

Classification of Open Directory Web Pages Using Artificial Neural Networks

Safak Kayikci (Corresponding author)
Bolu Abant İzzet Baysal University
Department of Computer Engineering, Bolu, Turkey
E-mail: safak.kayikci@ibu.edu.tr

Erhan Akyazi
Marmara University
Department of Informatics, Istanbul, Turkey
E-mail: eakyazi@marmara.edu.tr

Abstract

Users need get to access to the information in electronic environment quickly, correctly and appropriately, different methods of classification and categorization of data are strictly needed. In this study, multilayered perceptron artificial neural network model is used to classify the web sites in open directory project (DMOZ) according to the specified subjects. With the application developed, it is seen that artificial intelligence can be used for developing smart software on internet. Classification of web pages using neural networks enables to reach correct information quickly.

Keywords: Classification, Artificial Neural Networks

Açık Dizin Projesi Web İçeriklerinin Yapay Sinir Ağları ile Sınıflandırılması

Özet

Kullanıcıların elektronik ortamda sunulan bilgilere erişimlerindeki hız ve doğruluk gereksinimi nedeniyle, bu ortamlarda tutulan bilgileri sınıflandırma ve kategorilere ayırma yaklaşımlarına ihtiyaç duyulmaktadır. Bu çalışmada, açık dizin projesinde (DMOZ) bulunan web sayfalarının belirlenen konulara göre sınıflandırılabilmesi için, çok katmanlı yapay sinir ağı modeli kullanılmıştır. Geliştirilen örnek uygulama ile yapay zeka tekniklerinin internet üzerinde zeki yazılımlar oluşturmak için kullanılabilir olduğu görülmüştür. Sayfaların yapay sinir ağları ile sınıflandırılması, doğru bilgiye hızlı erişimin sağlanabilmesini olanaklı kılacaktır.

Anahtar Kelimeler: Sınıflandırma, Yapay Sinir Ağları

1. Giriş

Yirmibirinci yüzyıl bilişim teknolojileri alanında çok önemli gelişimlere tanık olmuştur. İletişim kavramı da bundan büyük ölçüde etkilenmiştir. İletişim süreci enformasyon iletimi olgusu çerçevesinde gelişmiş ve de kolaylaşmıştır. İletişim bilimleri çatısı altındaki bilişim yaklaşımı temellerini, iletinin karşıya sorunsuz iletilmesini sağlamayı amaçlayan Shannon ve Weaver'in kuramında bulur. Günümüzde internet aracılığıyla erişilebilen bilgi sistemlerinin sayısı hızla artmaktadır. Bu sistemler üzerindeki bilgi kaynakları da giderek çeşitlenmekte ve daha fazla yer kaplamaktadır. Yakın zamana dek bilgisayar ağları aracılığıyla çoğunlukla metin (text) türü bilgilere erişim sağlanabilirken, günümüzde grafik, ses, görüntü, canlandırma ve diğer görsel-işitsel veriler içeren çoklu ortam (multimedia) türü bilgiler de İnternet üzerinde hizmete sunulabilmektedir. Coğrafik bilgi sistemleri, elektronik kütüphaneler, film ve TV arşivleri bu tür bilgilere örnek olarak gösterilebilir.

2. İnternet Aracılıęıyla Erişilebilen Bilgi Kaynakları

Bilgisayar aęları aracılıęıyla erişilebilen bibliyografik veri tabanlarına ve bilgi kaynaklarına birkaç örnek tanımlanmıştır.

2.1. Çevrimiçi kütüphane katalogları

Çevrimiçi kütüphane katalogları kullanımı genellikle komut diline ya da mönüye dayalı, kolayca ve çoęu zaman ücretsiz erişilebilen bibliyografik veri tabanlarıdır. Bilgi aęları aracılıęıyla erişilebilen çevrimiçi kataloglar ulusal kaynaklar olarak giderek önem kazanmakta ve kütüphaneler arası işbirliğinin artmasına yol açmaktadır. Son yıllarda çevrimiçi kataloglar üzerinde belirli bir kütüphanede bulunan kaynakların bibliyografik bilgilerinin yanı sıra belli başlı özet ve dizin dergilerinin (örneğin, Current Contents, MEDLINE vb.) veri tabanları da yer almaya başlamıştır.

