

나이프 베이지안 환경에서 미분류 데이터를 이용한 성능향상

이 창 환[†]

요 약

많은 경우에 분류데이터의 생성은 사람의 시간과 노력에 의존하기 때문에 많은 비용과 시간을 요구한다. 이에 반하여 미분류 데이터는 거의 비용을 들이지 않고 무제한의 데이터를 쉽게 획득할 수 있다. 따라서 기계학습에 있어서 이러한 미분류 데이터를 이용하여 분류학습의 성능을 향상시킬 수 있는 준감독자(semi-supervised)학습 방법이 최근 관심을 끌고 있다. 본 논문에서는 미분류 데이터가 분류학습의 성능향상에 미치는 영향을 분석하기 위하여 나이프 베이지안의 환경에서 미분류 데이터를 이용한 학습방법을 제시하고 이를 이용하여 미분류 데이터의 효용성을 실험적으로 조사하였다. 미분류 데이터는 나이프 베이지안의 환경에서 분류데이터의 숫자가 적을 때 특히 많은 효과를 보임을 알 수 있었다.

키워드 : 기계학습, 준감독자 학습, 인공지능

Improving the Classification Accuracy Using Unlabeled Data : A Naive Bayesian Case

Lee, Chang-Hwan[†]

ABSTRACT

In many applications, an enormous amount of unlabeled data is available with little cost. Therefore, it is natural to ask whether we can take advantage of these unlabeled data in classification learning. In this paper, we analyzed the role of unlabeled data in the context of naive Bayesian learning. Experimental results show that including unlabeled data as part of training data can significantly improve the performance of classification accuracy. The effect of using unlabeled data is especially important in case labeled data are sparse.

Key Words : Machine Learning, Semi-supervised Learning, Artificial Intelligence

1. 서 론

현재 감독자학습(supervised learning)방법[1]에서 분류데이터의 샘플 크기가 적은 경우에는 감독자학습방법의 알고리즘이 좋은 정확도를 유지하기가 힘든 경우가 많다. 따라서 감독자 학습 방법에서는 만족할만한 분류의 성능을 얻기 위하여 많은 양의 분류(labeled) 데이터를 필요로 한다.

하지만 많은 경우에 분류데이터를 생성하기해서는 사람들의 노력이 필요하며 따라서 많은 비용과 시간을 요구한다. 예를 들어 문서를 분류하는 응용이나 패턴 인식 등에 있어서 데이터의 분류는 일일이 사람의 판단에 의해서 생성되며 따라서 많은 양의 분류데이터 생성을 위해서는 많은 양의 시간과 비용이 발생한다. 또한 분류의 기준이 바뀌게 되면 이전의 분류작업에 관계없이 다시 전체를 분류를 해야 하는 비효율성도 존재한다.

이에 반하여 미분류(unlabeled) 데이터는 거의 비용을 들이지 않고 무제한의 데이터를 쉽게 획득할 수 있다. 예를 들어 웹 페이지 분류 등의 응용에서 미분류 웹페이지 데이터는 비용을 들이지 않고 거의 무한대로 획득할 수 있다. 이와 같이 분류데이터의 숫자를 최소한으로 유지하는 상황은 흔하게 발생할 수 있는 사항이다. 따라서 기계학습에 있어서는 이러한 미분류 데이터를 이용하여 분류학습의 성능을 향상시킬 수 있는 방법들이 최근 관심을 끌고 있다.

이런 이유로 본 논문은 나이프 베이지안(naive Bayesian)의 환경에서 미분류 데이터를 이용하여 분류학습의 성능을 향상시키는 방법을 실험적으로 제시하는 연구이다. 미분류 데이터는 분류값이 없기 때문에 누락(missing) 데이터로 간주할 수 있으며 이러한 미분류 데이터가 어떻게 분류 학습의 성능을 향상시키는데 사용될 수 있는지 의문을 품을 수 있다. 즉 얼핏 생각하면 미분류 데이터는 목표값이 없기 때문에 쓸모없는 데이터로 생각할 수 있다. 하지만 미분류 데이터는 목표값은 없지만 목표값을 제외한 다른 속성들의 확률분포에 대한 정보를 제공하고 있다. 예를 들어서 원본에

[†] 종신회원 : 동국대학교 정보통신학과 부교수
논문접수 : 2005년 11월 4일, 심사완료 : 2006년 5월 30일

이터가 특정한 확률분포를 따른다고 가정하면 이러한 분포의 파라미터들을 정확히 예측할 수 있으면 더욱 정확한 분류학습을 할 수 있으며 미분류 데이터는 목표값은 없지만 데이터분포에 대한 파라미터의 예측을 정확히 하는데 도움을 준다.

