

군집화 기법을 이용한 B2B Marketplace 상의 최적 파트너 검색 시스템

김신영, 김수영

포항공과대학교 산업공학과

An Optimized Partner Searching System for B2B Marketplace Applying Clustering Techniques

Shin-Young Kim, Soo-Young Kim

Division of Mechanical and Industrial Engineering

Abstract

With the expansion of e-commerce, E-marketplace has become one of the most discussed topics in recent years. Limited theoretical works, however, have been done to optimize the practical use of e-marketplace systems. Other potential issues aside, this research has focused on this problem: 'the participants waste too much time, effort and cost to find out their best partner in B2B marketplace.' To solve this problem, this paper proposes a system which provides the user-company with the automated and customized brokering service.

The system proposed in this paper assesses the weight on the priorities of a user-company, runs the two-stage clustering algorithm with self-organizing map and K-means clustering technique. Subsequently, the system shows the clustering result and user guide-line. This system enables B2B marketplace to have more efficiency on transaction with smaller pool of partners to be searched.

Keywords: B2B marketplace; Partner search; Clustering

1. 서론

B2B(Business to Business)란, 개인과 기업을 모두 포함하는 다수의 판매자와 구매자가 인터넷을 통해 서로 상품이나 용역, 정보 등을 판매하고 교환하는 상거래를 의미하는 것으로, 그 상거래가 일어나는 장소를 B2B 마켓 플레이스라고 정의한다[11]. 기존 상거래에서 기업 간의 정보 교환이 폐쇄적이고 제한적이며 속도가 느렸던 것에 반해, 온라인을 통한 전자 상거래는 기업들의 정보 교환 과정을 단축함으로써 그 흐름을 가속화시키고 있다. 실시간 정보 교환을 통해 재고 관리가 용이해

지고 이로써 기업은 전체 공급망 관리, 추가 수요에 대한 탄력적 대처 등이 가능하게 되고 이는 생산성 향상과 거래 매출 증대로 연결된다. 그리하여 많은 기업들의 활동 중심이 B2B 마켓 플레이스로 이동될 것으로 전망되었는데, IBM 글로벌 파이낸싱은 2000년도 연구에서 전 세계 B2B 전자 상거래의 규모가 2002년에 2조 달러에 달할 것으로 예상하였다. 반면 낙관적인 전망과는 달리 실제 B2B 상거래는 전세계적으로 기대만큼의 빠른 성장을 가져오지는 못하고 있으며 오히려 국내의를 막론한 B2B 마켓 플레이스들의 경영 부진과 기대 이하의 저조한 시장 참여 현상 등은 이미 심각한 수준에 이르러 있어, 한 때 e-Business의 총아로 각광 받던 B2B에 거품론이 대두될 정도로 투자 열기가 식어가고 있는 실정이다[12].

상거래 부진의 여러 가지 원인 중, 본 연구에서는 B2B 시장에서 거래 파트너를 탐색하는 데의 비효율성에 초점을 맞추고자 한다. B2B 거래에서 이익을 얻기 위해서는, 주어진 정보를 활용하여 거래를 위해 마켓 플레이스에 모인 수많은 업체들 중 자사의 기호를 최대한 만족하는 상대 기업을 선별해 내는 작업이 우선적이다. 그러나 마켓 플레이스에 모인 모든 업체들을 검색하고 비교하는 과정은 비용이나 시간, 인력, 기술 등의 제약을 받을 뿐 아니라 자사의 우선 갖대와 무관한 기업을 탐색하는 데에 많은 시간과 비용을 소모하게 된다. 이러한 탐색 과정에서의 비효율성을 최소화하는 방법의 고안이 본 연구의 궁극적 목표이다. 제안된 모델은 각 기업이 우선적으로 관심을 두고자 하는 속성에 근거하여 전체 기업을 군집화(Clustering)하고 각 기업에게 군집의 결과에 대한 분석 자료를 서비스하는 것이다. 참여 기업들은 결과 분석 자료에 근거해 자사의 요

구를 만족시킬만한 소그룹을 선정함으로써 탐색 과정의 효율성에 기여하는 것이다.

우선 2장에서는 본 연구에 사용될 군집화 기법과 관련된 기존 연구 내용을 서술하겠다.

2. 군집화 기법

군집화 기법은 사전 분류 체계 없이 데이터의 유사성에 근거하여 상호 유사한 집단으로 묶으면서 묶여진 각각의 군집들을 서로 분리 시켜나가는 과정이라고 정의할 수 있다[4]. 이 기법은 수많은 데이터를 그들이 지닌 속성의 유사성에 기반하여 분리함으로써 데이터 구조를 간단히 나타낼 수 있고, 정보를 효과적으로 검색할 수 있다는 장점을 가지고 있어서 복잡한 시장 분석을 위한 도구로써 많이 사용되고 있다. 현재까지 다양한 군집화 기법들이 제시되어 왔는데 그 중 시장 분석의 연구에 적절하다고 평가받고 있는 K-평균법과 자기 조직화 신경망(Self-Organizing Map, SOM)에 대해 이 장에서 살펴보도록 하겠다. 아울러 군집의 유효성 정도를 파악할 수 있는 Davies-Bouldin Index(DBI)에 대해서도 알아보도록 하겠다.

