

Feature Impact Evaluation Based Pattern Classification System

Hyun-Sook Rhee*

Abstract

Pattern classification system is often an important component of intelligent systems. In this paper, we present a pattern classification system consisted of the feature selection module, knowledge base construction module and decision module. We introduce a feature impact evaluation selection method based on fuzzy cluster analysis considering computational approach and generalization capability of given data characteristics. A fuzzy neural network, OFUN-NET based on unsupervised learning data mining technique produces knowledge base for representative clusters. 240 blemish pattern images are prepared and applied to the proposed system. Experimental results show the feasibility of the proposed classification system as an automating defect inspection tool.

▶ Keyword: feature selection, cluster validity measurement, feature impact evaluation, unsupervised learning, pattern classification

I. Introduction

최근 컴퓨터 시스템은 우리 생활과 산업에 컴퓨터보조 진단 시스템이나 자동화시스템 등 지능형 시스템의 형태로 널리 활용되고 있다. 이러한 지능형 시스템에 공통적으로 포함되는 핵심기능은 주어진 응용분야의 데이터를 인식하고 판별하기 위하여 패턴들을 특정한 목적을 가지고 분류하는 패턴분류 작업이다. 패턴분류시스템은 일반적으로 주어진 응용분야의 데이터를 획득하는 전처리모듈, 특징선택모듈, 지식베이스구축모듈, 판단모듈 등의 4가지 주된 과정으로 이루어져 있다. 전처리과정은 원 데이터의 종류에 따라 처리과정이 다양하지만 보통 데이터를 분할하여 문제해결의 열쇠를 쥐고 있는 속성(특징)을 기계적으로 추출해 내는 과정으로 주어진 응용분야에 따라 여러 방법을 사용하고 있다. 이렇게 추출한 데이터를 가지고 패턴분류가 보다 효율적으로 이루어지도록, 패턴공간의 차원을 적절하게 감소시키는 역할을 하는 특징선택은 문제영역에서 관찰된 다차원 속성데이터 중에서 의사결정에 핵심적인 역할을 수행하는 속성데이터를 선택하는 작업이다[1,2]. 실제적으로 관찰가능한 모든 특징을 고려하면 분류과정이 복잡하거나 문제와 관련이 적은 특징의 개입으로 좋지 않은 결과를 나타내고 있다. 이와 같이 적합한 특징을 선택하여 실제 시스템에 적용하기 위하여

효과적인 응용데이터 집합을 구성하는 과정은 시스템의 성능향상에 중요한 구성요소로서 통계와 정보이론을 바탕으로 연구되어왔다. 또한 이와 같은 연구는 빅데이터 처리가 가능해지고 데이터 사이의 복잡도가 증가하면서 신경망, 퍼지이론 등을 이용한 방법이 적용되기도 하였다[3,4]. 그러므로 시스템의 목적이나, 응용 영역에 따라 다른 데이터의 특성을 반영하고 계속 생성되는 새로운 데이터에 대한 적응력을 향상시키는 연구가 필요하게 되었다. 이를 위하여 본 논문에서는 데이터가 가지는 특성을 바탕으로 새로운 데이터를 학습하고 일반화하는 개념을 활용하여 문제 영역에 따른 성능의 편차를 줄이기 위한 방법을 고안하였다. 이렇게 선택된 특징을 가지고 이미 알려진 해당분야의 데이터를 가지고 지식베이스를 구축한 후 판단모듈에서는 패턴 찾기 등의 분류를 기반으로 응용에 필요한 의사결정을 하게 된다. 이와 같은 과정에서 필요한 선택과 판단을 위하여 일반적으로 사용하는 방법은 이미 알려진 데이터를 학습데이터로 활용하여 적용하는 것이다. 이를 위하여 신경망, 유전자 알고리즘 등의 방법론을 이용한 매핑과 최적화 과정이 필요하며 이는 지능형시스템의 가장 중요한 구성요소인 것으로 알려져 있다. 최근 각광받고 있는 데이터마이닝의 연구도 이 매핑방법을 기

• First Author: Hyun-Sook Rhee, Corresponding Author: Hyun-Sook Rhee
*Hyun-Sook Rhee (hsrhee@dongyang.ac.kr), Dept. of Software Engineering, Dongyang Mirae University
• Received: 2018. 09. 12, Revised: 2018. 10. 30, Accepted: 2018. 11. 07.
• This work was supported by Dongyang Mirae University Research Grant.

