

지역 기반 분류기의 앙상블 학습

최성하⁰, 이병우¹, 양지훈¹, 김선호¹

LG전자 DM연구소⁰, 서강대학교 컴퓨터학과¹

shchoi@lge.com, elva1212@gmail.com, yangjh@sogang.ac.kr, shkim@lex.yonsei.ac.kr

Ensemble Learning of Region Based Classifiers

요약

기계학습에서 분류기들의 집합으로 구성된 앙상블 분류기는 단일 분류기에 비해 정확도가 높다는 것이 입증되었다. 본 논문에서는 새로운 앙상블 학습으로서 데이터의 지역 기반 분류기들의 앙상블 학습을 제시하여 기존의 앙상블 학습과의 비교를 통해 성능을 검증하고자 한다. 지역 기반 분류기의 앙상블 학습은 데이터의 분포가 지역에 따라 다르다는 점에 착안하여 학습 데이터를 분할하고 해당하는 지역에 기반을 둔 분류기들을 만들어 나간다. 이렇게 만들어진 분류기들로부터 지역에 따라 가중치를 둔 투표를 하여 앙상블 방법을 이끌어낸다. 본 논문에서 제시한 앙상블 분류기의 성능평가를 위해 UCI Machine Learning Repository에 있는 11개의 데이터 셋을 이용하여 단일 분류기와 기존의 앙상블 분류기인 배깅과 부스팅 등의 정확도를 비교하였다. 그 결과 기본 분류기로 나이브 베이즈와 SVM을 사용했을 때 새로운 앙상블 방법이 다른 방법보다 좋은 성능을 보이는 것을 알 수 있었다.

1. 서론

기계학습의 분야에서 앙상블 학습은 주어진 학습 데이터를 가장 잘 설명하는 가설 하나를 찾는 것이 아니라 가설들의 집합을 만들고 가설들의 집합들로 투표를 하여 예측하는 것이다[1]. 좀 더 형식적으로 표현하면, 가설들의 집합 $\{h_1, \dots, h_k\}$ 을 만들고 가중치 집합 $\{w_1, \dots, w_k\}$ 을 정하여 $H(X) = w_1 h_1(X) + \dots + w_k h_k(X)$ 인 분류기를 만드는 것을 앙상블 학습이라고 한다.

가장 잘 알려진 앙상블 학습으로 배깅(Bagging)[2]과 부스팅(Boosting)[3]이 있다. Breiman은 배깅이 결정트리와 같이 불안정한 분류기들의 성능을 향상시킨다고 하였다[2]. 반면에 배깅과 부스팅은 나이브 베이즈나 선형 판별식 분석을 사용하는 안정적인(Stable) 알고리즘에서는 큰 정확도 상승은 눈에 띄지 않는다고 밝혔다[4]. 또한 Dietterich도 결정트리를 포함하여 이보다 덜 표현적인 가설을 생성하는 알고리즘에서 부스팅이 추천된다고 하였다[5]. 그래서 대부분의 앙상블 학습과 관련된 논문들이 결정트리들의 앙상블에 초점을 맞추고 있다[6],[7],[8],[9].

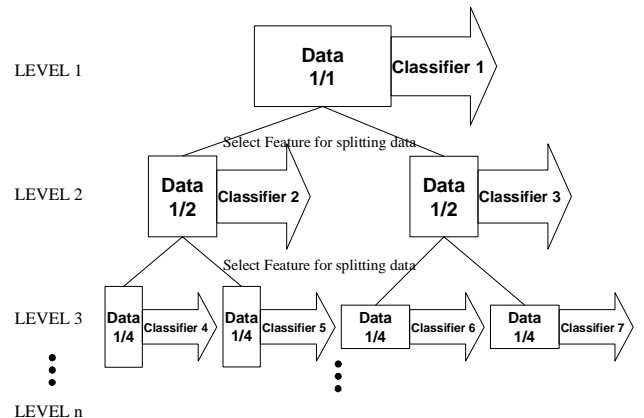
앙상블 분류기가 그것을 구성하는 각각의 단일 분류기보다 정확도가 더 높기 위해서는 각 분류기가 임의의 추측보다 정확해야 한다. 또한 입력 공간의 다른 부분에서 에러를 가져야 한다는 다양성을 만족해야 한다[10]. 이러한 기존의 앙상블 학습에 대해 다음과 같은 두 가지 문제점을 제시할 수 있다.