본 논문에서 이와 같은 미분류 데이터의 분류학습에서의 역할을 분석하고 나이브 베이지안 환경에서 미분류 데이터의 효용성을 실험을 통하여 분석하고자 한다.

2. 미분류 데이터의 분석

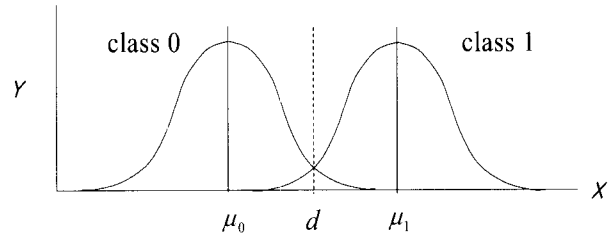
본 장에서는 미분류 데이터의 학습에서의 가치를 분석한다. 우선 분류데이터는 전혀 없이 미분류 데이터만 있는 상태에서의 학습은 임의추측(random guess)의 경우와 동일한 성능을 보인다[2, 3]. 즉 미분류 데이터만 있으면 전혀 학습을 할 수 없다는 의미이다. 하지만 이와 같은 사실이 미분류 데이터는 학습에 있어서 전혀 쓸모없는 데이터라는 결론을 의미하지 않는다.

미분류 데이터는 목적값을 제외한 나머지 값을 가지고 있기 때문에 목적값을 제외한 속성들의 결합 데이터분포(joint distribution)에 대한 정보를 가지고 있다[4]. 이러한 이유로 인하여 여러 분야에서 미분류 데이터는 분류 데이터와 같이 사용될 때에 분류학습의 정확도를 향상시키는데 사용될 수 있다[5, 6, 7, 8, 9]. 즉 미분류 데이터는 단독으로만 존재할 때는 의미가 없지만 분류데이터와 같이 존재할 때에 그 사용가치를 발하게 된다.

미분류 데이터의 효과분석을 위하여 다음의 가우시안 복합 모델(Gaussian mixture model)을 사용하여 설명 하고자 한다.

클래스의 값이 2개인 이진 분류(binary classification)의 경우를 가정하였으며 클래스의 값이 2개 이상인 일반적인 경우도 본 문제와 별 차이가 없다. 각 클래스의 값들은 (그림 1)과 같은 가우시안(Gaussian) 분포를 따른다고 가정하고 각 분포의 파라미터(평균값, 분산값 등)값은 알 수 없다고 가정한다. 이 경우에 최적의 분류자(optimal classifier)는 값 d 를 중심으로 구분하는 방법이다[1]. 즉 X 축의 값이 d 이하면 클래스 0을 의미하고 d 이상이면 클래스 1을 의미하는 분류방법이 주어진 상황에서 최적의 분류자이다. 이러한 가우시안 분포의 평균, 분산, 그리고 특정 클래스가 선택되는 선택 파라미터를 알게 되면 위의 d 값을 정확히 계산할 수 있게 된다.

(그림 1)에서 미분류 데이터가 무한대로 존재한다고 가정하면 미분류 데이터만으로 각 클래스의 분포들 모양을 정확히 결정할 수 있게 된다. 하지만 미분류 데이터가 무한개 있고 분류 데이터가 전혀 없으면 클래스의 분포는 정확히 알 수 있지만 어떤 분포가 클래스 0인지(혹은 1인지)를 알 수 없게 된다[10]. 따라서 데이터의 분포는 완전히 복원하더라도 실제 분류학습에는 도움이 되지 않게 된다. 따라서 이 경우에 분류된 데이터를 사용하여 어떠한 분포가 클래스 0



(그림 1) 가우시안 복합 모델

(혹은 1)에 속하는지를 정해주게 된다. 따라서 클래스의 값을 결정해주기 위해 충분한 분류된 데이터가 존재하면 미분류 데이터만으로 클래스 값의 분포를 정확히 복원할 수 있다.

