【발명의 설명】

【발명의 명칭】

시간단위 군집화 모델에 기반한 수요예측 방법 및 시스템{METHOD AND SYSTEM FOR FORECASTING BASED ON TIME UNIT CLUSTERING MODEL}

【기술분야】

본 발명은 시간단위 군집화 모델에 기반한 수요예측 방법, 장치 및 시스템에 관한 것이다.

【발명의 배경이 되는 기술】

데이터 수집과 처리 기술이 발달하면서 빅데이터를 분석하고 활용하려는 노력이 증가하고 있다. 물류산업에서도 데이터 수집 빈도가 높아지고, 종류가 다양해지며, 분석 또한 세분화되고 있다. 이에 데이터에서는 복잡한 계절성, 종류의 다양화, 세분화 예측을 위한 집단화와 같은 새로운 특성이 관찰되고 있으며, 정확한 예측을 위해 수요예측 모델은 이런 데이터 성질의 변화를 반영해야 하지만, 많은 모델이 구조적 한계를 가지고 있다.

데이터 수집장치와 저장장치의 성능이 증가하고 가격이 낮아지면서, 데이터의 수집 및 분석에 소요되는 비용은 크게 감소했다. 수집비용의 감소로 인하여 물류산업에서의 데이터는 두 가지 측면에서 변화된 모습을 보이는데, 이 중 하나는 데이터의 계절성이 복잡해졌다는 것이다. 이는 데이터 수집비용의 감소가 수집빈도의 증가로 이어졌기 때문으로, 데이터의 수집빈도가 높아질수록 계절성 주기는 길어진다.

기존의 시계열 분석 방법론은 복잡한 계절성, 데이터의 종류의 증가, 그리고 데이터의 세분화에 대응하기에 구조적 한계가 분명하다. 대표적인 시계열 분석 방법론인 ARIMA와 ETS는 1950년대 이전에도 사용되던 방법들이다. 이들이 안정적이고 신뢰할 수 있는 모델들임은 분명하나 반영가능한 최대 주기의 길이가 24이고 다중계절성을 반영하기 어려운 구조, 변수 증가 시 기하급수적으로 복잡해지는 구조, 집단 고유의 성질을 모델에 반영하기 어려운 구조 등의 한계가 있다. ARIMA(Autoregressive Integrated Moving Average), ETS(Exponential Smoothing)와 같은 기존 시계열 모델에서 계절성 주기가 길어지는 문제는 심각하다. 주기의 길이는 한 주기 전체를 관찰하기 위해 필요한 데이터 수로, 기존 시계열 모델에서 계절성 주기가 길어지면 변수 개수가 크게 증가해야 한다. 예를 들어, 1시간 단위로 데이터 수집 시 하루 24시간의 주기 표현을 위해 23개의 변수 지정이 필요한 반면, 1초 단위로 데이터 수집 시에는 하루에 해당하는 주기의 길이는 24×60×60이고, 해당 개수만큼의 변수가 필요하다. 그러나, 변수 개수의 증가는 모델 복잡성과 해석가능성의 측면에서 문제가 된다. 계절성 주기의 상한이 제한된 시계열 모델도 존재한다. 다중 계절성 역시 기존 시계열 모델로는 반영이 어려운 경우가 많으며(Hyndman & Athanasopoulos, 2018), 따라서, 기존 시계열 분석 방법들은 복잡해진 계절성 반영에 한계를 보인다.

수집비용의 감소로 인하여 물류산업에서의 데이터가 보이는 변화된 모습 중 다른 하나는, 데이터의 종류가 다양해졌다는 것이다. 과거에는 몇 개의 제품이 팔렸는지에 대한 데이터만을 수집했다면, 현재에는 몇 개의 제품이 어디에서 누구에게 팔렸는지까지 수집한다. 이는 예측모델에서 설명변수로 이용가능한 데이터가 많아지고 예측문제가 복잡해짐을 의미하기 때문에, 회귀 모델을 사용한다면 정확도와 해석가능성 면에서 유리할 수 있다.

분석비용 감소에 따른 가장 큰 변화는 예측의 세분화다. 분석비용이 컸던 과거에는 여러 집단이 혼합된 합(sum) 시계열의 예측만이 주로 이루어졌다면, 현재는 세분화된 집단별 예측이 경제적으로 가능하다. 30가지 품목을 유통하는 기업을 예로 들면, 과거에는 30가지 품목의 합인 총 유통량의 시계열만을 예측했지만, 현재는 30개의 품목별로 시계열을 예측한다. 집단 성질이 분명해지는 세분화 과정에서 집단 고유의 성질을 예측에 반영할 수 있는 시계열 모델은 높은 예측정확도를 가진다.

물류 데이터에서는 업무일정 등에 영향을 받는 계절성이 관찰된다. 물류는 수요와 공급을 결정하는 사람들의 영향을 많이 받는데, 이들이 업무일정에 따라 의사결정을 내리기 때문이다. 이 의사결정자들의 행동은 연, 분기, 월, 주 등 달력에서 관찰되는 주기성을 보인다. 대표적 예로, 평일이 휴일인 주말보다 물류이동량이 많다는 사실이 있다. 그러나, 기업과 산업마다 업무패턴이 다르므로, 한 데이터에서 관찰한 계절성 패턴을 다른 데이터에 그대로 적용하기는 어려우며, 해당 데이터로부터 계절성 패턴을 학습하는 등의 처리가 필요하다.

복잡한 계절성, 데이터 종류의 증가, 그리고 데이터의 세분화에 따라, 이러한 변화들에 대응할 수 있는 새로운 시계열 모델 개발이 필요한 실정이다. 따라서 새로운 시계열 분석 방법론들이 개발되어야 하며 산업 현장에서는 이런 기술의 변화를 빠르게 수용해야 한다. 그럼에도 불구하고 현실에서는 기본적인 회귀모델도 이용하지 않고 실무자의 직관을 기반으로 주먹구구식 예측을 하고 있다. 이에 따른 경제적인 손실은 측정할 수 없을 정도로 크다.

수요예측은 기업의 공급에 대한 의사결정의 근거가 되기 때문에 중요하다. 특히 시계열 데이터의 성분 중 계절성은 의사결정자에게 중요한 분석 결과이다(Terwiesch & Cachon, 2012). 물류산업의 많은 기업들이 다루는 공급망관리(Supply Chain Management) 분야를 예로 들어보면, 공급망관리의 목적은 불확실한 수요에 대응하여 이익을 최대화하고 비용을 최소화하는 공급계획을 세우는 것이다. 여기서 수요의 불확실성을 얼마나 줄이는가에 따라서 공급망관리 문제의 복잡도가 결정되기 때문에 수요예측이 매우 중요하다.

【선행기술문헌】

【특허문헌】

(특허문헌 1) KR101173823 B1

【비특허문헌】

(비특허문헌 1) Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013), Bayesian data analysis, Chapman and Hall/CRC.

【발명의 내용】

【해결하고자 하는 과제】

본 발명의 일 실시예는, 복잡화된 계절성, 풍부해진 설명변수 후보 데이터 및 세분화된 시계열 예측을 고려하여 예상 수요 적중률을 증대시킨 예측 모델을 구축하고, 구축된 예측 모델을 이용하여 매출 예측이 용이하도록 한 예측 모델 기반의 매출 예측 방법, 장치 및 시스템을 제공하는 것을 목적으로 한다.

본 발명의 일 실시예는 복잡한 계절성, 데이터 종류의 증가 및 데이터의 세분화에 대응할 수 있는 시계열 모델 및 이를 이용한 수요예측 방법, 장치 및 시스템을 제공하는 것을 목적으로 한다.

본 발명의 일 실시예는 세분화된 예측이 가능하고, 예측정확도가 우수하며, 해석가능한 결과를 제시할 수 있는 시계열 모델 및 이를 이용한 수요예측 방법, 장치 및 시스템을 제공하는 것을 목적으로 한다.

【과제의 해결 수단】

본 발명의 일 측면에 따르면, 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출단계; 상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계; 상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 및 상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델(Time Unit Clustering Model) 기반 수요예측 방법이 제공될 수 있다.

또한, 상기 요약통계량 산출단계는 상기 클러스터링 단계를 위하여 데이터를 분석하여 시간단위별로 데이터를 하위 세트(set)로 분할하고, 분할된 세트에 대한 평균과 분산의 계산이 수행되어 상기 하위 세트의 유형을 파악할 수 있다.

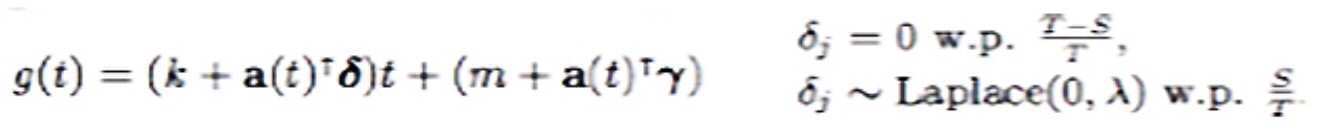
또한, 상기 클러스터링 단계는 상기 시계열 데이터의 값, 상기 요약통계량 산출 단계에서 계산된 시간단위별 계산값, 및 각 시간단위별 정보를 이용하여 각 군집을 라벨링하는 라벨링 단계; 및 상기 교차검증 단계를 위하여 최적 군집화 개수를 도출하기 위해 군집의 개수 후보군을 생성하는 군집생성단계; 를 포함할 수 있다.

또한, 상기 교차검증 단계는 상기 클러스터링 단계에서 생성된, 시간단위별로 구분되는 군집화 후보들에 대하여 최적 군집화 개수를 계산하는 군집 개수 결정 단계; 를 포함할 수 있다.

또한, 상기 최종예측 단계는 상기 각 시계열 데이터 및 상기 군집에 미치는 트렌드(trend) 영향력, 주기성(seasonality) 영향력 및 휴가(holiday)의 영향력을 계산하는 요소별 영향력 계산 단계; 및 상기 트렌드, 상기 주기성 및 상기 휴가의 영향력과 오차를 계산하여 최종 값을 도출하는 단계;를 포함할 수 있다.

또한, 상기 트렌드 영향력은 조각적 선형함수(Piecewise linear function) 또는 하기 수학식 3에 따라 계산될 수 있다.

[수학식 3]



또한, 상기 주기성 영향력은 푸리에 급수(Fourier series) 또는 하기 수학식 4에 따라 계산될 수 있다.

[수학식 4]



또한, 상기 휴가의 영향력은 지시 함수(Indicator function) 또는 하기 수학식 5에 따라 계산될 수 있다.

[수학식 5]



본 발명의 일 측면에 따르면, 샘플 데이터를 업로드하는 제 1단계; 상기 샘플 데이터로부터 예측 모델을 생성하는 제 2단계; 분석 대상인 입력 데이터(input data)를 업로드하는 제 3단계; 상기 제 2단계에서 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 4단계; 및 상기 최종 예측으로 얻어진 값을 제공하는 제 5단계;를 포함하고, 상기 제 2단계는 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출단계; 상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계; 및 상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 를 포함하여 수행되고, 상기 제 4단계는 상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 를 포함하여 수행되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법이 제공될 수 있다.

본 발명의 일 측면에 따르면, 제 1 서버에 샘플 데이터를 업로드하는 제 1단계; 상기 샘플 데이터로부터 예측 모델을 생성하는 제 2단계; 생성된 예측 모델을 제 2 서버에 제공하는 제 3단계; 분석 대상인 입력 데이터(input data)를 상기 제 2 서버에 업로드하는 제 4단계; 상기 제 2단계에서 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 5단계; 및 상기 최종 예측으로 얻어진 값을 제공하는 제 6단계;를 포함하고, 상기 제 2단계에서 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출단계; 상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계; 및 상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 가 수행되어 예측 모델이 생성되고, 상기 제 5단계에서 상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 가 수행되어 최종 예측이 수행되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법이 제공될 수 있다.

본 발명의 다른 측면에 따르면, 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부; 상기 요약 통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부; 상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 및 상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델(Time Unit Clustering Model) 기반의 수요예측 장치가 제공될 수 있다.

또한, 상기 요약통계량 산출부는 상기 클러스터링부를 위하여 데이터를 분석하여 시간단위별로 데이터를 하위 세트(set)로 분할하고, 분할된 세트에 대한 평균과 분산의 계산이 수행되어 상기 하위 세트의 유형을 파악할 수 있다.

또한, 상기 클러스터링부는 상기 시계열 데이터의 값, 상기 요약통계량 산출 단계에서 계산된 시간단위별 계산값, 및 각 시간단위별 정보를 이용하여 각 군집을 라벨링하는 라벨링부; 및 상기 교차검증부를 위하여 최적 군집화 개수를 도출하기 위해 군집의 개수 후보군을 생성하는 군집생성부; 를 포함할 수 있다.

또한, 상기 교차검증부는 상기 클러스터링부에서 생성된, 시간별로 구분되는 군집화 후보들에 대하여 최적 군집화 개수를 계산하는 군집 개수 결정부; 를 포함할 수 있다.

또한, 상기 최종예측수행부는 상기 각 시계열 데이터 및 상기 군집에 미치는 트렌드(trend) 영향력, 주기성(seasonality) 영향력 및 휴가(holiday)의 영향력을 계산하는 요소별 영향력 계산부; 및 상기 트렌드, 상기 계절성 및 상기 휴가의 영향력과 오차를 계산하여 최종 값을 도출하는 최종도출부;를 포함할 수 있다.

본 발명의 다른 측면에 따르면, 업로드된 샘플 데이터, 분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 데이터 웨어하우스부; 상기 샘플 데이터를 이용하여 예측 모델을 생성하고, 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 컴퓨팅 클러스터부; 상기 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 모델 스토리지부; 및 최종 예측 데이터를 제공할 수 있는 웹 어플리케이션부;를 포함하고, 상기 컴퓨팅 클러스터부는 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부; 상기 요약통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부; 상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 및 상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 장치가 제공될 수 있다.

본 발명의 또 다른 측면에 따르면, 업로드된 샘플 데이터, 분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 데이터 웨어하우스부; 상기 샘플 데이터를 이용하여 예측 모델을 생성하고, 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 컴퓨팅 클러스터부; 상기 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 모델 스토리지부; 및 최종 예측 데이터를 제공할 수 있는 웹 어플리케이션부;를 포함하고, 상기 컴퓨팅 클러스터부는 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부; 상기 요약통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부; 상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 및 상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델(Time Unit Clustering Model) 기반의 수요예측 시스템이 제공될 수 있다.