첫째, 기본 학습 알고리즘이 분류하기 어려운 데이터 일 경우 데이터를 재추출하더라도 역시 분류하기 어려울 수 있다. 이럴 때 배깅 방법을 사용할 경우 오분류율을

보이는 지역이 집중되며 이는 좋은 앙상블의 조건인 분류기의 다양성에 위배된다.

둘째, 각 지역에 따라 분포가 틀리지만 어떤 분류기가 그 지역에서 높은 정확도를 보이는지 알지 못한다. 예를 들어, 부스팅의 경우 선행되는 분류기가 잘못 분류하는 인스턴스에 대해 가중치를 두기 때문에 선행되는 분류기가 분류하지 못한 부분을 잘 분류하도록 학습된다. 하지만 실제로 분류할 때 그 부분이 어딘지 알지 못한다. 만약, 각 분류기 별로 잘 분류하는 인스턴스가 어떠한 것인지 안다면 보다 정확하게 분류할 수 있을 것이다.

2. 지역 기반 분류기의 앙상블 학습

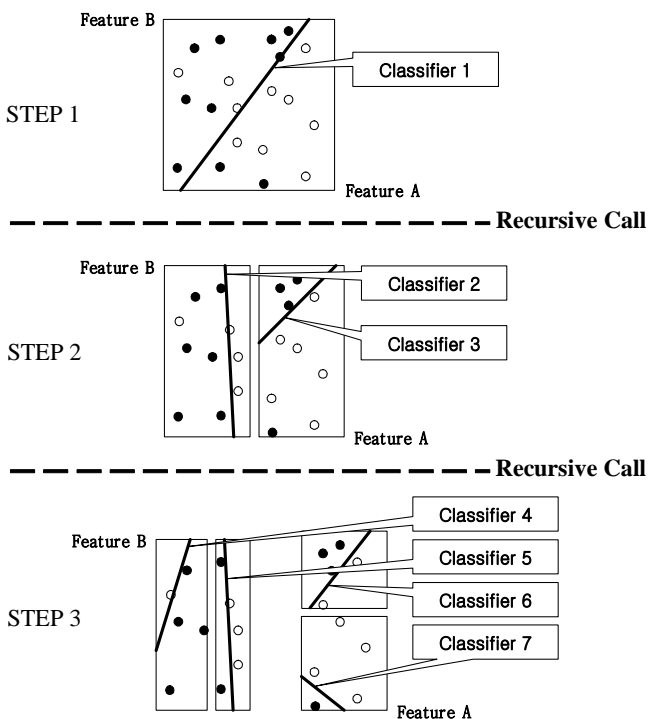


(그림 1) 지역 기반 분류기의 앙상블 학습에 대한 개괄도

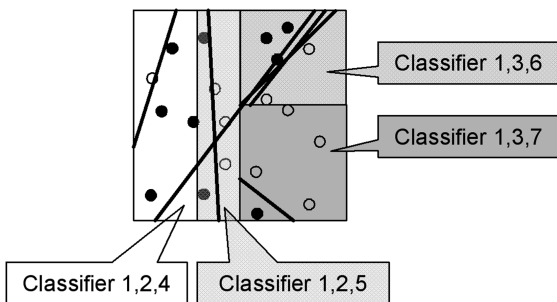
(그림 1)은 지역 기반 분류기의 앙상블 학습을 트리

형식으로 나타낸 것이다. 트리의 각 노드에 있는 분류기들을 지역 기반 분류기라 하고 특히 말단 노드에 있는 분류기들만을 사용하는 것을 RBC로 지칭한다. RBC는 입력 인스턴스가 속하는 말단 노드의 분류기 하나만이 분류에 사용된다. 그리고 이러한 모든 지역 기반 분류기들로 지역에 따라 가중치를 두어 여러 개의 분류기가 투표를 하는 앙상블 분류기를 지역 기반 분류기의 앙상블이라 하며 앞으로 RBE라 지칭할 것이다.

