Makine Teknolojileri Elektronik Dergisi
Cilt: 8, No: 4, 2011 (75-80)

Electronic Journal of Machine Technologies
Vol: 8, No: 4, 2011 (75-80)

TEKNOLOJİK ARAŞTIRMALAR

www.teknolojikarastirmalar.com
e-ISSN:1304-4141

Teknik Not
(Technical Note)

İstanbul Göztepe Bölgesinin Makine Öğrenmesi Yöntemi ile Rüzgâr Hızının Tahmin Edilmesi

Mustafa TİMUR[*], Fatih AYDIN[*], T.Çetin AKINCI[**]
[*]Kırklareli Üniversitesi Tek. Bil. M.Y.O., 39100 Kırklareli/TÜRKİYE
[**] Kırklareli Üniversitesi Mühendislik Fak., 39100 Kırklareli/TÜRKİYE
mustafatimur@kirklareli.edu.tr
Geliş Tarihi: 04.08.2011 Kabul Tarihi: 20.12.2011

Özet

Rüzgâr ölçümleri, birçok alan için gereklidir. Bunlara örnek olarak meteoroloji, iklim, tarım, endüstriyel uygulamalar ve bilimsel araştırmalar verilebilir. Ölçümlerdeki çok ufak gibi görünen bir fark bile yatırımın ekonomikliğini etkileyebilmektedir ve planlanan yatırım için risk getirebilmektedir. Bu amaçla yapılan çalışmaların ileriye yönelik ve çok hassas dengeler üzerine kurulması gerekmektedir.

Bu çalışmada makine öğrenmesi yöntemi kullanılarak İstanbul Göztepe bölgesi için rüzgâr tahminleri yapılmıştır. Tahminlerin yapılması için Bagging metodu esas alınarak, sınıflandırma işlemi için REPTree öğrenme ağacı kullanılmaktadır. Bagging sınıflandırıcının parametrelerindeki değişiklik ile diğer sınıflandırıcılara göre en yüksek öğrenmeyi gerçekleştirmektedir. Yapılan çalışmada korelasyon katsayısı 0.9114 ve Root mean squared error değeri 0.6591 olarak elde edilmiştir. Burada korelasyon katsayısının 0.5 ile 1.0 arasında olması aralarında yüksek bir ilişki olduğunu göstermektedir. Ayrıca root mean squared error değerinin 0'a yakın olması çok önemli hataların yapılmadığını göstermektedir.

Anahtar Kelimeler: Makine Öğrenmesi, Bagging, REPTree, Rüzgâr Hızı

1. GİRİŞ

Bilgi teknolojilerindeki gelişmeler sayesinde artık çok büyük miktarlarda veriyi kaydedebilmekteyiz. Çok büyük miktardaki verilerin manuel olarak işlenmesi, analizlerinin yapılması mümkün değildir. Bu problemlere çözüm bulmak amacıyla makine öğrenmesi metotları geliştirilmiş ve bu yöntemler geliştirilmeye devam etmektedir. Makine öğrenmesi metotları önceki verileri kullanarak veriye en uygun modeli bulmaya çalışır. Yeni gelen verileri de bu metoda göre analiz eder. Farklı uygulamaların analizlerden farklı beklentileri olmaktadır. Makine öğrenmesi metotlarını bu beklentilere göre sınıflandırmak mümkündür [1].