준으로 최적화 하는 방향으로 학습방법을 정립하는 것이 핵심 기술이다. 신경망을 이용한 클러스터링 방법론은 비교사 학습(Unsupervised Learning) 방법으로 유사한 데이터끼리 그룹을 나누어 각 그룹의 대표 데이터를 추출하여 다른 데이터를 처리하는 방법이다. 물론 새로운 데이터가 생기면 대표 데이터도 이동하면서 최적화된 방향으로 학습하는 방법론이 연구되고 있다 [5]. 클러스터링 방법론은 최근 대두되는 빅 데이터 처리에 적합한 방법이지만 기존의 연구는 대부분 주어진 데이터 사이의 분류 경계가 명확하다는 가정에서 각 데이터를 하나의 클래스에 소속시키는 방법으로 진행되었다. 그러므로 경계가 불명확한 실제계의 모호한 데이터를 반영하기 위하여 지지벡터기반의 퍼지신경망 모델과 진단을 위하여 필요한 퍼지클러스터링 지능형 에이전트 모델이 연구되어왔다[4,5].

본 연구에서는 [6]의 유방 x선 영상진단시스템을 위한 특징선택 방법에 대한 후속 연구로서 이를 확장하여 일반적인 패턴분류시스템의 모델을 제안한다. 제안된 모델은 준비된 데이터의 정량적인 퍼지 데이터 분석을 기본으로 문제 영역에 따른 성능의 편차를 줄이는 특징영향평가(Feature Impact Evaluation) 방법으로 특징을 선택한다. 그 다음 과정으로 비교사학습기반 데이터 마이닝 기법을 바탕으로 하는 OFUN-NET을 활용하여 지식베이스를 구축하고, 마지막으로 그 결과를 판단모듈에 적용하는 패턴분류시스템을 제안한다. 또한 제안된 모델을 자동 홈페이지 검사 시스템에 활용하기 위하여 [7]의 연구에서 준비한 240개의 홈페이지 영상에 대한 11개의 특징을 추출하여 적용하고자한다. 이렇게 추출된 특징의 영향 평가를 통하여 여러 조합의 선택된 특징 데이터 집합을 가지고 분류정확도를 측정하여 본 시스템의 자동화 도구로서의 가능성과 제안된 방법의 타당성을 보이고자한다.

본 논문의 구성은 다음과 같다. II장에서는 관련연구로서 그동안 연구되어 온 특징평가방법들과 패턴분류방법을 요약하고, III장에서는 본 논문에서 제안하는 특징 영향평가 기반 패턴 분류시스템의 구성도를 보여주며 주요 모듈을 설명한다. IV장에서는 준비된 홈페이지 영상을 제안된 시스템에 적용하여 홈페이지 검사시스템에 적용할 수 있는 타당성을 보이고자한다. V장에서는 본 논문의 결론으로서 연구내용을 요약하고 향후 연구방향을 제시한다.

II. Related Works

본 장에서는 본 연구와의 관련연구로서 그동안 연구되어 온 기존의 특징평가방법들과 패턴분류방법을 요약하고 고찰함으로써 연구배경을 서술하고자한다.

1. Feature Evaluation Methods

특징선택은 전체적인 분류시스템의 성능에 중요한 의미를 지니는 부분으로 각 클래스의 특징들을 완벽하게 추출해 내는 경우, 분류과정은 매우 단순해지며 시스템의 성능에도 가장 중요한 요소