본 발명의 또 다른 측면에 따르면, 제 1 장치부; 및 제 2 장치부;를 포함하고, 상기 제 1 장치부는 업로드된 샘플 데이터를 저장하는 제 1 데이터 웨어하우스부; 상기 샘플 데이터를 이용하여 예측 모델을 생성하는 제 1 컴퓨팅 클러스터부; 상기 제 1 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 제 1 모델 스토리지부; 및 상기 생성된 예측 모델을 상기 제 2 장치부로 제공할 수 있는 제 1 웹 어플리케이션부;를 포함하고, 상기 제 2 장치부는 분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 제 2 데이터 웨어하우스부; 상기 제 1 장치부로부터 제공된 예측 모델을 저장하는 제 2 모델 스토리지부; 상기 제 2 모델 스토리지부에 저장된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 2 컴퓨팅 클러스터부; 및 상기 제 2 컴퓨팅 클러스터부에서 얻어진 최종 예측 데이터를 제공할 수 있는 제 2 웹 어플리케이션부;를 포함하고, 상기 제 1 컴퓨팅 클러스터부는 시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부; 상기 요약통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부; 및 상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 를 포함하고, 상기 제 2 컴퓨팅 클러스터부는 상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 시스템이 제공될 수 있다.

【발명의 효과】

본 발명의 일 실시예는, 복잡화된 계절성, 풍부해진 설명변수 후보 데이터 및 세분화된 시계열 예측을 고려하여 예상 수요 적중률을 증대시킨 예측 모델을 구축하고, 구축된 예측 모델을 이용함으로써, 매출 예측의 정확도를 크게 향상시킬 수 있다.

본 발명의 일 실시예는 분석 대상 데이터를 시계열 분석에 기초한 시간단위 군집화 모델로 분석함으로써, 복합적인 계절성의 분석이 가능하며, 그 분석결과를 이용하여 수요 예측 데이터를 높은 정확도로 산출할 수 있다.

본 발명의 일 실시예는 복잡한 물류산업에서 다루는 복합적인 데이터를 분석함으로써, 정확도가 우수한 예측 방법으로 수요 예측이 가능하며, 이를 통하여 불확실한 수요에 대응하여 이익을 최대화하고 비용을 최소화하는 공급계획을 수립할 수 있다.

【도면의 간단한 설명】

도 1 내지 3은 본 발명의 일 실시예에 따른 수요예측 방법을 도시한 순서도,

도 4는 본 발명의 일 실시예에 따른 수요예측 장치의 구성도,

도 5는 본 발명의 일 실시예에 따른 수요예측 장치의 구성도,

도 6은 본 발명의 일 실시예에 따른 수요예측 장치에서 수행되는 수요예측 방법의 순서도,

도 7은 본 발명의 일 실시예에 따른 수요예측 장치의 구성도,

도 8은 본 발명의 일 실시예에 따른 수요예측 장치에서 수행되는 수요예측 방법의 순서도,

도 9는 본 발명의 일 실시예에 따른 수요예측 장치에서 제공되는 인터페이스를 나타낸 것,

도 10은 본 발명의 일 실시예에 따른 수요예측 장치에서 수행되는 수요예측 방법의 다이어그램,

도 11은 일반적인 기업 물류 구조를 도시한 것,

도 12는 파렛트 대여 산업의 물류 구조를 도시한 것,

도 13은 본 발명의 일 실시예에 따른 시간단위 군집화 모델의 진행단계를 도시한 것,

도 14는 본 발명의 일 실시예를 적용하기 위한 대상인 파렛트 이동량의 요일별 패턴을 도시한 것.

【발명을 실시하기 위한 구체적인 내용】

이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

본 발명을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

본 발명의 기술적 사상은 청구범위에 의해 결정되며, 이하의 실시예는 본 발명의 기술적 사상을 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 효율적으로 설명하기 위한 일 수단일 뿐이다.

또한 첨부된 도면은 본 발명의 기술적 사상을 보다 구체적으로 설명하기 위해 도시한 일 예에 불과하므로 본 발명의 기술적 사상이 첨부된 도면의 형태에 한정되는 것은 아니다.

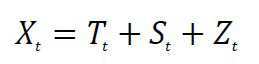
본 발명의 실시예의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 수단일 뿐이며, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되는 것은 아니다.

덧붙여, 명세서 전체에서, 어떤 부분이 다른 부분과 ‘연결’ 되어 있다고 할 때, 이는 ‘직접적으로 연결’ 되어 있는 경우뿐만 아니라, 그 중간에 다른 소자를 사이에 두고 ‘간접적으로 연결’ 되어 있는 경우도 포함한다. 또한, 어떤 구성요소를 ‘포함’ 한다는 것은, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.

본 발명의 일 실시예에 따른 수요예측 방법은 시계열 분석을 통해 수행될 수 있다.

ARIMA와 같이 전통적인 시계열 분석 기법으로 추세성과 계절성이 동시에 나타나는 시계열 데이터를 하기 수학식 1과 같이 계절차분, 일차 차분하여 정상시계열로 변환시킬 수 있다. 그러나 ARIMA 모델은 분기 혹은 월 단위처럼 계절성 주기(seasonal period, 한 주기성을 관찰하기 위해 필요한 데이터 수)가 짧은 경우에만 적용 가능하다는 단점이 있다. 이 단점은 최대 계절성 주기가 24로 제한된 ETS 모델에서도 발견된다.

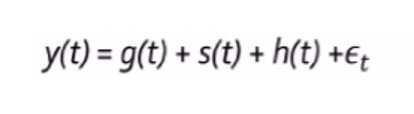
【수학식 1】



물류시스템의 디지털화와 데이터의 수집 구조 발달에 따라 시계열 데이터가 짧은 시간 단위로 저장되고, 이에 기존 ARIMA 모델과 ETS 모델 등의 전통적인 시계열 방법론으로는 높은 예측정확도를 기대하기 어렵다. 예를 들어 30분 단위로 수집되는 데이터는 48의 주기를 가지지만, ARIMA 모델과 ETS 모델은 긴 계절성 주기의 표현이 어렵다는 구조적 한계로 인해 1일 주기성이 누락된 예측값을 제공한다.

이의 대안으로 제안된 프로펫 예측모델(prophet forecasting model)은 유연성(flexibility), 결측치가 있는 시계열의 수용가능성, 빠른 적합(fast fitting), 해석 가능한 모수값과 같은 장점을 가진다. 이 모델은 Harvey & Shephard (1993)가 제안한 구조화된 시계열, 즉 분해 가능한 시계열 모델(decomposable time series model)을 기반으로, 시계열 y(t)를 하기 수학식 2와 같이 표현할 수 있다.

【수학식 2】



(상기 수학식 2에서, E는 epsilon값을 의미한다.)

상기 수학식 2에서, g(t)는 각 시계열 데이터 및 군집에 미치는 트렌드(trend) 영향력, s(t)는 각 시계열 데이터 및 군집에 미치는 주기성(seasonality) 영향력, h(t)는 각 시계열 데이터 및 군집에 미치는 휴가(holiday)의 영향력이고, Et는 오차이고, y(t)는 최종 값을 의미한다. 각 요소별 의미에 대하여는 후에 다시 설명하기로 한다.

시계열 회귀는 시계열 y가 다른 시계열 x와 선형적 관련성이 있다는 전제 하에 y를 예측하기 위해 다른 시계열 x를 이용하는 방식이다. 이때 y를 예측 변수(forecast variable), x를 설명 변수(explanatory variable)라 한다. 시계열 회귀를 이용하면 예측정확도를 높일 수 있다. 예를 들면, 파렛트 대여 산업의 물류에는 조달물류, 판매물류 및 회수물류가 포함될 수 있는데, 각 물류 흐름의 파렛트 이동량을 서로의 예측 변수로 이용하여 예측정확도를 높일 수 있다. 파렛트 대여 산업에 대한 적용 사례는 후에 다시 설명하기로 한다.

상기 시계열은 다중 주기성(multiple seasonality)을 가질 수 있다. 예를 들어, 1시간, 1일, 1주일과 같은 주기성을 가지는 시계열을 30분 단위로 수집한 데이터는 2, 2×24, 2×24×7과 같은 주기를 가질 수 있다. 본 사례에서, 이러한 주기들을 각각 P1<P2<P3,...로 표기하기로 한다. 현재 수요 시계열 데이터가 수집된 단위, 즉 시계열의 가장 작은 시간 단위를 'u'라 하고, 'u'가 'Pi'개 모인 시간 단위를 'u\_i'로 표기하기로 한다. 이러한 예시에서 30분이 u라면, P1=2, P2=2×24, P3=2×24×7이므로, u1=1시간, u2=1일, u3=1주일이 된다. u와 P는 상대적으로 정해질 수 있으며, 그 예시를 아래 표 1에 나타냈다.

【표 1】

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| u=week | - | - | - | - | 월  (month) | 분기  (quarter) | 년  (year) |
| P | - | - | - | - | 4 | 12 | 52/53 |
|  |  |  |  |  |  |  |  |
| u=second | 분  (minute) | 시  (hour) | 일  (day) | 주  (week) |  |  |  |
| P | 7 | 7×24 | 7×24×60 | 7×24×60×60 |  |  |  |

우선, P2가 P1의 배수라 가정하고, 두 주기성을 바탕으로 수행하기로 하면, 같은 방식으로 P3, P4, ...도 다층형태로 확장할 수 있다. 일반적인 시간 단위(time uits)들은 배수관계이므로, 이러한 가정에 큰 무리는 없는 것으로 보인다. 1달 내에 4주, 5주 혹은 1달내 28일 내지 31일과 같이, 정확한 배수관계가 아닌 시간 단위들이나, 특정 산업에서 관찰되는 전형적이지 않는 주기성이 있는 경우들은, missing value 처리를 통해, 후자의 경우 여러 형태의 시간 단위 계층구조를 앙상블 하여 해소 가능한 것으로 보인다. 종래 시계열 방법론들에는 전술한 것과 같은 문제가 있으므로, 본 발명의 일 실시예는 전체 시계열을 P1 개의 구별되는 군집(distinct group)으로 카테고리화하여 트렌드, 주기성 및 휴가 요소 성분에 해당하는 파라미터들을 부분풀링(partial pooling)하는 시간단위별 모델을 이용하여 예측을 수행할 수 있다.

본 발명의 일 실시예에 따른 수요예측 방법은 군집별 예측을 통해 수행될 수 있다.

군집별 예측은 일차적으로 모집단을 군집화 후 각 군집마다 독립적인 예측모델을 만드는 방식이다. 이를 군집 후 예측(cluster-then-predict) 혹은 군집별 예측(cluster-wise prediction)이라 한다. 이의 적용으로는, 예를 들면, 한 지역의 현금자동인출금기를 시계열 패턴을 이용하여 군집화한 후 군집별로 독립적인 예측을 수행할 수 있다.

군집별 예측 기술은 모든 군집이 독립적인 모수를 학습한다. 이에 비해, 다수준 회귀 기술은 군집 간 공통모수 학습(complete-pooling)에서부터 독립적인 모수(no-pooling) 학습까지의 넓은 스펙트럼을 제공한다. 현재 베이지안 추론(Bayesian Inference) 분야에서 활발히 연구되는 계층모델도 이 스펙트럼에 포함된다. 계층모델에서는 군집들이 모수의 분포를 결정하는 공통초모수를 학습(partial-pooling)한다. 따라서 군집별 예측 기술은 다수준 회귀 기술의 일부 형태라 볼 수 있다.

본 발명의 일 실시예는 계절성이 있는 시계열의 시간단위별 분리학습구조를 이용하여 수행될 수 있다. 상세하게는, 시계열을 한 계절성 주기 내 성분으로 이산화 모델링하고, 이산화된 집단별로 분리학습하여 수행될 수 있다. 보다 상세하게는, 먼저 시간 단위별 이산화 단계를 수행하고, 다음 단계에서 이산화된 집단의 분리 학습구조를 학습한 후에, 다음 단계에서 학습된 분리 학습구조에 따라 학습이 수행되고, 마지막으로 학습된 모델로 예측이 수행될 수 있다. 이 때, 시간단위는 주기성을 가지는 단위이면 되고 특별히 제한되지는 않으며, 예를 들면 분, 시간, 일(day), 주(week), 월(month), 분기(quarter) 및 년(year)으로 이루어진 군에서 선택되는 적어도 하나일 수 있다.

본 발명의 일 실시예에 따른 수요예측 방법은 시간단위 군집화(Time Unit Clustering) 모델에 기반하여 수행될 수 있다.

큐빅 스플라인(cubic spline)과 같은 비선형 추세 모델들은 데이터 적합은 잘 되지만, 실제 예측정확도가 떨어지므로 예측구간에서는 선형 추세 모델이 더 정확하다. 선형 추세 모델 중에는 구간별 선형 추세(piecewise linear trend) 방식이 효과적이다. 그러나 추세는 시간 단위의 영향을 받는다. 예를 들어 경기활성화로 소비가 증가해도 회사 밀집지역인 마포구 식당의 저녁 고객의 증가추세는 요일별로 차이가 있을 수 있다.

마찬가지로 계절성 성분은 푸리에 계열을 주로 이용하는데, 시간 단위별로 푸리에 계수가 다를 수 있음에도 모든 시계열 성분들이 동일한 푸리에 계수를 가지도록 모델링된다. 이는 앞서 설명한 다수준 모델에서 모델을 ‘군집 간 공통모수 학습(complete-pooling)’으로만 국한시킨다. 세부 군집들이 모여 하나의 큰 군집을 형성하는 구조에서 세부 군집간의 차이가 클 때 ‘군집 간 공통모수 학습’방식은 좋지 않으며, 이를 개선하기 위해 본 발명의 일 실시예에서는 시계열을 시간단위로 이산화한 후 유사한 요약통계량을 가지는 시간단위로 군집화한다. 그 후 군집별로 독립적인 모수를 학습한다.

