

문자인식 분야의 기계학습 기법

전남대학교 | 손화정 · 김수형*

1. 서론

컴퓨터 비전(computer vision)은 인간이 시각을 통해 영상의 어떠한 장면이나 특징들을 이해하는 것과 유사한 능력을 컴퓨터에게 부여하고자 하는 인공지능(artificial intelligence)의 한 분야를 말한다. 컴퓨터 비전을 실현하기 위한 다양한 기술 중 하나로, 인간의 시각 및 청각을 통해 받아들인 데이터를 인지하는 패턴인식 능력을 컴퓨터상에서 구현하고자 하는 것을 패턴인식 기술이라 한다. 이러한 기술은 문자인식, 음성인식, 도면인식, 공장에서의 자동 품질검사, 의료영상진단, 유전자 검사, 기후 예측, 주가 예측 등 수많은 분야에 활발히 이용되고 있다.

본고에서는 앞서 나열한 패턴인식 기술 중 문자가 포함된 다양한 형태의 문서들을 인식하는 문자인식(character recognition) 기술에 대한 개념, 응용 분야, 기계학습(machine learning) 알고리즘 등을 개괄적으로 고찰한다. 제 2절에서는 문자인식 기술의 정의를 살펴보고 문자의 입력 장치 또는 입력 방식에 따라 문자인식 기술을 분류한다. 제 3절에서는 문자인식 기술이 응용될 수 있는 예를 살펴보며 현재 상용화되어 있는 소프트웨어들을 제시함으로써, 문자인식 기술의 현 수준을 파악한다. 제 4절에서는 문자인식 분야의 대표적인 기계학습 알고리즘 몇 가지를 살펴본다. 제 5절에서는 현 기술에서의 한계점 및 앞으로 지향해야 할 연구의 방향 등을 고찰함으로써 결론을 맺는다.

2. 문자인식의 기술 분류

문자인식에 관한 연구는 패턴인식 분야에서 상당히 긴 역사를 가지고 있다. 그러나 초창기에는 문자인식의 어려움으로 인해 침체에 들어서게 되고 음성인식이나 영상인식 등 다른 새로운 분야로 연구 대상을 변경하게 되었다. 그러나 오늘날 PDA와 같은 하드웨어 기술이 발달하고 패턴인식 기술이 획기적으로 향

상됨에 따라 문자인식의 필요성은 더욱 증대되었으며, 이전에 불가능하였던 기술들이 가능하게 되어 보다 새로운 방법론들이 도입되기 시작하였다[1].

문자인식은 문자의 입력 형태에 따라 크게 온라인 문자인식과 오프라인 문자인식의 두 가지 범주로 분류한다(그림 1 참조). 온라인 문자인식은 전자펜, 터치스크린 등을 통해 입력받은 필기 문자를 인식하는 방법을 말하는데, 필기하는 방법에 따라 자유롭게 필기하는 방법과 제약을 가하여 필기하는 방법이 있다. 제약된 필기는 주어진 영역에 맞추어 쓰도록 하거나 또박 또박 띄어쓰기를 요구하는 등 필기 시 약간의 제약을 가하는 방법이며, 자유로운 필기는 어떠한 제한사항 없이 필기자 마음대로 쓰는 방법이다. 이러한 필기 문자는 사용자에게는 매우 편한 방법이지만 제약이 가해진 필기 문자에 비해 변형이 커지므로, 인식하는 것이 더 어렵게 된다.

오프라인 문자인식은 이미 쓰여진 문자영상을 카메라 및 스캐너를 통해 입력받아 인식하는 방법을 말하는데, 입력되는 문자의 형태에 따라 인쇄체 문자인식과 필기체 문자인식으로 구분한다. 인쇄체 문자인식은 타자기, 워드프로세서 등을 이용해 입력된 문자영상을 인식하는 것을 말하며, 필기체 문자인식은 사람이 직접 필기한 문자영상을 인식하는 것을 말한다. 오프라인 문자인식은 기존에 필기되거나 인쇄된 문자들을 입력받아 인식하므로 대량의 정보를 빠른 속도로 컴퓨터에 입력시킬 수 있다는 특성을 갖는 반면, 온라인

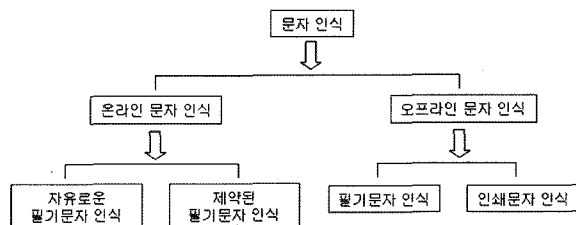


그림 1 문자인식의 기술 분류

* 중신회원

문자인식은 필기자가 진행 중에 인식을 수행하여 컴퓨터에 입력이 가능하므로 키보드를 이용하지 않고도 자연스럽게 편리한 입력이 가능하다는 특성을 갖는다.

3. 문자인식의 응용 분야

현재까지의 문자인식 기술은 우리 실생활의 다양한 분야에 적용되어 이미 널리 활용되고 있다. 이 절에서는 문자인식 기술이 응용될 수 있는 여러 분야를 살펴보고 현재 이러한 기술을 이용하여 상용화된 소프트웨어들을 나열한다.

3.1 온라인 문자인식

1950년대 말 전자 테블릿(tablet)의 등장으로 시작된 온라인 필기문자 인식에 관한 연구는 눈부신 하드웨어 기술의 발달과 함께 1980년대 중반부터 본격적으로 시작되었다. 오늘날 전자펜과 테블릿을 이용한 펜컴퓨터, PDA, 터치스크린 등과 같은 제품들이 보편화됨에 따라, 키보드를 통한 입력 방식이 아닌 펜이나 손동작을 이용한 자연스러운 입력 방식을 이용한 연구의 필요성이 증대되었다.