RBE는 학습알고리즘에 독립적인 앙상블 학습법(Meta learner)이며, 말단 노드뿐만 아니라 모든 노드에 분류기가 존재하며 결정트리를 구성해 나가는 것이 아니라 앙상블 분류기 생성을 위해 미리 정의된 분할 단계에 의해 데이터를 분할해 나간다는 특성을 갖는다.



(그림 2) RBE의 학습에 대한 구체적인 과정



(그림 3) RBE를 통한 분류

RBE의 학습은 입력으로 들어온 학습데이터로부터 학습을 통한 분류기 생성과 학습데이터의 분할을 반복하며 분류기를 생성시켜 나간다. 이러한 방식으로 여러 개의 분류기를 생성시킬 수 있다. RBE의 학습의 구체적인 절

차를 (그림 2)에 나타내었고 앙상블 분류기를 통한 분류는 (그림 3)에 나타내었다.

(그림 2)에서 보는 바와 같이 우선 전체 학습 데이터 D_0 을 선택되어진 학습 알고리즘으로 학습하여 분류기 C_0 을 생성시킨다. 그 후에 학습 데이터를 이등분한다. 여기서 이등분 하는 이유는 RBE의 목적이 앙상블 분류기를 생성하는 것이므로 한쪽에 학습 인스턴스가 물리는 것을 방지하기 위함이다. 여기서 데이터를 이등분하기 위해 각 속성에 대하여 이등분 되는 값을 중심으로 나누어 엔트로피를 계산하여 그 속성의 정보획득(Information Gain)[11]을 구한다. RBE에서는 각 속성별로 정보획득이 가장 큰 속성을 선택하여 데이터를 이등분하므로 클래스가 한쪽으로 치우치는 효과를 보게 된다. 이로 인해 데이터는 클래스별로 덜 섞인 분포를 갖게 되며 이는 결과적으로 데이터가 분할해 감에 따라 좀 더 분류하기 쉬울 것이라는 기대를 할 수 있다. 이렇게 선택된 속성을 기준으로 데이터를 이등분한다. 이렇게 이등분하여 나온 데이터에 위의 과정을 반복하여 분류기 C_1, C_2 을 생성해 나간다. 본 논문에서는 지역 분할의 단계를 3회로 설정하여 총 7개의 분류기(C_0, \dots, C_6)를 생성하였다. 새로운 패턴이 들어올 때 클래스를 결정하기 위하여 그 패턴이 속한 지역에 해당하는 분류기에 가중치를 두어 투표한다. 가중치를 두는 방법은 분류해야 할 인스턴스가 속하는 말단 노드와 그 직속 상위 노드들만으로 투표를 한다. 이러한 노드들을 제외한 나머지 노드들은 투표에 참여하지 않는다. 만약 투표에 참여하는 노드들이 짝수이어서 다수결로 동률이 되었을 때는 학습 데이터에 대해 더 높은 정확도를 갖는 말단노드의 결정에 따른다.

3. 실험 및 결과

3.1 실험 방법

본 논문에서 제시한 RBE가 학습 알고리즘에 독립적으로 작동하는 기법임을 보이기 위해 기본 학습 알고리즘으로서 나이브 베이즈 분류기뿐만 아니라, SVM을 구현한 SMO[12]와 C4.5[13]를 통한 결정트리를 사용하였다. 본 연구에서 SMO는 선형커널을 사용하였다. UCI Machine Learning Repository[14]에 있는 실제 데이터를 사용하였으며 실험은 11개의 데이터 셋에서 수행하였고 10-fold cross validation으로 검증하였다.

RBE의 성능을 검증하기 위해 RBE를 구성하는 분류기들을 생성시키는 데 사용된 기본 학습 알고리즘의 단일 분류기, 부스팅을 구현한 알고리즘인 Adaboost.M1[15]으로 7개의 분류기를 생성시킨 앙상블 분류기, 배경 기법으로 역시 7개의 분류기를 생성시킨 앙상블 분류기 그리고 2장에서 언급한 RBC를 비교 대상으로 하여 분류 정확도를 비교하였다. 본 논문에서 RBE로 학습하여 3레벨까지 데이터를 분할하여 7개의 분류기를 생성시켰기 때문에 공정한 비교를 위해 부스팅과 배경도 7개의 분류기를 생성하였다.