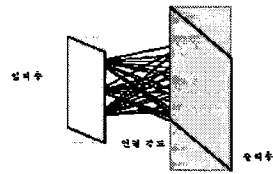
2.1 K-평균법(K-Means Clustering Technique)

K-평균법은 사전에 정해진 개수의 군집으로 주어진 데이터를 최적에 가깝게 군집화하는 것이다. 군집의 최종 개수가 k 로 정해지고 k 개의 군집 초기 중심치가 정해지면, k 개의 군집을 유지하면서 데이터들이 최적의 군집 유효성을 가지도록 군집화를 실행한다. 모든 데이터들에 대한 일차 군집화가 끝나면 새로운 군집의 중심이 계산될 것이고 그 중심 값을 가지고 다시 새로운 군집을 계산해 낸다. 이 과정의 반복을 통해 결과가 어느 정도 수렴하게 되면 그 때의 군집을 최종 군집으로 본다[2].

K-평균법은 군집수 k 를 정하는 것과, 초기치를 정하는 방법, 그리고 데이터 간의 유사성을 정하는 척도 등에 따라 그 결과가 달라지기 때문에 이러한 부분에 대한 연구가 많이 진행되어 왔다. 특히 군집수 k 를 정하는 일은 군집의 유효성과 직결된 문제이기 때문에 K-평균법에서 가장 중요한 요소라 여겨질 수 있으며 본 연구에서는 현재까지 가장 많이 이용되고 있는 DBI 지표를 이용하여 군집의 유효성을 측정하도록 한다.

2.2 자기 조직화 신경망(Self-Organizing Map, SOM)

SOM은 비교사 학습(Unsupervised learning)을 사용한 신경망의 일종으로 경쟁 학습 모형을 이용하고 있다[5]. 기대되는 출력 타입 없이 입력 데이터만 주어지는 비교사 학습에서는 유사한 입력 값이 들어올 경우, 그들을 같은 군집으로 묶기 위해 스스로 연결 강도를 조절해 나간다. 즉, 입력된 데이터 집합의 통계적 특성을 추출해서 유사한 데이터를 같은 군집으로 묶어 가는 학습 알고리즘이 진행되며, 이는 실제 데이터를 사용한 군집화, 패턴 인식, 낮은 차원에서의 image 추출 등의 많은 분야에서 효율성이 입증되어 널리 사용되고 있다[9].



[그림 1] SOM의 구성

분야에서 효율성이 입증되어 널리 사용되고 있다[9].

SOM은 [그림 1]과 같이 입력층과 출력층의 두 개의 층으로 구성되어 있다. 관측된

데이터 개수만큼의 뉴런을 가지고 있는 입력층은 각각의 뉴런을 통해 실제 데이터로부터 정보를 받아들이는 역할을 하며, 입력층에 존재하는 뉴런들은 출력층에 있는 뉴런 전체와 연결이 되어 있고 각각의 연결선은 연결 강도(weight)를 나타낸다. 학습을 통해 임의로 선정되어 있던 연결 강도가 조정되어 가는 데 그 과정은 다음과 같다.

입력된 데이터들이 출력층에 투영되면서, 출력층에 존재하는 뉴런들은 정해진 학습 알고리즘의 규칙인 유사성의 정도를 두고 서로 경쟁을 한다. 이 때 유사성의 척도로는 유클리디안 거리가 사용되며 그 결과로 승자 뉴런이 결정되면, 결정된 승자 뉴런과 연결된 연결 강도들은 입력 데이터에 대응되도록 조절된다. 이 과정의 반복을 통해 입력된 데이터들은 상대적으로 큰 유사성을 가지는 것들끼리 스스로 모이게 되고 그 결과 출력층은 특징적인 형상을 나타내게 된다. Kohonen은 이 과정에서 모든 뉴런들의 연결 강도를 조정하면 이전 신경망의 개념과는 달리, 정해진 승자 뉴런만이 유일하게 출력 신호를 보낼 수 있다는 ‘승자 독점(winner takes all)’의 원리를 제안하여 학습의 효율성을 높였다. 또한 이웃(neighborhood) 개념을 추가 사용하여 승자 뉴런의 주변 뉴런들이 지닌 연결 강도도 승자 뉴런의 연결 강도와 함께 조정되도록 하여 입력 패턴의 유사성을 더욱 잘 반영하는 특성 지도를 형성할 수 있도록 하였다[10].