온라인 문자인식은 필기하는 동안 필기된 문자를 인식하기 때문에 오프라인 방식과 달리 필기의 시간적, 공간적인 정보를 얻을 수 있으며, 획의 수, 획의 순서 및 필기 방향, 속도, 필압 등과 같은 인식하는데 있어 중요한 정보들을 얻을 수 있다. 또한 필기자에게 일정하게 주어진 영역 내에서만 필기하도록 제약을 준다거나 오인식된 경우에는 적절히 수정하고 편집할 수 있도록 하는 등 피드백이 가능한 이점이 있어, 오프라인 인식에 비해 훨씬 다양한 응용 소프트웨어들이 출시되어 있다.

PC 또는 PDA에서 전자펜이나 터치스크린을 통해 입력된 필기 문자를 인식하는 기술은 키보드를 대신한 문자 입력, 전자사전 검색, 온라인 서명검증 등에 응용 가능하다. 또한 최근에는 휴대폰에 탑재된 센서를 통한 동작인식 기술을 이용하여 전자펜이나 터치스크린과 같은 별도의 입력장치 없이 문자 입력이 가능한 기술도 개발되었다.

1) 온라인 필기체 인식 소프트웨어

온라인 필기체 인식 기술의 정확도는 매우 높은 수준까지 올라와있어, 이를 이용한 제품들이 국내 및 국외에서 많이 시판되었다. 국내 시판된 몇 가지 상용 소프트웨어를 예를 들면, 1999년 디오텍에서는 DioPen[2]이라는 한글, 영문, 한자 등 다양한 언어의 필기문자를 인식할 수 있는 온라인 필기인식 소프트웨어를 개

발하였다. 마이그룹에서는 2002년 한글과 영문에 대해 정자체로 필기할 때 박스 내에 입력할 경우에는 99%, 자유로이 입력할 경우에는 95%의 인식률을 보인다고 알려진 온라인 필기인식 소프트웨어 MMiPen[3]을 출시하였다.

2) 온라인 서명검증 소프트웨어

서명검증도 온라인과 오프라인 방식으로 구분할 수 있는데, 상용화된 제품으로는 금융권에서의 거래업무, 사용자인증 전자지불시스템, 사내 개인 출입인증 등 여러 분야에 활용할 수 있는 온라인 방식이 대부분이다. PassSign[4]은 2000년 패스사인에서 개발한 전자펜을 이용한 온라인 서명검증 소프트웨어로서, 국내 많은 유통업체 및 백화점의 사용자인증 전자지불시스템으로 널리 사용되고 있다. 2001년 마이그룹에서도 SignQ[3]를 출시하였다.

3) 연속동작 인식 휴대폰

삼성전자는 2004년 버튼을 누르거나 말을 해 음성을 인식시키지 않고 휴대폰을 움직이면서 필기된 글자를 인식하여 휴대폰의 기능을 작동시키는 연속동작인식 휴대폰 SCH-S310[5]을 개발하였다. 이는 공간에서 움직임을 정확히 인식하는 제스처 인식 기술과 인식된 제스처 즉, 필기 문자를 인식하는 문자인식 기술이 결합된 형태로, 삼성종합기술원과 공동 개발하여 특허 22건을 출원했고, 국내·외 유명 학회에 논문을 발표하여 그 기술의 우수성을 인정받았다.

3.2 오프라인 문자인식

오프라인 문자인식은 1930년대 독일과 미국을 시작으로 1950년대에 미국, 영국, 독일 등에서 상용화된 광학 문자 인식 소프트웨어(Optical Character Recognition: OCR)를 발표하면서부터 활발한 연구가 시작되었다. **처음에는 제한된 활자체의 인쇄문자를 대상으로 연구가 이루어졌으나, 1960년대에 이르러서는 필기체 인식으로 연구가 확대되었으며, 우편번호 인식용 OCR과 같은 실용적인 분야에의 응용도 가능하게 되었다.** 1970년대에는 필기체 숫자 및 한자, 가타가나 등 좀 더 어려운 문제를 해결하는 데에 연구가 이루어졌으며, 1980년대에 들어서는 97% 이상의 필기체 한자 인식 기술을 얻는 기술을 개발하는 등의 큰 성과를 이루었다. 미국이나 유럽과 같은 국가에서는 1970년대 후반에 이미 인쇄체 문자를 성공적으로 인식하는 OCR을 개발하였고, 그 이후에는 주로 일본에서 인쇄체 및 필기체를 중심으로 대부분의 연구가 이루어졌다. 이처럼 국외 여러 나라에서는 일찍부터 체계적이고 성공적인 연구가 이루어졌

음에 반해, 국내에서의 한글 문자인식에 관한 연구는 1980년대 말에 이르러서야 소수의 국가 연구소와 기업체를 중심으로 본격적으로 시작하였다.

오늘날 문자인식 기술은 우편물 자동 분류를 위한 우편번호 인식, 전표 또는 수표 인식을 통한 자동업무 처리, 필기체 문자인식을 통한 고문서 자동입력, 문서인식을 이용한 도서관에서의 문서영상 검색 등에 효과적으로 사용되고 있다.

1) 문서인식 소프트웨어

스캔된 인쇄 문서의 문자를 인식하는 대표적인 OCR 소프트웨어로 글눈과 아르미를 들 수 있다. 한국인식 기술에서 개발한 글눈[6]과 퍼셉컴에서 개발한 아르미[7]는 한글과 영어, 한자 등 다양한 문자에 대해 스캔 문서를 인식해주는 국내의 대표적인 문자인식 소프트웨어이다.

