

일별 굴 생산량의 예측 가능성에 관한 연구

성수진, 박성재^o, 차정원
창원대학교

20153057@changwon.ac.kr, tjdw01289@changwon.ac.kr, jcha@changwon.ac.kr

A Study on the Daily Predictability of Oyster Production

Su-Jin Seong, Seong-Jae Park^o, Jeong-Won Cha
Changwon National University

요약

농수산물의 생산량을 예측하기 위해 관련 요인을 찾고 다중회귀분석, 시계열분석, 인공신경망 등을 사용한 연구들이 진행되어왔다. 하지만 기존의 연구에서는 예측 범위가 연도별 또는 월별로 제한되어 본 연구에서는 일별 굴 생산량 예측 모형을 제안한다. 예측 모형의 구축하기 위해 인공신경망과 최소자승법, 단계별 선택법을 이용한 다중선회귀모형을 사용하였고, 학습 자료의 범위에 차이를 두었다. 평가 결과, 학습 자료의 수가 많을수록 좋은 성능을 나타냈고 단계별 선택법을 이용한 모형이 가장 정확한 예측 결과를 보였다.

1. 서론

1차 산업은 2차 산업과 달리 생산량을 예측할 수 있는 기법이 상대적으로 부족하다. 또한 생산량을 예측할 수 있는 요인이 부정확하여 요인 간의 연관관계를 찾기가 어렵다.

굴 산업은 9월부터 4월까지 조업이 이루어지는 시간 한정이 있는 산업이다. 생산이 바다에서 이루어지기 때문에 바다의 환경에 영향을 많이 받고 바람과 태양 등의 대기 환경 그리고 인근 육지 환경에 의해서도 영향을 많이 받는 산업이다. 따라서 이 요인들이 굴 생산에 어떠한 영향을 미치는지 조사하기 위해서는 오랜 시간과 넓은 공간에 대한 조사가 함께 이루어져야 하는 어려움이 있다. 기상청, 해안관리공단 그리고 지방자치단체 등에서 바다, 대기에 대한 정보를 제공하고 있다. 그렇지만 이들 주체가 제공하는 데이터의 형식과 측정 시점 등이 상이하여 연관관계를 찾기 위한 데이터를 수집하는데 어려움이 있다.

본 연구는 1차 산업에 속하는 굴 산업의 생산량에 영향을 미치는 각 요소들을 찾고 이들의 연관관계를 분석하여 굴의 일별 생산량을 예측하는 모델을 제시한다. 생산량의 일별 예측이 가능하면 굴 산업의 특성상 조업(채취시간, 횟수, 보관 등)에 대한 계획 설정이 가능해져 조업의 효율과 신선도를 증대시킬 수 있다.

2. 관련연구

농산물 생산량 예측 관련 연구로, [1]에서는 기상 요인을 고려하여 쌀, 보리, 콩 그리고 감자의 생산량에 대한 예측모형을 제시하였고, [2]에서는 시군으로 범위를 좁혀 쌀 생산량을 예측한 연구가 진행되었다. 또한 [3]에서는 쌀 생산량을 예측하기 위해 인공신경망을 사용하였다.

수산물 관련으로는 대구 어획량을 예측하기 위해 수온과 월별 대구 어획량을 가지고 회귀분석을 실시한 연구와[4] 멸치어업의 월별 어획량 예측에 대하여 ARIMA 모형을 연구가 진행되었고 [5], [6]에서는 일별 굴 생산량 예측을 위해 다중회귀모형, SARIMA 모형, VECM 모형을 이용하였다.

이러한 농수산물의 생산량을 예측하기 위한 연구가 진행되었지만 일별 생산량을 예측하는 연구는 아직 없었다. 이에 본 연구에서는 굴 생산량에 대해 다중회귀모형과 인공신경망을 이용하여 예측 모형을 구축하고 각 모형을 분석, 비교하여 일별 굴 생산량의 예측 가능성을 제시한다.

3. 자료 분석

본 논문에 사용된 굴 생산량(생굴 기준)과 작업 인원수 관측치는 116개로 2016년 11월부터 2017년 3월까지 직접 관측한 일별 자료를 이용하였다. 또한 생산량에 영향을 미칠 수 있는 환경 요소로 기온과 수온, 풍속을 고려하였고 이들 자료는 굴 생산량 자료를 관측한 통영 지역을 기준으로 하여 기상청의 자료를 이용하였다.