2.3 Davies-Bouldin Index (DBI)

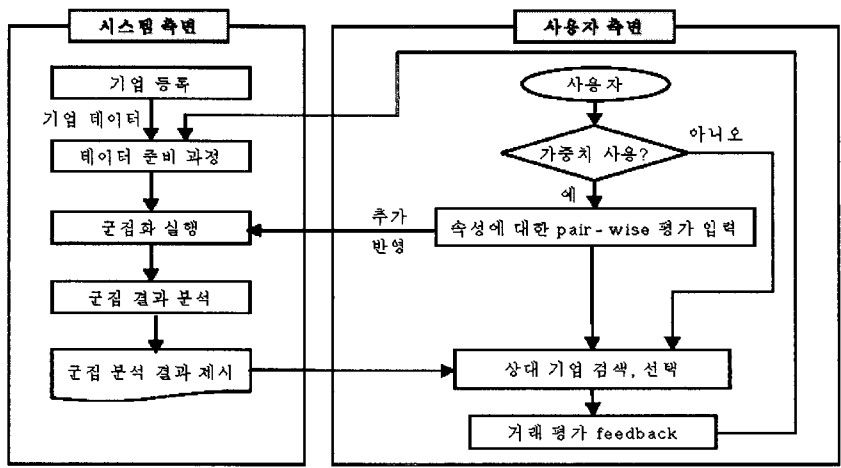
실제 데이터들을 몇 개의 군집으로 나누어 분석해야 가장 유효한 결과를 얻을 수 있을 것인가에 관한 문제는 오래 동안 연구되어온 매우 어려운 문제이다. 1985년, Hubert and Arabie의 연구를 비롯해서, 군집화의 유효성을 측정하기 위한 방안에 관한 많은 연구가 있어 왔는데, 그 문제는 다른 군집들과의 분리 정도는 크고, 같은 군집 내에서의 데이터들 사이에는 유사성을 많이 가져 서로 총총하게 모여 있어야 한다는 공통의 이론과 그에 따른 목적식을 가지고 있다[3]. 이 장에서는 현재 까지 가장 많이 이용되고 있는 Davies-Bouldin index(DBI)에 대해 알아보도록 한다.

DBI는 군집 내에서의 분산 정도 대비 각기 다른 군집 간의 분리 정도의 비율로 계산되는 함수 값으로, 존재하는 모든 두 개의 군집 쌍에 대해 각 군집의 크기 합을 그들의 중심 간 거리로 나눈 값으로써 표현된다. 이때, 각 군집의 크기는 해당 군집 중심과 각 데이터 값들과의 거리 차의 평균으로 계산할 수 있으며, 거리 차의 측정치에 사용되는 척도는 유클리디안 거리이다. 때문에 이 방법은 계산이 빠르고 쉬우며 일관성 있는 값을 제시해 준다는 장점을 가지고 있다[6]. 군집의 결과가 유효할수록 같은 군집내의 분산은 작아져 군집의 크기는 작아지고, 서로 다른 군집들 간의 거리는 커지므로 DBI 지표가 작은 값을 가질수록 유효한 결과를 갖는다고 할 수 있다[1].

3. 파트너의 효율적 검색을 위한 제안 시스템

본 연구의 목적은 B2B 전자 상거래의 첫 단계인 업체 선정을 위한 탐색 과정의 효율성을 높이는 데에 있으므로, 이를 위해 제안된 시스템은 군집화 기법을 이용해 마켓 플레이스에 참가한 전체 기업을 의미있게 구분하고, 그 결과의 분석 자료를 사용자에게 제공한다. 이때, 각 사용자마다 상대의 선정 과정에서 중요하게 여기는 요소가 다르다는 점을 반영하기 위해 각 속성에 사용자마다 서로 다른 가중치를 부여하여 개개의 사용자는 이 시스템으로부터 차별화된 서비스를 받을 수 있도록 한다. 이 시스템의 전체적인 프로세스는 다음의 [그림 2]와 같다.

[그림 2]의 순서도를 바탕으로 본 제안 시스템의 작동 과정을 살펴보면, 먼저 기업이 마켓 플레이스에 가입을 하면서 제공된 기업 데이터는 자동적으로 마켓 플레이스의 데이터 베이스에 저장이 된다. 저장된 데이터는 군집화의 준비 단계인 각 속성별 정규화 과정을 거치게 되고, 이 과정을 지난 후 사용자의 입력이 있을 때까지 대기 상태로 머무른다. 그 후 사용자가 이들 데이터 속성에 대해 각각 쌍대 평가(pair-wise evaluation)를 수행하여 이를 바탕으로 AHP(Analytic Hierarchy Process) 기법을 적용하여 시스템은 자동적으로 그 평가 결과를 적절한 가중치로 변환하여 군집화에 반영하게 된다. 이로써 시스템은 개별 기업에게 좀더 차별화된 군집 결과를 제시해 줄 수 있다. 군집화 기법으로는 제 2장에서 설명되었던 K-평균법과 SOM을 단계적으로 적용하도록 한다.