3. 관련 연구

미분류 데이터를 처리하는 가장 널리 사용되는 방법은 EM (Expectation-Maximization) 방법[11, 12]이다. EM 알고리즘은 데이터가 특정한 파라미터들로 결정되는 분포를 따른다고 가정하고 임의의 파라미터로부터 출발한다. EM 알고리즘은 구체적으로 expectation 단계와 maximization로 구분되어 있는데 expectation 단계에서는 예측된 파라미터를 이용하여 미분류 데이터의 값을 계산하는 단계이고 maximization 단계에서는 예측된 값들을 이용하여 데이터분포를 결정하는 파라미터를 예측하는 단계이다. 이러한 과정을 반복하여 파라미터의 예측치들이 거의 혹은 전혀 변동이 없을 때까지 진행한다.

또한 최근에 널리 사용되는 미분류 데이터 사용방법은 협력학습(co-training)방법[13, 14, 15, 16]을 들 수 있다. 협력학습에서는 두개의 학습자를 사용하고 입력데이터의 속성을 독립된 두개의 그룹으로 분할하여서 각자 독립적으로 학습을 진행한다. 그 후에 미분류 데이터에 대하여 각각의 분류자는 미분류 값을 분류하고 분류된 값을 다른 분류자에게 제공하여 서로의 분류성능을 향상시키는 방법이다.

아울러 액티브 학습(active learning,[17, 18, 19, 20])에서는 알고리즘이 미분류 데이터를 선택하고 이의 정확한 값은 전문가로부터 제공받는다. 제공받은 값을 사용하여 알고리즘의 가설을 수정하고 이러한 작업을 반복하게 되는데 액티브 학습에서의 중요한점은 분류를 요청하게 되는 미분류 데이터의 선정방법에 있다.

4. 준 나이브 베이지안(semi-naive Bayesian) 학습 방법

본 논문에서는 미분류 데이터의 분류학습에 미치는 영향을 분석하기 위하여 나이브 베이지안 방법[1]을 학습의 기본 방법으로 사용한다.

본 논문에서 사용하는 학습 알고리즘의 내용을 설명하면 다음과 같다. 본 연구에서 사용할 방법은 나이브 베이지안 방법을 기반으로 하고 있으며 나이브 베이지안 학습 방법을

미분류 데이터의 처리를 위하여 수정한 준 나이브 베이지안 방법을 사용하였다. 본 학습 방법에서 사용하는 학습 데이터에서 a_i 를 i 번째 속성이라고 하고 v_j 를 해당 클래스값 이라고 하면 예측치 v 를 구하는 기본적인 나이브 베이지안 학습 방법은 다음과 같이 계산된다.

$$v = \operatorname{argmax}_{v_j \in \{1\}} P(v_j) \prod_i P(a_i|v_j)$$

위의 식에서 보는것과 같이 나이브 베이지안 학습을 위해서는 $P(v_j)$ 와 $P(a_i|v_j)$ 의 파라미터에 대한 예측이 필요하다. 클래스 값은 m 개의 값을 가지고 있고 각 속성은 p 개의 값으로 구성되어 있다면 전체적으로 $mp+m$ 개의 파라미터의 값을 예측하여야 한다. 나이브 베이지안 학습의 핵심은 이와 같이 위의 파라미터들의 예측이 주된 내용이다.

파라미터 $P(a_i|v_j)$ 의 예측치를 \tilde{P}_{ij} 라고 하고 $P(v_j)$ 의 예측치를 $\tilde{\alpha}_j$ 라고 하자. 이때 $N(a_i \wedge v_j)$ 를 전체 데이터에서 조건 $a_i \wedge v_j$ 를 만족하는 데이터의 개수라고 정의하고, N 을 전체 데이터의 개수라고 할 때 파라미터 예측치 \tilde{P}_{ij} 와 $\tilde{\alpha}_j$ 는 다음과 같이 정의된다.

$$\tilde{P}_{ij} = \frac{\tilde{P}(a_i \wedge v_j)}{\tilde{P}(v_j)} = \frac{N(a_i \wedge v_j)}{N(v_j)}, \quad \tilde{\alpha}_j = \frac{N(v_j)}{N}$$

아울러 본 연구의 방법에서는 파라미터의 예측치 계산을 위하여 라플라스 유연화(Laplace smoothing) 방법[16]을 사용하였다. 라플라스 유연화는 제로 확률의 발생을 방지하는데 목적이 있다. 목적속성의 값 갯수를 C 라고 할 때 본 연구에서 라플라스 유연화를 이용한 파라미터 예측을 위한 최종 식은 다음과 같다.