2) 문서영상 검색

문서영상 검색은 웹 기반 디지털 도서관 구축을 위한 주요 기술 중의 하나로 문서 영상의 자동 색인 및 검색을 수행한다. 이를 위해, 문자나 자소 단위의 분할 기술과 OCR 기술을 결합하여 문서영상의 내용 전체를 텍스트 형태로 변환한 후 형태소 분석 등의 과정을 거쳐 색인하고 검색하는 방법과 문서영상을 띄어쓰기 단위로 단어영상의 집합으로 분할한 후, 검색하려는 주제어와 분할된 단어 영상을 매칭하는 과정을 통해 문서영상을 검색하는 방법이 있다[8]. 이전에는 OCR 기반 문자인식 방법이 주 연구 대상이었으나 요즘에는 OCR 기술의 한계를 극복할 수 있는 영상기반 문자인식 방법에 관한 연구가 주로 이루어지고 있다[9-12].

3) 고문서 인식

오프라인 문자인식은 문서 형태로 기록된 방대한 기존 지식의 데이터베이스화를 가능하게 하는 기술로서, 인쇄된 공공문서나 연구문헌뿐만 아닌 예전에 손으로 쓰여진 고문서도 포함할 수 있다. 이러한 기술은 조선실록, 승정원일기, 팔만대장경 등 대용량의 한국학 자료의 입력을 수작업이 아닌 자동화가 가능하도록 하므로, 기존에 전문가의 수작업에 의해 이루어졌던 방식에 비해 입력 시간 및 비용에 있어 최대 10배 이상의 효율을 높일 수 있다. 2004년 국내에서는 동방라이텍에서 개발한 고문서영상 자동인식 고속처리 시스템[13]을 일본, 중국 등에 수출함으로써 국내 한자인식 기술의 우수성을 입증하였다.

4) 우편물 자동분류

우편물 자동분류는 우체국에 들어오는 수많은 양의

우편물을 발송지역별로 분류하는 작업을 말한다. 사람의 손에 의해 분류되었던 이러한 작업을 문자인식 기술을 이용하여 자동화된 시스템으로 구축함으로써, 분류 작업의 속도 및 정확도를 향상시킬 수 있다. 우편물 자동분류 기술은 현재까지도 꾸준히 연구되고 있는데, 초기 우편물 자동분류기술은 OCR을 이용하여 우편번호를 인식하여 인식 결과를 바코드로 표기하여 일괄적으로 분류하는 방식이었다. 그러나 바코드를 별도로 인쇄해야하는 불편함과 바코드 판독기를 갖추어야하는 단점이 있어, 이러한 문제 해결을 위해 우편번호 및 주소를 동시에 인식하는 방법에 관해 연구[14-16]가 이루어졌다. 즉, 별도의 바코드 입력 및 인식 과정 없이 우편번호와 주소를 인식함으로써 해당 주소로의 일괄적인 분류가 가능하도록 하는 방법이다. 또한 한국전자통신연구소(ETRI) 우정기술 연구센터에서 2004년부터 RFID 기술을 활용한 제2세대 우편물류 시스템인 u-포스트(post)시스템 구축을 목표로 연구를 수행하고 있다[17].

5) 형식문서 인식

문자인식의 대상이 되는 문서 중에서 카드전표, 청약서, 고지서, Fax 문서, 각종 계약서 등 일정한 형식을 갖춘 서식문서를 인식하는 기술은 자유로이 필기되거나 인쇄된 문서를 인식하는 기술에 비해 상대적으로 쉬운 문제이다. 따라서, 높은 인식 성능을 갖춘 전표인식, 명함인식 등 다양한 상품이 출시되어 판매되고 있다. Inzi iForm[18]은 인지소프트에서 개발한 형식문서 인식 소프트웨어로서 은행, 카드회사, 보험회사, 공공기관 등의 데이터베이스 구축에 효과적으로 이용되고 있다. 또한 한국인식기술에서는 Hi-Name[6]이라는 97% 이상의 인식 성능을 갖는 명함인식기를 개발하여 명함 안에 있는 이름, 전화번호, 이메일, 주소 등을 자동으로 인식하여 휴대폰이나 컴퓨터에서 관리 가능하도록 하였다.

6) 자연 영상에서의 문자인식

지금까지 살펴본 문자 위주로 이루어진 일반 문서에서의 문자인식 기술 외에도 자연 영상에 포함된 문자를 인식하는 기술 또한 새로운 응용분야이다. 정지영상나 동영상에 인위적으로 포함된 텍스트들은 영상의 주요 내용을 함축적이고 간략하게 표현하는 중요한 정보라 할 수 있다. 이러한 정보들을 실시간에 추출하여 정확히 인식할 수 있다면 웹 이미지/비디오 검색 시스템, 차량번호판 인식 시스템, 스팸 메일 필터링, 표지판이나 간판을 읽어주는 시각장애인 보행안내시스템 또는 외국인인을 위한 자동번역시스템 등 다양한 분야에 유용하게 사용할 수 있다[19-21].

4. 기계학습 기법

문자 데이터를 입력패턴으로 하는 문자인식은 일반적으로 학습 단계와 일반화 단계로 구분되며, 진행 과정은 다음과 같다(그림 2 참조).

학습 단계에서는 입력된 인쇄체 또는 필기체 데이터에서 어떤 주어진 공통의 속성(특징)을 찾고 다양한 기계학습 알고리즘을 통해 추출된 특징들을 해당 클래스로 분류해줄 수 있는 판별함수(모델)를 생성한다. 온라인 데이터에서는 획의 순서, 방향, 필압 등과 같은 특징을 얻을 수 있으며, 오프라인 데이터에서는 메쉬, 윤곽선, 히스토그램 영상의 너비나 높이 등과 같은 특징을 추출할 수 있다.

