Journal of the Korean Data & Information Science Society
2009, **20**(3), 551–562

한국데이터정보과학회지

2단계 군집분석을 통한 해구별 조업정보의 유사성 분석

조용준[1]

수협중앙회 수산경제연구원

접수 2009년 2월 17일, 수정 2009년 5월 20일, 게재확정 2009년 5월 23일

요 약

수협의 어선조업정보는 국가 공식 통계가 가지고 있지 못한 위치별 조업정보를 가지고 있다는 장점이 있다. 위치별 조업정보는 해당 지역의 어업피해보상, 자원가치 산출 등을 추정할 수 있어 국가통계자료로의 가치가 매우 높으나 어업인들의 자기 정보의 노출에 대한 기피로 인해 신뢰성이 떨어지는 단점을 지니고 있다. 본 연구는 유용성분석을 통해 이러한 수협의 어선조업정보의 활용을 위한 방안을 제시하고 위치별 조업패턴의 특성을 분류하여 해구별 유사성의 정보를 산출을 목적으로 하였다. 분석결과 수협의 어선조업정보는 정부 생산통계대비 어획량의 약 33% 수준이나 유용성 분석에서 그 패턴과 상관관계가 밀접해 위치별 패턴파악에 유용한 것으로 나타났다. 이를 바탕으로 대해구별 2단계 군집분석을 통해 어획량, 조업일수, 조업척수에 대해 각각 최적의 군집을 구분하고 이를 종합하여 8개의 군집으로 패턴을 구분하였다.

주요용어: 2단계 군집분석, 분류분석, 어업조업정보, 유사성분석, 해구.

1. 서 론

우리나라의 공식적인 수산물 생산통계는 농림수산식품부 (농수부)의 어업생산통계이다. 하지만 여기서 제공되는 정보는 위치별 생산통계를 제시하고 있지 못하다는 단점이 있다. 어획자원은 특정 해구 또는 위치에 주로 분포하면서 계절에 따라 변동성이 생기는데 이에 대한 연구는 축적된 데이터가 아닌 표본조사에 의해 연구되고 있는 실정이다.

수협중앙회 (수협)의 어선조업정보 (조업정보) 시스템은 신 한·일 어업협정 체결 이후, 기초자료의 미흡과 어업자원산정의 어려움 등으로 교섭의 애로가 발생하여 1999년 구축되었고 지난 10년간 우리나라 연근해 및 EEZ 지역의 조업정보를 저장하고 있다. 수협의 조업정보시스템은 현재 2008년 11월 기준으로 약 43백만건의 출어선 조업정보가 저장되어 있는 등 엄청난 양의 조업정보를 보유하고 있다. 하지만 이에 대한 활용은 정부측에 자료를 제공하는 것 이외에는 미미한 실정이다. 그 이유는 어업인들이 어선의 위치, 조업실적에 대한 노출을 꺼려 신뢰성 있는 정보를 제공하지 않고 있기 때문이다. 그러나 수협의 조업정보는 위치별 조업정보를 보유하고 있으며 원하는 시점에서 실시간으로 조업정보의 확인이 가능하다는 장점을 지니고 있다. 대부분의 국가 통계는 수집되는 기간의 시차가 수개월이 걸려 시급한 연구에는 활용이 어렵다는 단점을 지니고 있다. 즉, 위치별 조업정보의 활용이 중요한 이유는 허베이스피리트호 사건 등과 같은 유류유출사고, 석유탐사피해보상, 국제간의 분쟁 등과 같은 피해 사례발생 시 피해액 산출을 위한 추정이 가능하며 어업정보가 정확하지 않다하더라도 어종별/어업별 생산량 패턴의 유사성이 확인되면 이에 대한 추정이 일정부분 가능하기 때문이다.

[1] (138-730) 서울 송파구 신천동 11-6, 수협중앙회 수산경제연구원, 수석연구원.
E-mail: ehclub@nate.com

따라서 본 연구는 조업정보에 대한 유용성 분석을 통해 활용 가능한 통계자료인지를 확인하여 활용방안을 제시하고 위치별 조업패턴의 특성을 분류하여 해구별 유사성 정보의 산출을 목적으로 한다.

이를 위한 방법으로 먼저 조업정보의 유용성 분석을 위하여 기존 농수부 어업생산통계와의 상관분석 및 시각적 방법에 의한 패턴분석을 통해 유용성을 분석하여 평가하였다. 해구별 조업패턴의 유사성분석을 위해서는 다차원 통계분석 방법인 군집분석 방법을 이용하였다. 군집분석 방법에는 K-평균 (means) 군집분석으로 대표되는 (Selim과 Ismail, 1984; Kovesi 등, 2001) 최적화 방법과 계층적 군집화 방법 (Hand, 1981), 2단계 군집화 (two step clustering) 방법, 자율학습에 의한 자기조직화 지도 (SOM: self-organization maps)의 코호넨 네트워크 (Kohonen, 1982) 등이 있는데 이중 슈와르츠 (Schwarz, 1978) 베이지언 정보기준 (Schwarz Bayesian Criterion: SBC)과 아카이케 정보기준 (Akaike information criterion: AIC, Akaike, 1976)을 통해 최적 군집수의 산출이 가능한 장점을 가지고 있는 2단계 군집분석을 선택하여 분석하였다. 이를 통해 해구별 유사성을 최적군집으로 구분하여 해구별 특성을 제시하고자 한다.

