

Yavaşlık İlkesini Kullanarak Öznitelik Seçimi: Alakalı Öznitelik Analizi

Using Slowness Principle for Feature Selection: Relevant Feature Analysis

Hande Çelikkanat, Sinan Kalkan
KOVAN Araştırma Laboratuvarı, Bilgisayar Mühendisliği Bölümü
Orta Doğu Teknik Üniversitesi
{hande, skalkan}@ceng.metu.edu.tr

Özetçe —Bu çalışmada, yavaşlık ilkesini yeni bir öznitelik seçme tekniği geliştirmek için kullanmayı öneriyoruz. Yavaşlık ilkesi, gerçek hayattaki fiziksel varlıkların maruz kaldığı değişimlerin yavaş ve sürekli değişimler olduğunu öne sürmektedir. Bu nedenle, dünyayı anlamlandırabilmek için, sensörlerimize ulaşan düzensiz ve hızlı değişen sinyallerin işlenerek, yavaş değişen ve daha fazla anlam taşıyan, üst seviye sinyallere dönüştürülmesi gerekmektedir. Yavaşlık ilkesi Wiskott ve Sejnowski'nin önceki çalışmalarında, biyolojik olarak sağlam temellere dayanan ve gürbüz bir nesne tanımayaya olanak sağlayan bir görme sistemi oluşturmak için kullanılmıştır. Bu çalışmada, yavaşlık prensibinin, çok sayıda öznitelikli uzaylarda sınıflandırma yapılırken alakalı öznitelikleri seçmek için de kullanılabilirliğini önermekteyiz. Elde ettiğimiz ilk sonuçları, bu alandaki önde gelen algoritmalarından ReliefF öznitelik seçme yöntemi, ve ayrıca Temel Bileşen Analizinin öznitelik seçmek için özelleştirilmiş bir versiyonu ile karşılaştırıyoruz. Bildiğimiz kadarıyla, bu çalışma yavaşlık prensibinin öznitelik seçme ya da sınıflandırma amacıyla kullanıldığı ilk çalışmadır.

Anahtar Kelimeler—alakalı öznitelik seçimi, yavaş öznitelik analizi

Abstract—We propose a novel relevant feature selection technique which makes use of the slowness principle. The slowness principle holds that physical entities in real life are subject to slow and continuous changes. Therefore, to make sense of the world, highly erratic and fast-changing signals coming to our sensors must be processed in order to extract slow and more meaningful, high-level representations of the world. This principle has been successfully utilized in previous work of Wiskott and Sejnowski, in order to implement a biologically plausible vision architecture, which allows for robust object recognition. In this work, we propose that the same principle can be extended to distinguish relevant features in the classification of a high-dimensional space. We compare our initial results with state-of-the-art ReliefF feature selection method, as well a variant of Principle Component Analysis that has been modified for feature selection. To the best of our knowledge, this is the first application of the slowness principle for the sake of relevant feature selection or classification.

Keywords—relevant feature selection, slow feature analysis

I. GİRİŞ

Çok boyutlu uzaylarda öğrenmenin zorluğu iyi bilinen bir problemdir. Robotik, biyoinformatik, görüntü işleme gibi verinin doğal olarak çok boyutlu olduğu alanlarda, herhangi bir sınıflandırma probleminde konu ile alakalı olan özniteliklerden ziyade, konu ile ilgili değişim göstermedikleri için gürültü etkisi yapan öznitelik sayısının katlarca daha fazla olması sık rastlanan bir durumdur. Bu ilgisiz öznitelikler sınıflandırma

problemini zorlaştırarak optimal sınıf ayrımlarının bulunmasını pratikte imkansız hale getirebilirler. Örneğin, Yuruten ve dig.'nin çalışmasında [1], robotların sıfatları (isimlerin aksine) direk görüntü verilerinden öğrenmesinin, sağlıklı verilerinden öğrenmesine kıyasla daha zor olduğu gözlenmiştir. Bunun nedeni, görüntü verilerinde çok sayıda bulunan ilgisiz özniteliklerdir. Bu öznitelikler gürültü gibi davranarak öğrenme problemini zorlaştırırlar. Önceden bu ilgisiz öznitelikler ayıklandığı takdirde öğrenme çok daha yüksek bir performansla gerçekleşebilir. Dolayısıyla, alakalı ve alakasız özniteliklerin birbirlerinden ayırt edilmesi, çok boyutlu uzaylarda öğrenme çalışmalarında önemli bir yer hak etmektedir.