이에, 본 발명의 일 실시예는 과거 요일별 시계열의 요약통계량을 바탕으로 군집화한 후 군집마다 다른 모수를 갖는 모델을 설정하는 방식을 적용하여 수행될 수 있다. 전체 데이터를 바탕으로 예측모델을 적합시키지 않고, 시계열을 시간 단위별로 이산화를 하여 군집화를 한 후 각 군집별로 독립적인 예측모델을 만들며, 이 방식을 ‘시간 단위 군집화(Time Unit Clustering; TUC)’라 정의한다.

시간단위 군집화 모델은 기본적인 토대를 Tayor & Letham (2017)의 프로펫 예측모델에 두고 있다. 프로펫 예측모델은 유연성, 결측치 수용가능성, 빠른 적합성, 모수의 해석가능성 등의 장점을 가진 시계열 분석 기법이며, 이 모델은 Harvey & Peters (1993)가 제안한 분해 가능한 시계열 모델을 따라 시계열 데이터를 추세, 계절성 등의 성분으로 분해한다. 해당 추세성분은 구간별 선형모델을, 계절성 성분은 푸리에 모델을 이용하며, 베이지안 방식을 이용하여 데이터로부터 추세와 계절성 성분을 결정하는 모수를 학습할 수 있다.

시간단위 군집화 모델은 기존 프로펫 예측모델에 ‘군집화 후 예측’ 방식을 적용할 수 있다. 프로펫 예측모델에서 모든 시계열 자료를 바탕으로 추세와 계절성 성분의 모수를 학습하였다면, 시간단위 군집화 모델은 시계열을 시간단위로 이산화하여 군집화한 후 각 군집들이 독립적으로 모수를 학습할 수 있다. 이런 군집화 선처리 과정이 필요한 이유는 시간 단위들 간의 성질이 다르기 때문이며, 이때 시간 단위란 ‘일’, ‘주’, ‘월’ 과 같이 시계열 데이터 생성에 영향을 주는 이산적인 시간 단위들을 말한다.

상기 시간단위 군집화 모델은 시간 단위 계층 모델(time unit hierarchical model) 일 수 있다. 상기 시간 단위 계층 모델(time unit hierarchical model)은 이산화된 집단들의 부분적 풀링 학습구조, 모수는 분리하여 학습하되 초모수를 공유하여 학습할 수 있으며, 계절성에 따라 여러 형태의 계층구조가 구성 가능하고, 다중 계절성이 있는 경우 시간 단위 층을 3개 이상의 층으로도 구성할 수 있다. 새로운 계절성이 생겼을 때 모델에 해당하는 층을 추가할 수 있고, 새로운 정보를 모델에 직관적으로 반영할 수 있으며 결과 모델의 해석가능성이 높다는 특징이 있다. 또한, 계층구조를 여러 개 만들어 앙상블(ensemble) 학습법의 수행이 가능하고, 각 계층의 사전분포 조정 시 각 계절성의 반영 강도 조절이 가능하며, 이를 통해 직관성 및 해석가능성이 향상될 수 있다.

본 발명의 일 실시예는 베이지안 시계열 방식(Bayesian time series model) 및 가법모형(generalized additive model)으로 이루어진 군에서 선택되는 적어도 하나를 더 적용하여 수행될 수 있다.

상기 베이지안 방법론(Bayesian method)은 생성 모델(generative model)로 입력 데이터뿐 아니라, 전체적인 모델의 구조와 모수 또한 함께 추정범위에 포함하기 때문에, 시간단위 분리학습 모형에 베이지안 시계열 방식을 적용하는 경우에 결측치 모델링이 용이할 수 있다. 상기 두 방식의 결합 시 기존 계절성을 비정수(non-integer)로 모델링했던 방식을 실제 데이터 생성기작(data generating mechanism)에 따라 정확히 표현할 수 있으며, 예를 들면, 1년을 365.25 주기로 표현했던 방식에서 365, 365, 365, 366의 방식으로 모델링할 수 있다. 이는 연속적(continuous)으로 표현된 값을 분리(discrete) 및 연속적(continuous)인 것으로 나타낸 것이라 할 수 있다.

상기 가법모형(generalized additive model)을 이용할 시에는, 새로운 설명 변수 성분의 추가가 용이한 장점이 있다. 예를 들어, 최종 값(y(t))을 도출하기 위한 수학식이 트렌드(g(t)), 주기성(s(t)) 및 epsilon(t)를 변수로 이용하는 식인 경우에, 그를 상기 수학식 2와 같이 모델링하여 새로운 요소 성분인 휴가(h(t))를 더하여, 새로운 요소 또한 시간 단위 별로 분리 학습이 가능하도록 할 수 있다. 이를 통하여 보다 세분화된 예측을 할 수 있는데, 예를 들면, 월요일의 크리스마스와 일요일의 크리스마스의 수요 변화는 큰 폭의 차이를 가지고 있는데, 이러한 변수를 고려한 예측을 수행할 수 있다.

도 1은 본 발명의 일 실시예에 따른 수요예측 방법의 흐름을 도시한 순서도이다.

도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 수요예측 방법은 요약통계량 산출단계(S1); 클러스터링 단계(S2); 교차검증 단계(S3); 및 최종예측 단계(S4)를 포함하여 수행될 수 있다.

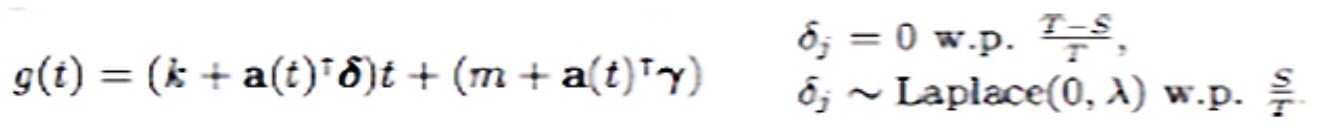
상기 요약통계량 산출단계(S1)는 시계열 데이터를 시간단위(time unit)별로 나누어 평균 또는 분산을 계산하여 수행될 수 있다. 본 단계(S1)에서, 입력된 시계열 데이터(input data)를 시간단위별로 나누어 평균과 분산의 계산이 함께 수행될 수도 있다. 또한, 본 단계(S1)에서는, 후술할 클러스터링 단계(S2)를 위하여, 데이터를 분석하여 시간단위별로 상기 분석된 데이터를 하위 세트(set)로 분할하고, 분할된 세트에 대하여 평균과 분산의 계산이 수행됨으로써 분할된 하위 세트의 유형의 파악이 수행될 수 있다.

상기 클러스터링 단계(S2)에서는 상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화가 진행될 수 있다. 상기 계산된 값은, 평균 또는 분산일 수 있다. 본 단계(S2)는 라벨링 단계(S21) 및 군집생성단계(S22)를 포함할 수 있다. 상기 라벨링 단계(S21)에서는 전 단계(S1)에서 입력된 상기 시계열 데이터의 값, 전 단계(S1)에서 계산된 시간단위별 계산값, 및 각 시간단위별 정보를 이용하여 각 군집의 라벨링이 수행될 수 있다. 상기 군집생성단계(S22)에서는 군집이 생성될 수 있으며, 상세하게는, 후술할 교차검증 단계(S3)를 위하여 최적 군집화 개수를 도출하기 위해, 군집의 개수 후보군의 생성이 수행될 수 있다.

상기 교차검증 단계(S3)에서는 상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수의 선택이 이루어질 수 있다. 본 단계(S3)는 군집화 후보들에 대하여 최적 군집화 개수를 계산하는 군집 개수 결정 단계(S31)를 포함할 수 있다. 이 때, 상기 군집화 후보들은 이전 클러스터링 단계(S2)에서 생성되되, 시간별로 구분되도록 생성될 수 있다.

마지막으로, 상기 최종예측 단계(S4)에서 상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측이 수행될 수 있다. 본 단계(S4)에서는 트렌드(trend), 주기성(seasonality) 및 휴가(holiday)로 삼분화된 각 요소별 영향력의 계산이 수행될 수 있으며, 계산된 각 요소별 영향력을 이용하여 최종 값의 도출이 수행될 수 있다. 이를 위하여, 본 단계(S4)는 요소별 영향력 계산 단계(S41) 및 최종 값을 도출하는 단계(S42)를 포함하여 수행될 수 있다. 상기 요소별 영향력 계산 단계(S41)에서는, 트렌드(trend), 주기성(seasonality) 및 휴가(holiday)에 대하여, 각 요소별 영향력을 계산할 수 있다. 이 때, 각 요소별로 별개로 적용되는 함수를 이용하여 영향력을 계산할 수 있다. 상세하게는, 조각적 선형함수(Piecewise linear function)를 이용하여 각 시계열 데이터 및 군집에 미치는 트렌드(trend) 영향력을 계산하고, 푸리에 급수(Fourier series)를 이용하여 각 시계열 데이터 및 군집에 미치는 주기성(seasonality) 영향력을 계산하고, 지시 함수(Indicator function)를 이용하여 각 시계열 데이터 및 군집에 미치는 휴가(holiday)의 영향력을 계산할 수 있다. 또한, 각 요소별로 별개로 적용되는 수학식을 이용하여 영향력을 계산할 수 있다. 상세하게는, 하기 수학식 3을 이용하여 각 시계열 데이터 및 군집에 미치는 트렌드(trend) 영향력을 계산하고, 하기 수학식 4를 이용하여 각 시계열 데이터 및 군집에 미치는 주기성(seasonality) 영향력을 계산하고, 하기 수학식 5을 이용하여 각 시계열 데이터 및 군집에 미치는 휴가(holiday)의 영향력을 계산할 수 있다.

【수학식 3】



【수학식 4】

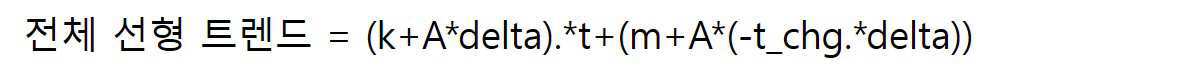


【수학식 5】



상기 트렌드 영향력은 트렌드 모델(trend model)을 이용하여 계산할 수 있다. 비선형(nonlinear) 트렌드 모델들은 데이터 피팅(fitting)은 잘 되지만, 실제 예측정확도는 떨어지는 문제가 발생할 수 있으므로, 예측구간에서는 선형(linear) 트렌드 모델의 정확도가 더 우수할 수 있다. 선형 트렌드 모델 중에서는 조각적 선형함수(Piecewise linear function) 방식이 선택할 수 있다. 기존 선형 트렌드 모델과 다르게, 조각적 선형함수는 트렌드 변화지점(trend change point)을 전후로 다른 기울기를 적용할 수 있다. 이를 하기 수학식 6에 나타냈다.

【수학식 6】



상기 수학식 6을 참조하면, 트렌드 변화지점을 indicator matrix(A[T,s]) 형태로 표현한 A에 의해 t의 기술기가 s번 변한다(T: # of time period, s: # of trend change point, 다른 표현으로 g(t)=(k+a(t)^T\*delta)\*t+(m-a(t)^T\*delta)도 가능하다).

그러나, 트렌드는 시간단위의 영향을 받는다. 예를 들어, 경기활성화로 소비가 증가해도 회사 밀집지역인 마포구 식당의 저녁고객의 증가추세는 요일별로 차이가 있을 수 있다. 월~금요일 저녁의 기술기 변화가 토, 일요일의 기울기 변화보다 더 클 것으로 예상된다. 불금 문화로 월~목보다도 금요일의 증가추세가 가장 크게 나타날 ㅅ 있다. 그러나, 기존 조각적 선형 트렌드 방식에서는 이런 구분 없이 이 경우 '일(day)' 단위에 해당되는 7개의 그룹들이 같은 트렌드 파라미터를 가지는 방식으로 모델링했다. 이와 유사한 문제점이 크로스 섹션(cross section)을 분석하는 계층적 모델(hierarchical model)에서 지적되었다. 다른 카테고리(도는 레벨)의 데이터임에도, 모든 카테고리의 데이터가 동일 파라미터를 공유하는 모델을 'complete pooling'방식이라 하는데, 이 경우 그룹간 차이를 반영하지 못하는 단점이 있다. 이의 개선을 위해 부분풀링(partial pooling) 방식이 제안되었는데, 본 발명에 따른 시간 단위 계층 모델(time unit hierarchical model)에서는 전체 시계열을 P1개의 distinct group으로 카테고리화하여 트렌드(trend), 주기성(seasonalith) 및 휴가(holiday) 성분에 해당하는 파라미터들을 부분풀링(partial pooling)하는 방식을 통해 위 문제를 개선할 수 있다. 다시 말하면, 시간단위별로 트렌드 관련 파라미터인 선형 트렌드의 기울기(k)와 기술기 변화량(delta)가 다르지만 독립적(different but dependent)일 수 있다. 시계열을 P1개 카테고리로 나누고, 트렌드 성분의 기울기(k)와 기울기 변화량(delta)를 부분풀링(partial pooling)하여 모델링할 수 있다.

주기성(seasonality) 영향력은 주기성 모델(seasonality model)을 이용하여 계산할 수 있다. 예를 들어, u=1달, u1=1분기, P1=4, u2=1년, P2=1인 경우에, 1분기의 첫번째 달인 1월의 2017년과 2018년의 차이가 3분기의 첫번째 달인 7월의 2017년과 2018년 차이와 같다고 할 수는 없는데, 현재 푸리에 급수(Fourier series)를 이용한 방법들에서는 전체 데이터에 연속함수를 적합시키는 방식으로 모든 P1개의 그룹이 푸리에 계수를 동일하게 모델링하는 단점이 있다. 이러한 문제를 TUHM에선 개선할 수 있다. 시간단위별로 주기성 파라미터인 푸리에 계수(coefficients of fourier series(beta))가 다르지만 독립적(different but dependent)일 수 있다. 시계열을 P1개 카테고리로 나누고, 주기성 성분의 푸리에 계수(coefficients of fourier series(beta))를 부분풀링(partial pooling)하여 모델링할 수 있다.