2. 해구별 유사성 분석을 위한 설계와 프로세스

해구별 유사성 분석을 위한 프로세스는 데이터 마이닝분석 방법론 중 하나인 CRISP-DM (Champman 등, 1999) 방법론에 입각하여 기획하였다. 분석을 위한 프로세스는 다음의 그림 2.1과 같다. 먼저 분석을 위한 데이터는 수협의 조업정보와 농수부 어업생산통계를 통해 어획량, 조업일수, 조업척수 데이터를 추출하여 분석용 데이터셋을 구축한다. 다음으로 데이터 유용성분석에서는 수협 조업정보와 농수부 어업생산통계간의 어획량 비교 및 시각화를 통한 패턴비교, 상관분석을 통해 수협 조업정보의 유용성을 상대적으로 분석하여 이용가능성을 제시한다. 다음으로 위치별 정보인 해구별 데이터에 대해 품질분석을 통해 통계적으로 유용한 해구를 도출하여 패턴파악을 한 후, 2단계 군집분석을 통해 해구별 유사성에 대하여 분류분석하여 유사한 해구를 제시하도록 하였다.

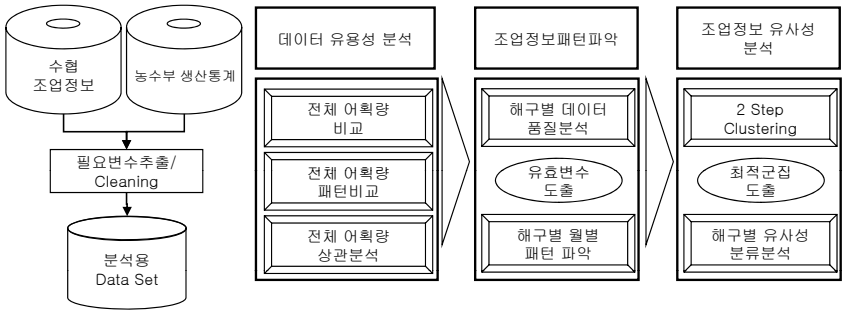

그림 2.1 분석 프로세스

분석을 위한 데이터는 조업정보의 유용성 분석에서는 연도별/월별 데이터를 이용하여 비교하였으며 해구별 조업정보의 유사성 분석에서는 조업정보의 특성상 계절성의 패턴이 중요하기 때문에 월별 데이터를 이용하였다. 사용 데이터의 기간은 2002년 1월부터 2008년 7월까지의 6년 7개월 (79개월)의 월별데이터를 추출하였다. 사용된 변수인 어획량의 단위는 데이터의 유용성분석에서는 M/T이고 유사성 분석에서는 kg이다. 조업일수는 일, 조업척수의 단위는 척이다. 해구별 자료는 크게 대해구와 소해구로 구분할 수 있는데 범주의 크기를 제한하기 위하여 대해구를 이용하였다. 이용된 대해구는 총 528개 대

해구이다.

3. 수협 조업정보의 유용성 분석

수협 조업정보를 활용하기 위해서는 서론에 언급하였던 것과 같이 수협 조업정보의 신뢰성에 대한 점검이 필수적이다. 따라서 데이터에 대한 유용성분석을 통해 축적되어 있는 수협 조업정보의 유용성 정도의 파악이 필요하다. 그런데 이러한 데이터 유용성분석은 신뢰성이 보장되는 비교 가능한 정보 또는 통계와의 비교를 통해 가능하다. 현재 조업과 관련된 비교 가능한 통계는 농수부의 어업생산통계이다. 유용성 분석을 위해서는 전체 어획량뿐만 아니라 업종별 또는 어종별 세부 비교를 통해 심층적 분석을 수행하여야 보다 정확한 결과를 산출할 수 있다. 따라서 전체 연구과정에서는 이러한 세부적 비교분석을 수행하였으나 본 연구의 결과에서는 범위의 방대함으로 인해 전체 어획량에 대한 비교분석 결과만을 제시하도록 하였다. 수협 조업정보가 연근해 및 EEZ지역 출어어선의 어획량 정보이기 때문에 농수부 어업생산통계의 일반해면어업 생산량과 비교하였다.

3.1. 전체어획량 비교

다음의 표 3.1은 수협 조업정보의 어획량과 농수부 일반해면어업 생산량과의 연도별 비교에 대한 표이다.