$$\tilde{P}_{ij} = \frac{N(a_i \wedge v_j)}{N(v_j)} \cong \frac{N(a_i \wedge v_j)}{N(v_j) + N}, \quad \tilde{\alpha}_j \cong \frac{N(v_j)}{N + C}$$

```

Input : L: labeled data
       U: unlabeled data

/* process labeled data */
calculate  $\tilde{P}_{ij}$  and  $\tilde{\alpha}_j$  using all data in L
while (no change in the values in  $\tilde{P}_{ij}$  and  $\tilde{\alpha}_j$ ) do
    /* process unlabeled data */
    for each data  $(a_1, \dots, a_t) \in U$  do
        calculate  $v$  satisfying  $v = \operatorname{MAX}_j \tilde{\alpha}_j \prod_i \tilde{P}_{ij}$ 
        assign label  $v$  to  $(a_1, \dots, a_t)$ 
    end-for
    calculate  $\tilde{P}_{ij}$  and  $\tilde{\alpha}_j$  using both L and U
end-do
    
```

(그림 2) 준 나이브 베이지안(semi-naive Bayesian) 분류 알고리즘

4.1 미분류 데이터의 처리 기능

본 연구에서는 나이브 베이지안 학습 알고리즘을 미분류 데이터의 처리를 위하여 기능을 수정하였다. 미분류 데이터를 처리하는 방법은 다음과 같다. 알고리즘은 우선 분류 데이터만을 사용하여 $P(v_j)$ 와 $P(a_i|v_j)$ 의 파라미터들을 예측한다. 그리고 여기서 예측된 파라미터를 이용한 나이브 베이지안 방법을 이용하여 미분류 데이터의 각각에 대하여 목적속성값을 예측한다. 그리고 미분류 데이터 각각을 예측된 목적속성값과 함께 분류데이터에 포함시켜서 다시 파라미터들을 예측한다. 이러한 과정을 반복적으로 수행하여서 파라미터의 변동이 없을때까지 수행하여서 미분류 데이터의 목적속성값을 최종적으로 결정한다. (그림 2)는 본 연구에서 사용한 준 나이브 베이지안 방법의 알고리즘을 기술한 내용이다.

5. 실험의 구성 및 결과

본 연구에서는 실험의 편의를 위하여 이진 데이터를 이용하여 실험하였다. 즉 목적속성을 포함한 각 속성들은 0/1의 값을 가지는 가상의 데이터를 이용하여 실험한다. 최적(optimal) 분류값의 계산을 가능하게 하기 위하여 (그림 3)의 알고리즘을 이용하여 데이터를 생성하였다. 데이터의 발생이 이진 나이브 베이지안 분포에 따라서 발생하기 때문에 최적의 분류기 설계가 가능해진다. 실험 데이터를 생성하는 알고리즘의 내용은 (그림 3)과 같다.

생성된 데이터에서 이론적인 최적의 분류정확도를 계산하는 방법은 (그림 4)의 방법으로 계산할 수 있는데 이는 생성된 데이터에 대하여 최적 분류기(optimal classifier)를 수행하였을 때 얻을 수 있는 최적의 분류정확도를 계산하는 방법을 기술하였다. 알고리즘의 내용은 최적의 분류기 알고리즘의 내용과도 유사하다. 본 실험에 사용된 데이터에서의 α , p_i , q_i 의 값은 다음과 같다.

```

Input :  $\alpha$ : class 0 의 값이 선택될 확률
        $p_i$ : 목적속성의 값이 0인 경우에  $i$ 번째의 속성값이 1일 확률
        $q_i$ : 목적속성의 값이 1인 경우에  $i$ 번째의 속성값이 1일 확률
       t: 속성의 개수, n: 전체 데이터의 개수, D: 데이터 집합
D := {};
repeat n times
    determine the class value  $v$  ( $v=0$  with probability  $\alpha$ )
    for j = 1 to t
        if  $v=0$  then
            determine the value of  $a_j$  ( $a_j=1$  with probability  $p_j$ )
        else if  $v=1$  then
            determine the value of  $a_j$  ( $a_j=1$  with probability  $q_j$ )
        end-if
    D := D  $\cup$   $(a_1, \dots, a_t, v)$ ;
end-repeat
return D
    
```