2.2. Ticari bibliyografik veri tabanları

Bu grupta Dialog, OCLC ve RLIN gibi ticari ya da kooperatif hizmetler veren kuruluşların bilgi aęları aracılıęıyla kullanıma sundukları bilgi kaynakları yer almaktadır. Knight-Ridder Information Service (eski Diyalog) üzerinde her konuda yüzlerce bibliyografik ve tam metin veri tabanı bulunmaktadır. OCLC (Online Computer Library Center) ve RLIN (Research Libraries Information Network) gibi kooperatifler ortaklaşa kataloglama hizmetlerinin vazgeçilmez bir parçası haline gelmişlerdir

2.3. CD-ROM'lar üzerindeki çevrimiçi veri tabanları

CD-ROM'lar daha çok yayın ve dağıtım için kullanılan bir ortamdır. Kütüphane kataloglarının ve basılı dergilerde yer alan makalelerin bibliyografik bilgilerinin ya da tam metinlerinin CD-ROM'lara kaydedilerek kullanıma sunulması oldukça eski bir uygulamadır (yaklaşık 10 yıl). CD-ROM üzerindeki veri tabanları son yıllarda bilgi aęları aracılıęıyla hizmete sunulabilmektedir. Örneğin, Kaliforniya Üniversitesi CD-ROM Bilgi Sistemi 1990 yılı nüfus sayımı sonuçlarını içeren yaklaşık 260.000 ciltlik yayımlanmış federal istatistiklerin yerini almıştır.

3. İnternette Bilgi Kaynaęı Keşfetme ve Erişimle İlgili Sorunlar

İnternet sağladığı kolaylıklar yanında birtakım zorlukları da beraberinde getirmektedir. Bu bölümde internet üzerinde bilgiye ulaşırken karşılaşılabilecek zorluklara değinilmiştir.

3.1. Çok Çeşitli Bilgi Sistemlerinin Varlığı

İnternet son derece hareketli bir bilgi erişim ortamıdır. İnternet aracılıęıyla her gün yeni bilgi kaynakları hizmete sunulmakta, bazılarının yerleri değişmekte ya da güncelliğini yitiren bilgiler kullanımdan çekilmektedir. Yeni hizmete sunulan bu tür kaynakları keşfetmek, ortadan kaybolan ya da yeri değişen bilgi kaynaklarını izlemek, İnternet aracılıęıyla erişilebilen bilgilerden yararlanmayı kısıtlayan etmenlerden birisidir. İnternet üzerinde bulunan ve sürekli değişen 'aktif' belgelerin kataloglanması ve sınıflanması konusunda çalışmalar yürütülmesine karşın, bu alanda büyük mesafe kaydedildiğini söylemek oldukça güçtür. Kütüphanecilikte basılı materyaller için kullanılan kataloglama ve sınıflama teknikleri elektronik ortamdaki bilgiler için yetersiz kalmaktadır. İnternet aracılıęıyla hizmete sunulan her bilgi sistemi ve veri tabanı birbirinden farklıdır. Kullanılan veri tabanı yönetim sistemleri ve veri yapıları birbirinden farklıdır. Veri tabanında yer alan bilgiler farklı anlayış ve sistemlere göre düzenlenmektedir. Her belgeye ilişkin ortak/asıgari erişim noktaları (yazar adı, belge adı vb. gibi) sistemden sisteme farklılıklar göstermektedir. Bu bilgi kaynaklarını düzenlemek için harcanan entelektüel çaba göz ardı edilemeyecek kadar büyüktür.

3.2. Giderek Büyüyen Kullanıcı Topluluęu

Günümüzde çeşitli İnternet hizmetlerinden her yaş ve mesleki gruptan yaklaşık 30-40 milyon insanın yararlandığı tahmin edilmektedir. Kullanıcı sayısı her yıl büyük bir hızla artmaktadır. Üniversite öğretim üyeleri ve araştırmacılardan ilkökul öğrencilerine, sanayici ve iş adamlarından yaşlılara ve çocuklara kadar her kesimden kullanıcılar İnternet aracılıęıyla elektronik bilgilere erişebilmekte, ticaret yapabilmekte ve eğlenebilmektedirler.

3.3. İnternet Aracılıęıyla Erişilebilen Bilgilerin Fiyatlandırılması

Kütüphaneciler bilgi aęlarını, kaynak paylaşımını ve bilgi erişimi geliştirmek için bir araç olarak görmektedirler. Bu nedenle kamu kaynakları harcanarak geliştirilen ve İnternet aracılıęıyla erişilebilen bilgi kaynaklarının ve sağlanan hizmetlerin (kütüphane katalogları, ftp arşivleri vb. gibi) birçoęu ücretsizdir. Ancak, eęer fiyatlandırmak gerekirse bu hizmetler için ne tür bir fiyatlandırma sistemi kullanılması gerektiği de henüz tam olarak açıklığa kavuşmuş değildir. İnternet aracılıęıyla erişilebilen

bilgi kaynakları için henüz ortak bir fiyatlandırma modeli geliştirilememiştir. Bazı şirketler bu tür bilgi kaynaklarını abonelik esaslarına göre hizmete sunmakta, bazıları şirket, üniversite ya da kamu kuruluşlarına lisans anlaşması karşılığı kiralamakta, bazıları ise erişilen makale başına belli bir ücret istemektedir.

