

신중한 분류기와 학습 예제 가중치 조정을 이용한 차량번호판인식시스템의 인식성능 향상 방안

Vehicle License Plate Recognition System Using the Cautious Classifier and the Weighted Instance Method

백남철* · 이상협** · 류광렬***

Baik, Nam Cheol · Lee, Sang Hyup · Ryu, Kwang Ryul

Abstract

Vehicle License Plate Recognition System reads information from vehicles license plate using image detection devices. Of many applications provided by Vehicle License Plate Recognition System, some, such as speed enforcing system, can be problematic when the system incorrectly scans letters or numbers from a vehicle's license plate. Using Cautious Classifier avoids such problems by discarding the scanned information when the confidence level is doubted to be low. This study develops the License Plate Recognition System using Cautious Classifier and investigates effectiveness of applying the Weighted Instance Method to improve the performance of Cautious Classifier.

Keywords : vehicle license plate recognition system, vehicle image, confidence level, cautious classifier, weighted instance method

요 지

차량번호판인식시스템은 무인 카메라 등의 영상장치를 통하여 입력된 차량 이미지로부터, 차량번호판 정보를 읽어내는 시스템이다. 이러한 차량번호판인식시스템의 응용 시스템 중 과속차량 단속과 같은 일부 응용 시스템은 번호판의 글자나 숫자를 다른 글자나 숫자로 잘못 인식할 경우 심각한 문제를 발생시킬 수 있다. 이러한 문제를 피하기 위해 우리는 인식 결과에 대한 신뢰도가 낮은 경우 인식을 포기 또는 위임하는 '신중한 분류기(Cautious Classifier)'를 이용하여 인식시스템을 구성하였다. 또한 학습 예제의 가중치를 조정하는 방법을 사용하여 이러한 신중한 분류기의 성능을 향상시켰다.

핵심용어 : 차량번호판인식시스템, 차량 이미지, 신뢰도, 신중한 분류기, 가중치부여방식

1. 서 론

자동차의 차량번호판인식시스템은 무인 카메라에 의해 촬영된 차량 영상으로부터 차량번호를 자동으로 인식하는 시스템이다. 차량번호판 인식은 크게 전체 이미지로부터 차량의 번호판을 찾고 이의 인식을 위해 글자나 숫자를 이미지로부터 분리해내는 문자 분리 과정과 분리해낸 문자나 숫자를 학습 알고리즘 등을 이용하여 인식하는 2가지 과정으로 이루어져있다. 본 연구의 초점은 학습 알고리즘을 이용한 번호판의 문자 인식에 맞추어져 있다.

번호판인식 기술을 사용하는 일부 응용 시스템은 번호판을 다른 번호판으로 잘 못 인식할 경우 심각한 문제를 초래하는 경우가 있다. 일례로, 과속 방지용 카메라의 경우 카메라에 잡힌 과속 차량의 번호판을 인식해 해당 차량 소유주에

게 벌금 통지서를 보내게 되는데, 만약 번호판 인식 과정에서 다른 번호판으로 잘 못 인식할 경우 과속을 하지 않은 다른 차량 소유주에게 벌금 통지서를 보내는 문제가 발생할 수 있다. 이러한 응용 시스템에서는 인식에 대한 신뢰도가 낮은 인식 결과에 대해서는 인식을 포기하고 인식을 다른 시스템이나 사람에게 위임하는 방법을 통해 잘 못 인식할 가능성을 최소화한다(김종배 등, 2001; Papageorgiou and Poggio, 1999). 본 연구에서는 이와 같이 잘못된 인식에 의해 심각한 문제가 발생하거나 매우 큰 비용이 발생하는 경우를 최소화하기 위해 번호판인식시스템을 신중한 분류기(cautious classifier)를 이용해 구성하는데 신중한 분류기의 구성을 위해 MLP(multi-layer perceptron)를 사용하고, 성능 향상을 위해 학습 예제의 가중치를 조정하는 방안을 제안한다.

*한국건설기술연구원 도로연구부 선임연구원 (E-mail : nc100@kict.re.kr)

**정회원 · 교신저자 · 감시원 평가연구원 (E-mail : infohi2@hanmail.net)

***부산대학교 컴퓨터공학과 (E-mail : krryu@pusan.ac.kr)

2. 차량번호판의 문자 인식 과정

자동차의 차량번호판으로부터 인식해야 하는 내용은 “서울”, “부산”, “경기” 등의 지역 명과 지역 명 외의 한글 그리고 숫자의 3부분으로 나눌 수 있다. 본 연구에서는 문자 분리 과정 후 각 글자, 숫자의 위치 정보를 활용하여 지역 명, 한글, 숫자로 나누고 각각의 경우에 대해 서로 다른 분류기를 사용하여 인식한다.