동일한 클래스 내의 패턴과 동일하지 않은 클래스 내의 패턴이 명확히 구분될 수 있다면, 이를 구분하는 분류 문제는 보다 간단해질 수 있다. 그러나 실생활에서 접할 수 있는 데이터로부터 명확히 구분 가능한 특징을 추출하는 것은 불가능하거나 매우 어렵다. 따라서 특징 추출 단계에서는 정보의 손실을 최소화할 수 있는 특징을 추출하는 것이 요구된다. 또한 추출된 특징을 이용하여 클래스를 명확하게 구분해 줄 수 있는 판별함수를 학습하는 기계학습 과정 역시 중요하다. 단순히 입력된 하나의 패턴에서 얻은 정보만을 이용해 판별하는 것은 합리적인 추론 방법이라 할 수 없다. 일반적인 경우에도 정확한 분류가 가능한 최적의 판별함수를 모델링하기 위해서는 여러 개의 데이터를 가지고 반복적으로 학습(learning) 또는 훈련(training)하는 최적화 과정이 필수적이다. 온라인이나 오프라인이냐에 따라 입력되는 데이터의 특성이 다르므로 분류를 위한 기계학습 기법도 서로 상이해지게 되는데, 대표적인 기계학습 방법에 관해서는 다음 장에서 간략히 서술하겠다.

마지막으로 훈련된 결정함수를 이용하여 미지의 입력 데이터가 어느 클래스에 해당하는지를 판단해주는

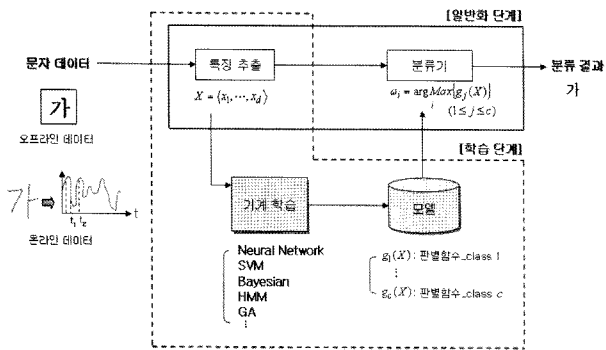


그림 2 문자인식 시스템의 구성도

과정이 바로 분류 단계이다. 입력 데이터로부터 추출된 특징 벡터를 X 라 하고, 이 값이 클래스 $\Omega = \{\omega_i, i=1, \dots, c\}$ 내에 포함된다고 가정할 때, 분류 단계에서는 훈련 과정에서 구축된 c 개의 판별함수를 이용하여 입력 데이터 X 가 가장 큰 값을 갖는 판별함수 $g(X)$ 를 찾고 데이터 X 를 클래스 ω_i 로 분류하는 과정을 거친다(그림 3 참조). 예를 들어, 다중 클래스를 분류하는 선형 판별함수의 경우에는 훈련을 통해 최적의 가중치 w 를 갖는 c 개의 판별함수

$$g_i(X) = w_i^T X + w_{i0}, \quad i = 1, \dots, c$$

를 계산하고, 이를 이용하여 모든 $i \neq j$ 에 대해 $g_i(X) > g_j(X)$ 이면 X 를 클래스 ω_i 에 할당한다.

다음은 문자인식 시스템의 구성 단계 중 기계학습에서 대표적인 알고리즘이라 할 수 있는 신경망, SVM, Bayes 결정규칙, HMM에서의 학습 과정을 간략히 서술한다. 이들 기계학습 알고리즘의 훈련 과정은 공통적으로 각 클래스별 판별함수를 결정짓는 가중치 벡터를 계산하고자 하는 것을 목적으로 한다. 단, 확률 통계적 방법에 속하는 HMM과 Bayes 결정규칙의 경우에는 신경망이나 SVM과는 달리, 판별함수의 값이 통계에 근거한 확률로 주어지기 때문에 입력 데이터 X 가 클래스 ω_i 에 속할 확률 $P(\omega_i | X)$ 가 앞서 말한 $g(X)$ 에 대치될 수 있다. 그리고 이 경우에는 확률밀도함수의 파라미터를 최적화하는 것이 바로 훈련 단계의 목적이다. 여기서는 문자인식 분야에서 이러한 기계학습 방법이 어떻게 적용되는지를 살펴보고자 하므로, 알고리즘에 관한 자세한 사항은 생략하였다.

4.1 신경망(Neural Net)

초기 신경망 분야의 연구에 커다란 영향을 미친 퍼셉트론(perceptron)은 1957년 Rosenblatt에 의해 처음 제안되었다[22]. 그러나 퍼셉트론은 선형 분리의 단순한 문제인 XOR도 해결할 수 없다는 단점이 존재하

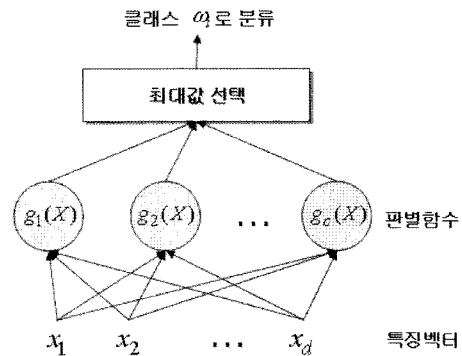


그림 3 클래스 분류 규칙

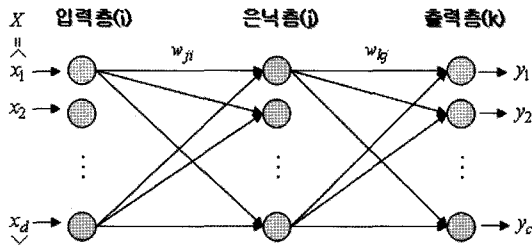


그림 4 다층 퍼셉트론의 구조

였다. 이에 입력층과 출력층 사이에 하나 이상의 중간층을 추가하고 오류역전파(backpropagation) 학습 알고리즘을 적용함으로써 선형 분리문제를 해결한 다층 퍼셉트론(Multi-Layer Perceptron: MLP)이 1980년대에 제안되었다. MLP의 기본 구조는 그림 4와 같다. 입력층, 은닉층, 출력층의 각 노드는 연결강도를 나타내는 가중치로 연결되어 있고, 출력층에서 입력층으로의 직접적인 연결은 존재하지 않는 전방향 네트워크이다.