[그림 2] 제안 시스템의 전체 흐름도

군집 결과에 따른 분석을 토대로 사용자는 자신이 우선적으로 원하는 조건을 갖춘 소규모의 거래 가능 기업 군집을 소개 받게 되고, 그 소규모 집합 가운데 세밀한 조사를 통해 거래를 위한 상대 기업을 선택하는 최종 의사결정을 내리면 된다.

군집 결과는 마켓 플레이스 내에 존재하는 전체 기업의 속성별 평균값과 군집 내 기업의 속성별 평균값의 비로 분석이 되는데, 기업 속성 전체에 대해 평균 대비의 분석 가이드 라인을 제시하는 것은 오히려 기업 내 의사 결정 담당자의 단기 기억에 부담을 주어 올바른 방향 경로를 선택하려는 그들의 의사결정을 어렵게 만들 수도 있다[8]. 따라서 본 연구에서는 군집화에 사용된 속성들 가운데 군집화에 기여도가 큰 속성들을 먼저 가려내어 그들을 중심으로 분석 결과를 제시한다.

군집화 분석은 궁극적으로 각기 다른 군집들과는 이질성을 높여주고 같은 군집 내의 데이터 간에는 동질성을 높여주기 위한 과정이다. 여기서 동질성이 높다는 것은 서로 조밀하게 모여있다는 것으로 해석을 할 수 있고, 그것은 곧 군집화를 실행할 경우, 그 전과 비교했을 때 분산이 더 작아져야 한다는 결론을 내리게 한다. 따라서 각각의 속성들 별로, 군집화를 실행하기 전과 후의 분산을 비교했을 때, 군집화 실행 후의 데이터 분산이 많이 작아질수록 의미 있는 군집화라고 보고, 그들 가운데 분산이 줄어든 정도가 가장 큰 속성이 그 군집화에 가장 큰 영향을 미쳤다고 판단을 하는 발견적(Heuristic) 기법을 제안한다.

마지막으로 거래가 성사 되었을 경우, 각 기업은 상대 기업과의 거래에 대한 만족도를 평가하여 입력함으로써 시스템에 피드백을 제공하고, 이것은 추후에 본 시스템을 이용할 사용자들에게 실제 경험에 근거한 기

업에 대한 신뢰도 지수로써 군집화에 사용되게 된다.

다음 장에서는 수집된 실제 데이터를 이용한 사례 연구를 통해 제안 시스템의 효율성을 검증하도록 하겠다.

4. 사례 연구

4.1 데이터 수집

본 연구를 위해 사용된 총 87개의 기업 데이터는, 각종 기업 정보를 제공하는 기업 정보 전문 포털 사이트(<http://www.bizpark.co.kr>)로부터 수집된 실제 데이터이며, 그 외에 추가적으로 마켓 플레이스 내 기업을 세분화하기에 필요하다고 여겨지는 일부 속성의 데이터 값은 임의로 생성하였다. 또한 본 연구에 사용된 각 기업은 현재 상장 기업이거나 코스닥에 등록중인 기업으로 제한하였으며 업종에 따라 각 속성별 차이가 있음을 고려하여, 통신기기 및 망송장비 제조업체로 업종에 제한을 두었다.

연구에 사용된 기업 데이터의 속성에는 Y.Luo [7]의 연구를 토대로 모두 아홉 가지가 선택되었으며, 이 아홉 가지는 종업원수, 총 매출액, 기업 수명, 단기 순이익, 자기 자본 이익률(ROE), 재무 우량도, 기업의 거래 수준에 대한 사용자 평가 피드백, 그리고 제품의 품질과 가격이다. 각각의 속성에 대한 설명은 [표 1]에 나타내었다. 위 [표 1]에서 알 수 있듯이 기업의 거래 수준에 대한 사용자 평가 피드백 및 제품의 품질, 가격은 임의로 생성된 데이터이며, 나머지는 실제 데이터를 사용하였다. 다만 재무 우량도는 해당 기업의 재무 상태나 경영 성과를 종합적으로 평가할 수 있도록, 개별 분석은 어렵지만 판별력이 높은 지표들을 중심으로 전문가에

[표 1] 사례 연구에 사용된 기업의 속성 분류

분류	데이터	설명
기업 속성	종업원수	기업의 규모
	총 매출액	기업의 규모
	기업 수명	기업의 시장에서의 경험 정도
	단기 순이익	기업의 수익성
	자기 자본 이익률	기업의 수익성
	재무 우량도	해당 기업에 대한 시장에서의 재무적 신뢰 정도
	사용자 평가(*)	기업의 거래 계약 이행 수준
제품 속성	품질 수준(*)	생산 제품의 품질 수준
	가격 수준(*)	생산 제품의 시장 가격 수준

(*) 속성에 대해서는 임의로 생성된 데이터를 사용함

의해 구간과 배점을 모형화하고 그것으로 우량도를 평가한 수치를 사용하였다.