이다. 그러므로 추출된 특징들의 연관성이나 클래스와 특징간의 연관성을 평가하므로 분류에 오류를 적게 하고 기여도가 많은 특징을 선택하게 된다. 이와 같은 특징선택과정에는 이미 추출된 특징과 준비된 학습데이터를 가지고 효율적인 분류가 이루어지도록 확실적인 분석, 기계학습, 퍼지이론, 신경망 등의 방법론을 적용하여 클래스간의 각 특징에 대한 의존도, 클래스의 패턴들이 각 특징에 대하여 가지는 분포의 무질서를 측정하여 특징공간의 차원을 줄이는 과정이 포함된다. 기존의 특징평가는 특징공간에서 클래스들의 분포가 밀집해 있는 정도를 측정된 각 특징이나 특징사이의 관계에 대한 통계적인 분석, 엔트로피, 거리 등을 이용하여 적용하였다. 이와 같은 연구는 빅 데이터 처리가 가능해지고 널리 활용되면서 데이터 처리의 복잡도가 증가하면서 신경망, 퍼지이론 등을 적용하기도 하였다. 그러므로 시스템의 목적이나, 응용 영역에 따라 다른 데이터의 특성을 반영하고 계속 생성되는 새로운 데이터에 대한 적응력을 향상시키는 연구가 필요하게 되었다. 구체적으로 이러한 특징선택을 위한 방법으로 [8]에서는 각 클래스들이 가지는 특징공간의 분포를 하나의 퍼지 집합으로 정의하고, 그 공간에서 얼마나 무질서하게 퍼져있는가를 나타내는 불확실성 정도를 계산함으로써 각 특징의 기여도를 평가하였다. 즉 무질서가 클수록 불확실성 정도는 커지며 그 특징의 기여도는 작아지게 된다. 이 과정은 주어진 패턴들을 학습시키면서 클래스별로 이루어지며 동시에 각 특징들이 클래스 형성에 미치는 기여도를 자동적으로 계산한다. 또한 [9]에서는 통계적인 분류방법에서의 특징평가 방법을 강조하면서 클래스들 간의 정보의 중복(redundancy)이 없는 소수의 특징들을 이용하는 것이 바람직한 것으로 연구되었다. 우선 가장 유용한 특징을 선택하고, 이 특징들과 조합했을 때 오차가 적은 방향으로 추가되는 특징을 선택하면서 특징들 간의 정보 중복을 최소화 하는 전략을 취하였다. 이와 같이 기존의 방법에서는 불확실성의 정도나 중복 여부 등을 평가할 때, 휴리스틱 방법이 적용되었기 때문에 문제영역에 따라 성능의 편차가 큰 것으로 분석되었다. 본 논문에서는 데이터가 가지는 특성을 적용하면서 새로운 데이터를 학습하고 일반화하는 개념을 활용하여 문제 영역에 따른 성능의 편차를 줄이기 위하여, 준비한 데이터의 정량적인 퍼지 특징 데이터 분석을 기본으로 하는 특징영향평가 방법을 도입하고자한다. 이와 같은 특징 영향평가를 위하여 클러스터 분석 결과가 주어진 데이터의 구조를 반영하는 정도를 측정하는 기준이 필요하게 된다. 이러한 연구분야는 "cluster validity problem" 이라고 정의하고 기존의 클러스터 타당성 측정 함수를 고안하였다[10]. 이와 같은 클러스터 타당성 측정 함수들은 클러스터 분석 방법에 의하여 형성된 결과가 얼마나 의사결정에 유용하게 활용될 수 있는지를 체크하는 과정과 특징을 서열화하기 위한 기준으로도 적용될 수 있다[6, 11].

2. Pattern Classification Methods

패턴 분류는 공장자동화나 로봇틱스 분야에서 사용되는 지능형 시스템의 핵심기능으로 주어진 영상에 대한 특징을 찾아내고 그 값을 미리 구축해 놓은 지식베이스의 값과 비교하여 해당 클래스로 판별해 내는 방법이 주로 사용되고 있다. 자동화를 이루기 위해서는

주어지는 환경을 인식하고 판단해야하며 인식과 판단의 기본 기능은 주어진 패턴을 영역별로 분류하는 일이다. 이러한 패턴분류 작업은 정량적인 측정결과 추출한 특징을 가지고 각 데이터에 대한 특징벡터를 만들어 낸 후 이를 알려진 각 영역으로 매핑 하여 처리하는 것을 기본으로 한다. 이러한 매핑과정을 인간의 개입 없이 자동적으로 처리하기 위해서 필요한 분류 방법은 준비된 입력패턴의 클래스가 알려져 있는가의 여부에 따라 교사학습방법과 비교사학습방법으로 나누어지며, 학습과정에 휴리스틱 방법을 적용한 경우와 결정함수를 설정하고 최적의 해를 찾아가는 방법 등이 많이 적용되어왔다. 또 다른 측면에서 방법론적으로 공분산 빈도수와 같은 통계적인 모델을 분류에 이용하기도 하였고, 주어진 패턴을 구조적인 표현모델로 변환하여 발견된 기본규칙에 따라 분류하기도 하였다. 각 클래스를 다변수 정규분포와 같은 함수로 정의하여 공분산, 평균 등을 구하여 분류에 활용하는 매개변수 방법과 각 클래스를 특징지을 수 있는 매개변수를 찾을 수 없는 경우 데이터 자체가 가지는 다양한 속성을 활용하는 비매개변수 방법으로 나눌 수 있다. 또한 퍼지이론이 특징평가나 판단과정 등의 분류시스템 구성에 활용되기도 하였다[12].