Bu çalışmada, çok sayıda öznitelikli uzaylarda sınıflandırma yapılırken, alakalı ve alakasız öznitelikleri (*ing.* feature) ayırt edebilmek amacıyla, yavaşlık ilkesini kullanmayı öneriyoruz. Yavaşlık ilkesi, görsel algı sisteminin nasıl çalıştığını açıklamak için kullanılan prensiplerden biridir [2]. Bu düşünce sistemine göre, izlediğimiz bir sahneyi anlamlandırmak aslında, retinamıza ulaşan çok sayıda ve hızla değişen sinyali işleyerek, bu sinyallere neden olan fiziksel olgulara ulaşmak demektir. İnsan gözüne ulaşan sinyaller, göz bebeğinin anlık genişliği ve ortamdaki ışık miktarı gibi faktörler nedeniyle zaman içerisinde hızlı ve düzensiz değişiklikler gösterirler. Oysa bu sinyallere neden olan fiziksel olgular, dünyanın fiziksel yasaları nedeniyle uzaysal ve zamansal bir süreklilik göstermesine bağlı olarak, çoğunlukla zaman içerisinde yavaş ve düzgün bir değişim gösterirler. Yavaşlık ilkesine göre, bir sahneyi anlamlandırmak için, düzensiz retinal sinyallerden bir şekilde bu fiziksel temelli, yavaş sinyalleri elde etmek gerekir, ve elde edilen bir sinyal ne kadar yavaşsa, dünya hakkında o kadar fazla bilgi içerir. İşte bu yavaşlık prensibi, bu çalışmada, çok öznitelikli uzayların sınıflandırılmasında, sınıflandırmayla alakalı özniteliklerin ilgisizlerden ayrılması ve bu yöntemle sınıflandırma performansının artırılması amacıyla kullanılmıştır. Yavaşlık ilkesi, çeşitli bilişsel alanlarda genel bir prensip olarak karşımıza çıksa da, sınıflandırma ve alakalı özniteliklerin ayrıştırılması amacıyla kullanılabileceği fikri literatürde ilk defa bu çalışmada önerilmektedir.

A. İlgili Çalışmalar

Alakalı özniteliklerin seçilmesi probleminde önde gelen yaklaşımlardan biri ReliefF metodu [3] olup, bu metod robotik öğrenme problemlerinde sıkça kullanılmaktadır (Örnek olarak bkz. [4]). Bu yöntem özniteliklerinin değerleri cinsinden birbirine yakın olan örneklerin sınıflarının aynı ya da farklı olmasına bakarak öznitelik ağırlıklarını güncelleme fikrine dayanır. Bu nedenle, sınıfları ayıran marjinleri maksimize

etmeye yönelik biri sayılır. Marjin-tabanlı yöntemler dışında, bilgi (*ing.* information) tabanlı, öznitelik ve sınıf arasındaki karşılıklı bilgiyi (*ing.* mutual information) maksimize etmeye yönelik yaklaşımlar (örn. mRMR [5]), öznitelik ve sınıf arasındaki korelasyona bakan yöntemler [6], ve tutarlılığı maksimize etmeye çalışan yöntemler [7] sayılabilir.

B. Bu Çalışmanın Katkıları

Bu çalışma, yavaşlık ilkesini alakalı öznitelikleri seçmek amacıyla kullanmayı öneren ilk çalışmadır. Yavaşlık ilkesi, Bölüm II'de incelediğimiz gibi, görsel algının temel prensiplerinden biridir, ve bu nedenle biyolojik açıdan gerçekçi ve kendi kendine örgütlenebilen görme sistemleri oluşturmak için kullanılabilir. Bununla birlikte, bu çok genel ilke, şimdiye kadar sadece görsel çalışmalar kapsamında kullanılmıştır. Bizim tezimiz, yavaşlık ilkesinin bundan çok daha genel geçer olduğu ve farklı bilişsel bağlamlarda karşımıza çıkabileceğidir. Nitekim, alakalı özniteliklerin seçimi ve sınıflandırma bağlamında incelediğimiz bu çalışmada ilk sonuçlar umut verici görünmektedir. Yavaşlık ilkesine bağlı ortaya çıkan alakalı öznitelikler, sistemde tümüyle kendi kendine örgütlenerek ortaya çıkmalarına rağmen, direk bu problem için tasarlanmış bir yaklaşım olan Relief yöntemiyle kıyaslanabilir bir başarıyı göstermektedir.