4. 연구방법

본 연구는 굴 생산량과 관련이 있다고 판단되는 요소를 분석하여 당일 굴 생산량 예측 모형을 구축하는 것을 목표로 한다. 모형을 구축하는데 사용되는 자료의 범위를 결정하기 위해 전체 자료를 월별로 나누고, 각 월별 자료를 평가 자료로 사용하였다. 학습 자료는 평가 자료를 기준으로 이전 1달에서 4달까지의 자료를 이용하여 각각의 모형을 구축하고 비교·분석하여 효율적인 자료의 범위를 정한다. 또한 효율적인 굴 생산량 예측 모형 구축을 위해 회귀 계수의 산정을 위해 최소자승법과 단계별 선택법을 사용한다.

4.1 회귀 계수 산정

4.1.1 최소자승법

모형 구축을 위해 당일 굴 생산량을 예측하기 위해 설명 변수로 오전 굴 생산량(AMPROD)과 오전, 오후 작업 인원수(AMWORKER, PMWORKER)를 포함해 작업자의 효율성에 영향을 줄 것이라 판단되는 기온(TEMP), 습도(HUM), 풍속(WINDS)을 고려하였고, 평가 자료의 범위를 제외한 모든 자료를 사용하

였다. 각 독립변수의 계수를 산정하기 위한 추정기법으로 최소자승법을 이용하였다. 종속 변수인 굴 생산량에 대한 독립변수의 영향력을 판단하기 위해 t-검정 값을 검토하였다.

표 1은 평가 자료를 제외한 모든 자료를 이용하여 굴 생산량에 대해 최소자승법으로 모형을 구축한 결과를 나타낸다. 표에는 11월 자료를 평가 자료로 사용하고 나머지 12월, 1월, 2월, 3월 자료를 학습 자료로 사용한 train(12,1,2,3) 모형부터 3월 자료를 평가 자료로 사용하고 11월, 12월, 1월, 2월 자료를 학습 자료로 사용한 train(11,12,1,2) 모형까지의 나타나 있으며, 각 열은 상수부터 각 변수에 대한 계수 값을 나타낸다. 표의 내용에 따르면 3월의 굴 생산량 예측 모형은 다음 식과 같이 표현된다.

$$\begin{aligned}
 TOTALPROD = & 0.0198 \\
 & + 0.8968 \times AMPROD \\
 & - 0.1415 \times AMWORKER \\
 & + 0.3345 \times PMWORKER \\
 & + 0.0934 \times TEMP \\
 & - 0.0636 \times HUM \\
 & + 0.0294 \times WINDS
 \end{aligned}$$

각 변수의 t-검정 값을 살펴볼 때 굴의 총 생산량 예측 모형에 가장 큰 영향을 미치는 독립변수는 t-검정 값이 22 이상인 오전 생산량임을 알 수 있다. 전체 모형에서 오전 생산량의 계수는 평균 0.85로, 이는 다른 조건이 일정할 시 오전 생산량이 1% 증가할 때 굴 생산량은 평균 0.85% 증가하는 것을 의미한다. 그 다음으로는 오후 작업 인원 수, 오전 작업 인원수가 종속변수에 대한 영향력이 높은 것을 알 수 있다.

표 1 최소자승법을 이용한 모형의 계수 및 t-검정 값

	train(12,1,2,3)		train(11,1,2,3)		train(11,12,2,3)		train(11,12,1,3)		train(11,12,1,2)	
	계수	t-검정	계수	t-검정	계수	t-검정	계수	t-검정	계수	t-검정
Intercept	-0.0098	-0.3315	0.0105	0.3932	0.0046	0.1699	-0.0371	-1.9023	0.0198	0.727
AMPROD	0.8826	22.4876	0.8515	22.5556	0.8622	23.9419	0.7794	23.0825	0.8968	23.8493
AMWORKER	-0.1194	-1.8138	-0.1183	-2.2976	-0.1312	-2.5353	-0.1045	-2.7485	-0.1415	-2.2583
PMWORKER	0.3001	3.9843	0.3398	8.0393	0.3333	7.667	0.3058	9.6228	0.3345	7.0331
TEMP	0.0375	0.9354	0.045	1.3973	0.0531	1.5453	0.0327	1.4087	0.0934	2.843
HUM	-0.0131	-0.3817	0.0025	0.0716	-0.0236	-0.8185	-0.0184	-0.8402	-0.0636	-1.976
WINDS	0.0326	1.1854	0.0272	0.9155	0.0187	0.6545	0.0282	1.4728	0.0294	1.1004