4.2 사례 연구 과정의 설계

앞에서도 밝힌 바와 같이 각 기업이 거래 파트너를 찾기 위해 탐색하는 과정은 각각의 선호도가 반영된 개별적인 과정이다. 때문에 본 사례 연구는 제안 시스템이 사용자의 우선순위를 제대로 반영하는가를 확인하기 위해 설계되었다. 가중치를 적용한 속성의 종류와 그 개수에 따라 크게 세 가지의 경우로 나누어지는데, <Case1>에서는 기업의 선호도를 반영하지 않은 본래의 데이터를 사용하여 다른 경우와 비교될 수 있도록 하였고, <Case2>에서는 임의로 선택한 두 가지 속성, 재무 우량도와 제품의 가격에 중요도를 부여한 데이터를 사용하였으며, <Case3>에서는 세 가지 속성, 총 매출액, 재무 우량도 및 제품의 가격에 각각 중요도를 부여한 데이터를 사용하여 군집화를 실행하였다. 각 경우에 선택된 속성들은 <Case1>의 군집화에서 비교적 낮은 중요도를 점한 것으로, 두 경우를 비교함으로써 사용자의 우선순위가 반영이 되는지를 확인할 수 있다.

각 사용자 기업의 선호도 평가 내역을 AHP 방식에 의해 적합한 가중치로 바꾸는 작업에는 Expert Choice2000 소프트웨어가 사용되었으며, 데이터의 군집화는 Matlab6.1을 기반으로 구현한 프로그램을 사용

하였다. 이어지는 장에서는 사례 연구의 각 단계별 결과에 대해 기술하도록 하겠다.

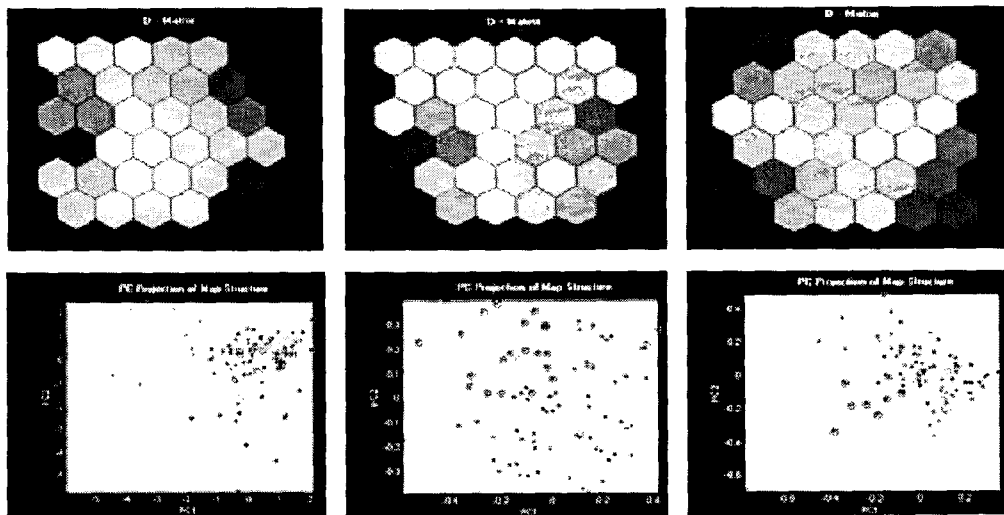
4.3 실험 결과

4.3.1 사용자의 선호도 반영

본 단계에서는 사용자가 입력한 쌍대평가로부터 각 기업의 선호도를 반영하는 가중치를 AHP기법을 적용하여 구해보았다. 구해진 각 경우별 속성에 대한 가중치는 다음의 [표 2]에 정리하였으며, 이 과정의 평가 기준은 Y.Luo의 연구 내용을 참고하여 임의로 선정되었다. 본 사례 연구에서는 <Case2>와 <Case3>에서 선택된 각각의 속성들에 대해 나머지 속성들 가중치의 두 배가 부가되었다.

[표 2] 경우에 따른 각 속성별 가중치

	<Case1>	<Case2>	<Case3>
종업원수	0.111	0.067	0.060
총 매출액	0.111	0.079	0.172
기업 수명	0.111	0.077	0.070
당기 순이익	0.111	0.094	0.085
ROE	0.111	0.086	0.077
재무 우량도	0.111	0.196	0.175
사용자 평가	0.111	0.094	0.085
품질 수준	0.111	0.090	0.081
가격 수준	0.111	0.216	0.194



[그림 3] 각 경우별 데이터의 군집화 결과, 좌측으로부터 상하방향으로 <Case1>의 SOM map과 K-평균법 결과, <Case2>의 SOM map과 K-평균법 결과, 그리고 <Case3>의 SOM map과 K-평균법 결과를 각각 나타낸다.