II. YAVAŞ ÖZİNTELİK ANALİZİ

A. Matematiksel Yaklaşım

Wiskott ve Sejnowski [2], yavaşlık ilkesinin, izlenen sahnedeki anlamlı değişikliklerin tespit edilmesinde kullanılabilirliğini göstermişlerdir. Yavaş Öznitelik Analizi (Slow Feature Analysis, SFA) ismi verilen bu yaklaşım, hızlı ve düzensiz değişen retinal sinyallerin, bunlara denk düşen en yavaş sinyallere çevrilmesinde, dolayısıyla sahneyi en iyi özetleyen sinyallerin bulunmasında kullanılmaktadır. Wiskott ve Sejnowski bunu bir optimizasyon problemi olarak ifade ederler:

I boyutlu bir girdi sinyalinin ($\vec{x}(t) = [x_1(t), x_2(t), \dots, x_I(t)]^T$) elimize ulaştığını varsayarsak, amaç $\vec{g}(x)$ olarak ifade edilen bir dizi girdi-çıkı fonksiyonu bularak, I boyutlu girdi sinyalini, J boyutlu çıktı sinyaline çevirmektir. $\vec{y}(t) := \vec{g}(\vec{x}(t))$ şeklinde ifade edilecek olan çıktı sinyalinin her bir bileşeni, mümkün olduğu kadar yavaş değişmeli, aynı zamanda da girdi hakkında olabildiği kadar fazla bilgi içermelidir. Bu durumda optimizasyon probleminin hedefi, $\Delta(y_j) := \langle \dot{y}_j^2 \rangle$, $\forall j \in \{1, \dots, J\}$ ifadesini minimize ederken, bir yandan da aşağıdaki kısıtları sağlamak olarak tanımlanmaktadır:

$$\langle y_j \rangle = 0 \text{ (ortalama = 0)} \quad (1)$$

$$\langle y_j^2 \rangle = 1 \text{ (kovaryans = 1)} \quad (2)$$

$$\forall j' < j : \langle y_{j'} y_j \rangle = 0 \text{ (ilintisizlik, ing. decorrelation)} \quad (3)$$

Üçgen parantezler zaman ortalamasını ifade etmektedir.

Kısıt 1, çözümleri dilemsel olarak 0 etrafına çekmekteyken, Kısıt 2'in amacı, $y_j(t) = \text{sabit}$ şeklindeki, bilgi değeri olmayan basit çözümden kaçmaktır. Kısıt 3 ise birden fazla çözümün birbirlerinin kopyaları olmamasını garantiler. Kısıt 3 aynı zamanda bir tümel sıralama (*ing.* total order) da getirmektedir: Çıktılar arasında indisi (j) küçük olanlar, büyük olanlara göre daha yavaş ve daha optimaldir.

Bu optimizasyon probleminin analitik olarak çözülmesi zor olduğundan, bir doğrusallaştırma (*ing.* linearization) tekniğiyle basitleştirilebilir. Bunun için, her bir çıktının, sonlu bir doğrusal olmayan fonksiyonlar kümesinin elemanlarının doğrusal kombinasyonları olduğu, yani $y_j(t) = g_j(\vec{x}(t)) := \vec{w}_j^T \vec{z}(t)$ şeklinde ifade edilebileceği varsayılabilir. Bu doğrusal olmayan $\vec{z}(t)$ fonksiyonları, girdi sinyalinin üzerine $\vec{h} = [h_1, \dots, h_K]$ şeklinde ifade edilebilecek bir girdi-çıkı fonksiyonları serisi uygulayarak, dolayısıyla girdi sinyalini genişleterek (*ing.* expanding) elde edilebilir. Bu genişletme işleminin ardından, problem genişletilmiş sinyal bileşenleri $z_k(t)$ 'ler üzerinde doğrusal kabul edilebilir. Bu teknik, doğrusal olmayan bir sınıflandırma problemini doğrusallaştırabilmek için girdiyi bir çekirdekle (*ing.* kernel) zenginleştirme tekniğinin aynısıdır.