기본 학습 알고리즘은 공개 데이터마이닝 소프트웨어인 Weka3[16]을 사용하였고 기타 앙상블 기법에 사용된 프로그램은 자체 구현하였다.

<표 1> RBE와 각 방법들과의 정확도 비교 (Naive Bayes)

DATA SET	Single	vs RBE	Ada-7	vs RBE	Bagging	vs RBE	RBC	vs RBE
BALANCE-SCALE	89.92	88.63 X	89.76	88.63	89.60	88.63	88.00	88.63
GLASS	47.45	56.32 O	47.92	56.32 O	48.85	56.32 O	61.06	56.32 X
IONOSPHERE	82.91	89.75 O	91.76	89.75 X	82.91	89.75 O	89.21	89.75
IRIS	96.00	96.67	96.67	96.67	95.33	96.67 O	95.33	96.67 O
PIMA	75.62	76.15	75.88	76.15	75.75	76.15	72.37	76.15 O
SATIMAGE	79.60	80.75 O	79.85	80.75	79.70	80.75	78.30	80.75 O
SEGMENT	77.04	89.63 O	77.04	89.63 O	76.67	89.63 O	89.51	89.63
SPAMBASE	79.50	79.87	79.50	79.87	79.90	79.87	79.39	79.87
VEHICLE	46.28	69.83 O	46.28	69.83 O	46.52	69.83 O	67.21	69.83 O
WAVEFORM	80.02	82.66 O	80.02	82.66 O	80.06	82.66 O	80.68	82.66 O
WDBC	93.33	93.50	95.96	93.50 X	92.98	93.50	94.03	93.50

<표 2> RBE와 각 방법들과의 정확도 비교 (SMO with Linear Kernel)

DATA SET	Single	vs RBE	Ada-7	vs RBE	Bagging	vs RBE	RBC	vs RBE
BALANCE-SCALE	87.52	89.28 O	87.52	89.28 O	87.84	89.28 O	88.16	89.28 O
GLASS	53.94	61.51 O	58.14	61.51 O	53.92	61.51 O	59.24	61.51 O
IONOSPHERE	88.03	88.91	87.78	88.91 O	87.47	88.91 O	87.77	88.91 O
IRIS	96.67	95.33 X	98.00	95.33 X	96.00	95.33	96.00	95.33
PIMA	77.19	76.81	77.19	76.81	76.80	76.81	74.19	76.81 O
SATIMAGE	85.05	86.40 O	85.30	86.40 O	84.95	86.40 O	86.15	86.40
SEGMENT	92.72	92.59	92.72	92.59	93.09	92.59	93.46	92.59 X
SPAMBASE	90.39	91.87 O	90.83	91.87	91.14	91.87	91.76	91.87
VEHICLE	74.55	77.16 O	74.43	77.16 O	73.95	77.16 O	77.40	77.16
WAVEFORM	86.60	86.56	86.60	86.56	86.34	86.56	85.74	86.56
WDBC	97.89	97.54	97.89	97.54	97.89	97.54	95.97	97.54 O

<표 3> RBE와 각 방법들과의 정확도 비교 (C4.5)