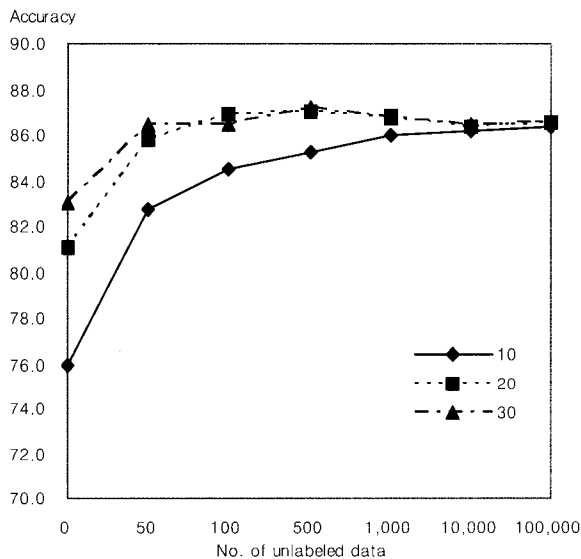
(그림 3) 실험 데이터의 생성방법

```

Input :  $\alpha$  : class 0 의 값이 선택될 확률
        $p$  : 목적속성의 값이 0 인 경우에  $i$  번째의 속성값이 1 일 확률
        $q$  : 목적속성의 값이 1 인 경우에  $i$  번째의 속성값이 1 일 확률
        $t$  : 속성의 개수,  $D$  : 전체 데이터의 집합

correct := incorrect := 0;
for each tuple  $(a_i, \dots, a_i, v) \in D$  do
  Sum_0 :=  $\alpha$ ; Sum_1 :=  $1-\alpha$ ;
  for  $i=1$  to  $t$ 
    if  $v=0$  then
      if  $a_i=1$  then  $\Delta_i = p$ , else  $\Delta_i = 1-p$ ;
      Sum_0 := Sum_0 *  $\Delta_i$ ;
    else if  $v=1$  then
      if  $a_i=1$  then  $\delta_i = q$ , else  $\delta_i = 1-q$ ;
      Sum_1 := Sum_1 *  $\delta_i$ ;
    end-if
  end-for
  if Sum_0 > Sum_1 then output:=0 else output:=1 end-if
  if output= $v$  then correct++ else incorrect++ end-if
end-for
return correct/(correct+incorrect);
    
```

(그림 4) 최적 분류 정확도의 계산 알고리즘



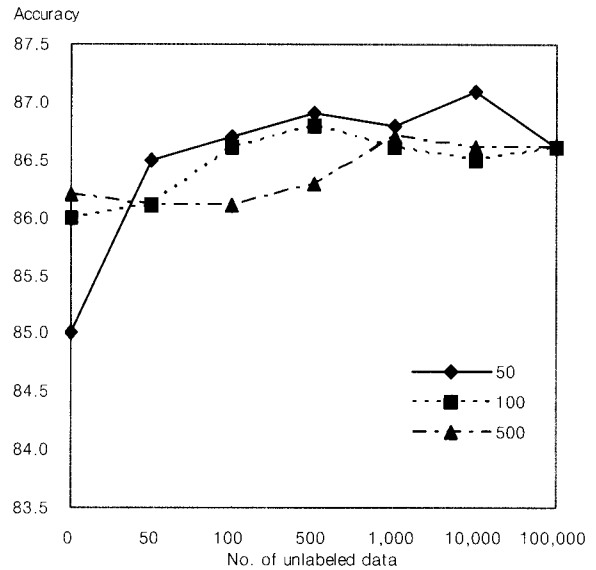
(그림 5) 미분류 데이터 개수에 따른 정확도의 변화(1)

$$\alpha = 0.5, p = (0.7, 0.3, 0.8, 0.3, 0.2, 0.1, 0.8),$$

$$q = (0.2, 0.6, 0.3, 0.3, 0.6, 0.7, 0.8)$$

이와 같은 실험 조건의 데이터에서 최적의 분류 정확도는 86.6%로 계산된다.