표 3.1 전체 어획량 비교

년도	농수부 일반해면어업생산량(A)	수협어획량(B)	B/A비율(%)
2002년	1,095,812	329,305	30.05
2003년	1,096,526	312,994	28.54
2004년	1,076,687	373,930	34.73
2005년	1,097,041	367,583	33.51
2006년	1,108,815	377,790	34.07
2007년	1,152,299	418,855	36.35
2008년 7월까지	537,260	162,972	30.33
전체평균(08년제외)	1,104,530	363,410	32.90
전체합계	7,164,440	2,343,430	32.71

위의 수협 조업정보의 어획량과 농수부 어업생산통계의 일반해면어업 생산량과의 비교결과 수협의 어획량은 농수부 어업생산통계의 약 33%정도의 수준인 것으로 파악되었다. (※수협 어획량과 농수부 일반해면어업 생산량간에는 자료의 구분에 있어서 차이가 날 수 있음.)

3.2. 전체 어획량 패턴비교 및 상관분석결과

그림 3.1의 시각적 분석에 의한 월별 어획량의 패턴의 비교결과, 두 데이터간에 거의 유사한 어획량 패턴을 보이는 것으로 나타났다. 또한 표 3.2의 두 데이터간 상관분석결과, 상관도가 0.929로 매우 유의한 수준의 상관도가 존재하는 것으로 파악되었다.

위의 분석결과를 종합하면 수협 조업정보가 농수부 어업생산통계의 일반해면어업 생산량 통계의 33%수준이지만 수협 조업정보를 통해 실제 어획량에 대한 패턴 파악이 가능할 것으로 판단된다. 즉, 수협 조업정보를 통해 위치별 조업패턴을 파악하기 위한 연구가 가능할 것으로 판단된다. 어종별 또는 어업별 위치에 따른 세부적 어획량 비교를 통해 산출된 각 구분별 유용성 결과를 가중치로 활용한다면 어

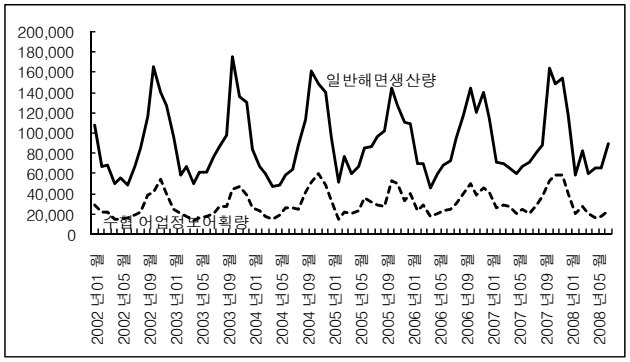

그림 3.1 전체 어획량 패턴 비교

표 3.2 어획량간 상관분석

		전체어획량_농수부	전체어획량_수협
전체어획량_농수부	상관계수	1	.929
	유의수준		0.000
	N	79	79

종별/어업별 어획량의 추정도 근사적으로 가능할 수 있을 것으로 파악되기 때문에 위치별 조업정보의 활용은 정보로서의 가치가 높을 것으로 판단된다.

4. 2단계 군집분석을 통한 대해구 유사성 분석

4.1. 대해구 조업정보의 품질분석 결과

먼저 대해구별 유사성 분석을 위해서는 분석 가능한 대해구별 조업정보의 추출이 중요하다. 조업정보가 거의 없는 해구 또는 결측이 많은 해구의 조업정보를 이용하여 분석할 경우, 분석결과의 신뢰성 및 품질이 크게 떨어질 수 있기 때문이다. 즉, 79개월 동안의 조업정보 중 몇 개월 안되는 해구정보로는 월별 조업의 특징을 산출해 낼 수 없기 때문이다. 따라서 대해구 조업정보의 품질분석을 수행하였다. 먼저 분석 대상이 되는 변수인 대해구별 어획량, 조업일수, 조업척수에 대한 변수의 수는 대해구 총 528개 해구, 1584개 변수, 그리고 전체 어획량 및 조업일수, 조업척수를 포함하여 1587개 변수를 대상으로 품질을 도출하였다. 품질분석 결과 100% 충실도 (0% 결측)인 해구변수가 753개 변수로 47.4%, 80% 이상의 충실도 (20%미만 결측)가 129개로 8.1%이다. 본 연구에서는 80% 이상의 충실도를 보이는 데이터를 분석대상으로 선택하였다. 따라서 882개 변수 (전체 및 293개 대해구)를 대상으로 기술통계량분석과 패턴분석을 수행하였으며 다음의 표 4.1은 품질분석 결과이다. (※ A: 100% 충실도, 0% 결측, B : 80% 이상 충실도, 20% 미만 결측, C : 50% 이상 충실도, 50% 미만 결측, D : 50% 미만 충실도, 50% 이상 결측.)

대해구별 유사성분석에 앞서 변수에 대한 기술통계분석과 월별/연도별 추세를 제시하여 해구에 대한 기초적 특성의 파악이 중요하다. 따라서 각 해구별 기술통계분석과 월별/연도별 추세를 도출하였으나 882개 변수 (294개 해구)에 대한 산출결과의 양이 방대하여 본 연구에서는 기본 품질분석결과만을 제시하였다.