Bu doğrusallaştırma aşamasından sonra problem, $\Delta(y_j) = \langle \dot{y}_j^2 \rangle = \vec{w}_j^T \langle \dot{\vec{z}} \dot{\vec{z}}^T \rangle \vec{w}_j$ ifadesini minimize edecek $\vec{w}_j = [w_{j1}, \dots, w_{jK}]^T$ ağırlık vektörlerini bulma problemine dönüşür.

h_k fonksiyonlarının, $\vec{z}(t)$ genişletilmiş sinyalinin ortalamasını 0 ve kovaryans matrisini birim matris yapacak şekilde seçildiğinden emin olabilirsek, kısıtlarımız aşağıdaki gibi olduğundan, ancak ve ancak ağırlık vektörlerinin bir birimdir küme (*ing.* orthonormal set) oluşturması durumunda sağlanabilir:

$$\langle y_j \rangle = \vec{w}_j^T \langle \dot{\vec{z}} \rangle = 0 \quad (4)$$

$$\langle y_j^2 \rangle = \vec{w}_j^T \langle \dot{\vec{z}} \dot{\vec{z}}^T \rangle \vec{w}_j = \vec{w}_j^T \vec{w}_j = 1 \quad (5)$$

$$\forall j' < j : \langle y_{j'} y_j \rangle = \vec{w}_{j'}^T \langle \dot{\vec{z}} \dot{\vec{z}}^T \rangle \vec{w}_j = \vec{w}_{j'}^T \vec{w}_j = 0 \quad (6)$$

Bu durumda $\langle \dot{\vec{z}} \dot{\vec{z}}^T \rangle$ matrisinin özvektörleri, bize bu koşulları sağlayan ağırlık vektörlerini verebilir. Eğer bu özvektörleri, en küçük özdeğerlere karşılık gelen özvektörler olarak seçersek, yani,

$$\vec{w}_j : \langle \dot{\vec{z}} \dot{\vec{z}}^T \rangle \vec{w}_j = \lambda_j \vec{w}_j \quad (7)$$

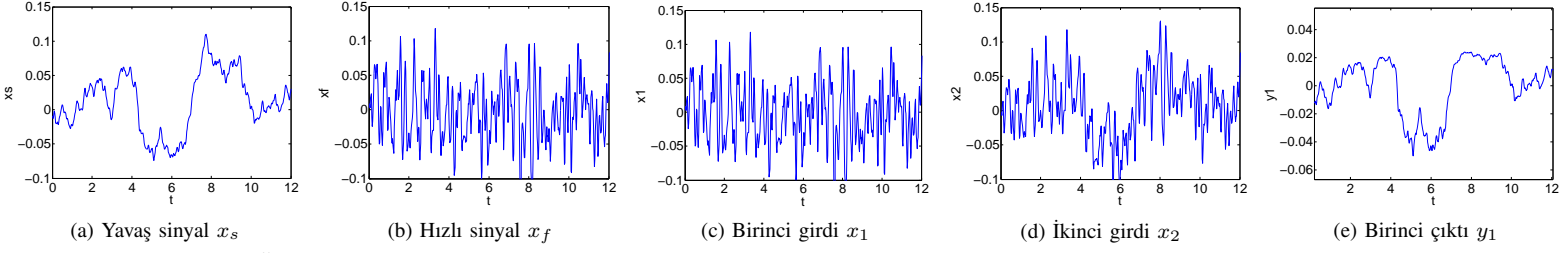
$$\text{with } \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_J \quad (8)$$

elde ettiğimiz girdi-çıkı fonksiyonları $g_j(x) = \vec{w}_j^T \vec{h}(x)$, aynı zamanda optimizasyon probleminin çözümü olan en yavaş sinyalleri de bize verir (bkz. Şekil 1).

Diğer bir deyişle, en yavaş sinyali bulmak için en küçük özdeğere karşılık gelen özvektör kullanılmalıdır, bu özvektör girdinin zaman türevinde en az varyansa sahip olan doğrultuyu belirtir. Bu doğrultu, girdi sinyalinin en az yavaş değiştiği doğrultudur. Birden fazla yavaş sinyal elde etmek için birdir doğrultular kullanılabilir. Bu diğer doğrultular gittikçe artan özdeğerlere karşılık gelen özvektörlerdir. Yani en yavaş J sinyalin ağırlık vektörlerini bulmak için, $\langle \dot{\vec{z}} \dot{\vec{z}}^T \rangle$ matrisine Temel Bileşen Analizi uygulamak ve en küçük J tane özdeğere karşılık gelen özvektörleri almak yeterlidir.