본 연구에서는 다양한 경우의 실험을 위하여 분류 데이터의 숫자를 10에서 10,000까지 9가지의 경우를 설정하였고 미분류 데이터는 0개에서 100,000개까지 7가지의 경우를 가정하였다. 따라서 전체적으로 63개의 실험 데이터 셋을 이용



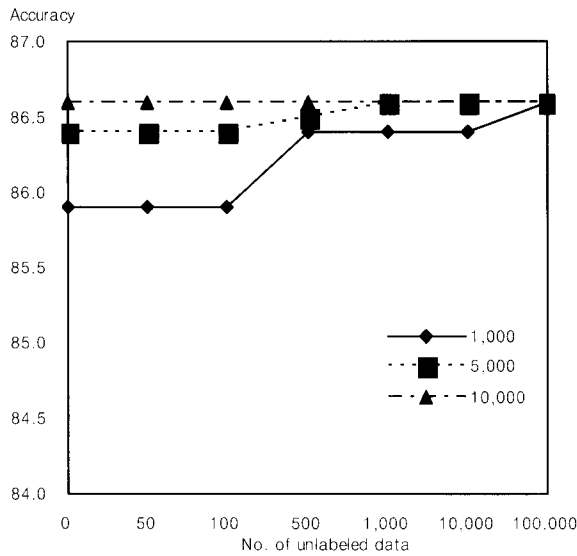
(그림 6) 미분류 데이터 개수에 따른 정확도의 변화(2)

하여 실험을 하였으며 정확도 측정을 위한 테스트 데이터는 학습데이터와는 별도로 1,000개의 데이터를 생성하여 정확도를 측정하였다.

먼저 (그림 5)는 분류 데이터의 개수가 10에서 30개의 경우에 미분류 데이터의 사용에 따라서 정확도의 변화내용을 보여주고 있다. 이 경우와 같이 분류 데이터의 개수가 작은 경우에는 실험결과에서 알 수 있듯이 미분류 데이터를 사용함에 따라서 정확도의 대폭적인 향상을 보여줌을 알 수 있다. 즉 분류 데이터의 개수에 관계없이 미분류 데이터의 개수가 10,000정도를 초과하게 되면 정확도는 이론적으로 얻을 수 있는 최적의 정확도에 근접함을 알 수 있었다. 이에 반하여 미분류 데이터를 사용하지 않는 나이브 베이저안의 학습의 경우에는 (그림 5)에서 미분류 데이터의 개수가 0인 경우를 의미하며 이 경우는 어떠한 경우에도 미분류 데이터를 사용하는 경우에 비하여 낮은 정확도를 보임을 알 수 있다.

(그림 6)은 분류 데이터의 개수가 50에서 500개의 경우에 미분류 데이터의 사용에 따라서 정확도가 변하는 내용을 보여주고 있다. 이 경우에도 (그림 5)와 같이 미분류 데이터를 사용함에 따라서 정확도의 향상을 보여줌을 알 수 있다. (그림 6)의 경우에는 분류 데이터의 숫자가 (그림 5)보다 많기 때문에 최적의 정확도에 좀더 일찍 수렴함을 알 수 있었다. 분류 데이터의 개수가 500인 경우에는 분류 데이터만으로 어느 정도의 정확도를 계산할 수 있었으며 따라서 미분류 데이터의 역할은 (그림 5)의 경우보다는 효과가 적음을 알 수 있다.

(그림 7)의 경우는 분류 데이터의 개수가 1,000에서 10,000개의 경우에 미분류 데이터의 갯수에 따라서 정확도가 변하는 내용을 보여주고 있다. 분류 데이터의 개수가 1,000개인 경우에는 미분류의 사용에 따라 정확도의 일부 향상을 가져오고 있지만 분류 데이터의 개수가 10,000개인 경우에는 이미 분류 데이터만으로 최적의 정확도에 도달하였기 때문에



(그림 7) 미분류 데이터 개수에 따른 정확도의 변화(3)

미분류 데이터를 추가하여도 정확도는 변화가 없음을 알 수 있었다.

이와 같이 본 실험은 대부분 경우에 미분류 데이터를 이용하면 정확도의 향상을 가져오며 미분류 데이터의 개수가 아주 많은 경우에는 초기 분류 데이터의 개수에 관계없이 최적의 정확도를 생성함을 알 수 있다. 또한 분류 데이터의 개수가 아주 작은 경우에는 다수의 미분류 데이터를 이용하여 정확도를 향상시킬 수 있었으며 따라서 미분류 데이터의 사용은 분류 데이터의 개수가 적을 때에 훨씬 효과적임을 알 수 있었다.