휴가(holiday) 영향력은 휴가 모델(holiday model)을 이용하여 계산할 수 있다. 종래 모형에서 휴가(holiday)는 dammy variable로 모델링될 수 있는데, 이는 indicator matrix(I[T,h])와 각 휴가(holiday)에 해당하는 계수벡터(coefficient vector(H[h,1]))를 구성하고, 전체 시계열 값에서 I\*H를 더하는 방식으로 수행될 수 있다(T: # of time period, h : # of holiday type). 그러나, 휴가 효과는 시간단위의 영향을 받는다. 예를 들어, 크리스마스가 수요일인 경우와 일요일인 경우 크리스마스 케이크 판매량의 증가정도가 다를 수 있다. 따라서, dummy variale의 계수는 요일에 따라 다르면서도 연관성 있게 모델링될 필요가 있다. 이를 반영한 다르지만 독립적(different but dependent)인 모델링은 부분 풀링(partial pooling)을 통해 가능하다. 시간단위별로 휴가 파라미터인 dummy variable coefficient(beta)가 다르지만 독립적(different but dependent)이므로, 시계열을 P1개 카테고리로 나누고, 휴가 성분의 dummy variable coefficient(beta)를 부분 풀링(partial pooling)하여 모델링할 수 있다.

상기 최종 값을 도출하는 단계(S42)는 상기 수학식 2를 이용하여, 상기 트렌드, 상기 주기성 및 상기 휴가의 영향력과 오차를 계산하여 수행될 수 있다.

도 2는 본 발명의 일 실시예에 따른 수요예측 방법이 하나의 서버 상에서 수행되는 경우의 일 예를 도시한 순서도이다. 이 때, 상기 서버는 클라우드 서버일 수 있으나 이에 한정되는 것은 아니다.

도 2에 도시된 바와 같이, 본 발명의 다른 일 실시예에 따른 수요예측 방법은, 서버에 샘플 데이터를 업로드하는 제 1단계; 예측 모델을 생성하는 제 2단계; 분석 대상인 입력 데이터(input data)를 업로드하는 제 3단계; 상기 제 2단계에서 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 4단계; 및 상기 최종 예측으로 얻어진 값을 제공하는 제 5단계;를 포함하여 수행될 수 있다.

이 때, 상기 제 2단계는 전술한 요약통계량 산출단계(S1); 클러스터링 단계(S2); 및 교차검증 단계(S3)를 포함하여 수행될 수 있다. 또한, 상기 제 4단계는 전술한 최종예측 단계(S4)를 포함하여 수행될 수 있다. 상기 요약통계량 산출단계(S1); 클러스터링 단계(S2); 교차검증 단계(S3); 및 최종예측 단계(S4);는 이전 실시예와 중복되는 구성인 바, 본 실시예에서는 그 상세한 설명을 생략하기로 하며, 이는 후술하는 다른 실시예에도 동일하게 적용한다.

도 3은 본 발명의 일 실시예에 따른 수요예측 방법이 두 개의 서버에서 연계 수행되는 경우의 일 예를 도시한 순서도이다. 이 때, 상기 서버는 클라우드(cloud) 또는 온 프레미스(on premise) 방식의 서버일 수 있으나 이에 한정되는 것은 아니다. 본 실시예에서는, 제 1 서버로 클라우드가, 제 2 서버로 온 프레미스 방식의 서버가 연계되는 환경인 것으로 설명한다.

도 3에 도시된 바와 같이, 본 발명의 일 실시예에 따른 수요예측 방법은, 클라우드 상에서 샘플 데이터를 업로드하는 제 1단계; 예측 모델을 생성하는 제 2단계; 생성된 예측 모델을 온 프레미스(On Premise; 구내형) 기반 환경으로 제공하는 제 3단계; 분석 대상인 입력 데이터(input data)를 상기 온 프레미스 기반의 환경에 업로드하는 제 4단계; 상기 제 2단계에서 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 5단계; 및 상기 최종 예측으로 얻어진 값을 제공하는 제 6단계를 포함할 수 있다. 이 때, 본 실시예에서 상기 제 1단계 내지 제 3단계는 클라우드 상에서 수행되고, 상기 제 4단계 내지 제 6단계는 온 프레미스 기반 환경 상에서 수행될 수 있다.

전술한 바와 같은 흐름으로 수행되는 본 발명의 일 실시예에 따른 수요예측 방법을 수행가능한 장치의 일 예를 도 4에 도시하였다.

도 4에 도시된 바와 같이, 본 발명의 일 실시예에 따른 수요예측 장치는 요약통계량 산출부(210); 클러스터링부(220); 교차검증부(230); 및 최종예측수행부(240)를 포함할 수 있다.

상기 요약통계량 산출부(210)에서는, 전술한 요약통계량 산출단계(S1)가 수행될 수 있다. 상기 클러스터링부(220)에서는, 전술한 클러스터링 단계(S2)가 수행될 수 있다. 상기 클러스터링부(220)는 라벨링부(221) 및 군집생성부(222)를 포함할 수 있으며, 상기 라벨링부(221)에서는 전술한 라벨링 단계(S21)가, 상기 군집생성부(222)에서는 전술한 군집생성단계(S22)가 수행될 수 있다. 상기 교차검증부(230)에서는, 전술한 교차검증 단계(S3)가 수행될 수 있다. 상기 교차검증부(230)는 군집 개수 결정부(231)를 포함할 수 있으며, 상기 군집 개수 결정부(231)에서는 전술한 군집 개수 결정단계(S31)가 수행될 수 있다. 상기 최종예측수행부(240)에서는, 전술한 최종예측 단계(S4)가 수행될 수 있다. 상기 최종예측수행부(240)는 요소별 영향력 계산부(241) 및 최종도출부(242)를 포함할 수 있으며, 상기 요소별 영향력 계산부(241)에서는 전술한 요소별 영향력 계산 단계(S41)가 수행될 수 있고, 상기 최종도출부(242)에서는 전술한 최종 값의 도출 단계(S42)가 수행될 수 있다.

본 발명의 일 실시예에 따른 수요예측 방법을 수행가능한 장치의 일 예를 도 5에 도시하였다. 본 실시예는 데이터의 저장이 가능한 적어도 1개의 장치일 수 있으며, 상기 장치는 클라우드 서버를 포함할 수 있다.

도 5에 도시된 바와 같이, 본 발명의 일 실시예에 따른 수요예측 장치는 데이터 웨어하우스부(100); 컴퓨팅 클러스터부(200); 모델 스토리지부(300); 및 웹 어플리케이션부(400)를 포함할 수 있다.

상기 데이터 웨어하우스부(100)는 업로드된 샘플 데이터, 분석 대상인 입력 데이터(input data) 및 산출된 최종 예측 데이터를 저장할 수 있다.

상기 컴퓨팅 클러스터부(200)에서는 상기 샘플 데이터를 이용하여 예측 모델을 생성하고, 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행될 수 있다. 상기 컴퓨팅 클러스터부(200)는 전술한 요약통계량 산출부(210); 클러스터링부(220); 교차검증부(230); 및 최종예측수행부(240)를 포함할 수 있다.

상기 컴퓨팅 클러스터부(200)에서 생성된 예측 모델은 상기 모델 스토리지부(300)에 저장될 수 있다.

상기 예측 모델을 이용하여 산출되는 상기 최종 예측 데이터는, 상기 웹 어플리케이션부(400)를 통해 제공될 수 있다.

본 실시예의 장치를 이용하여 수행되는 예측 방법의 일 예를 도 6에 도시하였다. 도 6을 참조하여 설명하면, 먼저 사용자가 샘플 데이터를 상기 데이터 웨어하우스부(100)에 업로드하고, 다음으로 상기 컴퓨팅 클러스터부(200)에서 모델링 엔진이 예측 모델을 생성하고, 다음으로 사용자가 분석대상 입력 데이터를 상기 데이터 웨어하우스부(100)에 업로드하면, 상기 컴퓨팅 클러스터부(200)에서 생성된 예측 모델을 이용하여 최종 예측이 수행되며, 얻어진 최종 예측 데이터는 상기 웹 어플리케이션부(400)를 통해 제공될 수 있다.

이상의 본 발명의 실시예들에 따른 예측 수요 예측 장치는 메모리와 각 동작을 처리하는 프로세서를 포함하는 독립적인 하드웨어 장치 형태로 구현될 수 있으며, 마이크로프로세서나 범용 컴퓨터 시스템과 같은 다른 하드웨어 장치에 포함된 형태로 구동될 수 있다. 또한, 본 발명의 실시예들에 따른 수요 예측 장치의 산출부(210); 클러스터링부(220); 교차검증부(230); 및 최종예측수행부(240)는 적어도 하나 이상의 프로세서(processor)로서 구현될 수 있다. 또한, 본 발명의 실시예들에 따른 수요 예측 장치는 시스템으로 구현될 수 있다.

본 발명의 실시예들에 따른 수요 예측 장치는 어느 하나의 상품군 또는 업종에 한정되는 것은 아니며, 수요량을 예측하고, 예측된 수요량에 기반하여 공급을 예측하기 위한 상품군 또는 업종이라면 얼마든지 적용가능하다.

본 발명의 일 실시예에 따르면, 적어도 두 개 장치가 연계되는 환경에 기반한 수요예측 방법이 수행될 수 있으며, 이의 수행이 가능한 시스템의 일 예를 도 7에 도시하였다.

도 7에 도시된 바와 같이, 본 발명의 일 실시예에 따른 수요예측 시스템은 제 1 장치부; 및 제 2 장치부;를 포함할 수 있다. 이 때 상기 제 1 장치부는 클라우드일 수 있으나 이에 한정되는 것은 아니고, 상기 제 2 장치부는 온 프로미세 연계 환경일 수 있으나 이에 한정되는 것은 아니다. 상기 제 1 장치부는 업로드된 샘플 데이터를 저장하는 제 1 데이터 웨어하우스부(100A); 상기 샘플 데이터를 이용하여 예측 모델을 생성하는 제 1 컴퓨팅 클러스터부(200A); 상기 제 1 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 제 1 모델 스토리지부(300A); 및 상기 생성된 예측 모델을 상기 제 2 장치부로 제공할 수 있는 제 1 웹 어플리케이션부(400A);를 포함할 수 있다. 상기 제 2 장치부는 분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 제 2 데이터 웨어하우스부(100B); 상기 제 1 장치부로부터 제공된 예측 모델을 저장하는 제 2 모델 스토리지부(300B); 상기 제 2 모델 스토리지부에 저장된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 2 컴퓨팅 클러스터부(200B); 및 상기 제 2 컴퓨팅 클러스터부에서 얻어진 최종 예측 데이터를 제공할 수 있는 제 2 웹 어플리케이션부(400B);를 포함할 수 있다.

이 때, 상기 제 1 컴퓨팅 클러스터부(200A) 및 상기 제 2 컴퓨팅 클러스터부(200B) 중 적어도 하나는 전술한 요약통계량 산출부(210); 클러스터링부(220); 교차검증부(230); 및 최종예측수행부(240)를 포함할 수 있다. 본 실시예에서는, 상기 제 1 컴퓨팅 클러스터부(200A)가 상기 요약통계량 산출부(210); 상기 클러스터링부(220); 및 상기 교차검증부(230)를 포함하고, 상기 제 2 컴퓨팅 클러스터부(200B)가 상기 최종예측수행부(240)를 포함하는 것으로 한다.

본 실시예의 두 개 장치 연계 시스템을 이용하여 수행되는 예측 방법의 일 예를 도 8에 도시하였다. 도 8을 참조하여 설명하면, 먼저 사용자가 샘플 데이터를 상기 제 1 장치의 상기 제 1 데이터 웨어하우스부(100A)에 업로드하고, 다음으로 상기 제 1 컴퓨팅 클러스터부(200A)에서 모델링 엔진이 상기 샘플 데이터를 이용하여 예측 모델을 생성하여 상기 제 1 모델 스토리지부(300A)에 저장하면, 사용자는 상기 제 1 장치에서 생성된 예측 모델을 상기 제 1 웹 어플리케이션부(400A)에서 제공받아 상기 제 2장치에 설치할 수 있다. 다음으로, 상기 제 2 장치의 상기 제 2 데이터 웨어하우스부(100B)에 입력 데이터를 업로드하면, 상기 제 2 모델 스토리지부(300B)에서는 저장된 상기 예측 모델을 이용하여 모델링 엔진이 최종 예측을 수행하며, 얻어진 최종 예측 데이터는 상기 제 2 웹 어플리케이션부(400B)를 통해 제공될 수 있다.

본 발명의 일 실시예에 따른 수요예측 장치 또는 시스템에서 제공되는 인터페이스의 일 예를 도 9에 나타냈다. 도 9를 참조하면, 제공되는 시계열 데이터의 형태의 일 예를 확인할 수 있다.

이하에서는, 본 발명의 일 실시예에 따른 수요예측 방법을, 구체적인 적용 사례를 예시로 들어 상세히 설명하기로 한다. 시간단위 군집화 모델의 다이어그램을 도 10에 나타냈으며, 이에 기반하여 본 실시 사례를 설명하기로 한다.

본 실시 사례에서 시간단위 군집화 모델의 예측정확도 평가에 사용된 데이터는 파렛트 대여 산업의 물류데이터이다. 파렛트는 물류라는 추상적인 흐름을 명확하게 해주는 운송 매개체이다. 매일 수십만 종류의 다양한 제품들이 파렛트 위에 실려 이동된다. 통상 파렛트를 하나의 제품이 아닌 다른 제품의 운송을 위한 수단으로 보지만, 본 실시 사례에서는 파렛트를 다른 제품을 이동시키는 수단이 아닌 관리되어야 하는 하나의 제품의 관점으로 파렛트의 물류를 분석한다.

실험대상 기업의 물류 구조는 일반적인 구조인 도 11과는 조금 다르며, 실험대상 기업의 물류 구조를 도 12에 도시하였다. 도 12에 도시된 구조에서도 각 물류들은 서로에게 영향을 주는 점에서 일반적인 구조와 크게 다르지 않다. 파렛트 대여산업의 기업(Pallet Center)은 생산공장(Manufacturer)은 물론 판매처(Vendor)와도 빈번한 교류가 있어야 한다. 파렛트 대여 산업의 물류에는 조달물류, 판매물류 및 회수물류가 포함될 수 있다. 이 때 상기 판매물류는 파렛트 센터에서 생산공장으로의 물류와, 생산공장에서 판매처로의 물류가 있다. 일반적인 산업의 물류 흐름과 차이가 나는 부분은 회수물류다. 일반 제품들의 회수물류는 판매처에서 생산공장으로 돌아오는 흐름이지만, 파렛트 대여산업에서 회수물류는 판매처에서 파렛트 센터로 돌아오는 흐름이다. 조달물류, 판매물류, 회수물류는 각각 도 12에서 Shipment, Delivery, Collection에 해당한다. 도 12의 물류 구조도에서 각 물류흐름의 파렛트 이동량을 서로의 예측변수로 이용하여 예측정확도를 향상시킬 수 있다.