B. Yaklaşımın Avantajları

Ortaya çıkan sistem, sadece en yavaş çıktıları bulmayı garantilemekle kalmayıp, aynı zamanda da oldukça gürbüzdür. Franzius ve dig. [8], bu sistemi ilgin dönüşümlere (*affine transformations*) maruz kalan (hareket eden, dönen ve/ya ölçeği değişen) bir nesneyi bu dönüşümlerden bağımsız olarak tanıyabildiğini göstermişlerdir. Sistem ilgin dönüşümlerden bağımsız olarak nesnelere tanıyabildiği için, daha önce gördüğü ve görmediği nesnelere de birbirinden görüş açısından bağımsız olarak ayırt edilmektedir. Bunun dışında, birden fazla ilgin dönüşümün aynı anda uygulanmasından da etkilenmemektedir.



Şekil 1: Yavaş Öznitelik Analizinin çalışmasına bir örnek. Rastgele şekilde elde edilen biri yavaş, diğeri hızlı iki sinyal (a ve b), birbirine karıştırılarak iyi ayrı girdi sinyali (c ve d) elde edilmiştir. Bu iki girdi sisteme verildiğinde, sistem girdilerin altında yatan yavaş sinyali hızlı sinyalden ayırarak başarılı bir biçimde yeniden oluşturabilmektedir (e).

Temel olarak çıktılar, uygulanmakta olan bütün ilgin dönüşümleri, nesne kimliğine bağımlı bir şekilde kodlamaktadır.

Sistemin bir başka yararlı özelliği, birden fazla sayıda SFA modülünün hiyerarşik bir mimari oluşturacak şekilde düzenlenebiliyor olmasıdır. Bu durumda her modül, kendisinden bir alt seviyede görüş alanında olan girdileri işleyerek, çıktısını kendisinden bir üst seviyede olan ve görüş alanı içerisinde bulunduğu modüllere iletir. Bu hiyerarşik yapının çok sayıda avantajı bulunmaktadır. Öncelikle girdi matrislerinin büyüklüğünü sabit tutarak gerçek görüntülerle çalışabilmemize olanak sağlar. Bunun yanında her bir modül birbirinden bağımsız çalışabildiği için sistem paralel çalıştırılabilir. Son olarak da, bu hiyerarşik mimari, insan görme sisteminin biyolojik özelliklerine benzer bir şekilde kendini adapte eder [9]. Örneğin, alt seviyelerde kalan modüller, birincil görsel korteksin basit (*ing. simple*) hücreleri gibi pozisyona duyarlı yönelim seçiciliği göstermeye başlarken, üst seviyelerdeki modüller pozisyonundan bağımsız yönelim seçiciliği gösterecek şekilde adapte olurlar. Böylece yaklaşım, gerçek hayatta nesnelere tanımak için biyolojik olarak gerçekçi, matematiksel olarak analitik analizi mümkün [10] bir sistem ortaya koyar.

III. METOD

Bu çalışmada, yavaşlık ilkesini, çok öznitelikli uzayların sınıflandırılmasında, alakalı öznitelikleri seçmek için kullanmayı önermekteyiz. Bu yaklaşımdan makalenin devamında Alakalı Öznitelik Analizi (Relevant Feature Analysis, RFA) ismi ile bahsedilecektir.

Yavaş Öznitelik Analizi, orijinal halinde, girdi olarak bir sensörden, örneğin bir kameradan gelen, I boyutlu bir vektörü, zamana bağlı olarak her $t = t_i$ anında almaktadır. Bunun yerine, bu çalışmada, her $t = t_i$ anında, sisteme yeni bir örneğe ait öznitelik vektörü girdi olarak verilmektedir. Sisteme önce belli bir sınıfa ait olan örnekler gösterilir. Bu örnekler bittikten sonra diğer sınıfa ait olan örneklere geçilir. Bu gösterim sisteminde amaç, sistemin en yavaş sinyal olarak, sınıflar arasındaki geçişi algılamasını sağlamaktır. Nitekim, aşağıda bahsedilen CorrAL-100 veri kümesine ait örnekler sınıf sıralamasıyla sisteme verildiğinde Şekil 2'deki en yavaş çıktı sinyali elde edilmektedir. Bu sinyalin değeri örnekleri %100 başarıyla sınıflandırabilmektedir.