MLP의 학습은 각 층의 노드들을 연결하는 가중치 벡터를 결정하는 과정으로, 입력 데이터가 입력층의 각 노드에 제시되면 이 신호는 각 노드에서 변환되어 중간층에 전달되고 최종적으로 출력층으로 나오게 된다. 이 출력값과 원하는 출력값을 비교하여 그 차이를 감소시키는 방향으로 연결강도를 조정하고, 상위층에서 역전파하여 하위층에서는 이를 근거로 다시 자기층의 연결강도를 조정해나간다. 가장 대표적인 학습 알고리즘으로 오류역전파 학습 알고리즘이 있는데, p 번째 학습 패턴이 주어진 경우 노드 i 에서 노드 j 로의 연결강도의 변화량 $\Delta_p w_{ji}$ 는 식 (1)과 같이 표현 가능하다.

$$\Delta_p w_{ji} = -\eta \frac{\partial E_p}{\partial w_{ji}} = \eta \delta_{pj} o_{pi} \quad (1)$$

여기서 η 는 가중치 변화의 상대적인 크기를 나타내는 학습률을 의미하고 E_p 는 패턴 p 에 대한 오차의 제곱으로 다음과 같다.

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2 \quad (2)$$

여기서, t_{pj} 는 p 번째 목표출력 패턴의 j 성분, o_{pj} 는 p 번째 입력패턴으로부터 네트워크가 계산한 출력의 j 성분을 나타낸다.

실제 네트워크 상에서의 변화 과정을 살펴보면, 패턴 p 를 입력층의 노드 i 에 제시하여 출력되는 값을 o_{pi} 라 하고, 은닉층의 노드 j 의 출력과 출력층의 노드 k 의 출력을 o_{pj} 라 할 때, 이 값들은 다음과 같이 계산한다.

$$o_{pj} = f\left(\sum_i w_{ji} o_{pi} + \theta_j\right) \quad (3)$$

$$o_{pk} = f\left(\sum_j w_{kj} o_{pj} + \theta_k\right) \quad (4)$$

여기서, f 는 활성화 함수(activation function)로서 주로 시그모이드(sigmoid) 함수가 이용되며, 오프셋(offset 또는 bias) θ_j , θ_k 는 처음에는 아주 작은 임의의 수(일반적으로 $-0.5 \sim 0.5$ 사이의 값 사용)를 사용하고 반복과정을 거치면서 점차 수정해 나간다. 각 노드를 연결하는 가중치는 다음과 같이 수정해나간다.

$$w_{kj} = w_{kj} + \Delta_p w_{kj} = w_{kj} + \eta \delta_{pk} o_{pj} \quad (5)$$

$$w_{ji} = w_{ji} + \Delta_p w_{ji} = w_{ji} + \eta \delta_{pj} o_{pi} \quad (6)$$

여기서, 시그모이드 함수를 이용할 경우 은닉층과 출력층 내 해당 노드의 오차는 다음과 같다.

$$\delta_{pk} = (t_{pk} - o_{pk}) o_{pk} (1 - o_{pk}) \quad (7)$$

$$\delta_{pj} = \sum_k \delta_{pk} w_{kj} o_{pj} (1 - o_{pj}) \quad (8)$$

위의 훈련 과정을 통해 네트워크의 가중치가 결정되면, d -차원의 미지의 입력 데이터 X 에 대한 출력층의 노드 i 에 대한 결정함수 $g_i(X)$ 는 다음과 같이 정의할 수 있다. 은닉층의 노드의 수는 n_H 라 가정한다.

$$g_i(X) = f\left(\sum_{j=1}^{n_H} w_{jk} f\left(\sum_{i=1}^d w_{ji} x_i + w_{j0}\right) + w_{ik0}\right) \quad (9)$$

4.2 서포트 벡터머신(Support Vector Machine: SVM)

1995년 Vapnik[23]과 벨 연구소팀에 의해 제안된 SVM은 기본적으로 두 범주를 갖는 객체들을 분류하는 방법으로, 선형분리 문제를 해결하는 방법이다(그림 5 참조). 또한 비선형분리 문제의 경우에는 슬랙 변수(slack variable)와 커널 함수(kernel function)를 사용하여 입력 데이터의 공간을 더 높은 차원의 특징 공간으로 사상(mapping)시킴으로써 선형분리 문제로 근

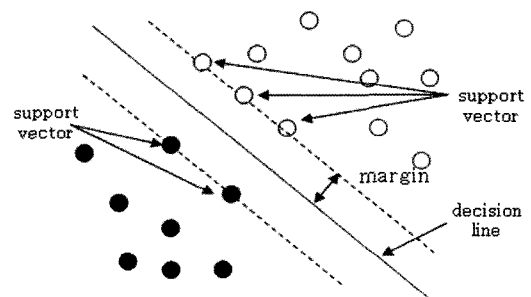


그림 5 SVM의 결정결계면

사화하여 해결 가능하다.

그림 5에서처럼 SVM의 학습 과정은 마진(margin)이 최대가 되도록 하는 결정경계면(판별함수)을 찾는 문제라 할 수 있다. 이때의 최적의 결정경계면은 가중치 w 와 바이어스 b 를 갖는 $y = w^T x + b$ 로 표현가능하다. 여기서 w 는 결정경계면의 법선 벡터로, 원점에서 이 초평면까지의 수직거리는 $|b|/\|w\|$ 가 된다.