표 4.1 품질분석 결과

범주	N	%
A	753	47.4
B	129	8.1
C	195	12.3
D	510	32.1
전체	1587	100.0

4.2. 대해구 조업정보 유사성분석 목적 및 프로세스

조업정보 유사성분석은 유사한 구분끼리 묶어 일종의 그룹으로 분류하고, 그 특성이 비슷한 집단으로 분류하는 것을 말한다. 이것은 원래 세분화 또는 분류분석으로 정의되는데 본 연구에서는 연구의 목적상 조업정보의 유사성분석으로 정의하였다. 이러한 세분화 또는 분류분석을 하는 이유는 많은 구분들에 대해 1:1로 그 특성을 파악하는 것이 어렵기 때문이며 세분화를 통해 보다 효과적 관리가 가능해져 정보의 유용성을 높일 수 있기 때문이다. 즉, 조업정보의 구분과 같이 다양한 구분의 경우, 세분화를 통해 보다 효과적인 정보의 제공이 가능해진다. 이러한 세분화는 최근 경영분야에서 많이 활용되나 금번 연구에서 정보제공의 효율성을 높이기 위한 방법으로 세분화분석을 수행하였다. 즉, 본 연구에서의 해구별 조업정보의 유사성 분석은 대해구별로 어획량, 조업일수, 조업척수에 대하여 유사한 특성을 가지고 있는 해구별로 분류하고 그 특성을 제시하여 우리나라 연근해에서 중요 해구를 도출하기 위한 분석이다.

이러한 세분화분석 방법은 목적변수(종속변수)의 유무에 따라 달라진다 (Berry와 Gordon, 1997). 목적변수가 존재할 때 사용되는 알고리즘에는 신경망분석 방법, 의사결정나무분석, SVM (Support Vector Machine), 로지스틱 회귀분석, 다중회귀분석, 판별분석 등이 있다 (Cho와 Ko, 2008). 이러한 분석방법은 독립변수와 목적변수의 관계를 통해 각종 인과성과 예측을 위한 방법들이다(Cho와 Hur, 2006).목적변수가 없을 때 사용되는 알고리즘은 군집분석으로 한정된다. 군집분석의 장점은 목적이 뚜렷하지 않거나 기준이 되는 목적변수가 없을 때 사용할 수 있기 때문에 실무에서 그 활용도가 매우 높다고 할 수 있다. 군집분석의 방법에는 서론에서 제시한바와 같이 K-평균 군집분석, 계층적 군집화 방법, 코호넨 네트워크, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies: Zhang 등, 1996) 알고리즘을 발전시킨 2단계 군집 (2 Step Cluster)분석 (Chiu 등, 2001) 등이 있다.

본 연구의 조업정보 유사성분석에는 목적변수가 존재하지 않기 때문에 군집분석방법을 통해 세분화하였고 군집분석방법은 2단계 군집분석방법을 고려하였다. 2단계 군집분석방법은 범주형 변수와 연속형 변수 모두 사용이 가능하며 SBC와 AIC를 통해 최적 군집수의 산출이 가능한 장점을 가지고 있기 때문이다.

유사성 분석의 프로세스는 먼저 293개 대해구에 대한 월 평균 어획량, 조업일수, 조업척수 데이터를 대상으로 SPSS프로그램에서 제공되는 2단계 군집분석 방법을 선택하고 최적군집 추출방법으로 SBC기준을 선택하여 분석을 수행하였다. 분석결과 SBC 정보량이 가장 적은 군집을 최적군집의 수로 선택하였다. 산출된 결과를 바탕으로 군집 프로파일 및 군집별 월별 특성을 각각 도출한 후, 이를 종합적으로 파악하여 최종결과를 제시하는 순으로 분석을 수행하였다. 다음의 그림 4.1은 유사성 분석에 대한 프로세스를 도표화한 것이다.

4.3. 대해구 조업정보 유사성분석 결과

4.3.1. 대해구 군집 특성

다음의 표 4.2는 각 군집 수별 SBC 값이다. 분석결과 군집의 수가 2개일 때. 가장 SBC값이 가장 작

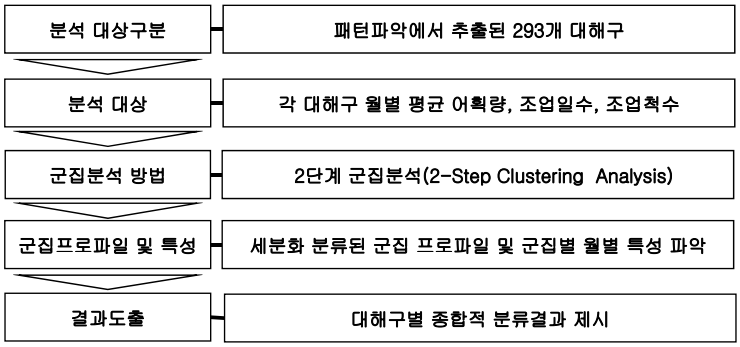

그림 4.1 2단계 군집분석을 이용한 결과도출 프로세스

아 모형적합에 대한 정보의 수준이 가장 높은 것을 알 수 있다. 따라서 어획량, 조업일수, 조업척수 모두 2개의 군집으로 분류하는 것이 적정한 것으로 판단된다.