Bu makaleye atıf yapmak için
Timur M[*]., Aydın F[*]., Akıncı T.Ç[**]., "İstanbul Göztepe Bölgesinin Makine Öğrenmesi Yöntemi ile Rüzgâr Hızının Tahmin Edilmesi" Makine Teknolojileri Elektronik Dergisi 2011, 8(4) 75-80

How to cite this article
Timur M[*]., Aydın F[*]., Akıncı T.Ç[*]., "The Prediction Of Wind Speed Of Goztepe District Of Istanbul Via Machine Learning Method" Electronic Journal of Machine Technologies, 2011, 8(4) 75-80

2. MALZEME ve METOT

2.1 Makine Öğrenmesi

Veri üzerinde örüntü bulma bilgi keşfi için önemli bir süreçtir. Bilgi keşfi süreci, veri madenciliği olarak da adlandırılır [2] diğer bir ifadeyle veri üzerinde makine öğrenmesi metotlarının uygulanması veri madenciliği olarak adlandırılır [1]. Makine öğrenmesi, bilgi keşfi sırasında kullanılan tümevarımsal algoritmaların uygulanması sürecini tanımlamak için çok yaygın bir biçimde kullanılan bilimsel bir çalışma alanıdır [2].

2.2 Bagging

Makine öğrenmesi alanında, tüm eğitim verileri üzerinde yeterli öğrenme gerçekleştirebilen tek bir öğrenme algoritması mevcut değildir. Bundan dolayı algoritma seçimi deneme yanılma yoluyla yapılır [4]. Aynı zamanda algoritmaların eğitim verisi üzerinde oluşturdukları model ya da hipotez eğitim verisine bağlı olarak da değişmektedir. Bu durum en iyi modelin seçimi konusunu gündeme getirmektedir.

Bagging [5] metodu, bir eğitim verisinin farklı kombinasyonları oluşturularak elde edilen zayıf eğitim verilerinin temel-öğreniciler tarafından öğrenilmesi sonucu oluşan modellerin sonuçlarının karıştırılması yöntemine dayanır. Bu anlamda bagging bir oylama metodudur [1]. Bagging'de eğitim verisinin farklı kombinasyonlarının oluşturulma süreci bootstrap [6] metoduna dayanır. Bu metod cross-validation'a [1, 2, 3] benzemektedir ve onun bir alternatifidir. Amaç, bir tek örnekten birden fazla örnek oluşturmaktır. Bunu yaparken orijinal örnekten yer değiştirme ile yeni örnekler oluşturmaktır. Bootsrap örnekleri cross-validation örneklerinden çoğu ile örtüşebilir [1]. Bundan dolayı onların tahminleri birbirlerine bağımlıdır. Bootstrap'da, N adet gözlemden oluşan veri kümesi yer değiştirilerek 1/N kadar olasılıkla bootstrap örnek veri kümesi oluşturulur [7]. Bu örnek kümelerinin sayısı N adettir.

Weka'da [9] bagging tekniği meta.bagging olarak adlandırılmaktadır. Bagging tekniği bir öğrenme ağacı algoritması olan REPTree [10] ile birlikte kullanıldı.

3. İSTATİSTİKSEL DEĞERLENDİRME KRİTERİ

3.1 Korelasyon Katsayısı

Korelasyon Katsayısı (Correlation Coefficient) gerçek değer ile tahmini değer arasındaki istatistiksel ilişkiyi ölçer. CC değeri -1 ve +1 arasında değişir. CC'nin pozitif değeri iki ilişkinin birbirleriyle aynı yönde olduklarını belirtir. Negatif değerler ise ilişkinin zıt yönde olduğunu gösterir. Eğer CC değeri sıfır ise, o zaman iki değer arasında herhangi bir ilişki olmadığı söylenebilir. CC değeri için detaylı açıklama Cohen [11] tarafından Tablo 1'de gösterilmektedir.

Tablo 1. Cohen'in Korelasyon Tablosu

Korelasyon	Negatif	Pozitif
Düşük	-0.29 / -0.10	0.10 / 0.29
Orta	-0.49 / -0.30	0.30 / 0.49
Yüksek	-0.50 / -1.00	0.50 / 1.00

3.2 Ortalama Karesel Hataların Karekökü

Bir öğrenme algoritmasının yaptığı tahminler ile gerçek değerleri arasındaki farkın karesinin ortalaması ortalama karesel hatayı (Mean Squared Error) verir. MSE değerinin karekökünün alınmasıyla ortalama karesel hatanın karekökü (Root Mean Squared Error) hesaplanır.