4.3.2 데이터의 군집화

본 단계에서는 정규화된 데이터에 계산된 가중치를 반영하고 군집화를 실행시키는 작업을 한다. 각 경우별로 먼저 SOM을 이용해 이상치(outlier)의 제약으로부터 자유로울 수 있게 하고, 그 결과를 바탕으로 K-평균법을 단계적으로 적용하였다.

이어지는 [그림 3]은 각각의 경우에 알맞은 데이터의 군집화를 실행한 결과이다. 상단의 그림들은 각 경우별 데이터에 대해 6×6 map을 가지는 SOM을 적용시켜 구해진 결과이고, 하단의 그림들은 상단의 결과를 다시 K-평균법으로 군집화한 결과이다. 이 때 K-평균법을 진행하기 위한 k 는 DBI 값을 최소화하는 값으로 정함으로써, 군집화의 결과가 최적의 유효성을 갖도록 하였다. 각 경우별로 적절하다고 판단되는 k 는 <Case1>의 경우 6개, <Case2>의 경우 4개, 그리고 <Case3>의 경우 5개이다.

4.3.3 속성별 중요도 계산

다음은 군집화의 결과를 바탕으로 사용자에게 가이드 라인을 제시하기에 앞서 먼저 중요도가 높은 속성을 골라내는 과정이다. 앞서 설명한 것처럼 이 과정에서는 각 속성별 분산의 변화를 파악하기 위해 간단히 전체 제곱합의 개념을 이용한 발견적 방법을 제안하고 그것을 적용해 보았다. 제안 방법에 따라 구한 각 경우마다의 속성들 중요도 순위를 [표 3]에 정리하였다.

이 중요도를 본 사례 연구에 적용시키기 앞서, 먼저 제안 방법의 타당성에 대한 검증은 해야 하는데, 이는 <Case2>와 <Case3>에서의 속성별 순위 비교에 의해 가능하다. 본 사례 연구에서 <Case2>는 군집화 이전에 재무 우량도와 가격 수준을 나타내는 속성에 가중치를 부여한 데이터이고 <Case3>는 재무 우량도, 가격 수준, 그리고 총 매출액에 각기 가중치를 부여한 데이터이다.

따라서 수학적 이론에 의하면 이들이 다른 속성들 보다 군집화에 많은 영향을 미쳐야 하고, 그런 과정을 거쳐 군집화가 되었다면, 본 제안 방법에 따라 계산되었을 때 각기 중요도가 높은 속성으로 나와야 할 것이다. 그런데 제안 방법에서의 군집화 기여도 결과를 정리한 [표 3]을 참고하면, 이러한 가정이 모두 만족되고 있다는 것을 알 수가 있고, 따라서 본 제안 방법은 현실적으로 쉽고 빠르게 계산될 수 있으며, 동시에 타당성을 가진 방법이라 할 수 있겠다. 즉, 군집화 분석 결과를 위해서는 먼저 [표 3]을 참조해 중요도가 높은 속성들을 고르고, 그들을 중심으로 분석하여 사용자에게 가이드 라인을 제시하도록 할 것이다.

[표 3]에서 알 수 있듯이 가중치를 고려하지 않은 <Case1>의 경우에는, 사용자 평가, 종업원수, 재무 우량도, 품질 수준, ROE, 총 매출액, 당기 순이익, 기업 수명, 그리고 가격 수준의 순으로 각 속성들이 군집화에 기여를 한 것으로 보인다. 그리고 <Case2>의 경우는 재무 우량도, 사용자 평가, 가격 수준, ROE, 당기 순이익, 품질 수준, 종업원수, 총 매출액, 그리고 기업 수명의 순으로, <Case3>의 경우는 재무 우량도, 사용자 평가, 가격 수준, 총 매출액, ROE, 종업원수, 품질 수준, 당기 순이익, 그리고 기업 수명의 순으로 군집화 실행에 있어 높은 중요도를 점했다고 할 수 있다.

4.3.4 각 경우별 분석 결과 제시

이 단락에서는 각 속성별로 산정한 중요도를 바탕으로 군집화 결과를 분석해 보고자 한다. 이 때의 분석은 전체 평균대비 군집의 평균값으로 진행이 되며 중요도가 높은 순서대로 각 경우 4개의 속성에 관해서 가이드 라인을 제시할 것이다.