Söz konusu sınıflandırmayla alakalı öznitelikler, en yavaş sinyalin ağırlık vektörüne bakılarak ayrıştırılabilir. Bölüm II'de anlatıldığı üzere, bu sinyal, genişletilmiş girdi sinyalinin bir ağırlık vektörüyle çarpılmasıyla elde edilmektedir:

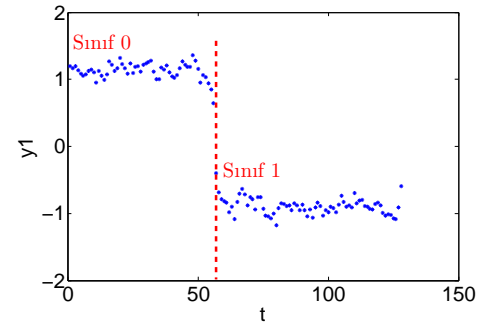
$$y_1(t) := \vec{w}_1^T \vec{z}(t) = w_{11} * z_1(t) + w_{12} * z_2(t) + \dots w_{1I} * z_I(t)$$

Alakalı Öznitelik Analizi şu şekilde çalışmaktadır: Girdi sinyalini doğrusal olmayan genişletme işleminden geçirmediğimizi varsayarsak¹, \vec{w}_1 ağırlık vektöründeki her bileşen, doğrudan $\vec{x}(t)$ girdisindeki kendisine karşılık gelen öznitelik ağırlığını, yani çıktı sinyaline ne kadar katkı vereceğini göstermektedir. Yani x_i özniteliklerinin önem derecesi, $abs(w_{1i})$ değeri ile doğru orantılıdır. Öznitelikler,

$$x_i \succ x_j \iff abs(w_{1i}) \succ abs(w_{1j}) \quad (9)$$

tümel sıralamasına göre dizilirse her öznitelik kendisinden sonra gelen özniteliklere göre daha alakalı olacaktır.

Problemimizi tanımlamak gerekirse, girdilerimiz sınıflandırılacak uzaydaki elemanlar iken, çıktı olarak bu uzaydaki özniteliklerden söz konusu sınıflandırma ile alakalı olanlarını almayı beklemekteyiz. Örnek olarak, deneylerde kullandığımız veri kümelerinden CorrAL-100 veri kümesinde [12], her biri 100 öznitelige sahip (x_1, \dots, x_{100}) 128 adet örnek bulunmaktadır. Örneklerden 56 tanesi 0, 72 tanesi 1 olarak sınıflandırılmıştır. Her örneğin ait olacağı sınıf, ilk dört özniteligine bakılarak, $(x_1 \oplus x_2) \wedge (x_3 \oplus x_4)$ (\oplus XOR operatörünü göstermektedir) formülüyle hesaplanmıştır. Beşinci öznitelik, x_5 , tümüyle rastgele hesaplanmıştır. Altıncı öznitelik x_6 ise örneğin sınıfıyla %75 ihtimalle aynı, %25 ihtimalle farklıdır. Geri kalan 94 öznitelik tamamen rastgele belirlenmiştir. Bu durumda iyi bir algoritmanın, 1.-4. öznitelikleri seçmesi gerekmektedir, çünkü bu öznitelikler örneğin sınıfını kesin olarak bilebilmek için yeterlidir.



Şekil 2: CorrAL-100 veri kümesindeki örnekler sisteme verildiğinde birinci çıktının zamanla değişimi. Bu sinyalin değeri örnekleri %100 başarıyla sınıflandırabilmektedir.

¹Bu çalışma özelinde elde ettiğimiz giriş seviye sonuçlarda, genişletilmiş, doğrusal olmayan bir sinyal hesaplamak yerine, girdi sinyalini değiştirmeden ($\vec{z}(t) = \vec{x}(t)$) kullanıyoruz. İlerki çalışmalarda girdi sinyalinin genişletilerek kullanılması ve bu yöntemle öznitelikler arasındaki doğrusal olmayan bağlanların da hesaba katılması planlanmaktadır.

