대한 예측이 필요한 경우 여러 제품에 대해 파이프라인을 병렬로 실행할 수 있습니다. 자세한 내용은 SageMaker AI 설명서의 MLOps 구현을 참조하세요.

화물 수요 예측을 위한 데이터

모든 ML 모델이 의미 있는 예측 및 예측을 하기 위해서는 고품질 데이터가 필수적입니다. 수요 예측의 경우 데이터세트는 최종 수요에 영향을 미칠 수 있는 모든 관련 데이터로 구성됩니다. 이 데이터는 다양한 출처에서 가져올 수 있습니다. 이 데이터를 내부 데이터와 외부 데이터의 두 범주로 분류할 수 있습니다.

내부 데이터

내부 데이터는 비즈니스에서 생성된 유기적 데이터입니다. 이 데이터는 일반적으로 Amazon Redshift 와 같은 데이터 웨어하우스에 저장됩니다.

관심 제품의 과거 볼륨이 포함된 데이터 웨어하우스의 테이블에서 대상 출력 값을 직접 생성하거나 추출할 수 있습니다. 운송 회사의 경우 출력 또는 목표 값은 해상 운송의 경우 전체 컨테이너 적재 단위 또는 항공 화물의 총 중량 단위일 수 있습니다.

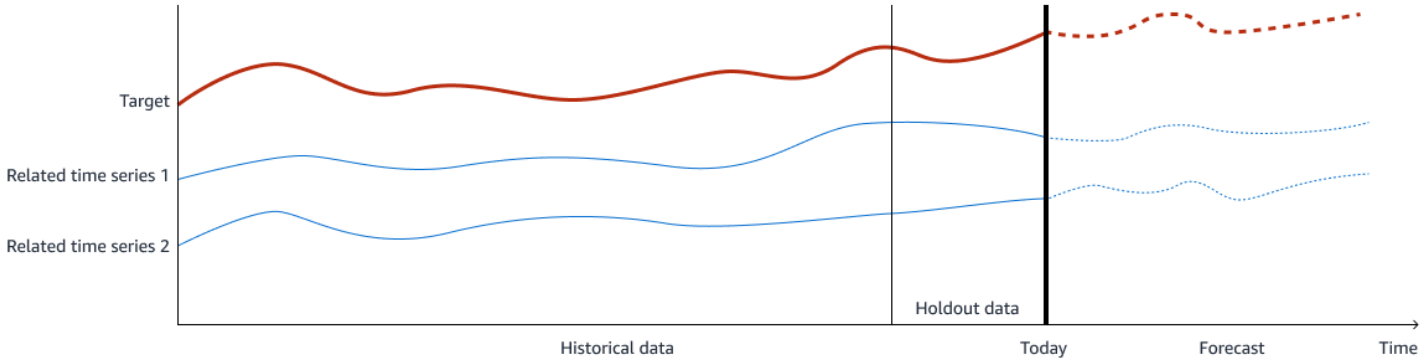
또한 다양한 과거 비즈니스 지표를 생성할 수 있습니다. 수요를 예측할 때 머신 러닝 모델의 기능으로 사용할 수 있습니다. 특징의 예로는 과거 가격, 비용, 용량, 재고 등이 있습니다.

외부 데이터

외부 데이터 소스를 추가 기능으로 사용하여 예측 정확도를 개선할 수 있습니다. 외부 데이터 소스의 예로는 날씨 데이터, 거시 경제 데이터, 산업 데이터, 시장 데이터 등이 있습니다. 이러한 요인은 물류 및 운송 산업에 직간접적인 영향을 미쳐 수요에 영향을 미칠 수 있습니다. 예를 들어, 시장 운임은 글로벌 화물 시장의 벤치마크를 제공하며, 이는 궁극적으로 회사별 수요에 영향을 미칩니다. 주요 경제국의 수입 및 수출 데이터와 같은 거시 경제 데이터도 시장 활동의 척도로 사용될 수 있습니다. 이러한 외부 데이터 소스를 통합하기 위해 다양한 API를 사용하여 데이터를 수집할 수 있습니다. 예를 들어 St. Louis Fed 는 거시 경제 데이터에 액세스할 수 Federal Reserve Economic Data (FRED) API있도록 하고 National Oceanic and Atmospheric Administration (NOAA) 는 전 세계 날씨 데이터에 액세스할 Climate Data Online (CDO) API수 있도록 합니다.

화물 수요 예측을 위한 기계 학습 모델

다음 이미지는 훈련 데이터의 예를 보여줍니다. 대상은 예측하려는 대상이고, 관련 시계열 1 및 2는 대상 예측과 관련된 입력 기능입니다. 기록 데이터는 훈련 및 검증에 사용되며 모델 검증을 위해 기록 데이터의 기간을 보류합니다.