이 때, 본 연구에서 제안한 시스템을 이용할 경우, 파트너 검색 과정에 있어 얼마만큼의 효율성이 증가될 수

[표 3] 각 경우별 속성들의 군집화 기여도

중요도	<Case1>	<Case2>	<Case3>
1	사용자 평가	재무 우량도	재무 우량도
2	종업원수	사용자 평가	사용자 평가
3	재무 우량도	가격 수준	가격 수준
4	품질 수준	ROE	총 매출액
5	ROE	당기 순이익	ROE
6	총 매출액	품질 수준	종업원수
7	당기 순이익	종업원수	품질 수준
8	기업 수명	총 매출액	당기 순이익
9	가격 수준	기업 수명	기업 수명

[표 4] 제안 시스템을 이용할 경우 검색 노력의 감소

	<Case1>	<Case2>	<Case3>
군집 1	19.54	28.74	25.29
군집 2	10.34	21.84	26.44
군집 3	16.09	27.59	6.9
군집 4	19.54	21.84	21.84
군집 5	11.49	-	19.54
군집 6	22.99	-	-

[표 5] 각 경우별 속성들의 군집화 기여도

중요도	<Case1>	<Case2>	<Case3>
1	사용자 평가	재무 우량도	재무 우량도
2	종업원수	사용자 평가	사용자 평가
3	재무 우량도	가격 수준	가격 수준
4	품질 수준	ROE	총 매출액
5	ROE	당기 순이익	ROE
6	총 매출액	품질 수준	종업원수
7	당기 순이익	종업원수	품질 수준
8	기업 수명	총 매출액	당기 순이익
9	가격 수준	기업 수명	기업 수명

있는가에 대한 검증을 함께 하도록 하는데, 효율성의 척도로는 같은 결과를 얻기 위해 검색해야 하는 상대자 집단의 기업 수를 사용하도록 한다. 아래의 [표 4]는 각각의 경우, 중요한 4가지 속성들을 바탕으로 평균에 근거한 분석 결과를 정리한 것이며, [표 5]는 사례 연구를 통해 설정한 각 경우 별로, 파트너 선정 시 들어드는 검색 상대 기업 수의 비율을 나타낸 것이며, 제안 시스템을 이용하지 않은 전체 탐색의 경우를 100으로 설정하였다.

따라서 <Case1>에서 제공하는 가이드 라인에 따르면, B2B 마켓 플레이스에 참여한 기업 중 기업의 규모와는 상관없이 기존 거래 내역이 충실하고 신뢰할 만하며 재무적으로도 안정적인 기업과 거래하고자 하는 기업은, 전체 기업을 탐색할 필요 없이 본 제안 시스템에 의해 추천되는 군집 1에 포함되는 기업들을 대상으로 입찰 공고를 하고 제안서를 받으면 된다. 이 경우, 전체 검색을 할 때에 비해 19.54%의 노력과 시간만으로 같은 결과를 얻을 수 있게 된다. 또한, 다른 어떤 속성보다 재무 우량도가 높고 경영진이 주주의 자본을 효율적으로 운용하고 있는 기업을 상대 기업으로 맞고자 한다면 이 사용자 기업에게는 군집 3이 추천될 것이고 그 기업은 이 속에서 가장 최적의 상대 기업을 선택하면 되는 것

이다. 이 경우에는 16.09%의 검색 노력만으로 전체 검색과 같은 결과를 얻게 된다. 게다가 <Case2>에서와 <Case3>에서는 추가적으로 사용자의 선호도가 반영된 맞춤형 가이드 라인을 제공할 수 있다는 부가적 장점이 있다. 이를 확인하기 위해 <Case2>에서는 <Case1>에서 군집화에 별다른 영향력을 미치지 못한 속성으로 판별된 가격 수준이라는 요소에 관심을 두는 기업이 있다고 가정, 그 부분의 중요도를 높여 보았다. 그 결과, <Case2>에서 가격 수준은 군집화에 미친 기여도 정도가 3순위로 올랐고, 그로 인해 윈 데이터로 군집화를 했던 <Case1>의 경우에 중요도가 낮아 분석이 안 되었던 부분에 대해서도 보완이 됨을 확인할 수 있었다. 또한 <Case3>에서는 총 매출액이라는 속성에 가중치를 추가적으로 부여했고, 그 결과 <Case2>에서와 마찬가지로의 맞춤형 가이드 라인을 제공할 수 있었다.

5. 결론 및 토의

향후 전자 상거래의 대표적 모습이라고 여겨지는 B2B 마켓 플레이스가 제 모습을 찾기 위해서는 그 속에서 많은 기업들이 모여 서로 활발히 거래를 주고 받아야 한다. 따라서 B2B 관련 연구는 시장의 활성화를 위해 마켓 플레이스가 제공해야 할 제반 서비스에 대해

고민하고 진행되어야 할 것이다. 이에 본 연구는 마켓 플레이스 참여 기업들의 거래 과정에 가장 기본이 되는, 서로 간의 탐색 과정에 효율성을 제고하고 그를 바탕으로 각종 비용을 감소시킬 수 있는 모델의 프레임워크를 제안하였다.