Tablo I: 5 ayrı veri kümesinde ReliefF, Temel Bileşen Analizi (öznitelik seçme versiyonu [11]) ve Alakalı Öznitelik Analizi kullanılarak elde edilen öznitelikler, ve bu özellikler kullanılarak elde edilen SVM sınıflandırma başarımı. Her bir veri kümesinin tanımı gereği bulunması gereken alakalı özellikler ayrı bir sütunda gösterilmiştir. Alakalı öznitelikler ayrıca bütün sütunlarda kalın harflerle belirtilmiştir. 3 nokta (...) gösterimi söz konusu algoritmanın çıkardığı sıralamada öznitelikler arasında belirgin bir kesim noktası olabilecek doğal bir sınırın bulunmadığını göstermektedir.

Veri Kümesi İsmi	Alakalı Öznitelikler	ReliefF	Temel Bileşen Analizi	Alakalı Öznitelik Analizi
CorrAL	1,2,3,4	1,2,3,4 [SVM: 100%]	1,6,3,5 [SVM: 79%]	1,3,6,2,4 [SVM: 98%]
CorrAL-100	1,2,3,4	7,98,3,19,4,... [SVM: 54%]	75,52,67,... [SVM: 51%]	3,2,1,4 [SVM: 100%]
XOR-100	1,2	50,1,2 [SVM: 82%]	35,28,41,... [SVM: 30%]	21,19,1,... [SVM: 54 %]
Monk 3	2,4,5	2,4,5 [SVM: 93%]	4,2,3,6 [SVM: 78%]	2,5,6 [SVM: 93%]
Parity3+3	1,2,3,4,5,6	4,1,5,2,6,3 [SVM: 48%]	2,1,7,8,3,5,9... [SVM: 42%]	1,4,3,6 [SVM: 36%]

IV. DENEYSEL SONUÇLAR

Bu çalışmada önerdiğimiz öznitelik değerlendirme sisteminin ilk sonuçları sunulmaktadır. Bu ilk sonuçlarda 5 yapay veri kümesi üzerindeki öznitelik değerlendirme sonuçları, ReliefF [3] ve Temel Bileşen Analizi'nin öznitelik seçmek için özelleştirilmiş bir versiyonu [11] ile, ayrıca idealde bulunması gereken alakalı özniteliklerle karşılaştırılmıştır. Ayrıca, bu öznitelikler kullanılarak bir SVM sınıflandırması yapıldığında elde edilen sınıflandırma başarımı da gösterilmiştir.

Karşılaştırmalarda aşağıdaki veri kümeleri kullanılmıştır:

CorrAL veri kümesi [13] 6 öznitelikle tanımlanmaktadır (x_1, \dots, x_6). Her bir öznitelik 0 ya da 1 değerini alabilir. Yukarıda bahsedildiği gibi, sınıflar, örneklerin ilk dört öznitelikleriyle tanımlanmıştır: $(x_1 \oplus x_2) \wedge (x_3 \oplus x_4)$. Beşinci öznitelik x_5 tümüyle rastgeledir. Altıncı öznitelik x_6 ise örneğin sınıfıyla %75 ihtimalle aynı, %25 ihtimalle farklı değerdedir.

CorrAL-100 veri kümesi [12], CorrAL veri kümesine 94 rastgele öznitelik eklenerek oluşturulmuştur.

XOR-100 100 öznitelikle tanımlanmaktadır (x_1, \dots, x_{100}). Her bir öznitelik 0 ya da 1 değerini alabilir. Sınıflar ilk iki öznitelikle belirlenir: $(x_1 \oplus x_2)$. Diğer 98 öznitelik rastgeledir.

Monk 3 veri kümesi [14] yapay bir robotik senaryosuna dayanmaktadır. Her bir robot 6 öznitelikle tanımlanır (x_1, \dots, x_6). Sınıflandırma ikili olup, her örneğin sınıfı $(x_5 = 3 \vee x_4 = 1) \wedge (x_5 \neq 4 \vee x_2 \neq 3)$ ifadesiyle belirlenmektedir. 122 örnek arasından %5'i yanlış sınıflandırılarak veri kümesine yapay gürültü eklenmiştir.

Parity3+3 veri kümesi [15] 12 öznitelik içermektedir (x_1, \dots, x_{12}). Her bir öznitelik 0 ya da 1 değerini alabilir. Özniteliklerden ilk üçü arasında tek sayıda 1 varsa, örneğin sınıfı 1, değilse 0 olarak belirlenir. 4.-6. öznitelikler 1.-3. özniteliklerin kopyasıdır. 7.-12. öznitelikler rastgeledir.