DATA SET	Single	vs RBE	Ada-7	vs RBE	Bagging	vs RBE	RBC	vs RBE
BALANCE-SCALE	77.62	79.37 O	79.54	79.37	82.26	79.37 X	79.05	79.37
GLASS	69.39	70.82 O	74.70	70.82 X	74.70	70.82 X	65.71	70.82 O
IONOSPHERE	88.05	90.34 O	92.62	90.34 X	90.63	90.34	88.06	90.34 O
IRIS	94.67	94.67	94.00	94.67	95.33	94.67	93.33	94.67 O
PIMA	73.53	74.57	73.14	74.57 O	76.40	74.57 X	72.09	74.57 O
SATIMAGE	85.20	87.95 O	88.20	87.95	88.90	87.95	86.05	87.95 O
SEGMENT	96.17	97.04	97.90	97.04	96.17	97.04	95.80	97.04 O
SPAMBASE	93.05	94.31 O	95.26	94.31	94.35	94.31	93.00	94.31 O
VEHICLE	75.02	74.79	76.80	74.79 X	76.21	74.79 X	73.37	74.79 O
WAVEFORM	75.36	77.18 O	81.56	77.18 X	80.78	77.18 X	75.10	77.18 O
WDBC	92.80	94.03	96.49	94.03 X	94.56	94.03	93.68	94.03

3.2 실험 결과

<표 1>부터 <표 3>는 각 기본 학습알고리즘 별로 단일 분류기, 다른 앙상블 방법들(배깅, 부스팅) 그리고 RBC 각각에 대해 지역 기반 분류기의 앙상블(RBE)과 정확도 비교를 한 표이다. 정확도가 더 높은 곳에 기울임체로 표시하였고 그 차이가 두드러지게 RBE가 좋은 곳에는 O표시를 하였고 나쁜 곳에는 X표시를 하였다.

단일 분류기와의 정확도 비교에서는 기본 학습알고리즘에 상관없이 대체적으로 정확도가 높다는 것을 알 수 있다. 이는 RBE가 기본 학습 알고리즘에 독립적으로 앙상블 분류기로서의 역할을 하고 있음을 나타내고 있다.

다른 앙상블 방법인 배깅과 부스팅과의 정확도 비교를 살펴보자. 여기서 부스팅은 다중 클래스에 대해서도 적용된 AdaBoost.M1 알고리즘을 사용하였으며 부스팅, 배깅 모두 7번의 반복 파라미터를 적용하여 7개의 분류기

들을 생성하였다. 이는 지역 기반 분류기의 앙상블을 3레벨까지 분할하여 7개의 분류기를 생성하였기 때문이다. 표에서 보는 바와 같이 나이브 베이즈와 SVM을 기본 학습알고리즘으로 사용한 앙상블들의 비교에서는 RBE가 대체적으로 정확도가 높음을 알 수 있다. 특히 일부 데이터에 대해서는 정확도의 차이가 두드러지게 성능이 좋았으며, 그 외에도 다른 앙상블 방법보다 크게 성능이 떨어지는 경우는 거의 없었다. C4.5를 기본 학습 알고리즘으로 선택하였을 때는 부스팅과 배깅의 앙상블 기법이 RBE보다 성능이 대체적으로 좋은 것으로 나타났다. 이는 단일 분류기와의 실험에서 살펴본 것처럼 RBE도 정확도를 높였지만 부스팅과 배깅이 그보다 더 정확도를 높였기 때문에 이와 같은 결과가 나타난 것이다. 위의 실험 결과들을 종합하여 볼 때, 나이브 베이즈와 SVM과 같은 안정적인 알고리즘에서는 RBE가 성능이

더 좋았음을 알 수 있고 C4.5와 같은 불안정한 알고리즘에서는 기존의 앙상블 방법인 부스팅과 배깅이 더 좋다는 것을 알 수 있다.

이제 RBC와의 비교를 살펴보면 표에서 보는 바와 같이 기본 학습 알고리즘과 상관없이 대부분의 데이터 셋에서 RBE가 정확도가 더 높았으며, 일부 데이터에 대해서는 주목할 만한 정확도 차이를 보였다.

4. 결론 및 향후과제

지금까지 새로운 앙상블 방법으로서 지역 기반 분류기의 앙상블 학습을 제시하였다. 이 방법은 기존의 앙상블 방법과 비교했을 때 안정적인 알고리즘(Stable Algorithm)을 기본 학습기로 사용한 경우에 더 뛰어난 정확도를 보였다.