MSE ve RMSE'nin hesaplanmaları aşağıda gösterilmektedir. Buna göre p, yapılan tahmini değerleri; a ise gerçek değerleri ifade etmektedir.

$$MSE = \frac{(p_1 - a_1)^2 + ... + (p_n - a_n)^2}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{(p_1 - a_1)^2 + ... + (p_n - a_n)^2}{n}} \quad (2)$$

$$MSE = Var(noise) + bias^2 + Var(p_i) \quad (3)$$

(1) eşitliğinden de görüleceği gibi yüksek oranda bias ve varyans MSE'nin değerini arttırmaktadır [12].

4. SINIFLANDIRMA SONUÇLARI VE TARTIŞMA

Bir meta sınıflandırıcı olan Bagging'in öğrenme sürecinde kullanılan eğitim verilerinin sayısı 7400'dür. Eğitim verilerinin giriş nitelikleri olarak tarih, sıcaklık ve basınç seçildi. Ayrıca yapılan ölçümler saatlik olarak yapılmaktadır. Yani 1 gün içerisinde 24 adet ölçüm sonucu kaydedilmektedir.

Farklı öğrenme tekniklerini kullanan sınıflandırıcıların eğitim sonundaki sonuçları Tablo 2'de gösterilmektedir. Tablo 2'de seçilen sınıflandırıcılar ile bagging sınıflandırıcı arasındaki fark açık bir biçimde görülmektedir. Cross validation için seçilen k değeri 5 ve 10 olarak belirlendi.

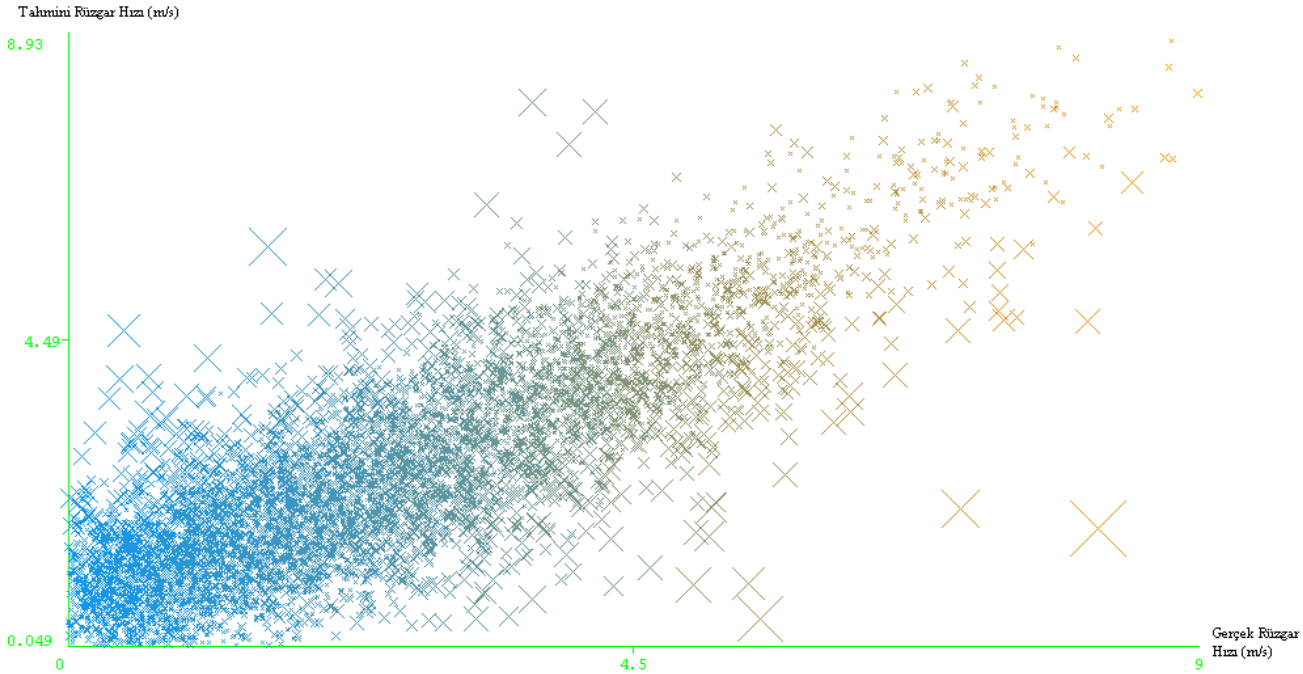
Tablo 2. Sınıflandırma algoritmalarının aldıkları parametrelere göre hata oranı ve korelasyon değerlerinin karşılaştırılması