표 2 단계별 선택법을 이용한 모형의 계수 및 t-검정 값

	train(12,1,2,3)		train(11,1,2,3)		train(11,12,2,3)		train(11,12,1,3)		train(11,12,1,2)	
	계수	t-검정	계수	t-검정	계수	t-검정	계수	t-검정	계수	t-검정
Intercept	-0.0178	-0.6459	0.0102	0.3866	0.012	0.4595	-0.0357	-1.8357	0.0208	0.7636
AMPROD	0.8878	22.9754	0.8549	23.2655	0.849	24.9828	0.7721	23.3528	0.8998	23.9619
AMWORKER	-0.1315	-2.0449	-0.1246	-2.462	-0.1285	-2.5178	-0.1058	-2.7891	-0.1498	-2.4059
PMWORKER	0.3088	4.2169	0.3443	8.3414	0.3383	7.9373	-	-	0.3407	7.2032
TEMP	-	-	0.0376	1.4531	-	-	-	-	0.0854	2.6626
HUM	-	-	-	-	-	-	-	-	-0.0634	-1.9667
WINDS	-	-	-	-	-	-	0.3106	9.8624	-	-

4.1.2 단계별 선택법

표 2는 최소자승법을 이용한 모형과 같이 평가 자료를 제외한 모든 자료를 학습 자료로 이용하여 단계적 선택법으로 계수를 산정한 결과를 나타낸다.

4.2 인공신경망 분석

본 연구에서는 인공신경망 모형을 구축하기 위해 오전 생산량, 오전 작업자 수, 오후 작업자 수, 기온, 습도, 풍량으로 입력 노드를 구성하였고, 은닉 노드는 3개, 출력 노드는 총 굴 생산량으로 구성하였다. 입력 자료는 모두 0과 1사이의 값을 가지도록 정규화 과정을 통해 변환하였다.

5. 결과 및 성능 비교

다중회귀분석모형에 사용된 모든 변수 간 다중공선성 유무를

점검해 본 결과, 변수와 상수의 VIF(분산 팽창 인수, Variance Inflation Factor)가 모두 10이하로 다중공선성 문제는 발생하지 않았다. 구축된 모형을 이용하여 평가 자료에 대한 예측을 수행하였고, 각각의 모형을 비교·평가하기 위해 수정 결정계수(adjusted R-square, adj.R)와 평균제곱근오차(Root Mean Square Error, RMSE), Nash-Sutcliffe 효율계수(Nash-Sutcliffe Efficiency Coefficient, NSE)를 이용하였다.

5.1 최소자승법

모형을 구축하기 위해 기준을 test month, 즉 11월, 12월, 1월, 2월, 3월로 두고, 각 기준에서 이전 1달에서 4달까지의 자료를 평가 자료로 사용하였다. 예를 들어, test month가 11월이고 학습 자료로 이전 2달의 자료를 사용할 경우, 2월, 3월 자료가 학습 자료로 이용된다. 표 3은 각 자료로 모형을 구축한 뒤

최소자승법을 이용하여 회귀 계수를 산정한 모형의 통계지표를 나타낸다. 모형을 구축하는데 이전 자료를 많이 사용할수록 예측 정확도가 높아짐을 알 수 있다. 하지만 평가 자료로 2월 자료를 사용하고 학습 자료로는 12월과 1월의 자료를 사용한 모형이 RMSE는 429.4, NSE는 0.924로 성능이 가장 좋게 평가되었다. 2월의 생산량이 이상치를 많이 포함하고 있어 많은 범위의 자료를 참고할수록 오차가 커진다고 판단된다.

표 3 최소자승법을 이용한 모형 통계지표

범위	통계지표	test month				
		11	12	1	2	3
1	adj.R	0.983	0.914	0.938	0.926	0.983
	NSE	0.674	0.807	0.782	0.740	0.396
	RMSE	315.7	275.2	268.3	795.5	597.3
2	adj.R	0.953	0.961	0.916	0.936	0.971
	NSE	0.766	0.788	0.908	0.924	0.526
	RMSE	267.1	288.5	174.3	429.4	529.1
3	adj.R	0.956	0.952	0.950	0.924	0.964
	NSE	0.784	0.902	0.917	0.916	0.903
	RMSE	257.1	196.5	165.9	451.8	238.9
4	adj.R	0.960	0.951	0.955	0.946	0.955
	NSE	0.793	0.916	0.922	0.905	0.925
	RMSE	251.4	181.2	160.8	480.4	210.4

5.2 단계별 선택법

최소자승법과 동일한 방식으로 모형을 구축하고, 단계별 선택법을 통해 종속변수 예측에 기여하는 정도를 판단하여 유의한 독립변수를 추가, 제거하여 회귀 계수를 산정하였다.