정보획득에 의해 속성을 선택하고 그 속성을 기준으로 데이터를 분할해 나가면서 데이터의 일부만 보고 학습을 하기 때문에 분류기의 다양성 측면에서 다른 앙상블 방법보다 더 좋을 것이라고 기대된다. 이는 좋은 앙상블의 조건 중에 하나인 분류기 간에 서로 독립적으로 입력공간의 다른 곳에서 에러를 가져야 한다는 것과 관련되어 있다. 또한 데이터를 분할하여 학습을 하면 그 지역을 제외한 다른 데이터에 대해서는 성능이 떨어지기 때문에 그 분류기가 정확하게 분류할 수 있는 지역에만 투표를 함으로써 다른 앙상블 방법보다 더 좋을 것이라고 기대된다. 이는 또 다른 좋은 앙상블의 조건은 각 분류기는 정확해야 한다는 것과 관련되어 있다.

위와 같은 두 가지 특성으로 인해 지역 기반 분류기의 앙상블 학습이 SVM과 나이브 베이즈와 같은 안정적인 알고리즘에서 다른 앙상블 방법보다 더 높은 성능 증가를 가져온 것으로 생각된다. 이러한 방법을 가능하게 하는 것은 각 분류기가 어느 인스턴스에서 높은 성능을 갖는지 알기 때문에 데이터에 큰 변화를 줄 수 있는 것이다. 이처럼 지역 분할 앙상블 학습은 기본 분류 알고리즘의 성질에 좀 더 독립적으로 작동할 수 있다는 점에서 그 의미가 있다고 할 수 있다.

본 논문에서는 RBE가 데이터 분할을 3단계까지만 진행하여 각 말단 노드의 깊이가 같은 3단계의 전 이진 트리(Complete Binary Tree)를 구성하도록 하였다. 이 트리는 총 7개의 분류기를 생성한다. 이를 확장하여 더 높은 분할 단계에서 성능이 어떻게 증가하는지 살펴 볼 필요가 있고 이산 값을 갖는 데이터에 대해서도 데이터를 분할할 수 있는 기준을 마련해야 한다. 이는 향후과제로 남겨둔다.

5. 참고문헌

[1] Dietterich, T.: Ensemble method in Machine learning. In J. Kittler and F. Roli (Ed.) First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, 1-15, 2000.
 [2] Breiman, L.: Bagging Predictors. Machine Learning, 24(2), 123-140, 1996.

[3]. Freund, Y., & Schapire, R.: Experiments with a new boosting algorithm. In Proc. of the 13th International Conference on Machine Learning, 148-156, 1996.
 [4] Breiman, L.: Bias, Variance, and Arcing Classifiers. Technical Report TR, 460, UC Berkeley, 1996.
 [5] Dietterich, T.: Ensemble Learning. In The Handbook of Brain Theory and Neural Networks, Second edition. The MIT Press, 405-408, 2002.
 [6] Bauer, E., & Kohavi, R.: An Empirical Comparison of Voting Classification Algorithm: Bagging, Boosting, and Variants. Machine Learning, 36(1-2), 105-142, 1999.
 [7] Dietterich, T.: An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Machine Learning, 40(2), 139-157, 2000.
 [8] Friedman, J., Hastie, T., & Tibshirani, R.: Additive Logistic Regression: a Statistical View of Boosting. Annals of Statistics, 28(2), 337-374, 2000.
 [9] Quinlan, J.: Bagging, Boosting, and C4.5. In Proc. of the Thirteenth National Conference on Artificial Intelligence, 725-730, 1996.
 [10] Hansen, L., & Salamon, P.: Neural Network Ensembles, IEEE Transaction on Pattern Analysis and Machine Intelligence, 12, 993-1001, 1990.
 [11] Quinlan, J.: Induction of Decision Tree. Machine Learning, 1(1), 81-106, 1986.
 [12] Platt, J.: Fast Training of Support Vector Machines using Sequential Minimal Optimization, chapter 12, The MIT Press, 185-208, 1999.
 [13] Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
 [14] Blake, C., & Merz, C.: UCI Repository of Machine Learning <http://www.ics.uci.edu/~mlern/MLRepository.html>, 1998.
 [15] Freund, Y., & Schapire, R.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Science, 55, 119-139, 1997.
 [16] Witten, I., & Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation, Second edition. Morgan Kaufmann, 2005.