Bu veri kümelerinin ReliefF, Temel Bileşen Analizi (öznitelik seçme versiyonu [11]) ve Yavaş Öznitelik Analizi ile incelenmesiyle seçilen öznitelikler, ve bu öznitelikler kullanıldığında elde edilen SVM başarısı Tablo I'de gösterilmiştir. Görüldüğü gibi, Alakalı Öznitelik Analizi, alakalı öznitelik seçiminde, ReliefF ile kıyaslanabilir (bazı veri kümeleri için daha kötü, bazı veri kümeleri için daha iyi) bir başarımla göstermekte, ve bütün denemelerde Temel Bileşen Analizi'nin alakalı öznitelik modifikasyonundan daha iyi çalışmaktadır.

V. SONUÇ

Bu çalışmada, yavaşlık ilkesinin ve yavaş öznitelik analizinin, çok boyutlu uzaylarda sınıflandırma yaparken, alakalı öznitelikleri ilgisiz olanlardan ayırt etmek için kullanılabilirliğini gösterdik. Bu yaklaşımı farklı veri kümelerinde, ReliefF

ve özvektörler üzerinde bir analiz gerçekleştirmesi bakımından bizim yöntemimize benzeyen Temel Bileşen Analizinin ilgili versiyonu ile karşılaştırdık. İlk sonuçlar, Alakalı Öznitelik Analizi (Relevant Feature Analysis, RFA) yönteminin, ReliefF ile kıyaslanabilir ve Temel Bileşen Analizi versiyonundan daha iyi bir başarımla gösterdiğini ortaya koymaktadır. Bu sonuçlar yavaşlık ilkesinin, sadece görsel algının değil, başka bilişsel problemlerin de altında yatan, incelenmeye değer prensiplerden biri olduğunu göstermesi açısından önemlidir.

VI. TEŞEKKÜR

Bu çalışma, 111E287 numaralı proje kapsamında TÜBİTAK tarafından desteklenmektedir.

KAYNAKÇA

- [1] O. Yürüten, E. Şahin, and S. Kalkan, "The learning of adjectives and nouns from affordance and appearance features," *Adaptive Behavior*, vol. 21, no. 6, pp. 437–451, 2013.
- [2] L. Wiskott and T. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comp.*, vol. 14, pp. 715–770, 2002.
- [3] I. Kononenko, "Estimating attributes: analysis and extensions of relief," in *ECML*, 1994, pp. 171–182.
- [4] E. Uğur and E. Şahin, "Traversability: A case study for learning and perceiving affordances in robots," *Adaptive Behavior*, vol. 18, no. 3-4, pp. 258–284, 2010.
- [5] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [6] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, 1999.
- [7] H. Almuallim and T. G. Dietterich, "Learning boolean concepts in the presence of many irrelevant features," *Artificial Intelligence*, vol. 69, no. 1, pp. 279–305, 1994.
- [8] M. Franzius, N. Wilbert, and L. Wiskott, "Invariant object recognition and pose estimation with slow feature analysis," *Neural Comp.*, vol. 23, pp. 2289–2323, 2011.
- [9] P. Berkes and L. Wiskott, "Slow feature analysis yields a rich repertoire of complex cell properties," *Journal of Vision*, vol. 5, pp. 579–602, 2005.
- [10] L. Wiskott, "Slow feature analysis: A theoretical analysis of optimal free responses," *Neural Comp.*, vol. 15, pp. 2147–2177, 2003.
- [11] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *ICSEM*, vol. 1, 2010, pp. 27–30.
- [12] G. Kim, Y. Kim, H. Lim, and H. Kim, "An mlp-based feature subset selection for hiv-1 protease cleavage site analysis," *Journal of Artificial Intelligence in Medicine*, vol. 48, pp. 83–89, 2010.
- [13] G. H. John, R. Kohavi, K. Pfleger et al., "Irrelevant features and the subset selection problem," in *ICML*, vol. 94, 1994, pp. 121–129.
- [14] S. B. Thrun et al., "The monk's problems: A performance comparison of different learning algorithms," 1991.
- [15] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and information systems*, vol. 34, no. 3, pp. 483–519, 2013.