지역명과 숫자의 경우 그 모양의 차이가 비교적 뚜렷하고, 그 종류가 비교적 적기 때문에 문자 분리만 잘 이루어진다면 잘 인식되기 때문에 특별한 과정을 거치지 않고 인식한다. 이에 비해 지역 명 외의 한글의 경우 그 종류가 무척 많고 경우에 따라 모음과 자음을 분리해야 하기 때문에 다소 복잡한 인식 과정을 거치게 된다. 본 연구에서는 그림 1에 보이는 비와 같이 한글 인식을 2단계에 걸쳐서 하게 된다. 먼저 각 한글의 자음을 무시하고 모음을 인식한 후 각각의 모음에 대해 서로 다른 인식기를 사용해 자음을 인식한다(이광원 et al., 2002).

예를 들어 1단계에서는 ‘가’, ‘나’, ‘다’ 등은 모두 동일하게 ‘ㅏ’로 인식하며, 그 후 각 글자의 자음을 인식한다. ‘ㅏ’, ‘ㅓ’, ‘ㅡ’와 같이 모음과 자음이 충분한 간격을 가진 경우에는 모음과 자음을 분리한 후 분리된 모음에 대해서만 인식하고, ‘ㄱ’, ‘ㅇ’와 같이 모음과 자음이 거의 붙어 있어 분리가 힘든 경우에는 분리하지 않고 모음을 인식하게 된다. 번호판에는 ‘ㅏ’, ‘ㅓ’, ‘ㅇ’, ‘ㅓ’, ‘ㅡ’의 총 5종류의 모음만이 사용되었기 때문에, 모음 분류를 위해서 1개, 각 모음에 대해 자음을 분류하기 위해 각 1개씩, 총 6개의 분류기를 학습해야 한다.

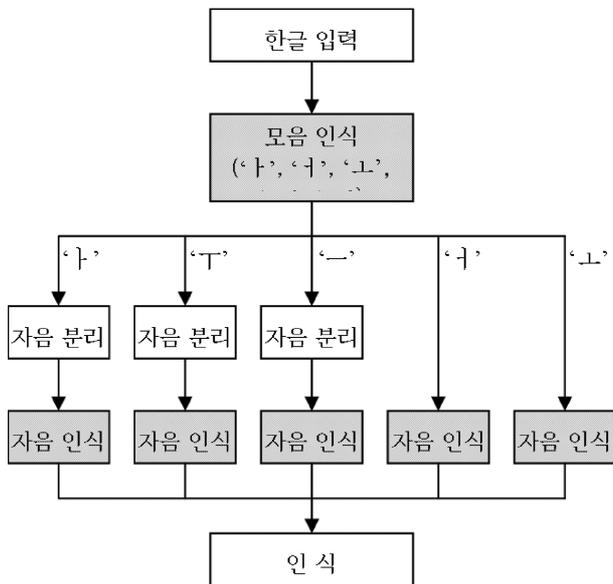


그림 1. 자동차 차량번호판의 한글 인식 과정

3. “신중한 분류기”를 이용한 차량번호판인식시스템의 성능 향상을 위한 학습 예제 가중치 조절

일반적인 차량번호판인식시스템이 모든 입력 번호판에 대해서 가능한 한 정확하게 인식하는 것을 목표로 하는 반면 본 연구에서 대상으로 하는 시스템에서는 일부 입력에 대해

서는 “모르겠다.”라고 답하는 경우가 생기더라도 잘못 인식하는 번호판이 생기지 않도록 하는 것이 목적이다. 이를 위해 본 연구에서는 신중한 분류기(cautious classifier)를 이용하여 번호판인식시스템을 구성한다. 신중한 분류기란 입력 전체에 대해 답을 하는 것이 아니라 일정 수준 이상의 높은 신뢰도(confidence)를 가지는 경우에 대해서만 답을 내어 놓고 그렇지 않은 경우에 대해서는 결과를 내어 놓지 않는 분류기이다. 이러한 신중한 분류기는 결과와 함께 결과에 대한 신뢰도를 내어놓을 수 있는 학습 알고리즘이라면 기존의 어떤 학습 알고리즘을 통해서도 생성할 수 있다(Draghici, 1997; 박창석 등, 2002). 분류기가 내 놓은 결과와 신뢰도를 가지고 일정 이상의 신뢰도인 경우 그 결과를 그대로 알려 주고 그렇지 않은 경우 분류를 포기하는 다음과 같은 규칙을 추가해 주기만 하면 되는 것이다. $fCLASS(e)$ 는 분류기가 예제 e 에 할당한 클래스이며, $fCONFIDENCE(e)$ 는 이 예측에 대한 신뢰도이다.