패턴분류 시스템의 구체적인 사례로 [13]에서는 주어진 영상의 부분적인 조직을 특징짓는 작은 기본 단위들을 정사각형 형태의 화소로 정의하고 가운데 화소와 주변 화소와의 강도 비교로 기본 단위를 표현하고 이 값의 분포함수인 텍스처 스펙트럼을 활용하여 영상을 분류하였다. [14]에서는 일정한 배경위에 나타나는 복잡한 형태의 흠들을 판별하기 위하여 베이지안 분류기와 밀집도를 이용한 후 다음 단계에서는 흠의 면적을 이용하였다. 이 결과를 이용하여 마지막 단계에서는 흠의 조직 특성을 적용하여 분류하였다. 또한 [15]에서는 균일한 형태의 조직영상들에 대하여 서로 다른 해상도에서 각 특징들의 특징 빈도수 행렬을 구하였으며 빈도수 분포를 특징짓는 분포함수의 모멘트를 구하여 분류한 사례도 있다.

본 논문에서 사용한 클러스터링 알고리즘은 비교사 학습방법으로 준비된 데이터의 비매개변수 방법으로 각 데이터의 속성을 정의하고 각 클래스에 대한 사전지식 없이도 패턴분류가 가능한 퍼지 클러스터링 비교사학습 방법으로 결정함수를 가지고 최적화 과정을 포함하는 방법으로 접근하고 있다. 또한 제안한 방법을 자동 흠패턴 검사시스템에 활용하기 위하여 마련된 흠패턴 영상에 적용하는 과정을 보였다.

III. Feature Impact Evaluation Based Pattern Classification Model

패턴분류는 얼굴인식이나 음성인식과 같은 인공지능 사용자 인터페이스 구현의 기초적인 단계이며 또한 로봇비전시스템이나 제품의 생산라인에서 품질 검사 등과 같은 자동화 영역에 필수적인 구성요소이다. 자동화를 이루기 위해서는 연속적으로 주어지는 환경을 인식하고 판별해야하며 인식을 위해서는 다양

한 패턴들을 특정한 영역별로 분류하는 일이 반드시 필요하다.

본 논문에서는 문제 영역에 따른 성능의 편차를 줄이기 위하여, 준비된 데이터의 정량적인 퍼지 특징 데이터 분석을 기본으로 하는 특징영향평가(Feature Impact Evaluation) 방법을 도입하여 특징을 선택하고 비교사학습기반 데이터 마이닝 기법을 바탕으로 하는 OFUN-NET을 활용하여 지식베이스를 구축하고 그 결과를 판단모듈에 적용하는 패턴분류시스템을 [그림 1]과 같이 제안한다.

준비된 각 특징값의 클러스터링 결과를 가지고 각 특징데이터의 특정 클러스터에 대한 밀집도와 분리도를 측정하여 이를 기준으로 특징을 선택하는 특징영향평가 모듈을 제안한다. n 개의 준비된 데이터는 m 개의 각 특징 데이터에 대하여 퍼지 클러스터 분석(FCA)[5,6] 방법에 적용하여 만들어진 클러스터의 대푯값을 가지고 그 특징데이터가 클러스터 형성에 얼마나 영향을 주는지를 측정한다. 이 과정에서 클러스터 중심점과 소속함수값을 사용되며 이 측정 모형을 기준으로 주어진 특징데이터의 밀집도 측정자(compactness index) G 와 분리도 측정자(separation index) S 를 정의하였다[6]. 형성된 클러스터링의 결과는 같은 클러스터에 속하는 데이터 사이에는 밀집도가 강하고 다른 클러스터에 속하는 데이터 사이에는 분리도가 강한 것을 잘 분류된 것으로 판단할 수 있다. 즉 G 의 값이 작을수록, S 의 값이 클수록 각 특징이 분류에 적합한 속성임을 알 수 있다. 이를 이용하여 다음과 같은 정의를 바탕으로 각 특징데이터의 영향평가 측정자 $I(d)$ 를 밀집도 측정자 G 에 대한 분리도 측정자 S 의 비율로 정의하였다.