파렛트 대여산업 데이터를 실험데이터로 선정한 이유는 해당 데이터가 물류산업 데이터를 대표할 수 있다고 판단했기 때문이다. 전자산업, 식품산업, 비료산업 등을 비롯한 다양한 산업에 속하는 기업의 유통데이터가 모두 파렛트 데이터에 포함된다. 또한, 데이터 특성 변화인 복잡한 계절성, 데이터의 다양화 및 세분화 역시 파렛트 데이터에서 관찰할 수 있었다. RFID (Radio-Frequency Identification)가 내장된 파렛트의 개발로 데이터 수집빈도가 높아졌고, 설명변수로 이용가능한 데이터들이 많아졌다. 특히 도 12의 구조에서 각 흐름(Shipment, Delivery, Collection)들 간 연관성이 존재하므로, 이들도 서로의 설명변수로 활용될 수 있다. 또한 기존에 파렛트 주문량의 총 합만을 예측했다면, 현재는 파렛트 유형별, 산업군별, 기업별로 세분화된 예측이 경제적으로 가능하다.

본 발명의 실시예에 따른 시간단위 군집화 모델을 국내 한 파렛트 대여기업의 파렛트 발주, 입고, 회수량(각각 도 12의 Shipment, Delivery, Collection에 해당) 예측에 적용했다. 해당 기업은 실무 담당자들의 경험에 따라 수요를 예측하고 있었으며, 이의 평균 오차율은 약 30%로, 파렛트의 생산, 운송, 회수계획 수립에 크게 도움이 되지 않을 정도의 예측정확도인 것으로 나타났다. 도 13에 도시한 본 발명의 일 실시예에 따른 시간단위 군집화 모델을 적용한 실험의 결과에서는, 평균 오차율이 약 7%로 감소하는 것으로 나타났다. 이하에서는 이의 분석실험에 대하여 보다 상세히 설명한다.

<각 요일별 평균 계산>

파렛트 대여산업의 한 물류업체(이하 A사)가 2012년 01월 01일부터 2018년 01월 31일까지 약 6년 동안 언제 어디에서 어디로 어떤 유형의 파렛트가 몇 개 이동되었는지를 일 단위로 수집한 자료를 이용하였다. 고객의 정보보호를 위해 파렛트와 생산공장 및 판매처 등에 대한 정보는 코드로 암호화하여 표기한다.

도 14는 네 가지의 유형에 따른 파렛트 이동량의 요일별 평균을 나타낸 것이다. 한 파렛트를 기준으로 보면 요일간 이동량의 차이를 확인할 수 있다. 휴일인 일요일과, 일요일이 아닌 요일의 평균이 대략 1:200 정도로 차이가 나며, 각 요일 간에도 분명한 차이가 존재하는 것을 확인할 수 있다. 또한, 이런 요일별 차이의 패턴이 파렛트의 유형마다 차이가 있는 것으로 나타난다. A사의 고객은 식품기업, 제관기업, 비료기업 등 다양한데, 분석결과 파렛트가 사용되는 산업과 주요 고객의 주문 방식 등에 따라 요일별 파렛트 이동량의 패턴이 달라졌다.

<시간단위 군집화 모델>

실험에 이용된 물류 파렛트 이동량 시계열 데이터를 예로 들면, 평일의 이동량과 일요일의 이동량은 그 크기가 200배 이상 차이 난다. 이에 본 실시 사례의 실험에서는 시간 단위 ‘요일’에 대해, 시계열을 요일을 기준으로 7개로 이산화했다. 그 후 k-means 군집화 방식을 적용하였다. 7개의 요일을 각각 1개에서 7개 군집개수(k)에 따라 군집화하여 7개의 군집화 집합을 생성했다. 그 후 교차검증을 통해 가장 교차검증 오차가 적은 군집의 개수(k)를 선정하여 최종 예측모델로 삼았다.

<예측모델을 통한 수요예측>

Taylor & Letham(2017)은 프로펫 모델이 ARIMA, 지수평활법, SNAIVE 모델, TBATS 모델보다 예측력이 뛰어남을 실험을 통해 보인 바 있다. 따라서 프로펫 모델과 시간단위 군집화 모델의 예측정확도만 비교해도 충분하다. 프로펫 모델의 수리적 시계열 구조는 시간단위 군집화 모델과 유사하다. 차이점은 군집별 예측이다. 프로펫 모델에서는 모든 요일이 동일한 추세와 계절성 모수를 공유하는 반면에 TUC 모델은 동일 군집에 속하는 요일들만 추세와 계절성 모수를 공유한다.

예측 정확도의 측정 기준으로 보편적으로 사용되는 것으로는 MAP(Mean Absolute Error), MAPE(Mean Absolute Percentage Error), MSE(Mean Squared Error), MSPE(Mean Squared Prediction Error)가 존재한다. MAP는 실제값의 크기와는 상관 없이 오차값의 크기가 중요한 경우에 사용하는 지표이고, MAPE는 오차값의 크기뿐만 아니라 실제값과의 상대적인 비율이 중요한 경우에 사용하는 지표이다. MSE는 오차값의 크기가 커질수록 비용이 증가하는 경우에 사용하는 지표이며 훈련 데이터셋(training dataset)를 기준으로 측정된다. MSPE는 MSE와 마찬가지로 오차값의 크기가 커질수록 비용이 증가하는 경우에 사용하지만 시험 데이터셋(test dataset)을 기준으로 측정한다.

물류산업의 경우 오차값의 크기가 중요하고, 이에 따른 비용은 추정하기 어려워 MAP가 적합하다 판단했다. 따라서 MAP를 기준으로 모델들의 예측정확도 우위를 판별하고자 한다.

전체 실험 데이터는 2012년 01일 01일부터 2017년 1월 31일까지 약 6년간 언제 어디에서 어디로 어떤 유형의 파렛트가 몇 개 이동되었는지를 일 단위로 수집한 자료이다. 1815, 1041, 1627 코드로 암호화된 세 가지 유형의 파렛트를 실험 데이터로 사용했다. 이들은 각각 A사 파렛트 총 주문량의 약 60%, 6%, 2%를 차지하는 유형들이기에 이동량 규모에 따른 모델의 적합도도 비교할 수 있게 설계되었다.

학습 데이터셋(train dataset)은 2012년 01일 01일부터 2017년 12월 31일까지의 데이터를 이용했으며 교차검증 데이터셋(cross-validation dataset)은 2017년 01월 01일에서 12월 31일까지의 12개월을 12등분으로 나누어 이용했다. 일반적인 시계열 교차검증방법을 따랐다(Arlot & Celisse, 2010). 교차검증을 통해 최적의 군집개수(k)를 선택한 후 시험 데이터셋(test dataset)인 2018년 01월 01일부터 2018년 01월 31일까지를 예측했다. 프로펫 모델과 시간단위 군집화 모델의 비교는 시험 데이터의 예측오차 비교를 통해 이루어졌다.

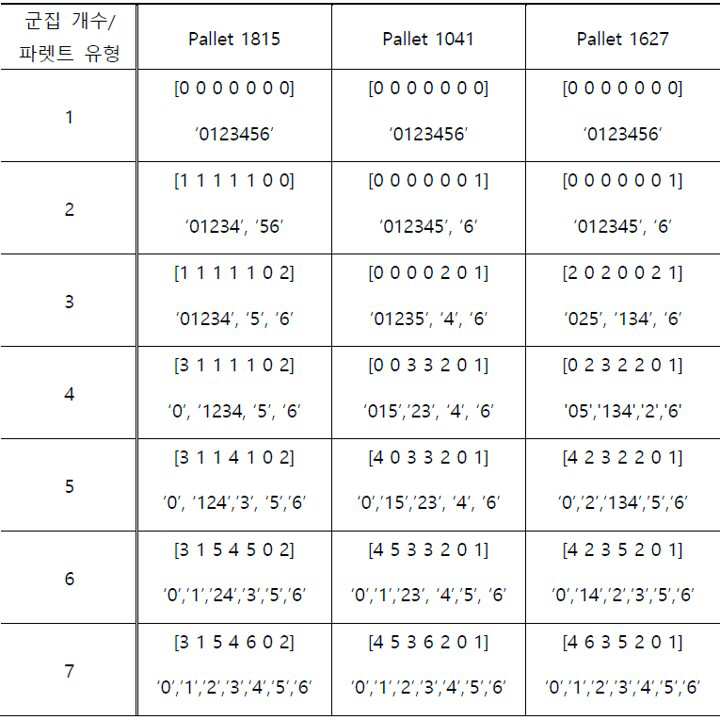
파렛트 3개 유형(1815,1041,1627)의 요일별 평균 이동량을 하기 표 2에 나타냈다. 일요일이 가장 작은 것은 공통적이지만, 각 요일별 비율이 유형별로 조금씩 다름을 알 수 있다. 이 차이는 이후 군집화 단계에서 다른 군집화 결과로 이어진다.

【표 2】

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 파렛트 종류 | 월 | 화 | 수 | 목 | 금 | 토 | 일 |
| 1815 | 35,107 | 33,339 | 33,516 | 33,533 | 33,839 | 25,831 | 951 |
| 1041 | 5,252 | 5,590 | 5,946 | 5,938 | 6,912 | 5,501 | 6 |
| 1627 | 1,070 | 1,499 | 1,239 | 1,386 | 1,474 | 954 | 1 |

상기 표 2의 요약통계량을 바탕으로 군집 개수가 1일때부터 7일때까지의 k-means 군집화를 7번 진행하였다. k-means 군집화는 scikit-learn라이브러리를 이용했으며, 랜덤시드 1을 이용했다. 그 결과는 표 3에 나타냈다. 표 3에서, 월요일에서 일요일을 각각 0에서 6으로 표시했다. 예를 들어, 파렛트 1041의 경우 군집 개수가 4개일 때 군집화 결과는 월화토, 수목, 금, 일이 된다.

【표 3】



이후 3개 파렛트 각각에 대해 군집 개수별 군집결과를 이용하여 예측모델 7개를 만들고, 예측정확도가 높은 군집개수 모수(k)를 선정하기 위해 교차검증을 진행하였다. 각 군집 결과를 바탕으로 교차검증을 한 결과값은 표 4에 나타냈다. 교차검증오차가 적은 군집결과가 가장 예측정확도가 높다. 오차는 MAE방식을 이용했다. 파렛트 1815는 군집 개수가 4개, 파렛트 1041은 군집 개수가 2개, 파렛트 1627은 군집 개수가 4일때 가장 적은 교차검증오차를 보였다.

【표 4】

|  |  |  |  |
| --- | --- | --- | --- |
| 군집 개수 | Pallet 1815 | Pallet 1041 | Pallet 1627 |
| 1 | 7,123 | 1,276 | 457 |
| 2 | 5,214 | **1,153** | 421 |
| 3 | 4,453 | 1,172 | 416 |
| 4 | **4,441** | 1,176 | **412** |
| 5 | 7,949 | 1,163 | 413 |
| 6 | 4,577 | 1,160 | 414 |
| 7 | 4,575 | 1,173 | 416 |

표 5에서는 교차검증 결과 최종 선택된 군집 개수 모수에 대응되는 군집화 집합(표 3에서 확인가능)을 바탕으로 예측모델을 만들었다. 2018년 1월 1일부터 1월 31일까지의 시험데이터셋에서 기존 프로펫 모델의 예측오차를 비교한다.

【표 5】

|  |  |  |  |
| --- | --- | --- | --- |
| 파렛트 종류 | 선택된 군집화 집합 | 프로펫 모델 | 시간단위 군집화 모델 |
| 1815 | '0', '1234', '5', '6' | 5,281 | 3,597 |
| 1041 | '012345', '6' | 1,056 | 1,054 |
| 1627 | '05','134','2','6' | 503 | 453 |

상기 표 2 및 3에서는 각 파렛트마다 요일별 패턴이 다르며, 따라서 군집화를 했을 때 모두 다른 형태로 군집화가 진행됨을 알 수 있다. 패턴이 파렛트 유형별로 다르다는 것은 일반적인 군집화를 모든 시계열에 적용하기 어려우며, 본 실시 사례에서 진행했듯이 데이터로부터 군집 패턴을 학습해야함을 알려준다. 표 5에서 알 수 있듯이 TUC 모델은 기존 프로펫 모델보다 약 10%~20% 예측오차의 개선이 있다. 이 개선은 실제 현업의 운영을 크게 효율화시킬 수 있다.

데이터를 저장하고 처리하는 기술의 발달로 인하여 물류산업에서 다루는 시계열 데이터의 복합적인 계절성이 나타나고, 다양화와 세분화되는 변화가 관찰된다. ARIMA, ETS와 같은 전통적인 시계열 모델은 장기적인 계절성과 다중 계절성을 분석하기 힘들다는 한계를 지니고, 이 한계를 개선한 프로펫 모델은 모수 학습을 공통모수 구조로 국한한다는 한계를 지닌다. 그러나 본 발명의 일 실시예에서 제공하는 다수준 회귀의 관점에서 시간단위 군집화 모델은 이러한 한계점을 개선할 수 있다.

이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

【부호의 설명】

100 : 데이터 웨어하우스부

200 : 컴퓨팅 클러스터부

210 : 요약통계량 산출부

220 : 클러스터링부

221 : 군집 라벨링부

222 : 군집생성부

230 : 교차검증부

231 : 군집 개수 결정부

240 : 최종예측수행부

241 : 요소별 영향력 계산부

242 : 최종도출부

300 : 모델 스토리지부

400 : 웹 어플리케이션부

【청구범위】

【청구항 1】

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출단계;

상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계;

상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 및

상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델(Time Unit Clustering Model) 기반 수요예측 방법.