다음의 두 초평면 H_1, H_2 를 고려할 때, 선형 SVM은 H_1 과 H_2 사이의 마진 $M = 2/\|w\|$ 을 최대화하는 문제로 귀착된다.

$$H_1 : y = w^T x + b = 1 \quad (10)$$

$$H_2 : y = w^T x + b = -1 \quad (11)$$

이를 마진의 역수를 최소화하는 문제로 바꾸면 다음과 같이 제약조건을 갖는 최적화 문제로 정의할 수 있다.

$$\begin{cases} \text{minimize} : \frac{1}{2} \|w\|^2 \\ \text{subject to} : d_i(w^T x_i + b) \geq 1 \quad i = 1, \dots, N \end{cases} \quad (12)$$

w 와 b 만 주어진다면 최적 결정경계면을 계산할 수 있으므로, 모든 데이터의 점이 정확히 어느 클래스에 속하는지를 판별할 수 있고 마진의 폭도 계산할 수 있다. 모든 데이터 점에 적합한 가장 넓은 마진을 이루는 최적의 w 와 b 를 찾아내는 방법은 라그랑지안 최적화(Lagrangian optimization) 기법을 이용한다. 즉, 위의 제약식을 결합하여 라그랑제 승수 α_i 를 포함한 라그랑지안 함수를 유도하고 최종적으로 이차 계획법(quadratic programming)을 이용하여 L_D 를 최대화하는 최적해 w 와 b 를 구한다. 비선형 문제도 해결할 수 있는 커널함수와 슬랙 변수를 추가하였을 경우, 다음과 같이 변형할 수 있다.

$$\begin{cases} \text{maximize} : L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{subject to} : \sum_{i=1}^N \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \quad \forall i \end{cases} \quad (13)$$

여기서 C 는 penalty 변수로, 선형의 경우에 비해 좀더 엄격한 제약을 주기위한 변수이다. 최종적으로 미지의 입력 데이터 X 에 대해 클래스를 분류할 수 있는 결정함수는 식 (14)와 같이 정의되는데, $g(X)$ 가 0보다 크면 클래스 1로 분류하며 그렇지 않을 경우에는 -1로 분류한다.

$$g(X) = w^T \Phi(X) + b = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \quad (14)$$

4.3 Bayes 결정규칙(Bayes Decision Rule)

Bayes 결정규칙은 확률 모델을 통해 패턴을 분류하는 단순한 기계학습 방법 중 하나로, 구조는 간단하지만 SVM이나 신경망과 같은 복잡한 분류 방법에 비해 비교적 좋은 성능을 보여주기 때문에 의사결정 함수로 많이 쓰이고 있다. 전술한 바와 같이 Bayes 결정규칙은 판별함수 자체가 확률값으로 표현되며, 관련된 사전확률이 알려져 있다는 가정 하에 입력 데이터에 해당하는 클래스가 명확히 구분될 수 있도록 가정한 확률 모델의 파라미터를 추정하는 과정이 바로 학습 단계에서 이루어진다.

어떤 연속형 데이터 X 가 클래스 ω_i 에 해당될 사전 확률 $P(\omega_i)$ 와 클래스 ω_i 에서 X 가 발생할 조건부확률 밀도 함수 $p(X|\omega_i)$ 를 안다고 가정할 때, 클래스 ω_i 에 속해있으면서 X 가 나타날 확률(결합확률)은 다음과 같이 표기 가능하다.

$$P(X, \omega_i) = P(\omega_i|X)p(X) = p(X|\omega_i)P(\omega_i) \quad (15)$$

위 식을 이용하면 X 가 클래스 ω_i 에 속할 사후확률 $P(\omega_i|X)$ 은 식 (16)과 같이 유도 가능하며, 이를 Bayes 공식이라 한다.

$$P(\omega_i|X) = \frac{p(X, \omega_i)}{p(X)} = \frac{p(X|\omega_i)P(\omega_i)}{p(X)} \quad (16)$$

여기서 $P(\omega_i)$ 와 $p(X|\omega_i)$ 는 사전에 주어진 값이며, 전 확률 $p(X)$ 는 다음과 같다.

$$p(X) = \sum_{i=1}^c p(X|\omega_i)P(\omega_i) \quad (17)$$

식 (16)에서 알 수 있듯이, Bayes 공식은 사전확률 $P(\omega_i)$ 와 조건부확률 $p(X|\omega_i)$ 로부터 계산할 수 있는데, 여기서 $p(X|\omega_i)$ 를 계산하기 위해서는 주어진 데이터가 어떤 분포를 갖는지에 대한 가정이 필요하다. 즉, 식 (18)과 같이 $p(X|\omega_i)$ 를 계산하기 위해서는 분자에 해당하는 $p(X)$ 에 대한 분포가 미리 알려져 있어야 한다.

$$p(X|\omega_i) = \frac{p(X, \omega_i)}{P(\omega_i)} = \frac{p(X)}{P(\omega_i)}, \quad X \in \omega_i \quad (18)$$

데이터 X 가 임의의 분포를 따른다고 가정한다면, 학습 단계에서는 이 분포가 갖는 파라미터를 추정한다. 예를 들어, X 가 다변량 가우시안(Gaussian) 확률분포를 따른다고 가정하면 d -차원 확률밀도함수는

$$p(X) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (X-\mu)^T \Sigma^{-1} (X-\mu)\right]$$

로 표기할 수 있다. 여기서, μ 는 d -차원의 평균 벡

터, Σ 는 $d \times d$ 공분산 행렬, $|\Sigma|$ 와 Σ^{-1} 은 각각 공분산의 행렬식과 역을 나타낸다. 가우시안 함수의 파라미터는 μ 와 Σ 로 이 값들을 훈련 단계에서 추정한다. 최종적으로 위 식을 식 (18)에 대입하여 Bayes 공식을 유도하면, 가우시안 분포를 따르는 Bayes 결정 규칙을 모델링할 수 있다.