표 1. 신중한 분류기의 의사결정 규칙

Decision Rule for a Cautious Classifier with Threshold τ :
IF $fCONFIDENCE(e) > \tau$ THEN PREDICT $fCLASS(e)$
ELSE ABSTAIN

이러한 신중한 분류기에 의해 분류되는 데이터는 할당한 클래스가 실제와 동일한지 여부, 예측 신뢰도가 임계치보다 높은지 여부에 따라 다음의 4가지 경우로 나눌 수 있다.

1. 할당한 클래스가 실제와 같으며, 신뢰도가 임계값 보다 높은 경우
2. 할당한 클래스가 실제와 같으며, 신뢰도가 임계값 보다 낮은 경우
3. 할당한 클래스가 실제와 다르며, 신뢰도가 임계값 보다 높은 경우
4. 할당한 클래스가 실제와 다르며, 신뢰도가 임계값 보다 낮은 경우

이들 중 1, 4의 경우에는 문제가 되지 않지만 2, 3의 경우에는 너무 높거나 너무 낮은 신뢰도를 가지고 예측된 예제들이다. 우리가 원하는 학습 결과는 예측 값이 실제와 같은 경우에는 가능한 한 높은 신뢰도를 가지고, 예측 값이 실제와 다른 경우에는 가능한 한 낮은 신뢰도를 가지기를 원한다. 이를 위해 우리는 문제가 되는 2, 3의 경우의 학습 데이터에 대해 가중치를 증가시키고, 1, 4의 경우에 대해서는 가중치를 감소시키는 방법을 통해 성능을 향상시키고자 한다.

가중치 조절을 위해 가중치가 조절되지 않은 원 학습 데이터로 생성된 분류기에 학습 데이터 자체를 테스트해서 학습 데이터에 대한 예측이 옳게 되는지 여부와 신뢰도를 결정한다. 결정된 예측 결과와 신뢰도를 바탕으로 학습 예제의 가중치를 조정한다. 아래와 같이 자체 분류 결과가 좋지 않은 학습 예제에 대해서는 가중치를 높여준다.

1. 할당한 클래스가 실제와 다르며, 신뢰도가 τ_1 (임계치)보다 높은 예제
 2. 할당한 클래스가 실제와 같으며, 신뢰도가 τ_2 (임계치)보다 낮은 예제
- 가중치를 높이는 방법으로는 간단하게 학습 데이터 내에

해당 학습 예제의 수를 10배로 늘여준다. 자체 테스트 통해 아주 좋은 결과를 보인 학습 데이터에 대해서는 가중치를 낮추는데 이를 위해서 예측 결과가 실제 클래스와 동일하면서 신뢰도가 매우 높은 학습 데이터의 가중치를 '0'으로 조정한다. 가중치를 '0'으로 조정하는 방법은 간단하게 학습 예제 집단으로부터 해당 학습 예제를 제거하기만 하면 된다. 제거되는 학습 예제의 수는 여기서는 가중치 조정 과정에서 추가된 학습 예제의 수만큼 제거한다. 결과적으로 학습 예제의 가중치를 조정하기 전과 조정한 후의 전체 학습 예제의 수는 동일하게 된다.

4. 실험 결과

실제 차량 1만대의 번호판이 실험을 위해 사용되었다. 문자 분리를 거친 각각의 문자는 10×10의 이미지로 변환 후 속성을 추출하였다. 결과적으로 각각의 문자는 100개의 속성과 1개의 클래스 값을 가지게 된다. 본 실험에서는 2가지 평가 기준을 사용하여 비교실험을 수행하였다.

첫번째는 신뢰도로 평가 데이터를 정렬한 후 분류한 결과에 대해서는 정확성이 100%가 되도록 신뢰도를 설정했을 경우 분류기에 의해 거부되지 않고 분류된 데이터의 비율이며 두번째는 이진 분류 문제에서 분류 결과의 순위를 평가하기 위해 자주 사용되는 AUC(Area Under Curve)와 유사한 값을 사용한다. 이진 문제 상에서 AUC는 참(true)인 예제에 대해서는 순위가 높을수록, 거짓(false)인 경우에 대해서는 순위가 낮을수록 값이 높아진다. 이와 유사하게 본 연구의 대상 문제에서는 신뢰도로 정렬한 상황에서 옳게 분류된 데이터의 경우에는 순위가 높고 잘못 분류된 데이터에 대해서는 순위가 낮을수록 좋다.

$$AUC = \frac{S_o - n_o(n_o + 1)/2}{n_o n_1}$$

기존 AUC의 경우에는 S_o 는 참인 예제의 순위의 합이며, n_o 는 참인 예제의 수이며, n_1 은 거짓인 예제의 개수이지만, 여기서는 S_o 는 옳게 분류된 예제의 순위의 합이며, n_o 는 옳게 분류된 예제의 개수, n_1 은 잘못 분류된 예제의 개수이다.