Sınıflandırıcılar	k değeri / Diğer Parametreler	RMSE	CC
Linear Regression	k=10 / default	1.5616	0.2208
Linear Regression	k=5 / default	1.563	0.2171
kNN	k=10 / default	1.8671	0.2817
kNN	k=5 / default	1.871	0.2841
Bagging	k=10 / default	0.7382	0.8895

Bagging	k=5 / default	0.7653	0.8809
Bagging	k = 10 / REPTree { minNum = 1.0, noPruning = True, numFolds = 5}	0.6591	0.9114
DecisionTable	k=10 / default	1.3462	0.5441
DecisionTable	k = 5 / default	1.3509	0.5402
REPTree	k = 10 / default	0.8981	0.8302
REPTree	k = 5 / default	0.926	0.8191
REPTree	k = 10 / minNum = 1.0, noPruning = True, numFolds = 5	0.8306	0.8653

Tablo 2'teki sınıflandırma algoritmaları incelendiği zaman en iyi sonucu Bagging sınıflandırıcının verdiği görülmektedir. Bagging sınıflandırıcı sınıflandırma işlemi için temel öğrenici olarak REPTree öğrenme ağacını kullanmaktadır. Bagging sınıflandırıcının parametrelerindeki değişiklik ile birlikte diğer sınıflandırıcılara göre en yüksek öğrenmeyi gerçekleştirmektedir. CC değeri 0.9114 ve RMSE değeri 0.6591 olarak elde edilmiştir. Bu sonuçlara en yakın sonuç REPTree öğrenme ağacının parametrelerindeki değişiklik sonucu elde edilen sonuçlardır. Bu sonuçlara göre CC değeri 0.8653 ve RMSE değeri 0.8306 olarak elde edilmiştir. Yalnız Bagging sınıflandırıcı, temel sınıflandırıcı olarak REPTree'yi kullanmasına rağmen yöntem olarak bootstrapping işlemi yapması ve her bootstrap için elde etmiş olduğu alt sonuçların bir karışımını sonuç olarak sunmasından dolayı en iyi sonucu vermiştir. Diğer bir nokta ise k değerinin 5 ya da 10 yapılması sonucu çok değiştirmemektedir. Bazı sınıflandırıcılarda k=5 olması olumlu sonuç verirken bazılarında ise kötü sonuç vermektedir. Burada önemli olan nokta cross validation kullanılarak eğitim verilerinin öğrenme için yeterli olup olmadığının belirlenmesidir. CC değerinin 0.9114 olması ve RMSE değerinin 0.6591 olması ile öğrenmenin yeterli olduğunu söyleyebiliriz. Çünkü CC değerinin 0.5 ile 1.0 arasında olması yüksek (high) bir ilişki olduğunu göstermektedir. Aynı zamanda RMSE değerinin 0'a yakın olması çok önemli hataların yapılmadığını göstermektedir.