4.4 은닉마르코프모델(Hidden Markov Model: HMM)

1980년대 이후 음성인식 분야에 가장 많이 사용된 HMM[24]은 유한상태의 마르코프 체인과 마르코프 체인의 함수인 심블 출력 프로세스로 구성된다. 이 모델은 음성인식뿐만이 아닌 다양한 형태를 가진 온라인 필기문자도 모델링할 수 있어 온라인 필기인식 분야에서 많이 이용되고 있다. 기본적으로 HMM을 이용한 문자인식은 주어진 관측열에 대해 최적의 확률값을 지니는 상태열을 찾는 것으로, 모델을 구성하고 있는 상태들 간의 전이가 특정한 확률값을 가정하여 이루어진다. HMM에서의 훈련과 인식 알고리즘을 간략히 설명하면 다음과 같다.

먼저, 상태 공간 S 의 원소수가 N 이고 출력 종류의 수가 L 일 때의 이산형 HMM에서의 출력치(observation) 생성에 대한 확률 구조를 정의하면 다음과 같다.

$$\begin{cases} \pi_{t+1} = \pi_t A \\ Q_t = \pi_t B \end{cases} \quad (19)$$

여기서 π_t 는 시점 t 에서의 상태확률분포를 나타내고, $A=(a_{ij})$ 는 전이확률행렬, Q_t 는 t 시점에서의 출력치확률분포, $B=(b_j(q_k))$ 는 출력확률행렬을 나타낸다. $b_j(q_k)$ 는 상태가 j 일때 출력치 q_k 가 생성될 확률이다. 일반적으로 출력치만 관측되고 상태의 흐름은 관측되지 못한 경우에 사용하므로 은닉 마르코프 모형이라 부른다. 출력치 생성을 위해서는 초기값 π_0 가 필요하므로, 이산형 HMM은 $\lambda=(A, B, \pi_0)$ 으로 표현된다.

훈련 단계에서는 관측열 $O=o_1, o_2, \dots, o_T$ 에 대해 주어진 모델 $\lambda=(A, B, \pi_0)$ 를 이루고 있는 모수들을 $F(O|\lambda)$ 를 최대를 하는 최대우도방법으로 추정한다. 일반적으로 EM 알고리즘에 의한 반복 계산으로 모수를 재추정해가는 Baum[25]의 재추정 알고리즘을 사용한다. (O, λ) 가 주어진 경우 (t, i) 에서 $(t+1, j)$ 로 이동하는 확률과 시점 t 에서의 상태 i 일 확률을 구하면 다음과 같다.

$$\begin{aligned} \gamma_t(i, j) &= P(s_t = i, s_{t+1} = j | O, \lambda) \\ &= \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) / \sum_{k \in S_t} \alpha_T(k) \end{aligned} \quad (20)$$

$$\gamma_t(i) = P(s_t = i | O, \lambda) = \sum_j \gamma_t(i, j) \quad (21)$$

위의 두 함수를 이용하여 (O, λ) 가 초기조건으로 주어진 경우 재추정된 모수값들은 다음과 같이 계산된다.

$$\hat{a}_{ij} = \sum_{t=1}^{T-1} \gamma_t(i, j) / \sum_{t=1}^{T-1} \gamma_t(i) \quad (22)$$

$$b_j(\hat{k}) = \sum_{t \in \{t | o_t = q_k\}} \gamma_t(j) / \sum_{t=1}^T \gamma_t(j) \quad (23)$$

$$\hat{\pi}_0^{(i)} = \gamma_1(i) \quad (24)$$

Baum-Welch 재추정 알고리즘은 다음과 같은 단계로 이루어진다.

[단계 1] HMM의 초기모형 $\lambda^{(0)}$ 선택한다.

[단계 2] $\gamma = 0, 1, 2, \dots$ 로 변화시켜가며 주어진 모형 $\lambda^{(\gamma)}$ 에서의 출력열 O 와 식 (22)-(24)를 이용하여 $\hat{\lambda}$ 을 구한 후 이를 $\lambda^{(\gamma+1)}$ 놓고 반복 수행한다.

[단계 3] 모수 추정치들의 변화량이 원하는 값보다 작아지면 반복을 중단하고, 이때 추정된 모형을 주어진 출력열 O 에 대응되는 HMM으로 결정한다.

인식 단계에서는 출력열과 HMM이 (O, λ) 로 주어진 경우 최적의 상태열을 찾음으로써 문자를 인식하는 과정으로, 일반적으로 다음과 같은 Viterbi 알고리즘을 이용한다.

[단계 1] 모든 $i \in S$ 에 대해, $\delta_1(i) = \pi_0^{(i)} b_i(o_1)$ 와 $\psi_1(i) = 0$ 을 계산한다.

[단계 2] $t=1, 2, \dots, T$ 로 변화시켜가면서,
 $\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] b_j(o_t)$,
 $\psi_t(j) = \arg \max_i [\delta_{t-1}(i) a_{ij}]$ 를 계산한다.

[단계 3] $P^* = \max_s \delta_T(s)$ 와 $s_T^* = \arg \max_s \delta_T(s)$ 를 계산한다.

[단계 4] $t=T-1, \dots, 1$ 로 변화시켜가며,
 $s_t^* = \psi_{t+1}(s_{t+1}^*)$ 을 이용해 경로역추적방법으로 최적 상태열 탐색한다.

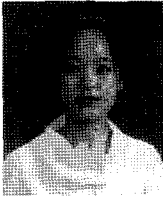
5. 결론

본고에서는 문자인식 기술에 대한 개괄적인 소개와 이러한 기술이 문자인식의 다양한 자동화 분야에 효과적으로 응용될 수 있음을 살펴보았다. 그리고 문자인식 시스템 구축을 위한 대표적인 기계학습 알고리즘의 학습 방법에 대해 간략히 서술하였다.