【청구항 2】

청구항 1에 있어서,

상기 요약통계량 산출단계는

상기 클러스터링 단계를 위하여 데이터를 분석하여 시간단위별로 데이터를 하위 세트(set)로 분할하고, 분할된 세트에 대한 평균과 분산의 계산이 수행되어 상기 하위 세트의 유형을 파악하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

【청구항 3】

청구항 1에 있어서,

상기 클러스터링 단계는

상기 시계열 데이터의 값, 상기 요약통계량 산출 단계에서 계산된 시간단위별 계산값, 및 각 시간단위별 정보를 이용하여 각 군집을 라벨링하는 라벨링 단계; 및

상기 교차검증 단계를 위하여 최적 군집화 개수를 도출하기 위해 군집의 개수 후보군을 생성하는 군집생성단계; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

【청구항 4】

청구항 3에 있어서,

상기 교차검증 단계는

상기 클러스터링 단계에서 생성된, 시간단위별로 구분되는 군집화 후보들에 대하여 최적 군집화 개수를 계산하는 군집 개수 결정 단계; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

【청구항 5】

청구항 1에 있어서,

상기 최종예측 단계는

상기 각 시계열 데이터 및 상기 군집에 미치는 트렌드(trend) 영향력, 주기성(seasonality) 영향력 및 휴가(holiday)의 영향력을 계산하는 요소별 영향력 계산 단계; 및

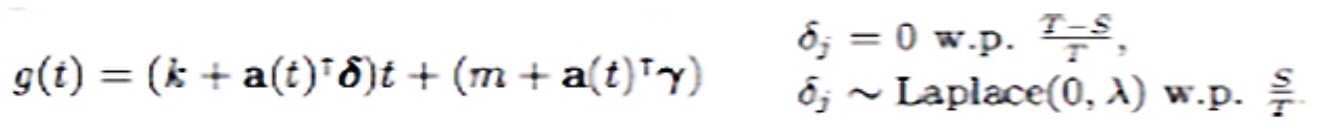
상기 트렌드, 상기 주기성 및 상기 휴가의 영향력과 오차를 계산하여 최종 값을 도출하는 단계;를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

【청구항 6】

청구항 5에 있어서,

상기 트렌드 영향력은 조각적 선형함수(Piecewise linear function) 또는 하기 수학식 3에 따라 계산되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

[수학식 3]



【청구항 7】

청구항 5에 있어서,

상기 주기성 영향력은 푸리에 급수(Fourier series) 또는 하기 수학식 4에 따라 계산되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

[수학식 4]



【청구항 8】

청구항 5에 있어서,

상기 휴가의 영향력은 지시 함수(Indicator function) 또는 하기 수학식 5에 따라 계산되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

[수학식 5]



【청구항 9】

샘플 데이터를 업로드하는 제 1단계;

상기 샘플 데이터로부터 예측 모델을 생성하는 제 2단계;

분석 대상인 입력 데이터(input data)를 업로드하는 제 3단계;

상기 제 2단계에서 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 4단계; 및

상기 최종 예측으로 얻어진 값을 제공하는 제 5단계;를 포함하고,

상기 제 2단계는

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출단계;

상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계; 및

상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 를 포함하여 수행되고,

상기 제 4단계는

상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 를 포함하여 수행되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

【청구항 10】

제 1 서버에 샘플 데이터를 업로드하는 제 1단계;

상기 샘플 데이터로부터 예측 모델을 생성하는 제 2단계;

생성된 예측 모델을 제 2 서버에 제공하는 제 3단계;

분석 대상인 입력 데이터(input data)를 상기 제 2 서버에 업로드하는 제 4단계;

상기 제 2단계에서 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 5단계; 및

상기 최종 예측으로 얻어진 값을 제공하는 제 6단계;를 포함하고,

상기 제 2단계에서

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출단계;

상기 요약통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계; 및

상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 가 수행되어 예측 모델이 생성되고,

상기 제 5단계에서

상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 가 수행되어 최종 예측이 수행되는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반 수요예측 방법.

【청구항 11】

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부;

상기 요약 통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부;

상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 및

상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델(Time Unit Clustering Model) 기반의 수요예측 장치.

【청구항 12】

청구항 11에 있어서,

상기 요약통계량 산출부는

상기 클러스터링부를 위하여 데이터를 분석하여 시간단위별로 데이터를 하위 세트(set)로 분할하고, 분할된 세트에 대한 평균과 분산의 계산이 수행되어 상기 하위 세트의 유형을 파악하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 장치.

【청구항 13】

청구항 11에 있어서,

상기 클러스터링부는

상기 시계열 데이터의 값, 상기 요약통계량 산출 단계에서 계산된 시간단위별 계산값, 및 각 시간단위별 정보를 이용하여 각 군집을 라벨링하는 라벨링부; 및

상기 교차검증부를 위하여 최적 군집화 개수를 도출하기 위해 군집의 개수 후보군을 생성하는 군집생성부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 장치.

【청구항 14】

청구항 11에 있어서,

상기 교차검증부는

상기 클러스터링부에서 생성된, 시간별로 구분되는 군집화 후보들에 대하여 최적 군집화 개수를 계산하는 군집 개수 결정부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 장치.

【청구항 15】

청구항 11에 있어서,

상기 최종예측수행부는

상기 각 시계열 데이터 및 상기 군집에 미치는 트렌드(trend) 영향력, 주기성(seasonality) 영향력 및 휴가(holiday)의 영향력을 계산하는 요소별 영향력 계산부; 및

상기 트렌드, 상기 계절성 및 상기 휴가의 영향력과 오차를 계산하여 최종 값을 도출하는 최종도출부;를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 장치.

【청구항 16】

업로드된 샘플 데이터, 분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 데이터 웨어하우스부;

상기 샘플 데이터를 이용하여 예측 모델을 생성하고, 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 컴퓨팅 클러스터부;

상기 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 모델 스토리지부; 및

최종 예측 데이터를 제공할 수 있는 웹 어플리케이션부;를 포함하고,

상기 컴퓨팅 클러스터부는

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부;

상기 요약통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부;

상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 및

상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 장치.

【청구항 17】

업로드된 샘플 데이터, 분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 데이터 웨어하우스부;

상기 샘플 데이터를 이용하여 예측 모델을 생성하고, 생성된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 컴퓨팅 클러스터부;

상기 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 모델 스토리지부; 및

최종 예측 데이터를 제공할 수 있는 웹 어플리케이션부;를 포함하고,

상기 컴퓨팅 클러스터부는

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부;

상기 요약통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부;

상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 및

상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델(Time Unit Clustering Model) 기반의 수요예측 시스템.

【청구항 18】

제 1 장치부; 및 제 2 장치부;를 포함하고,

상기 제 1 장치부는

업로드된 샘플 데이터를 저장하는 제 1 데이터 웨어하우스부;

상기 샘플 데이터를 이용하여 예측 모델을 생성하는 제 1 컴퓨팅 클러스터부;

상기 제 1 컴퓨팅 클러스터부에서 생성된 예측 모델을 저장하는 제 1 모델 스토리지부; 및

상기 생성된 예측 모델을 상기 제 2 장치부로 제공할 수 있는 제 1 웹 어플리케이션부;를 포함하고,

상기 제 2 장치부는

분석 대상인 입력 데이터(input data) 및 최종 예측 데이터를 저장하는 제 2 데이터 웨어하우스부;

상기 제 1 장치부로부터 제공된 예측 모델을 저장하는 제 2 모델 스토리지부;

상기 제 2 모델 스토리지부에 저장된 예측 모델을 이용하여 상기 입력 데이터에 대한 최종 예측이 수행되는 제 2 컴퓨팅 클러스터부; 및

상기 제 2 컴퓨팅 클러스터부에서 얻어진 최종 예측 데이터를 제공할 수 있는 제 2 웹 어플리케이션부;를 포함하고,

상기 제 1 컴퓨팅 클러스터부는

시계열 데이터를 시간단위(time unit)별로 나누어 평균 및 분산으로 이루어진 군에서 선택되는 적어도 하나를 계산하는 요약통계량 산출부;

상기 요약통계량 산출부에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링부; 및

상기 클러스터링부에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증부; 를 포함하고,

상기 제 2 컴퓨팅 클러스터부는

상기 교차검증부에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측수행부; 를 포함하는, 계절성이 있는 시계열 대상의 시간단위 군집화 모델 기반의 수요예측 시스템.

【요약서】

【요약】

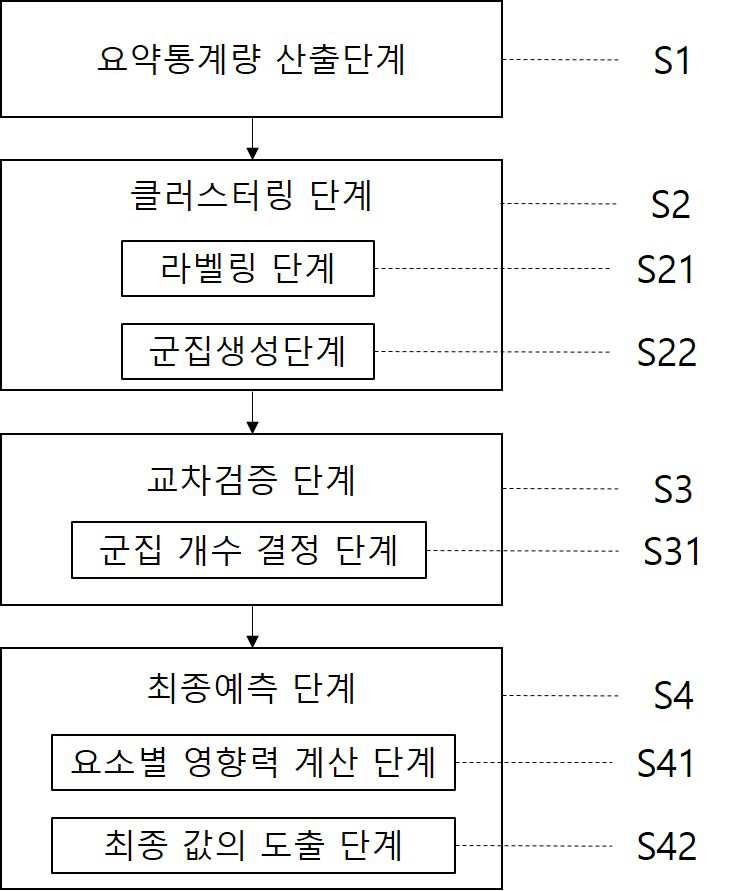
본 발명은 시간단위 군집화 모델에 기반한 수요예측 방법, 장치 및 시스템에 관한 것으로, 상세하게는 시계열 데이터를 시간단위(time unit)별로 나누어 평균 또는 분산을 계산하는 요약통계량 산출단계; 상기 요약 통계량 산출단계에서 계산된 값을 기준으로, 각 시간단위에 대하여 맞춤 군집화를 진행하는 클러스터링 단계; 상기 클러스터링 단계에서 생성된 군집에 대하여, 최적 군집화 개수를 선택하는 교차검증 단계; 및 상기 교차검증 단계에서 선택된 최적 군집화 개수에 따라 예측을 수행하는 최종예측 단계; 를 포함함으로써, 복잡화된 계절성, 풍부해진 설명변수 후보 데이터 및 세분화된 시계열 예측을 고려하여 예상 수요 적중률을 증대시킨 예측 모델을 구축하고, 구축된 예측 모델을 이용함으로써, 매출 예측의 정확도를 크게 향상시킬 수 있는, 시간단위 군집화 모델에 기반한 수요예측 방법, 장치 및 시스템에 관한 것이다.