수요 예측에서 출력(또는 대상)은 예측하려는 수요 볼륨입니다. 입력 기능은 출력과 관련된 시계열 데이터입니다. 수요량을 정확하게 예측하도록 ML 모델을 훈련하려면 솔루션에 두 가지 기계 학습 모델이 필요합니다. 첫 번째 모델은 내부 및 외부 데이터를 포함하여 입력 기능에 대한 시계열 예측을 수행합니다. 두 번째 모델은 모든 기능을 사용하여 최종 수요를 예측합니다. 이 두 모델을 함께 사용하면 시계열 추세와 대상과 입력 간의 관계를 효과적으로 캡처할 수 있습니다.

입력 기능 예측을 위한 ML 모델

입력 기능에는 내부 및 외부 과거 시계열 데이터가 모두 포함됩니다. 각 기능에 대한 예측을 수행하려면 1차원(1D) 시계열 모델을 사용할 수 있습니다. 다양한 알고리즘을 사용할 수 있습니다. 예를 들어, Prophet는 강력한 계절 효과와 여러 계절의 과거 데이터가 있는 시계열에 가장 적합합니다. 각 개별 기능에 대해 예측이 생성됩니다.

대상 변수 예측을 위한 ML 모델

출력의 ML 모델 또는 수요 볼륨은 모든 기능과 출력 간의 관계를 캡처하도록 구축되었습니다. , lasso, 및와 같은 다양한 지도 회귀 모델을 사용할 수 ridge regression random forest있습니다XGBoost. 모델을 구축하고 최상의 파라미터와 하이퍼파라미터를 찾을 때 홀드아웃 데이터를 사용할 수 있습니다. 홀드아웃 데이터는 레이블이 지정된 기록 데이터의 일부로, 기계 학습 모델을 훈련하는 데 사용되는 데이터 세트에서 보류됩니다. 홀드아웃 데이터를 사용하여 예측을 홀드아웃 데이터와 비교하여 모델 성능을 평가할 수 있습니다.

화물 수요 예측을 위한 사용자 입력

대부분의 기능의 미래는 알 수 없지만 비즈니스에서 제어할 수 있는 일부 기능이 있을 수 있습니다. 예를 들어 가격은 일반적으로 수요량과 밀접한 관계가 있는 특징 중 하나입니다. 업체에서 가격을 책정하기 때문에 가격 인상 시기 또는 할인 또는 프로모션 시기를 알 수 있습니다. 또한 영업팀의 규모도 수요량에 영향을 미칠 수 있으며 기업은 영업팀의 규모를 통제합니다. 비즈니스에서 관리하는 이러한 데이터 포인트에 대해 사용자 입력을 제공할 수 있습니다. ML 모델링 단계에서 1D 시계열 모델은 각 기능에 대한 예측을 제공하므로 사용자는 이러한 예측 값을 검토하고 사용자 입력으로 덮어쓸 수 있습니다. 그런 다음 덮어쓴 이러한 입력은 모델에서 최종 예측을 수행할 때 사용됩니다.

이 사용자 입력 단계는 1D 시계열 모델 예측이 기능의 향후 동작에 대한 비즈니스 기대치와 일치하지 않는 상황에서 중요할 수 있습니다. 이러한 예측값을 덮어써서 전체 출력 예측을 개선할 수 있습니다.

화물 수요를 예측하는 ML 모델의 모범 사례

이러한 모범 사례를 따르면 화물 수요 예측을 위한 기계 학습 모델의 정확성, 신뢰성 및 해석 가능성을 향상하여 궁극적으로 의사 결정 및 운영 효율성을 개선할 수 있습니다.