6. 결론

본 연구는 분류 학습에서 미분류 데이터의 사용이 학습의 정확도에 어떠한 영향을 미치는지를 나이브 베이지안 환경에서 검증하였다. 기존의 나이브 베이지안 학습 방법을 미분류 데이터의 처리를 위하여 수정한 준 나이브 베이지안 학습방법을 이용하였으며 실험을 위하여 다양한 갯수의 분류 데이터와 미분류 데이터 조합을 사용하여 실험을 진행하였다. 실험의 결과로서 거의 대부분의 경우에 미분류 데이터의 사용은 분류 학습의 성능을 향상 시키고 있음을 확인할 수 있었고 이러한 향상의 정도는 초기 학습에 사용된 분류 데이터의 개수가 적을 수록 더욱 많은 효과를 보여줌을 알 수가 있었다. 본 연구의 추후과제로서는 나이브 베이지안 외의 다른 학습 환경에 대해서도 미분류 데이터의 활용도를 검증하는 실험이 필요할 것이다.

참 고 문 헌

[1] R. Duda, et al. "Pattern Classification" 2nd edition. John

Wiley&Sons, 2001.

[2] T. Zhang "Some Asymptotic Results Concerning the Value of Unlabeled Data" NIPS 99 Workshop on Using Unlabeled Data for Supervised Learning, 1999.

[3] Y. Zhou and S. Goldman "Enhancing Supervised Learning with Unlabeled Data" 17th Int'l Conf. On Machine Learning, pp.327-334, 2000.

[4] Vittorio Castelli and Thomas M. Cover, "On the Exponential Value of Labeled Samples" Pattern Recognition Letters, Vol.16, pp.105-111, 1995.

[5] B. Shahshahani and D. Landgrebe "The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon" IEEE Trans. On Geoscience and Remote Sensing, Vol.32, No.5, 1087-1095, 1994.

[6] F. De Comite et al. "Positive and Unlabeled Examples Help Learning" Tenth Int'l Conf. on Algorithmic Learning Theory, pp.219-230, 1999.

[7] Sally Goldman and Yan Zhou "Enhancing Supervised Learning with Unlabeled Data" ICML, 2000.

[8] T. Hofmann "Text Categorization with Labeled and Unlabeled Data: A Generative Model Approach" NIPS 99 Workshop on Using Unlabeled Data for Supervised Learning, 1999.

[9] K. Nigam, A.K. McCallum, S. Thrun, and T. Mitchell "Text Classification from Labeled and Unlabeled Documents Using EM" Machine Learning, Vol.39, pp.103-134, 2000.

[10] T. Mitchell "The Role of Unlabeled Data in Supervised Learning" 6th Int'l Colloquium on Cognitive Science, 1999.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin "Maximum Likelihood from Incomplete Data via the EM Algorithm" Journal of Royal Statistical Society, Vol.39, pp.1-38, 1977.

[12] Xing Yi; Changshui Zhang; Jingdong Wang, "Multi-view EM algorithm and its application to color image segmentation" IEEE International Conference on Multimedia and Expo., 2004.

[13] Avrim Blum and Tom Mitchell "Combining Labeled and Unlabeled Data with Co-Training", COLT, 1998.

[14] Kamal Nigam and Rayid Ghani "Analyzing the Effectiveness and Applicability of Co-training", CIKM, 2000.

[15] A. Levin and P. Viola and Y. Freund, "Unsupervised improvement of visual detectors using co-training" the Ninth IEEE International Conference on Computer Vision, 2003.

[16] Rong Yan and Naphade, M. "Multi-Modal Video Concept Extraction Using Co-Training" IEEE International Conference on Multimedia and Expo., 2005.

[17] D. Cohn et al "Active Learning with Statistical Models" Journal of Artificial Intelligence Research, Vol.4, pp.129-145, 1996.

- [18] R. Liere and P. Tadepalli "Active Learning with Committees for Text Categorization" 14th National Conf. on Artificial Intelligence, pp.591-596, 1997.
- [19] Tur, G., Schapire, R.E., and Hakkani-Tur, D. "Active learning for spoken language understanding" IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.
- [20] Riccardi, G. and Hakkani-Tur, D. "Active learning: theory and applications to automatic speech recognition" IEEE Transactions on Speech and Audio Processing, 2005.
- [21] Tom Mitchell, "Machine Learning" McGraw Hill, 1997



이 창 환

e-mail : chlee@dgu.ac.kr

1982년 서울대학교 계산통계학과(학사)

1988년 서울대학교 계산통계학과(석사)

1994년 University of Connecticut

공학박사

1982년~1987년 한국기계연구소

1994년~1995년 AT&T Bell Laboratories

1996년~현재 동국대학교 정보통신학과 부교수

관심분야 : 기계학습, 데이터마이닝