[정의 1 : 밀집도 측정자] n 개의 데이터가 c 개의 클러스터를 형성한 경우, 임의의 두 데이터, f_j 와 f_k , 사이의 거리의 평균으로서 다음의 식 (1)과 같이 정의한다. 이때 ω_1 은 임의의 두 데이터가 같은 클러스터에 속할 가능성 정도를 나타낸다

$$G = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^n (f_j - f_k)^2 \omega_1 \quad (1)$$

$$, \text{ where } \omega_1 = \min_i \{u_{ij}, u_{ik}\}$$

[정의 2 : 분리도 측정자] 서로 다른 클러스터에 속하는 임의의 두 데이터 사이의 거리의 평균으로서 다음의 식 (2)과 같이 정의한다. $u_{i,j}$ 는 f_j 가 가장 큰 소속정도를 가지는 클러스터를 i_1 이라고 하고 f_k 가 i_1 이 아닌 가장 큰 소속정도를 가지는 클러스터를 i_2 라고 하자. 이때 ω_2 은 임의의 두 데이터가 서로 다른 클러스터에 속할 가능성 정도를 나타낸다.

$$S = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n (f_j - f_k)^2 \omega_2 \quad (2)$$

$$, \text{ where } \omega_2 = \min \{u_{i_1j}, u_{i_2k}\}$$

[정의 3 : 영향평가 측정자] m 개의 속성을 가지는 n 개의 특징데이터에 대하여, 각 특징 $d = 1, \dots, m$ 의 밀집도 측정자 G 에 대한 분리도 측정자 S 의 비율, $I(d) = S/G$ 를 영향평가 측정자로 정의한다.

[그림 1]은 본 논문에서 제안하는 특징영향평가기반 패턴 분류시스템(PIEP)의 기본구성도로 실제계로부터 수집된 n 개의 데이터 집합으로부터 m 개의 퍼지 클러스터 분석기(FCA_1, \dots, FCA_m), 영향평가측정자 $I(d)$ 를 계산하는 FIE, 그리고 이를 통하여 특징을 선택하는 특징선택 모듈로 구성되어 있음을 보여 주고 있다.

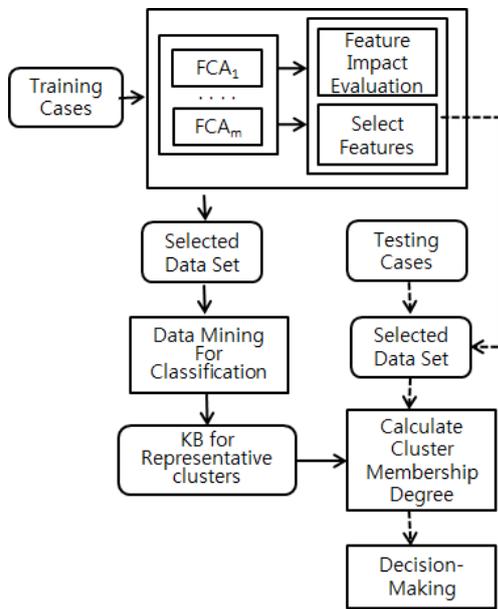


Fig. 1. Feature Impact Evaluation Based Pattern Classification Model

이렇게 정제된 특징 데이터 집합은 데이터 분류를 위하여 클러스터링 결과가 만들어내는 거리의 합을 계산해 주는 퍼지 소속 함수 식) 3의 J 를 최소로 하도록 학습을 진행하는 비교사 학습기반 데이터 마이닝 기법을 적용한 퍼지신경망 OFUN-NET을 사용한다[6]. OFUN-NET은 다음의 식 (3)과 (4)를 반복적으로 적용하여 클러스터링의 결과가 만들어 내는 오류 값을 요약하는 J 의 값이 최소가 되도록 최적화하는 방향으로 클러스터의 대푯값 (c_1, c_2, \dots, c_p)과 각 데이터의 클러스터 대표값과의 유사도를 측정하는 소속함수값 u_{ij} 를 학습해 간다.