표 4.2 군집수별 SBC 값

군집수	어획량		조업일수		조업척수	
	SBC	BIC변경	SBC	BIC변경	SBC	BIC변경
1	4566.685		5086.806		5002.089	
2	5460.311	893.626	6116.436	1029.630	5984.093	982.004
3	6624.426	1164.115	7423.051	1306.615	7243.082	1258.989
4	7845.417	1220.992	8820.139	1397.088	8592.544	1349.462
5	9149.262	1303.844	10234.664	1414.525	9962.215	1369.671
6	10503.338	1354.076	11764.836	1530.173	11444.761	1482.546
7	11857.414	1354.076	13295.009	1530.173	12927.308	1482.546
8	13211.490	1354.076	14825.182	1530.173	14412.627	1485.319
9	14565.566	1354.076	16355.355	1530.173	15897.946	1485.319
10	15986.184	1420.618	17903.883	1548.528	17408.212	1510.266
11	17406.802	1420.618	19500.598	1596.715	18957.301	1549.089
12	18827.421	1420.618	21097.313	1596.715	20506.389	1549.089
13	20248.039	1420.618	22694.028	1596.715	22055.478	1549.089
14	21668.657	1420.618	24290.743	1596.715	23604.566	1549.089
15	23089.275	1420.618	25887.457	1596.715	25153.655	1549.089

다음의 표 4.3은 대해구의 어획량, 조업일수, 조업척수에 대한 군집프로파일 결과이다. 대해구에 대한 2단계 군집분석 결과 어획량, 조업일수, 조업척수에 대해 모두 2개의 군집으로 분류되었다. 3개의 군집분석 결과, 상호 유사한 분포로 잘 분류되어 있음을 파악할 수 있다.

표 4.3 대해구 각 구분별 군집 프로파일

구분	어획량		조업일수		조업척수	
	N	퍼센트	N	퍼센트	N	퍼센트
군집 A	28	9.6%	29	9.9%	46	15.7%
군집 B	265	90.4%	264	90.1%	247	84.3%
전체	293	100.0%	293	100.0%	293	100.0%

다음의 표 4.4는 대해구 어획량에 대한 군집특성 결과이다. 군집분석결과 군집A는 군집B에 비해 전

체적으로 12.7배 어획량이 많은 해구이며 전체평균에 비해서도 어획량이 6배 정도 많은 대해구의 특성을 가지고 있다. 따라서 군집A를 대형어획량해구, 군집B를 일반어획량해구로 명칭을 제시한다.

표 4.4 대해구 어획량 군집특성

구분	중심평균 군집 A	군집 B	전체(C)	A/B (%)	A/C (%)
1월	2,238,141	122,249	324,451	18.3	6.9
2월	1,554,901	76,575	217,849	20.3	7.1
3월	1,470,435	97,065	228,309	15.1	6.4
4월	1,132,941	81,544	182,019	13.9	6.2
5월	1,294,506	73,368	190,064	17.6	6.8
6월	1,667,355	90,741	241,407	18.4	6.9
7월	1,424,291	111,492	236,947	12.8	6.0
8월	1,373,729	155,086	271,544	8.9	5.1
9월	1,613,021	219,704	352,854	7.3	4.6
10월	2,682,773	258,487	490,160	10.4	5.5
11월	3,101,959	244,147	517,249	12.7	6.0
12월	2,620,096	213,053	443,078	12.3	5.9
평균	1,847,846	145,293	307,994	12.7	6.0

다음의 표 4.5는 대해구 조업일수에 대한 군집특성 결과이다. 군집A는 군집B에 비해 전체적으로 14.0배 조업일수가 많은 해구이며 전체평균에 비해서도 조업일수가 6.1배정도 많은 대해구의 특성을 가지고 있다. 따라서 군집A를 대형조업일수해구, 군집B를 일반조업일수해구로 명칭을 제시한다.

표 4.5 대해구 조업일수 군집특성

구분	중심평균 군집 A	군집 B	전체(C)	A/B (%)	A/C (%)
1월	1,666	144	295	11.5	5.6
2월	1,330	99	221	13.4	6.0
3월	1,702	130	286	13.1	6.0
4월	1,737	129	288	13.5	6.0
5월	2,132	148	345	14.4	6.2
6월	2,528	128	365	19.8	6.9
7월	2,654	139	388	19.1	6.8
8월	2,319	182	393	12.7	5.9
9월	2,333	171	385	13.6	6.1
10월	2,677	200	445	13.4	6.0
11월	2,303	179	390	12.8	5.9
12월	2,113	168	361	12.6	5.9
평균	2,124	152	347	14.0	6.1

다음의 표 4.6은 대해구 조업척수에 대한 군집특성 결과이다. 군집분석결과 군집A는 군집B에 비해 전체적으로 조업척수의 경우, 9.9배 조업척수가 많은 해구이며 전체평균에 비해서도 조업척수가 4.1배 정도 많은 대해구의 특성을 가지고 있다. 따라서 군집A를 대형조업척수해구, 군집B를 일반조업척수해구로 명칭을 제시한다.