앞으로 고도화된 산업 기술의 발전 추세에 따라 문자인식 기술의 응용 분야도 더욱 다양해질 것이고, 이에 따라 보다 효과적인 기계학습 알고리즘의 필요성이 증대될 것이다. 또한 현재까지는 단순히 전자펜, 디지털 카메라, 스캐너 등과 같은 입력장치에서의 문자인식에 관한 연구가 대부분이었으나, 앞으로는 유비쿼터스 컴퓨팅 환경에서의 센서 등을 비롯한 새로운 입력장치가 계속적으로 개발될 것이므로 이에 빠르게 대응할 수 있도록 학문적인 연구뿐만이 아닌 실용적인 측면에서의 연구자들의 끊임없는 노력도 계속되어야 할 것이다.

참고문헌

- [1] C.C. Tappert, C.Y. Suen and T. Wakahara, "The State of the Art in On-Line Handwriting Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 8, pp.787~808, 1990.
- [2] DioPen, 디오텍, <http://www.diotech.co.kr>
- [3] MMiPen, SignQ, 마이그룹, <http://www.mmigroup.net>
- [4] PassSign, 패스사인, <http://www.passsign.com>
- [5] 연속동작인식 휴대폰, 삼성전자, <http://www.sec.co.kr>
- [6] 글눈, Hi-Name, 한국인식기술, <http://www.hi-name.net>
- [7] 아르미, 퍼셉컴, <http://www.perceptcom.com>
- [8] 오일석, 김수형, 유태웅, 박희규, "문서 영상 처리 기술과 디지털 도서관," 정보과학회지, 제20권, 제8호, pp.24~34, 2002.
- [9] D. Doermann, "The Indexing and Retrieval of Document Images: A Survey," Computer Vision and Image Understanding, Vol. 70, No. 3, pp.287~298, 1998.
- [10] 이준호, 이충식, 한선화, 김진형, "문자 인식에 의해 구축된 한글 문서 데이터베이스에 대한 정보 검색," 한국정보처리학회논문지, 제6권, 제4호, pp. 833~840, 1999.
- [11] J. Zhu, T. Hong, and J.J. Hull, "Image-based Keyword Recognition in Oriental Language Document Images," Pattern Recognition, Vol. 30, No. 8, pp.1293~1300, 1997.
- [12] C.L. Tan, W. Huang, Z. Yu and Y. Xu, "Imaged Document Text Retrieval Without OCR," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 6, pp.838~844, 2002.
- [13] 고문서영상 자동인식 고속처리 시스템, 동방라이텍, <http://www.dblitech.com>
- [14] E.A. Yonekura and J. Facon, "Postal Envelope Segmentation by 2-D Histogram Clustering through Water-shed Transform," Proceedings of the 7th International Conference on Document Analysis and Recognition, Vol. 1, pp.338~342, 2003.
- [15] V. Govindaraju and S. Tulyakov, "Postal Address Block Location by Contouring Clustering," Proceedings of the 7th International Conference on Document Analysis and Recognition, Vol. 1, pp.429~432, 2003.
- [16] 박상철, 정창부, 손화정, 김수형, "Virtual ID 사용을 위한 유사한 우편 영상 추출 방법," 2004 한국정보처리학회 추계학술발표논문집, 제11권, 제2호, pp. 817~820, 2004.
- [17] 한국전자통신연구소(ETRI) 우정기술 연구센터, <http://www.etri.re.kr>
- [18] Inzi iForm, 인지소프트, <http://www.inzisoft.co.kr>
- [19] J. Hoya, A. Shio and S. Akamatsu, "Recognizing Characters in Scene Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, pp.67~82, 1995.
- [20] R. Lienhart and W. Effelsberg, "Automatic Text Segmentation and Text Recognition for Video Indexing," ACM/Springer Multimedia Systems, Vol. 8, pp.69~81, 2000.
- [21] 김재광, "자동차 번호판 자동인식을 위한 문자 추출에 관한 연구," 명지대학교, 석사학위논문, 1996.
- [22] 김대수, "신경망 이론과 응용(I)," 하이테크정보, 1989.
- [23] V.N. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, 1995.
- [24] B.H. Juang and L.R. Rabiner, "Hidden Markov Models for Speech Recognition," Technometrics, Vol. 33, No. 3, pp.251~272, 1991.
- [25] L.E. Baum, T.Petrie, G. Soules and N. Weiss, "A Maximization Technique Occuring in the Statistical Analysis of Probabilistic Functions of Markov Chains," The Annals of Mathematical Statistics, Vol. 41, No. 1, pp.164~171, 1970.



손화정

2001 전남대학교 통계학과(학사)
 2004 전남대학교 전산학과(석사)
 2004~현재 전남대학교 전산학과 박사과정
 관심분야 : 패턴인식, 문서영상 정보검색
 E-mail : sonhj@iip.chonnam.ac.kr



김수형

1986 서울대학교 컴퓨터공학과(학사)
 1988 한국과학기술원 전산학과(석사)
 1993 한국과학기술원 전산학과(박사)
 1990~1996 삼성전자 멀티미디어 연구소(선임 연구원)
 2000~2002 캐나다 Concordia대학 CENPARMI 연구소 (방문교수)

1997~현재 전남대학교 전자컴퓨터공학부(교수)
 관심분야 : 인공지능, 패턴인식, 문서영상검색, 유비쿼터스컴퓨팅
 E-mail : shkim@chonnam.ac.kr

제17회 통신정보 융합기술대회(JCCI 2007)

- 일 자 : 2007년 5월 2일~4일
- 장 소 : 휘닉스파크
- 내 용 : 학술발표 등
- 주 최 : 정보통신연구회
- 상세안내 : <http://mobile.ajou.ac.kr/jcci2007>