본 연구가 제안한 모델은, 먼저 각 기업을 종업원수, 총 매출액, 기업 수명, 당기 순 이익, ROE, 재무 우량도, 사용자 평가, 품질 수준, 그리고 가격 수준의 아홉 가지 속성들의 유사성에 의거해 군집화를 실행한다. 이 때, 각 기업이 갖고 있는 회사 특성상의 우선 고려 사항을 반영하게 되며, 그 결과를 바탕으로 마켓 플레이스 참여 기업에게 적정 수준의 가이드 라인을 제공해 준다. 그 가이드 라인을 제공 받은 기업은 분석 결과에 의거해 자사의 요구에 부합될 것 같은 군집을 선택하고 전체 기업에서 탐색 과정을 거치는 대신 그 선택된 군집 내에서만 탐색 과정을 거치면 되는 것이다.

기존의 전체 탐색 과정을 거쳐야 했던 마켓 플레이스에서, 구매자는 모든 제안 업체를 비교 분석하고 업체별 접점을 찾아내는 데에 많은 시간과 인력을 소비하였으며 제한된 시간과 정보 수집의 한계로 인해 합리적이고 효과적인 탐색을 하기 어렵다. 또한 판매자 측에도 항상 공정한 기회가 주어지는 것이 아니었기 때문에 불필요한 제안서 작업이 많았고 그와 관련된 영업 비용을 추가로 소비할 수 밖에 없었다. 그러나 이 제안 모델을 이용하면, 각 기업은 자사가 탐색해야 할 공간 중 불필요한 부분을 일차적으로 제거할 수 있게 되고 그로 인해 거래 비용의 많은 부분을 차지하고 있는 '탐색 비용'을 줄일 수 있을 뿐더러, 검색 대상이 되는 기업은 최적의 거래 상대자가 될 가능성이 상대적으로 큰 기업들이므로 그들 가운데 상대자를 선택했을 때 일어날 수 있는 파트너 선정의 오류는 확률적으로 줄어들게 된다.

이렇듯 본 연구는 B2B 마켓 플레이스를 활성화시키기 위한 방안을 담은 연구가 거의 전무한 상황에서 의의를 가질 수 있으며, 또한 추후 연구에 동기 부여가 될 것으로 보인다. 향후 필요한 연구 과제 중 하나는 군집화에 대한 것이다. 군집화 알고리즘 자체는 비교사 학습이기 때문에 알고리즘의 한계에 의해 잘못된 결과도 출되더라도 그것을 인지할 가능성이 희박하다. 이 때 제공되는 잘못된 가이드 라인에 따라 거래를 하는 경우 발생하는 손실은 B2B 마켓 플레이스에 치명적일 수도 있다. 따라서, 이러한 문제를 극복하기 위해 잘못된 결과를 사전에 인지하는 보완 장치가 가미되어야 하는데

이러한 방법론에 대한 연구가 추가적으로 진행되어야 할 것이다.

이 논문은 2002년도 두뇌한국 21사업에 의하여 지원되었음

참고 문헌

- [1] Davies, D.L., Bouldin, D.W., "A cluster separation measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(1979), pp.224-227
- [2] Jain, A.K., & Dubes, R.C., *Algorithms for clustering data*, Prentice-Hall, Englewood Cliffs, N.J.1988
- [3] Jain, A.K., & Moreau, J.V., "Bootstrap Technique in Cluster Analysis," *Pattern Recognition*, 20(1987), pp.547-568
- [4] Jiang, M.F., Tseng, S.S., Su, C.M., "Two-phase clustering process for outlier detection", *Pattern Recognition Letters*, 22(2001), pp.691-700
- [5] Kohonen, T. "The Self Organizing map," *In Proc. IEEE* Vol.78, No.9(1990), pp.1464-1480
- [6] Laine, S. "Finding the variables interest," *Minerals Engineering*, Vol.15, No.3(2002), pp.167-176
- [7] Luo, Y. "Joint Venture Success in China: How Should We Select a Good Partner?," *Journal of World Business*, Vol.33, No.2(1998), pp.145-166
- [8] Proctor, R. W., & Van Zandt, T., *Human factors in simple and complex system*, New York, Wiley, 1994
- [9] Serrano-Cinca, C., "Self organizing neural networks for financial diagnosis," *Decision Support Systems*, Vol.17, No.3(1996), pp.227-238
- [10] Vesanto, J., & Alhoniemi, E. "Clustering of the Self-Organizing Map," *IEEE Transactions on Neural networks*, vol.11, No.3(2000)
- [11] http://searchcio.techtarget.com/sDefinition/0,,sid19_gc i214411,00.html
- [12] <http://www.e-bizgroup.com/paper/paper.asp>