위의 표 4.4~표 4.6의 결과를 종합하면, 어획량, 조업일수, 조업척수에 대해 크게 대형과 일반으로 구분할 수 있다. 어획량이 많은 대형 어획량 해구는 28개 해구로 전체의 9.6%이고 조업일수가 많은 대형 조업일수 해구는 29개 해구로 전체의 9.9%이다, 조업척수가 많은 대형 조업척수 해구는 46개 해구로 전체의 15.7%를 차지하고 있다.

표 4.6 대해구 조업척수 군집특성

중심평균				A/B	A/C
구분	군집 A	군집 B	전체(C)	(%)	(%)
1월	317	32	77	9.8	4.1
2월	238	26	59	9.1	4.0
3월	265	33	70	7.9	3.8
4월	267	32	69	8.3	3.9
5월	307	34	77	9.0	4.0
6월	344	26	76	13.3	4.5
7월	368	33	85	11.3	4.3
8월	369	45	96	8.2	3.9
9월	428	47	107	9.1	4.0
10월	466	42	109	11.1	4.3
11월	419	40	100	10.4	4.2
12월	412	34	93	12.1	4.4
평균	350	35	85	9.9	4.1

4.3.2. 대해구 군집할당 결과

대해구에 대한 어획량, 조업일수, 조업척수에 대한 군집분석결과, 대형해구와 일반해구로 분류되었다. 즉, 어획량별 대형, 일반으로 구분되며 조업일수, 조업척수도 마찬가지로 분류되었다. 따라서 대해구에 대한 총 군집의 수는 23인 8개의 군집으로 분류되었다. 즉, 어획량, 조업일수, 조업척수가 모두 대형인 군집부터 어획량은 대형이나 조업일수와 조업척수는 일반인 군집 등과 같이 8개의 군집으로 대해구를 구분할 수 있다. 분석결과 어획량, 조업일수, 조업척수 모두에서 대형으로 분류된 해구는 87, 97, 99, 104, 110, 163, 174, 222, 231, 234, 243해구, 총 11개로 293개 대해구 중 가장 어획량도 많으며 조업일수, 조업척수가 가장 많은 해구의 특성을 나타내고 있다. 어획량은 대형에 속하진 않지만 조업일수와 조업척수에서 대형에 속하는 해구는 48, 55, 62, 63, 66, 69, 70, 76, 92, 93, 98, 105, 154, 184, 193, 194, 203, 214, 234해구, 총 19개 해구로 어업 노력이 가장 많은 해구이다. 또한 조업일수는 대형해구는 아니나 어획량과 조업척수에서 대형인 해구는 94, 173, 223, 224, 232, 233, 242해구이다. 다음의 표 4.7은 종합적인 군집할당 결과, 대형으로 구분된 해구에 대한 결과표이다. 여기에 제시되지 않은 해구는 일반해구로 본 연구의 결과에서는 지면의 한계로 제시하지 않는다. (※굵은 글씨는 모두 대형인 해구, 이태릭 글씨는 조업일수와 조업척수에서 대형인 해구, 밑줄된 글씨는 어획량과 조업척수에서 대형인 해구임.)

4.3.3. 조업패턴이 대형인 대해구의 위치

군집분석을 통해 산출된 조업패턴 중 2개 이상이 대형인 해구를 지도에 표시하였다.

먼저, 대해구 군집결과, 어획량, 조업일수, 조업척수가 모두가 대형으로 분류된 대해구의 위치는 다음의 그림 4.2와 같다. 산출결과 동해안 1개 해구, 서해안 2개 해구, 남해안 8개 해구로 주로 남해안에 집중되어 있는 것으로 판단된다. 특히, 제주도 인근 4개 해구가 모두 대형인 해구로 나타나 제주도 인근에 중요 해구가 밀집된 것으로 파악되었다.

다음으로 어획량/조업일수/조업척수 중 2개 이상이 대형인 대해구 위치는 아래의 그림 4.3과 같다. 제주도 인근의 해구는 거의 대부분 모두 중형이상의 해구로 이루어져 있으며 앞의 결과와 합하여 보면 제주도 부근 해구는 모두 어획량이 풍부한 해구로 나타났다. 동해안과 서해안은 주로 어획량에 비해 조업 노력량이 많은 것으로 판단된다. 다음 그림의 표시 중 흰 글씨/검은바탕은 조업일수와 조업척수가 상위인 조업 노력량이 많은 해구이며 검은 글씨/흰 바탕은 어획량과 조업척수가 상위인 해구이다.