표 4는 단계별 선택법을 이용하여 회귀 계수를 산정한 모형의 통계지표를 나타낸다. 표에서 나타는 것과 같이 단계별 선택법을 통해 독립변수를 선별하여도 학습 자료의 범위 크기와 예측 정확도는 비례하였다. 평균제곱근오차의 경우에는 test month가 11월인 모형을 제외하고 최소자승법과 비교해 줄어들어 단계별 선택법 모형의 성능이 보다 개선되었음을 보인다.

표 4 단계별 선택법을 이용한 모형 통계지표

범위	통계지표	test month				
		11	12	1	2	3
1	adj.R	0.985	0.922	0.950	0.952	0.988
	NSE	0.756	0.820	0.660	0.669	0.458
	RMSE	272.8	266.0	334.9	898.0	565.9
2	adj.R	0.957	0.963	0.921	0.938	0.971
	NSE	0.687	0.802	0.908	0.938	0.434
	RMSE	309.2	278.4	174.5	387.2	578.6
3	adj.R	0.958	0.954	0.951	0.925	0.965
	NSE	0.711	0.912	0.912	0.906	0.915
	RMSE	297.4	185.9	170.7	479.9	223.9
4	adj.R	0.961	0.953	0.956	0.946	0.956
	NSE	0.74	0.920	0.922	0.906	0.930
	RMSE	282.1	177.7	160.7	477.4	203.9

5.3 인공신경망 분석

인공신경망 모형과 최소자승법 모형을 비교하기 위해 두 모형 모두 test month를 제외한 모든 자료를 학습에 사용하였다. 인

공신경망을 이용한 모형과 다중회귀분석방식을 이용한 모형의 통계지표는 각각 표 5와 표 6에서 확인할 수 있다.

표 5 인공신경망을 이용한 [0-1] 모형 통계지표

범위	통계지표	test month				
		11	12	1	2	3
4	NSE	0.862	0.854	0.872	0.890	0.901
	RMSE	205.6	229.4	188.1	423.8	222.0

표 6 최소자승법을 이용한 [0-1] 모형 통계지표

범위	통계지표	test month				
		11	12	1	2	3
4	NSE	0.901	0.882	0.903	0.882	0.896
	RMSE	173.6	206.1	163.7	439.9	227.9

인공신경망을 이용한 모형이 더 높은 성능을 보일 것이라는 예상과는 달리 다중회귀분석 모형과 비슷하거나 낮은 성능이 보였다. 이는 모형 구축에 사용된 입력 자료의 수가 100개 미만으로 학습을 진행하기에는 몹시 부족한 수이기 때문이라고 사료된다.

6. 결론

본 연구는 굴 생산량의 일별 예측 모형을 만드는 것을 목표로 하였다. 굴의 향후 일별 생산량에 영향을 주는 요소를 t-검정과 단계별 선택법을 이용해서 찾아내고 이들의 관계를 분석하여 예측 모형을 만들었다.

최소자승법, 인공신경망을 이용해 만든 모형을 이용해서 예측 성능을 측정하였다. 자료의 부족으로 성능 향상에 한계가 있음을 확인하였고 추가적인 데이터를 수집할 예정이다.

향후 추가적인 데이터 수집을 통해 학습 데이터를 확장할 예정이며 심층신경망과 추가적인 모델을 기법을 이용하여 보다 정확한 모델을 만들 예정이다.

이 논문은 2017년 미래창조과학부의 재원으로 SW융합기술고도화 사업의 지원을 받아 수행된 연구임(No. S0177-16-1056).

참고문헌

- [1] 장석환, 주요 식량작물의 생산량 예측 모형에 관한 연구, 한국데이터정보과학회지, 11권, 1호, 47-55, 2000
- [2] 정은정, 통계적 방법을 이용한 쌀 생산량 예측 방법에 대한 고찰, 학위논문(석사), 2012
- [3] 박진기, 인공신경망을 이용한 쌀 생산량 예측 및 기후변화 시나리오 적용, 학위논문(박사), 2014
- [4] 박종연, 여수 해역 대구(*Gadus macrocephalus*) 어획량과 수온의 상관관계 및 추정, 학위논문(석사), 2013
- [5] 박해훈, 윤갑동, 한국 멸치어업의 어획량 분석과 예측-ARIMA 모델 및 스펙트럼 해석-, 한국수산과학회지, 29권, 2호, 143-149, 1996
- [6] 남종오, 노승국, 시계열모형을 이용한 굴 생산량 예측 가능성에 관한 연구, Ocean and Polar Research, 34권, 2호, 185-195, 2012