3.4 Bilgilerin Doğruluğu ve Güvenliği

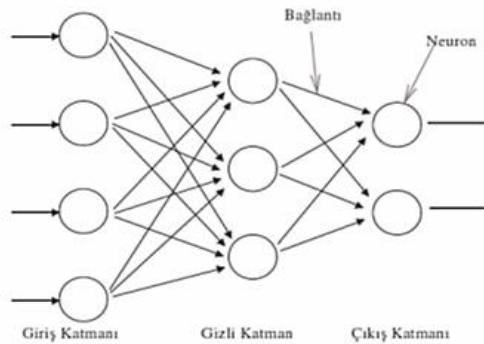
İnternet aracılığıyla erişilen bilgi kaynakları ile ilgili en önemli sorunlardan birisi de bilgilerin doğruluğu ve güvenliğinin sağlanmasıdır. İnternet'ten sağlanan bilgilerin doğrulanabilmesi ve aslına uygun olup olmadığının araştırılması (authentication), elektronik ortamdaki bilgilerin aradan geçen zaman içinde güncellenmesi ya da transfer anında kolayca değiştirilebilmesi (yanlışlıkla ya da kötü niyetle) veri bütünlüğü ve veri güvenliği ile ilgili sorunları da beraberinde getirmektedir. Bu tür sorunlara çözüm bulmak ve elektronik belgelerin aslına uygun olup olmadığını belirlemek amacıyla çeşitli algoritmalar ve elektronik damgalama (time stamping) teknikleri geliştirilmektedir.

4. Sınıflandırma Kavramı ve Web Sayfalarının Sınıflandırılması

İnsan beyni her gün binlerce enformasyon ile karşı karşıya kalmaktadır. Bu aşırı bilgi yüküyle baş etmenin en iyi yolu sınıflandırma yapmaktır. Sürekli değişen dış çevreye uyum sağlamak zorunda olan bireyin, değişik kaynaklardan kendisine gelen tüm bilgilere dikkat edip, onları algılamak, çözümlemek, saklamak ve gerektiğinde kullanmak için yeterli kapasitesi yoktur. Birey bu yoğun, karmaşık ve hızlı bilgi akışı ile "sınıflandırma" yardımıyla baş etmektedir. Sınıflandırma, nesnelere ya da insanları, belirli bir takım ortak niteliklerini temel alarak gruplara ya da sınıflara ayırma sürecidir. Sınıflandırma, aşırı bilgi yükünden kurtulmanın bir yoludur. Sınıflandırma problemi, yeni karşılaşılan bir nesnenin, özelliklerinin incelenmesi ve önceden tanımlanmış sınıflardan birine atanmasından oluşur.

5. Yapay Sinir Ağlarının Özellikleri

Genel anlamda yapay sinir ağları (YSA), beynin bir işlevi yerine getirme yöntemini modellemek için tasarlanan bir sistem olarak tanımlanabilir. YSA (Şekil 1), yapay sinir hücrelerinin birbirleri ile çeşitli şekillerde bağlanmasından oluşur ve genellikle katmanlar şeklinde düzenlenir. Donanım olarak elektronik devrelerle ya da bilgisayarlarda yazılım olarak gerçekleştirilebilir. Beynin bilgi işleme yöntemine uygun olarak YSA, bir öğrenme sürecinden sonra bilgiyi toplama, hücreler arasındaki bağlantı ağırlıklarını ile bu bilgiyi saklama ve genelleme yeteneğine sahip paralel dağılmış bir işlemcidir.