표 4.7 대해구 군집결과 할당

어획량		조업일수		조업척수	
대해구	군집결과	대해구	군집결과	대해구	군집결과
HG87	대형	*HG48*	대형	*HG48*	대형
HG94	대형	*HG55*	대형	*HG55*	대형
HG97	대형	*HG62*	대형	*HG62*	대형
HG98	대형	*HG63*	대형	*HG63*	대형
HG99	대형	*HG66*	대형	*HG66*	대형
HG104	대형	*HG69*	대형	*HG69*	대형
HG106	대형	*HG70*	대형	*HG70*	대형
HG110	대형	*HG76*	대형	*HG76*	대형
HG162	대형	**HG87**	대형	HG82	대형
HG163	대형	*HG92*	대형	HG83	대형
HG172	대형	*HG93*	대형	**HG87**	대형
HG173	대형	**HG97**	대형	HG88	대형
HG174	대형	*HG98*	대형	*HG92*	대형
HG182	대형	**HG99**	대형	*HG93*	대형
HG220	대형	**HG104**	대형	HG94	대형
HG221	대형	*HG105*	대형	**HG97**	대형
HG222	대형	**HG110**	대형	*HG98*	대형
HG223	대형	*HG154*	대형	**HG99**	대형
HG224	대형	**HG163**	대형	**HG104**	대형
HG230	대형	**HG174**	대형	*HG105*	대형
HG231	대형	*HG184*	대형	**HG110**	대형
HG232	대형	*HG193*	대형	*HG154*	대형
HG233	대형	*HG194*	대형	**HG163**	대형
HG234	대형	*HG203*	대형	HG164	대형
HG241	대형	*HG214*	대형	HG173	대형
HG242	대형	**HG222**	대형	**HG174**	대형
HG243	대형	**HG231**	대형	*HG184*	대형
HG557	대형	**HG234**	대형	*HG193*	대형
HG47	일반	**HG243**	대형	*HG194*	대형
HG48	일반	HG47	일반	*HG203*	대형
HG49	일반	HG49	일반	HG210	대형
HG50	일반	HG50	일반	HG211	대형
HG55	일반	HG56	일반	HG212	대형
HG56	일반	HG57	일반	HG213	대형
HG57	일반	HG58	일반	*HG214*	대형
HG58	일반	HG59	일반	HG220	대형
HG59	일반	HG60	일반	**HG222**	대형
HG60	일반	HG61	일반	HG223	대형
HG61	일반	HG64	일반	HG224	대형
HG62	일반	HG65	일반	**HG231**	대형
HG63	일반	HG67	일반	HG232	대형
HG64	일반	HG68	일반	HG233	대형
HG65	일반	HG71	일반	**HG234**	대형
HG66	일반	HG72	일반	HG242	대형
HG67	일반	HG73	일반	**HG243**	대형
HG68	일반	HG74	일반	HG253	대형

5. 결론 및 보언

본 연구는 수협 조업정보에 대한 유용성 분석과 위치별 조업패턴의 특성을 분류하여 해구별 유사성 도출을 목적으로 수행되었다. 데이터의 유용성 분석 결과, 농수부 어업생산통계대비 약 33%정보의 수준

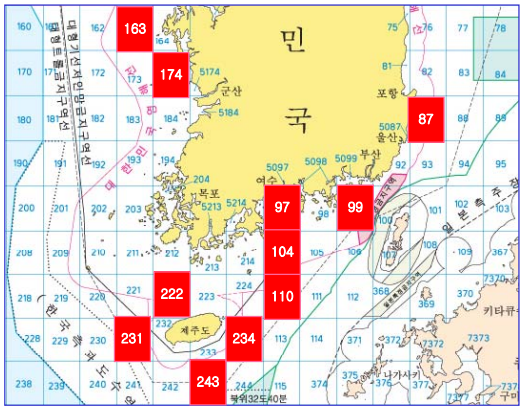

그림 4.2 어획량/조업일수/조업척수 상위 대해구

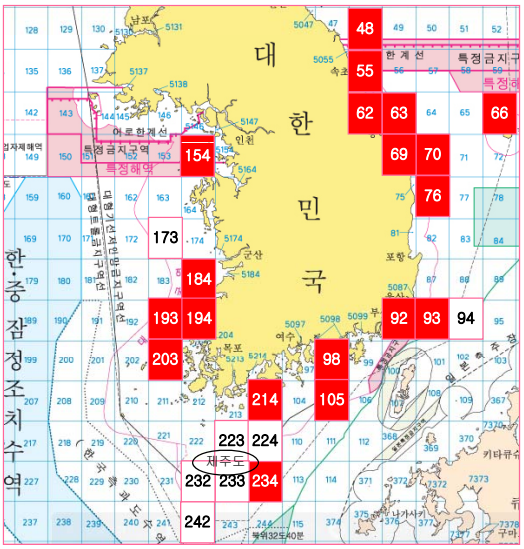

그림 4.3 어획량/조업일수/조업척수중 2개 상위 대해구

이었으나, 어획량 패턴 및 상관분석결과 거의 유사한 패턴을 지니고 있어 실질적 정보의 활용이 가능할 것으로 판단된다. 어업별 또는 어종별 세부 비교를 통해 유용성 퍼센트를 가중치로 활용한다면 다른 통계가 제시하지 못하는 위치별 어획량 추정이 일정 수준에서 가능할 것으로 판단된다. 본 연구에서는 수협 조업정보의 가치를 일정부분 확인하였으나 향후 더 다양한 방법의 유용성 비교가 필요할 것으로 판단된다.