그림 2는 모음 'ㄱ'에 대해 자음 14개를 인식하는 데이터

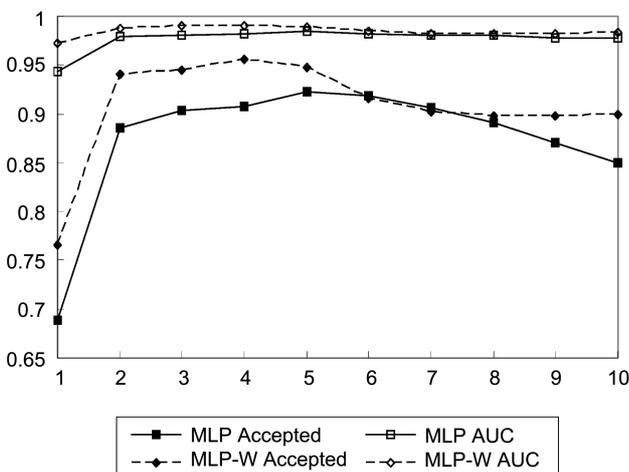


그림 2. 모음 'ㄱ'에 대한 자음 인식 결과

에 대한 실험 결과이다. MLP는 가중치 조정을 하지 않은 신경 회로망이며, MLP-W는 가중치 조정을 수행한 신경 회로망이다. 이 실험 결과에 의하면 학습이 진행되는 동안 가중치를 조정된 경우가 더 좋게 나타남을 알 수 있다. 또한 AUC와 인식을 모두 학습 초기에 상승하다가 일정 시점 후에 감소하는 것을 확인할 수 있다.

표 2는 각각의 모음에 대해 자음 인식을 수행한 결과인데 대부분의 경우에 가중치를 조정하는 방법이 성능 향상에 도움이 됨을 확인할 수 있다.

표 2. 자동차 차량번호판 인식 결과

	MLP		Weighted-MLP	
	AUC	accept rate	AUC	accept rate
ㅏ + 자음	98.3%	69.2%	99.2%	95.1%
ㅑ + 자음	98.4%	92.3%	99.0%	95.6%
ㅓ + 자음	96.0%	72.0%	95.8%	86.2%
ㅕ + 자음	98.3%	90.3%	98.9%	90.8%
ㅗ + 자음	96.9%	85.7%	98.7%	90.2%

5. 결론 및 향후 연구

실제 실세계의 기계 학습(machine learning)이나 데이터 마이닝의 응용 시스템에서 잘못된 인식이 큰 문제를 일으키거나 큰 비용을 발생시키는 경우가 많다. 이런 경우 판단에 대한 신뢰도가 낮을 경우 판단을 보류하여 사람에게 맡기거나 판단할 수 없으므로 결정하는 것이 더욱 바람직할 것이다. 이에 본 논문에서는 신중한 분류기를 이용해 차량번호판 인식시스템을 구성하였으며, 학습 예제의 가중치를 조정하는 방법을 통해 인식 성능 향상을 이루었다. 또한 실제 차량의 번호판을 인식하는 실험을 통해서 우리가 제안한 방법이 인식을 향상에 도움이 됨을 보여주었다.

본 연구에서 제안한 방법이 다른 학습 알고리즘과 데이터들에 대해서도 좋은 결과를 보여주리라 생각된다. 향후 우리는 다양한 알고리즘과 데이터에 대해 본 연구에서 제안한 방법을 적용해 보고자 한다.

참고문헌

김중배, 김갑기, 김광인, 박민호, 김항준(2001) 학습 기반의 자동차 번호판 인식 시스템, 한국신호처리·시스템학회, 한국신호처리·시스템학회, 제2권 제1호.
 박창석, 김병만, 이광호, 최조천, 오득환(2002) 모듈라 신경망을 이용한 자동차 번호판 문자인식, 한국정보과학회 학술발표논문집.
 이광원, 고미애, 김영모(2002) 통계적 특징점을 이용한 자동차 번호판의 단계별 문자인식, Joint Conference on Communication & Information.
 Draghici, S. (1997) A neural network-based artificial vision system for license plate recognition, *International Journal of Neural Systems*, Vol. 8, No. 1.
 Papageorgiou, P. and Poggio, T. (1999) A Trainable Object Detection System: Car Detection in Static Image, Massachusetts Institute of Technology Artificial Intelligence Laboratory.

(접수일: 2004.12.20/심사일: 2005.1.14/심사완료일: 2005.1.14)