【대표도】

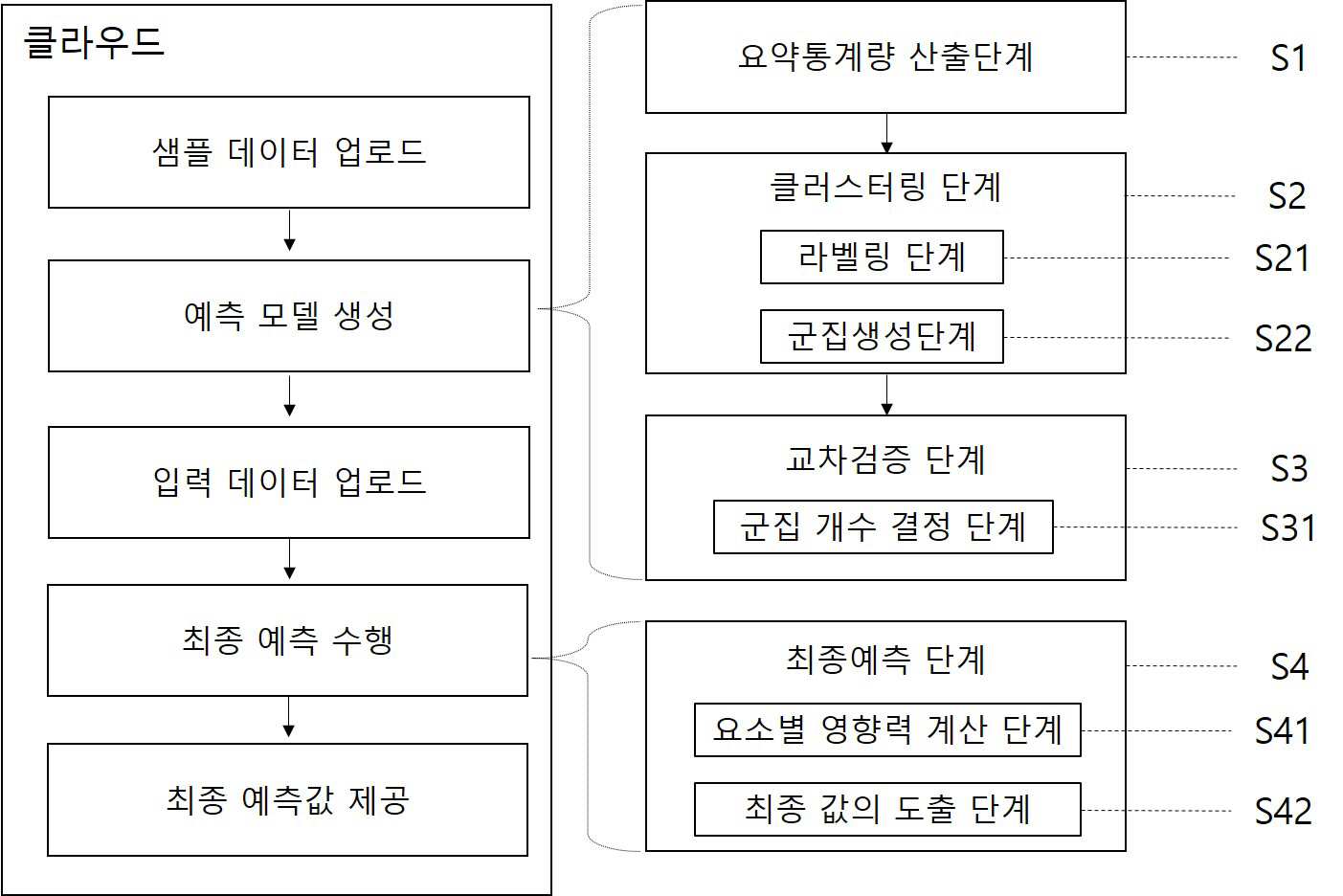
도 10

【도면】

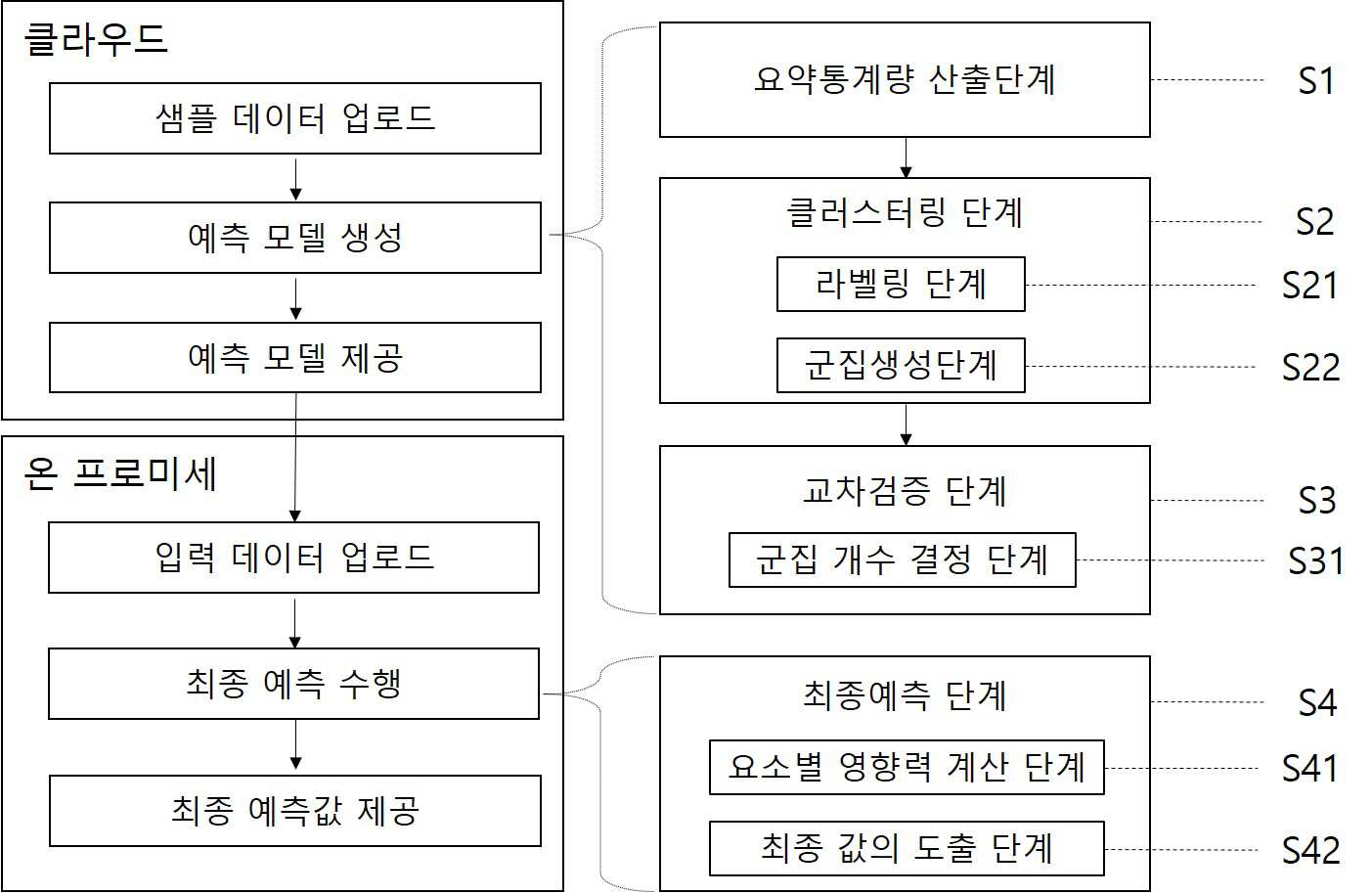
【도 1】



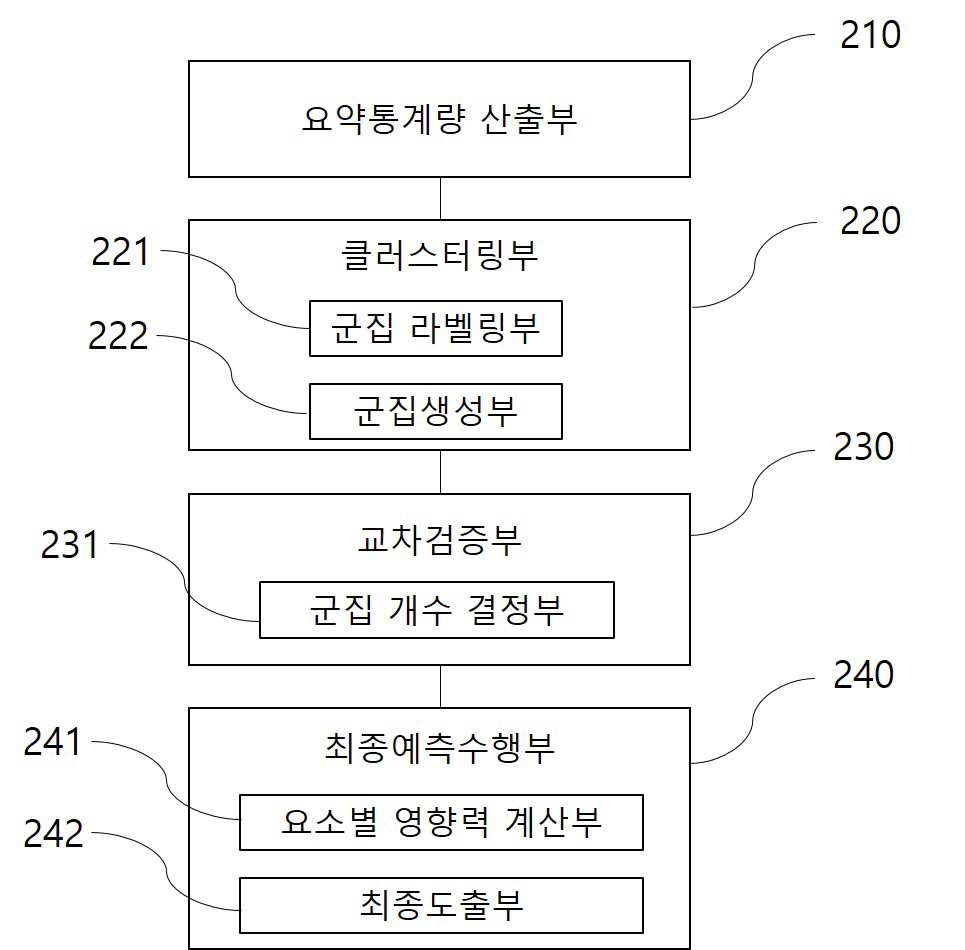
【도 2】



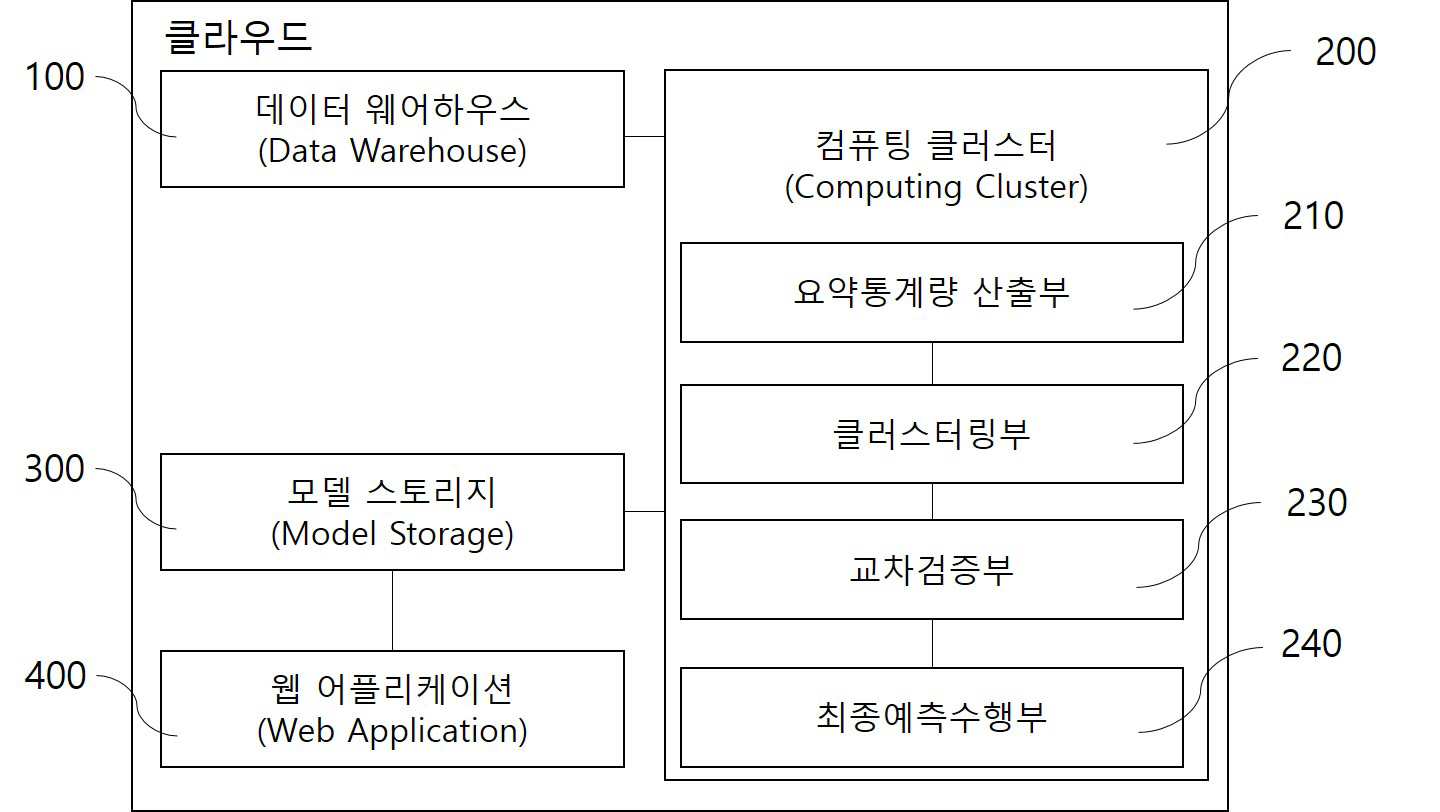
【도 3】



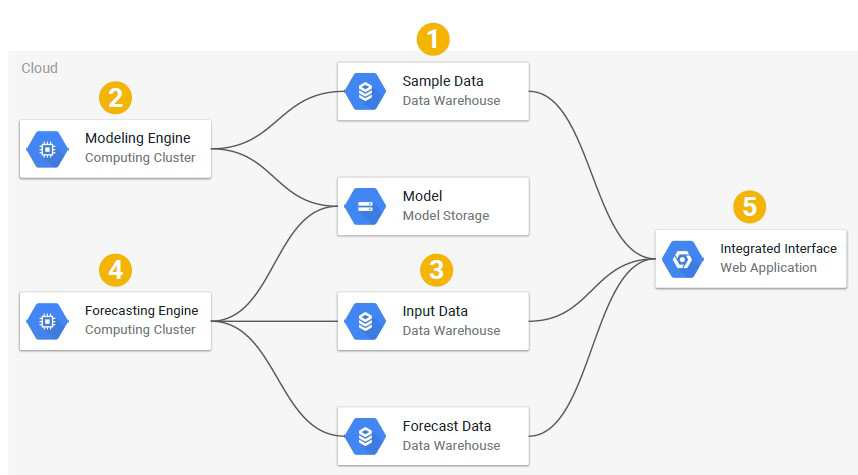
【도 4】



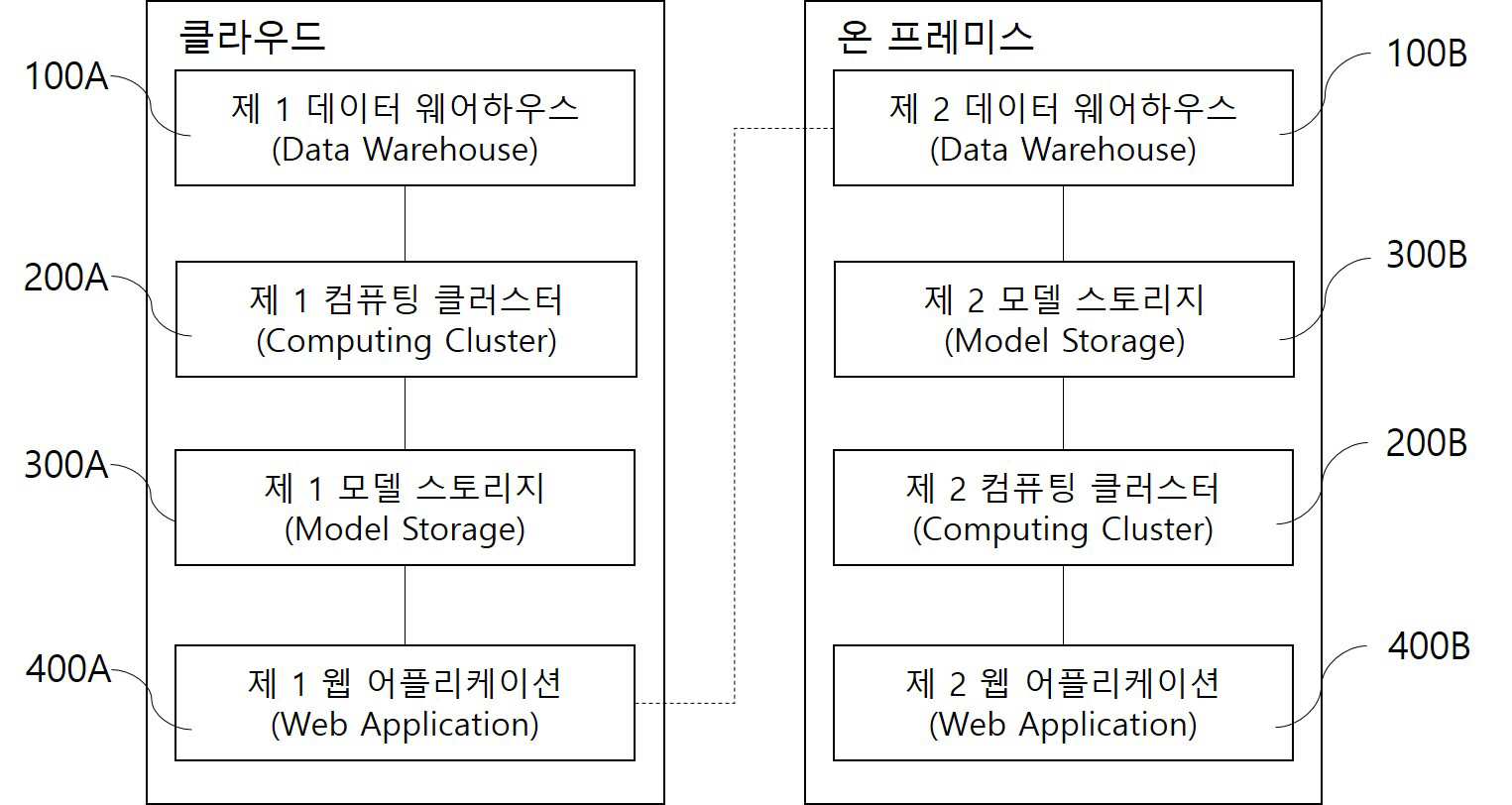
【도 5】