위치별 조업패턴의 유사성분석을 위한 대해구별 조업정보의 군집분석은 어획량, 조업일수, 조업척수에 대해 각각 2개의 군집으로 특성을 분류하여 제시하였다. 이를 통해 어획량, 조업일수, 조업척수 모두 대형인 즉, 가장 많은 수산자원을 보유하고 있으며 이에 대한 어업 노력량이 많은 11개 해구를 도출하여

제시하였다. 향후, 대해구별 유사성 분석이 아니 소해구 또는 더 세밀한 해구로의 유사성 분석을 한다면 보다 더 세밀한 위치별 특성분류가 가능해져 실질적이고 유용한 연구가 될 수 있을 것으로 예상된다. 본 연구에서는 향후 이러한 연구를 위한 기초연구로써의 가치를 제시할 수 있을 것으로 판단된다.

본 연구는 그동안 축적된 수협 조업정보의 유용성에 대한 검증을 수행하여 유용한 활용이 가능할 수 있도록 제시하였다는 의의가 있다. 수협 조업정보는 국내 수산관련 통계에서 제시하고 있지 못한 위치별 조업정보의 파악이 가능하기 때문에 세부적 데이터 유용성분석 결과를 활용하여 위치별 어획량 등을 추정한다면 해당 지역의 어업피해보상, 자원가치 산출 등을 추정할 수 있는 근거 자료로 활용이 가능할 것으로 판단된다. 본 연구에서는 위치별 조업패턴을 바탕으로 대해구별 군집분석을 통해 대해구를 유사한 특성으로 분류한 국내 최초의 연구라는데 의의가 있다고 판단된다.

향후 수협 조업정보의 어획량에 대한 신뢰성을 더욱 확보할 수 있다면 수산업뿐만 아니라 국가적 차원에서 가치가 높은 정보이기 때문에 이에 대한 방안모색의 연구가 필요할 것으로 예상된다.

참고문헌

Akaike, H. (1976). *Canonical correlation analysis of time series and the use of an information criterion, in system fieldentification: Advanced and case studies*, R. Mehra and D. G. Laniotis, eds., Academic Press, New York and London.

Berry, M. J. A. and Gordon, L. (1997). *Data mining techniques*, John Wiley & Sons Inc., New York.

Champman, P., Clinton, J., Khabaz, T., Reinartz, T. and Wirth, R. (1999). *The CRISP-DM process modeling*, CRISP-DM Consortium.

Chiu, T., Fang, D., Chen, J., Wang, Y. and Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263-268.

Cho, Y. J. and Hur, J. (2006). A study on improving the predict accuracy rate of hybrid model technique using error pattern modeling: Using logistic regression and discriminant analysis. *Journal of the Korean Data & Information Science Society*, **17**, 269- 278.

Cho, Y. J. and Ko, S. G. (2008). Segmentation of cooperatives' mutuality bank for effective risk management using factor analysis and cluster analysis. *Journal of the Korean Data & Information Science Society*, **19**, 831- 844.

Hand, D. J. (1981). *Discrimination and classification*, John Wiley & Sons, New York.

Kohonen, T. (1982). Self-organized formation of topologically collect feature maps. *Biological Cybernetics*, **43**, 59-69.

Kovesi, B., Boucher, J. M. and Saoudi, S. C. (2001). Stochastic K-means algorithm for vector quantization. *Pattern Recognition Letters*, **22**, 603-610.

Selim, S. Z. and Ismail, M. A. (1984). K-means type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 81-87.

Schwarz, G. W. (1978). Estimating the dimension of model. *The Annals of Statistics*, **6**, 462-464.

Zhang, T., Ramakrishnan, R. and Livny M. (1996). BIRCH: An efficient data clustering method for very large databases. *Proceedings of the ACM SIGMOD Conference on Management of Data*, 103-114, Montreal, Canada.

Journal of the Korean Data & Information Science Society
2009, **20**(3), 551–562

한국데이터정보과학회지

The similarities analysis of location fishing information through 2 step clustering

Yong Jun Cho[1]

National Federation of Fisheries Cooperatives, Fisheries Economic Institute

Received 17 February 2009, revised 20 May 2009, accepted 23 May 2009

Abstract

In this paper, I would present a using method for The Fishing Operation Information(FOI) of National Federation of Fisheries Cooperatives(NFFC) through the availabilities analysis and put out the similarities by the section of the sea through classifying characteristics of fishing patterns by their locations. As a result, although the catch of FOI is nothing more than 33% level to National Fishery Production Statistics(NFPS), FOI data is useful in understanding the patterns of fishing operation by the location because both patterns and correlation were very similar in the usability analysis, comparing the FOI data with NFPS. So I classified optimal clusters for catch, the number of fishing days and the number of fishing vessels through 2 step cluster analysis by the big marine zone and divided fishing patterns.

Keywords: 2 step cluster analysis, classification, fishing information, marine zone, similarity analysis.

[1] Research Fellow Ph.D, Fisheries Economic Institute, National Federation of Fisheries Cooperatives, Seoul 138-730, Korea. E-mail: ehclub@nate.com