$$J = \sum_{j=1}^n \sum_{i=1}^p (u_{ij})^2 \| \mathbf{f}_j - \mathbf{c}_i \|^2 \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^p \left\{ \frac{\| \mathbf{f}_j - \mathbf{c}_i \|^2}{\| \mathbf{f}_j - \mathbf{c}_k \|^2} \right\}^2} \quad (4)$$

이러한 학습 알고리즘은 기준 함수를 잘 설계하면 대부분의 문제가 가지는 최적화알고리즘에 적용가능하며 입력과 출력 사이의 관계를 기술하기 어려운 경우도 쉽게 처리하는 비교사 학습 신경망의 장점도 활용할수 있다. OFUN-NET의 학습 결과는 지식베이스에 저장되어 적용하고자하는 입력 데이터에 대한 각 클러스터의 소속 함수 값을 계산하여 패턴분류에 적용할 수 있다.

IV. Experiment Results

본 장에서는 III장에서 설계한 특징영향평가기반 패턴분류시스템을 자동 홈페이지 검사시스템에 활용하기 위하여 240개의 홈페이지 영상을 준비하여 11개의 특징을 추출하여 적용한다. 11개의 추출된 특징의 영향 평가를 통하여 선택된 특징데이터셋을 가지고 분류정확도를 측정하여 본 시스템의 자동화 도구로서의 가능성과 제안된 방법의 타당성을 보이고자한다.

1. Preparation of Experimental Data

제안된 시스템에 적용하기 위하여 8개 클래스의 각각 30장, 총 240장의 홈페이지 영상을 256×256 픽셀로 준비한다. 준비된 각 영상으로부터 [표 1]과 같은 11개의 특징을 추출하여 240×11 특징벡터를 획득한다. 이와 같은 전처리과정은 본 연구의 대상은 아니지만 실험을 위하여 [7]의 기존연구에서 사용한 특징을 활용하였다.

Table 1. Feature information

Features	Feature Descriptions
F_0	가로 방향으로 화소 값이 1인 연속된 화소(PE)들의 묶음 중 가장 긴 길이
F_1	세로 방향으로 화소 값이 1인 연속된 화소들의 묶음 중 가장 긴 길이
F_2	왼쪽 위에서 오른쪽 아래로 기울어진 대각선 방향으로 연속된 화소들의 묶음 중 가장 긴 길이
F_3	오른쪽 위에서 왼쪽 아래로 기울어진 대각선 방향으로 연속된 화소들의 묶음 중 가장 긴 길이
F_4	가로 방향으로 연결된 세 개의 화소 묶음들의 개수 백분율
F_5	세로 방향으로 연결된 세 개의 화소 묶음들의 개수 백분율
F_6	왼쪽 위에서 오른쪽 아래로 기울어진 대각선 방향으로 연결된 세 개의 화소 묶음들의 개수 백분율
F_7	오른쪽 위에서 왼쪽 아래로 기울어진 대각선 방향으로 연결된 세 개의 화소 묶음들의 개수 백분율
F_8	값이 1인 화소들이 차지하는 백분율
F_9	영상을 피라미드 형태로 축소했을 때 값이 1인 화소들이 줄어드는 비율
F_{10}	영상을 64x64의 16개의 정사각형으로 나눈 각 영역의 1인 화소 수들의 편차