- 데이터 품질 및 사전 처리 - 모델 훈련에 사용되는 데이터의 품질이 우수하고 오류, 누락된 값 및 불일치가 없는지 확인합니다. 누락된 값 처리, 이상치 감지, 특성 엔지니어링과 같은 데이터 사전 처리 단계는 모델의 정확도를 개선하는 데 중요한 역할을 합니다.
- 충분한 기록 데이터 - 패턴, 추세 및 계절성을 캡처하려면 충분한 기록 데이터를 확보하는 것이 필수적입니다. 그러나 기록 데이터의 관련성과 적시성을 고려하는 것도 중요합니다. 시장, 비즈니스 운영 또는 외부 요인에 상당한 변화가 있는 경우 이전 데이터는 현재 시나리오를 나타내지 않을 수 있습니다. 이 경우 최신 데이터에 더 높은 가중치를 부여합니다.
- 기능 선택 및 엔지니어링 - 관련 기능을 식별하고 기존 데이터에서 새 기능을 엔지니어링하면 모델의 성능을 크게 개선할 수 있습니다. 도메인 전문가와 긴밀히 협력하여 적절한 기능을 선택할 때 지식과 인사이트를 활용합니다. 또한 기능 중요도 분석을 수행하여 가장 영향력 있는 기능을 식별하고 중복되거나 관련이 없는 기능을 잠재적으로 제거하는 것이 좋습니다.
- 모델 앙상블 - 단일 모델에 의존하는 대신 여러 모델의 예측을 결합하는 앙상블 기술을 사용하는 것이 좋습니다. 앙상블 모델은 개별 모델을 능가하고 보다 강력하고 정확한 예측을 제공할 수 있습니다.
- 모델 평가 및 검증 - 평균 제곱 오차(MSE), 평균 절대 백분율 오차(MAPE) 또는 기타 도메인별 지표와 같은 적절한 지표를 사용하여 모델의 성능을 정기적으로 평가하고 검증합니다. 교차 검증 또는 홀드아웃 검증을 사용하여 모델의 일반화 기능을 평가합니다.
- 지속적인 모니터링 및 재훈련 - 경제 상황, 시장 역학 또는 비즈니스 운영 변화와 같은 다양한 요인으로 인해 시간이 지남에 따라 화물 수요 패턴이 변경될 수 있습니다. 모델의 성능을 지속적으로 모니터링하고 최신 데이터로 주기적으로 재학습하여 정확도와 관련성을 개선합니다.
- 설명 가능한 AI - 수요 예측 모델은 해석 가능하고 설명 가능해야 하며, 특히 이해관계자가 예측의 근거를 이해해야 하는 경우 더욱 그렇습니다. 특징 중요도 분석, 부분 종속성 도표, Shapley Additive 설명(SHAP)과 같은 기법은 모델의 결정을 설명하는 데 도움이 될 수 있습니다.
- 도메인 지식 통합 - 도메인 전문가 및 비즈니스 이해관계자와 긴밀히 협력하여 지식과 통찰력을 모델링 프로세스에 통합합니다. 도메인 전문 지식은 잠재적 편향을 식별하고, 결과를 해석하고, 예측을 기반으로 정보에 입각한 결정을 내리는 데 도움이 될 수 있습니다.
- 시나리오 분석 및 가상 시뮬레이션 - 시나리오 분석 및 가상 시뮬레이션을 수행하는 기능을 예측 솔루션에 통합합니다. 이를 통해 이해관계자는 다양한 비즈니스 결정 또는 외부 요인이 수요 예측에 미치는 영향을 탐색하여 정보에 입각한 의사 결정을 내릴 수 있습니다.

- 자동화되고 확장 가능한 파이프라인 - 데이터 수집, 전처리, 모델 훈련 및 배포를 위한 자동화되고 확장 가능한 파이프라인을 구축합니다. 이는 특히 여러 제품 또는 리전을 처리할 때 예측 프로세스를 일관되고 효율적으로 실행합니다.

다음 단계

이 수요 예측 솔루션을 구현하기 전에 해결하려는 문제를 평가하는 AWS것이 좋습니다. 비즈니스 소유자와 데이터 사이언티스트를 한데 모아 ML 모델에서 문제를 해결할 수 있는지 여부를 브레인스토밍하는 것이 좋습니다. 보유한 데이터 세트와 사용 가능한 과거 데이터의 길이를 이해하는 것이 중요합니다. 또한 비즈니스 소유자는 데이터 과학자와 협력하여 도메인 지식을 제공하고, 유용한 기능을 식별하고, 이러한 기능을 만드는 데 도움을 주는 것이 중요합니다. 모델의 신뢰성은 생성할 수 있는 관련 기능의 수에 따라 증가하여 보다 정확한 예측을 제공합니다.

이 아키텍처를 구축하려면 AWS먼저 데이터 스토리지용 Amazon Simple Storage Service(Amazon S3) AWS 계정 및 기계 학습 모델 훈련용 Amazon SageMaker AI와 같은 필수 서비스를 설정하고 프로비저닝합니다. 그런 다음 예측 모델의 입력 기능으로 사용할 내부 및 외부 데이터 소스를 식별하고 수집합니다. 이 데이터를 Amazon S3에 저장하고 SageMaker AI의 데이터 처리 기능을 사용하여 모델 훈련을 위해 데이터를 사전 처리하고 준비합니다. SageMaker AI에서는 자동 모델 튜닝 및 분산 훈련 기능을 사용하여 예측 모델을 훈련하고 최적화합니다. AWS Step Functions 또는 AWS 서비스 와 같은 AWS Lambda 를 사용하여 예측 모델을 최신 데이터로 주기적으로 재훈련하는 파이프라인을 설정할 수도 있습니다. 재훈련 후 SageMaker AI에서 배치 변환 작업을 시작하여 Amazon S3에 저장하는 예측 결과를 생성합니다. Amazon Quick을 사용하여 배치 변환 작업에서 생성된 예측 결과를 시각화하고 모니터링합니다.

리소스

Amazon Quick 설명서

- Amazon Quick이란 무엇입니까?
- 데이터 소스 작업

Amazon SageMaker AI 설명서

- Amazon SageMaker AI란 무엇인가요?
- 데이터 준비
- 자동 모델 튜닝 수행
- 배치 변환 사용

AWS 코드 샘플

- Amazon SageMaker AI 예제 리포지토리(GitHub)

문서 기록

아래 표에 이 가이드의 주요 변경 사항이 설명되어 있습니다. 향후 업데이트에 대한 알림을 받으려면 RSS 피드를 구독하면 됩니다.

변경 사항	설명	날짜
최초 게시	—	2024년 5월 14일

기계 번역으로 제공되는 번역입니다. 제공된 번역과 원본 영어의 내용이 상충하는 경우에는 영어 버전이 우선합니다.