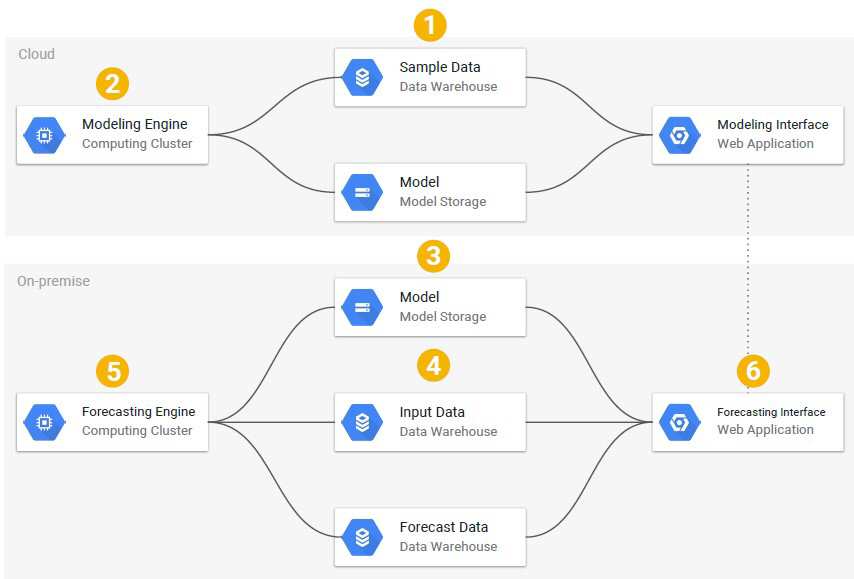
【도 6】



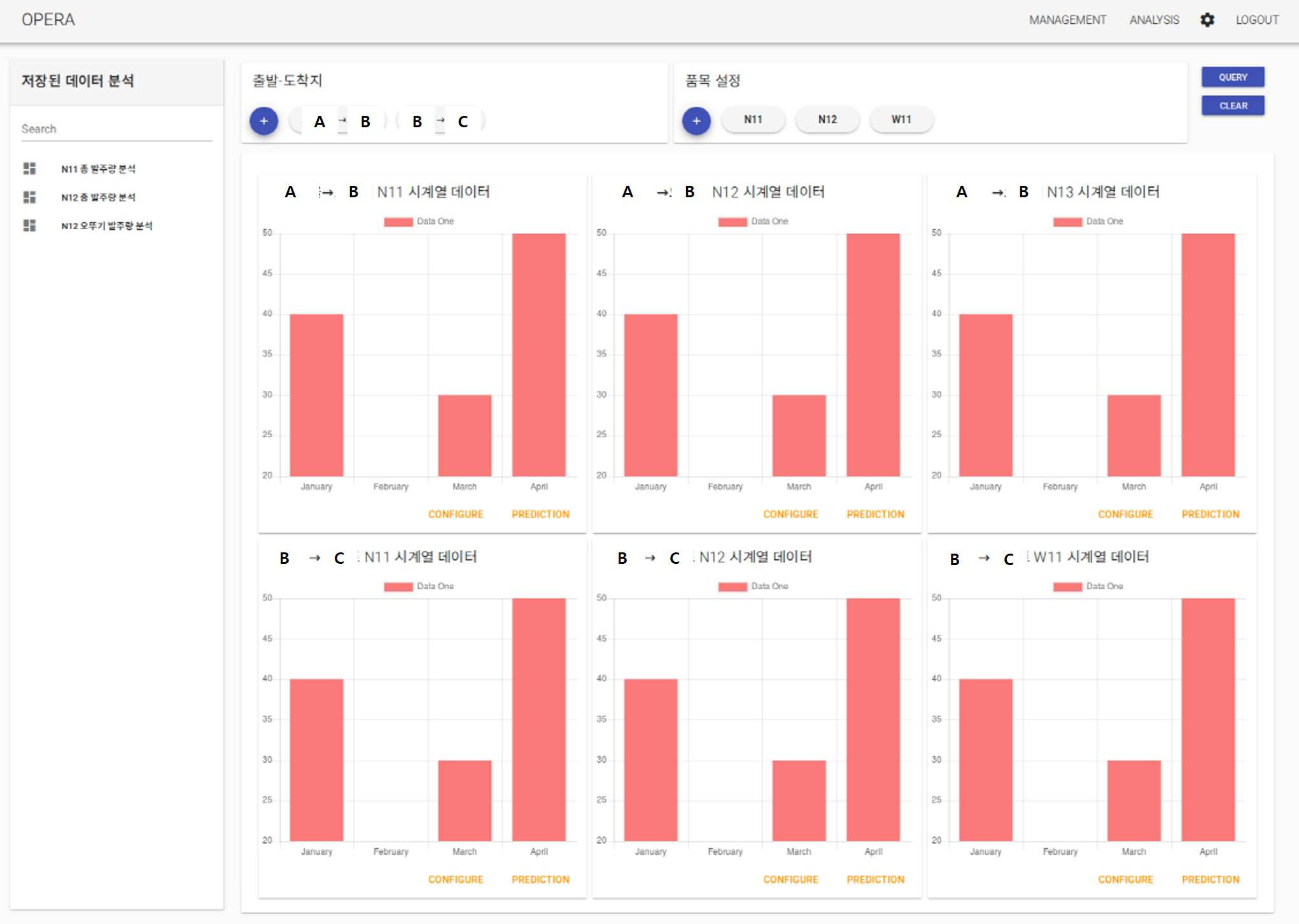
【도 7】



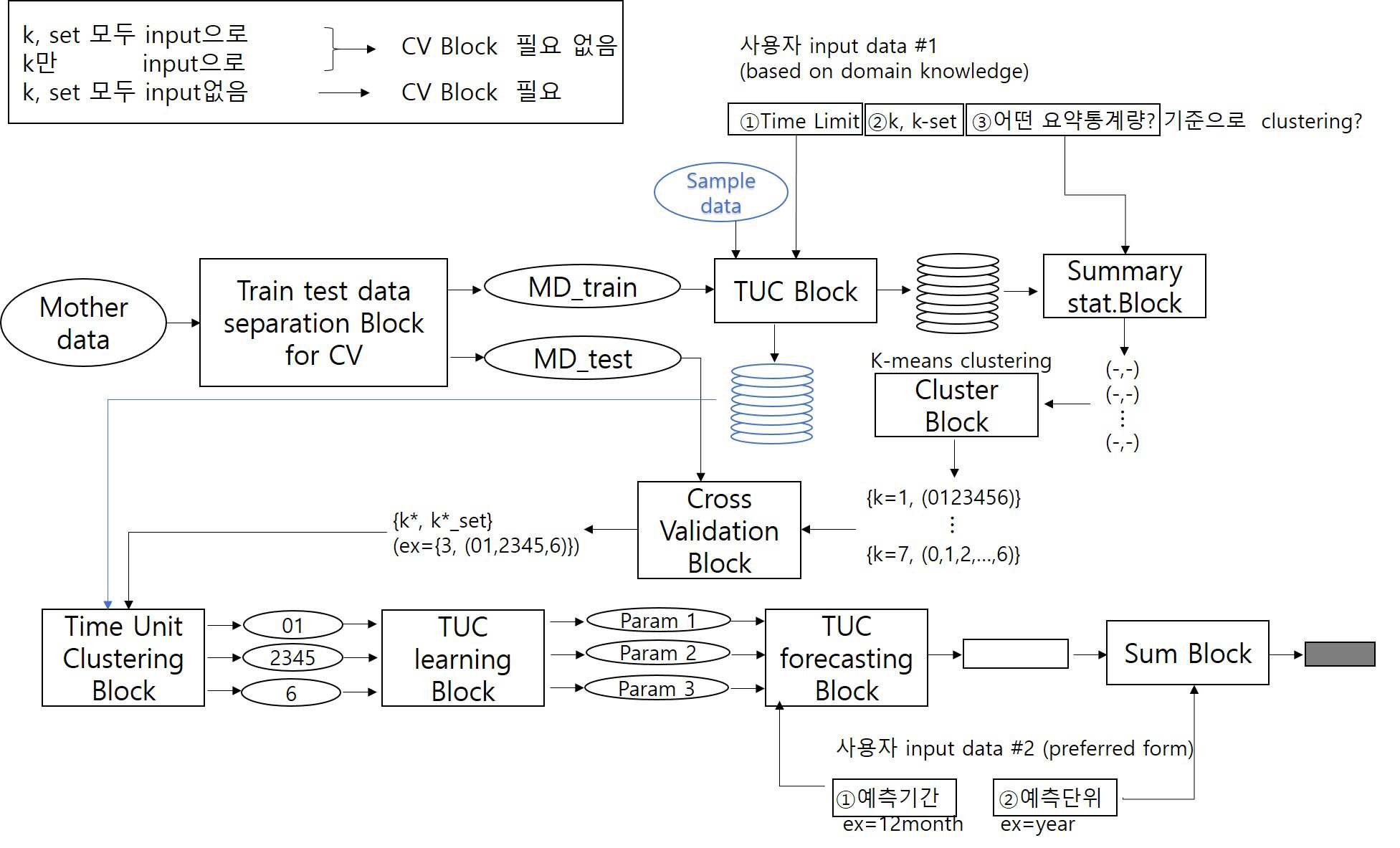
【도 8】



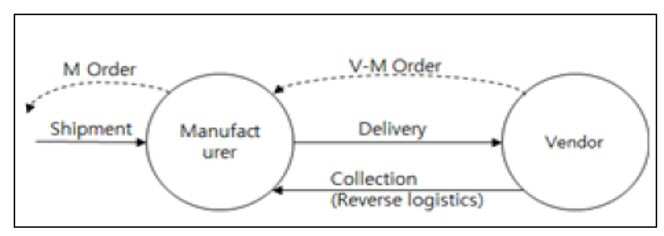
【도 9】



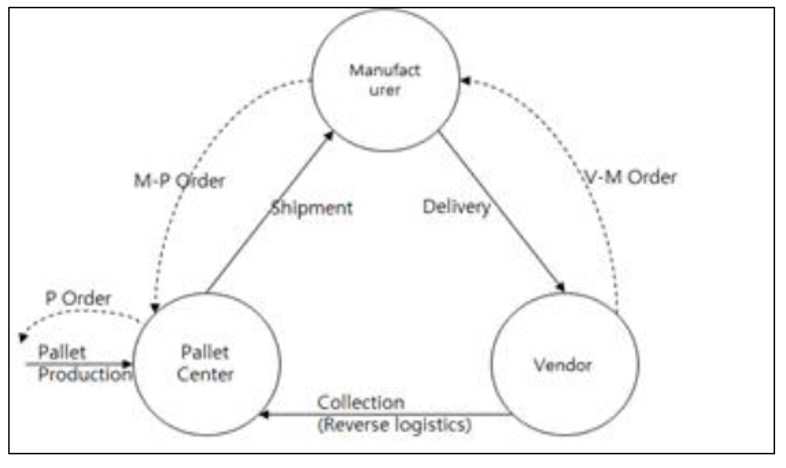
【도 10】



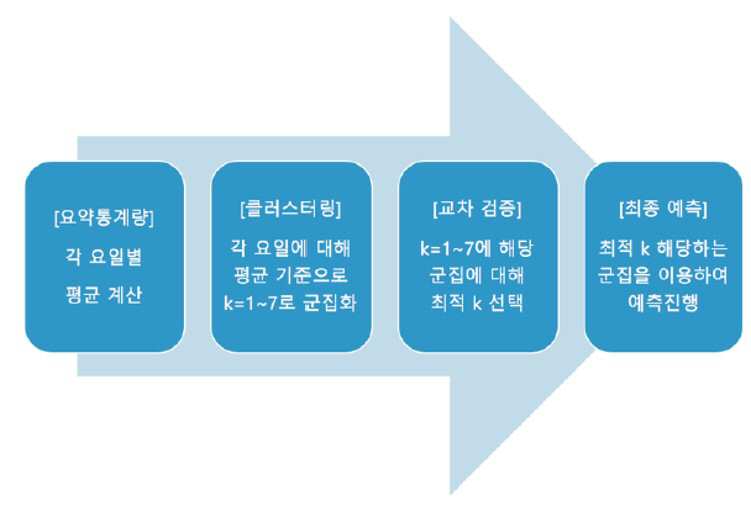
【도 11】



【도 12】



【도 13】



【도 14】