Şekil 1'de gerçek rüzgâr hızı değeri ile tahmini rüzgâr hızı değeri arasındaki ilişki gösterilmektedir. Buna göre gerçek değer ile tahmini değer arasındaki fark arttıkça çarpı işaretinin büyüklüğü de artmaktadır. Şekil 1 incelendiği takdirde çok fazla büyük çarpı işaretinin olmadığı görülmektedir. Genel olarak küçük çarpı işaretlerinin yaygın olduğu görülmektedir. Bu durum RMSE değerinin küçük olmasının bir sonucudur.



Şekil 1. x-ekseni gerçek rüzgar hızı, y-ekseni ise tahmini rüzgar hızını göstermektedir. Çarpı işaretinin büyümesi gerçek değer ile tahmini değer arasındaki farkın büyük olduğunu göstermektedir.

5. SONUÇLAR

Bu çalışmada, makine öğrenmesi yöntemi ile anemometer cihazına bağlı kalınmadan da istenilen sonuçlara ulaşılacağı belirlenmiştir. Makine öğrenmesi yöntemlerinden olan Bagging metodu kullanılarak çeşitli değerlere göre rüzgâr hızı tahmin edilmiştir. Tahmin edilen rüzgar hızı ile anemometreden elde edilen rüzgar hızı arasında ki korelasyon katsayısı 0.9114'dur. Bu da gerçek değer ile tahmin edilen değer arasında mükemmel bir ilişki olduğunu göstermektedir. Aynı zamanda değerler arasında ki hata oranlarının sıfıra çok yakın olması (RMSE=0.6591) bu iki değer arasındaki farkın oldukça düşük olduğunun bir göstergesi olup sonucun oldukça başarılı olduğu anlamına gelmektedir. Sonuç olarak kullanılan yöntem ile maliyet ve zaman faktörü ortandan kaldırılmıştır. Aynı zamanda bagging metodunun iyi sonuçlar vermesi makine öğrenmesinin rüzgâr hızı tahminlerinde yaygın bir biçimde kullanılmasının önünü açacaktır.

6. KAYNAKLAR

1. Alpaydın, E., 2004, Introduction to Machine Learning, The MIT Press, Printed and bound in the United States of America. ISBN 0-262-01211-1.

2. Glossary of Terms, Machine Learning, 1998, 271-274.

3. Ian H.Witten, Eibe Frank., 2005, Data mining : practical machine learning tools and techniques – 2nd ed. p. cm. – Morgan Kaufmann series in data management systems. ISBN: 0-12-088407-0.

4. Amasyalı, M.F., 2008, Yeni Makine Öğrenmesi Metotları ve İlaç Tasarımına Uygulamaları, Thesis (Phd). Yıldız of Technical University.

5. Breiman, L., 1996, Bagging Predictors, Machine Learning, 24, 123–140.

6. Efron, B., Tibshirani, R., 1993, An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC.

7. Yakupoğlu, Ç., Atıl, H., 2006, A Study on Bootstrap Method and It's Application II. Confidence Interval, Hypothesis Testing and Regression Analysis with Bootstrap Method, Ege Üniv. Ziraat Fak. Derg., 43(2):63-72 ISSN 1018-8851.

8. Mitchell, T.M. 1980, The need for biases in learning generalizations. CBM-TR 5-110, Rutgers University, New Brunswick, NJ.

9. WEKA, Available: http://www.cs.waikato.ac.nz/ml/weka/ [Erişim Tarihi: 08 Mart 2011].

10. WEKA, Class REPTree, http://weka.sourceforge.net/doc/weka/classifiers/trees/REPTree.html [Erişim Tarihi: 01 Haziran 2011].

11. Cohen, J. 1988, Statistical power analysis for the behavioral sciences (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

12. Kılıçaslan, Y., Güner E.S., and Yıldırım, S., 2009, "Learning-based pronoun resolution for Turkish with a comparative evaluation" Computer Speech & Language Volume 23, Issue 3, Pages 311-331.