2. Experiment and Analysis

이렇게 준비된 데이터는 11개 각 특징에 대하여 퍼지 데이터 분석기에 적용한 후 영향평가 측정자를 통해 $I(d) = S/G$ 의

값을 바탕으로 클래스 분류에 영향을 많이 끼치는 특징을 선택한다. 각 특징의 $I(d) = S/G$ 값을 가지고 서열화 한 결과는 [표 2]와 같다. 선택된 특징을 적용하여 클래스당 25장, 즉 200장의 정제된 학습 데이터셋은 OFUN-NET을 통하여 학습이 이루어지고 지식베이스를 구축하게 된다. 또한 시스템의 테스트를 위하여 학습에 사용된 각 클래스 당 5장의 데이터와 학습에 사용되지 않은 5장의 데이터 즉 모두 80장의 테스트 데이터셋 TDS1과 학습에 사용된 각 클래스 당 15장의 데이터와 학습에 사용되지 않은 5장의 데이터 즉 모두 160장의 테스트 데이터셋 TDS2에 대하여 분류정확도를 측정한 결과는 [표3]과 같다. 본 시스템에서 제안한 데이터 기반의 특징선택방법의 타당성을 확인하기 위하여 선택된 특징의 수를 상위 3, 4, 5, 6, 7, 8, 9개에 대하여 실험하였다. 실험결과는 학습에 참여한 데이터를 더 많이 가지고 있는 테스트 데이터 셋 TDS2의 분류정확도가 TDS1보다 높은 실험결과를 보여주고 있다. 이는 본 연구에서 제안한 특징영향평가방법과 시스템의 기본 학습능력이 타당함을 보여주고 있다. 하지만 특징의 수와 분류 정확도가 비례하는 것은 아니라는 것을 알 수 있다. 이 실험의 결과 주어진 데이터 셋에 대한 특징의 수는 [표2]에 제시된 상위 6개의 특징을 선택한 경우 최적인 것을 확인할 수 있다. 이와 같은 실험결과는 제안된 방법이 흠패턴 분류에 적용 가능하다는 타당성과 자동화 검사 시스템 도구로서의 응용 가능성을 확인하였다.

Table 2. Features Rank Based on Feature Impact Evaluation

rank	features
1	F ₇
2	F ₄
3	F ₈
4	F ₃
5	F ₆
6	F ₅
7	F ₉
8	F ₁₀
9	F ₁
10	F ₀
11	F ₂

Table 3. Comparison of Classification Accuracy

	TDS1	TDS2
Top3 selected	86.1	87.8
Top4 selected	88.7	89.2
Top5 selected	89.2	90.8
Top6 selected	90.6	91.4
Top7 selected	90.2	91.2
Top8 selected	89.6	90.3
Top9 selected	88.9	89.7

V. Conclusion

컴퓨터 기술의 발전은 계속적으로 변화하는 환경에서 데이터를 인식하고 적절한 조치를 취하는 다양한 응용분야의 지능

형시스템 분야에 적용되고 있다. 이러한 시스템의 핵심기능은 주어진 응용분야의 패턴들을 특정한 목적을 가지고 분류하는 작업이다. 이와 같은 패턴분류시스템은 일반적으로 주어진 응용분야의 데이터를 획득하는 전처리모듈, 특징선택모듈, 지식베이스 구축모듈, 판단모듈 등의 4가지 주된 과정으로 이루어져 있다. 점차 처리할 데이터의 양이 많아짐에 따라 분류에 유용한 데이터를 선택하는 방법과 계속 변화하는 새로운 데이터에 대한 학습능력이 중요한 연구 분야로 대두되고 있다. 본 논문에서는 데이터가 가지는 특성을 반영하면서 새로운 데이터를 학습하고 일반화하는 개념을 활용한 패턴 분류시스템을 제안하였다. 제안된 시스템은 데이터의 정량적인 퍼지 특징 분석을 기본으로 하는 특징영향평가(Feature Impact Evaluation) 방법을 도입하여 특징을 선택하고 비교사학습기반 데이터 마이닝 기법을 바탕으로 하는 OFUN-NET을 활용하여 지식베이스를 구축하고 그 결과를 판단모듈에 적용할 수 있다. 또한 240개의 흠패턴 영상을 제안된 시스템에 적용하여 자동 흠패턴 검사 시스템에 활용하는 과정을 보였다. 이와 같은 실험결과는 타당한 분류정확도를 통하여 본 시스템의 자동화 도구로서의 가능성과 제안된 방법의 타당성을 보였다.

본 연구는 준비된 데이터를 기반으로 하기 때문에 예외적인 데이터가 있다면, 이에 의존하는 성능의 편차가 생길 수 있는 가능성이 있으므로 이를 보완하기 위한 보다 일반적인 데이터 분석방법이 연구되어야한다. 또한 성능 향상을 위하여 특징선택과정에서 시스템의 성능을 예측하여 동적으로 학습모듈을 구성하는 방법도 함께 연구되어야한다. 각 모듈에 포함된 알고리즘에 대한 면밀한 분석을 바탕으로 다양한 데이터베이스에서 수집된 데이터를 가지고 여러 형태의 특징조합을 만들어 다양한 실험을 시도하는 과정에서 일반적인 모듈과 응용도메인에 의존하는 모듈로 나누어 연구하여 체계적인 지능형 시스템의 모형으로 발전시켜 나아가야할 것이다.

REFERENCES

- [1] Isabelle Guyon and Andre Elisseeff, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research* 3, pp.1157-1182, 2003.
- [2] Huan Liu, "Evolving Feature Selection", *IEEE Intelligent Systems and Their Applications* Vol. 20, Issue 6, pp.64-76, Nov.-Dec. 2005.
- [3] Gupta, M. M., Jin, L., and Homma, N., "Static and Dynamic Neural Networks : From Fundamentals to Advanced Theory", Wiley-IEEE Press, pp.634-752, April 2003.
- [4] Chin-Teng Lin, Chang-Mao Yeh, Shen-Fu Liang, Jen-Feng Chung and Nimit Kumar, "Support-Vector-Based Fuzzy Neural Network for Pattern Classification", *IEEE Trans. on Fuzzy System*, Vol. 14, No. 1, pp.31-41,

Feb. 2006.

- [5] Z. Vlad, M. D. Ofelia, and T-A. Maria, "Fuzzy Clustering in an Intelligent Agent for Diagnosis Establishment", Scientific Bulletin of the Petru Maior University of Tirgu Mures Vol. 6, pp.41-45, 2009.
- [6] H. S. Rhee, "Computer Aided Diagnosis System based on Performance Evaluation Agent Model", Journal of The Korea Society of Computer and Information Vol. 21 No. 1, pp.9-16, January 2016.
- [7] H. S. Rhee, "The Database Construction of a Classification System Using an Optimal Cluster Analysis Model", The Korean Institute of Communications and Information Sciences(KICS) Vol. 23, No. 4, pp.1045-1050, 1998.
- [8] M. K. Chakraborty, A. Skowron, M. Maiti, S. Kar, "Facets of Uncertainties and Applications", ICFUA, Kolkata, India, pp.33-50, Dec. 2013.
- [9] Milos R., Mohamed G., Nenad F., and Zoran O., "Minimum Redundancy Maximum relevance feature selection approach for temporal gene expression data", BMC Bioinformatics, Vol. 18, No. 9, pp.1-14, 2017.
- [10] Qinpei Zhao, "Cluster Validity in Clustering Methods", Dissertations in Forestry and Natural Sciences, Publications of University of Eastern Finland, pp.1-189, June, 2012.
- [11] Csaba L., Sandor, J. and Attila B., "Cluster Validity Measurement Techniques", proc. of the 5th WSEAS International Conference on AI, KE and DB, pp.1-11, 2006.
- [12] Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", 2nd Edition, John Wiley & Sons, Inc., pp.161-214, 2001.
- [13] C. Hung, M. Pham, S. Arasteh C. Kuo, and T. Coleman, "Image Texture Classification Using Texture Spectrum and Local binary Pattern", IEEE Trans. on Pattern Recognition, pp.2739-2742, 2006.
- [14] Mohd Shah Hairol Nizam, Sulaiman Marizan, Shukor Ahmad Zaki and Ab Rashid Mohd Zamzuri, "Vision based Identification and Classification of Weld Defects in Welding Environments: A Review, Indian Journals of Science and Technology, Vol. 19, No. 20, pp.1-15, May, 2016.
- [15] Mohsen Zand, Shyamala Doraisamy, Alfian Abdul Halin, Mas Rina Mustafa, "Texture Classification and Discrimination for Region-based Image Retrieval", J. Vis. Commun. Image R., Vol. 26, pp.305-316, 2015.

Authors



Hyun-Sook Rhee received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Sogang University, POSTECH, Sogang University, Korea, in 1989, 1991 and 1997, respectively. Dr. Rhee joined the faculty of the Department of

Software Engineering at Dongyang Mirae University, Seoul, Korea, in 1997. She is currently a Professor in the Department of Software Engineering at Dongyang Mirae University. She is interested in intelligent information system, data mining, and software development methodology.