Şekil 1: Yapay Sinir Ağı Modeli

Katmanların değişik şekilde bir birleriyle bağlanmaları değişik ağ mimarilerini doğurur. YSA'lar üç katmandan oluşur. Bu katmanlar sırasıyla;

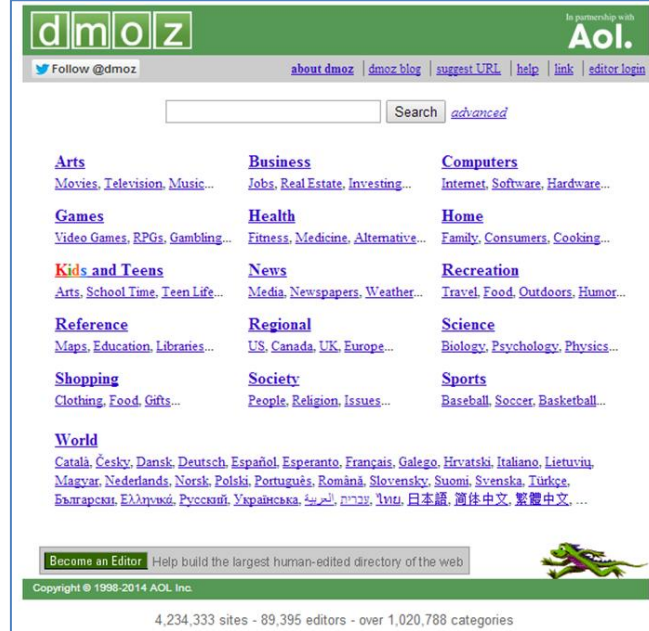
- Girdi Katmanı: Bu katmandaki proses elemanları dış dünyadan bilgileri alarak ara katmanlara transfer ederler. Bazı ağlarda girdi katmanında herhangi bir bilgi işleme olmaz.
- Ara Katman (Gizli Katman): Girdi katmanından gelen bilgiler işlenerek çıktı katmanına gönderilirler. Bu bilgilerin işlenmesi ara katmanlarda gerçekleştirilir. Bir ağ içinde birden fazla ara katman olabilir.
- Çıktı Katmanı: Bu katmandaki proses elemanları ara katmandan gelen bilgileri işleyerek ağına girdi katmanından sunulan girdi seti için üretmesi gereken çıktıyı üretirler. Üretilen çıktı dış dünyaya gönderilir.

6. Uygulanan Model ve Metodolojisi

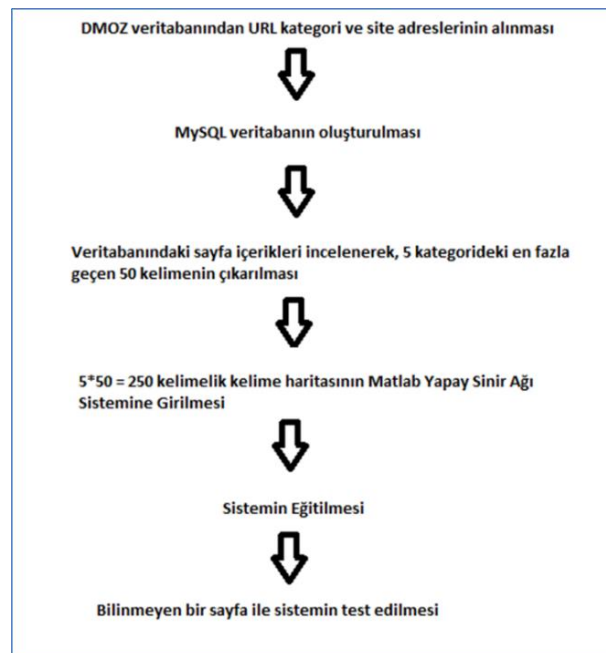
Bu çalışmada web sayfalarının kategorilere ayrılmasında, kelime içeriklerine göre yapay sinir ağları ile

öğrenen bir sistem üzerinde durulmuştur. İnternet üzerinde belli kategorilerdeki site adresleri, Açık Dizin Projesi üzerinden alınmıştır. Bu siteler içerisindeki kelimelerin kullanılma sayılarını ölçen bir program geliştirilmiş, frekans sayılarına ve kategorilere göre tablolar oluşturulmuştur. Bu tablolar daha sonra Matlab Yapay sinir ağıları modülünde kullanılarak web sayfalarını sınıflandırmayı amaçlayan bir araç geliştirilmiştir (Şekil 2).

Oluşturulan bu eğitim kümesi YSA'ya sunulmuş ve Levenberg-Marquardt öğrenme algoritması kullanılarak eğitim yapılmıştır. Eğitim sırasında nöronlar arasındaki ağırlıklara ilk değer olarak [-1,+1] arasında değerler atanmıştır ve ara katmanda 20 adet nöron kullanılmıştır. Daha önceden belirlenen beş kategoriden yirmişer adet sayfa için toplamda yüz girişli bir eğitim kümesi oluşturulmuştur. Çıkış değerleri eğitim seti için sayfanın ait olduğu kategori için 1, diğer kategori değerleri için 0 verilmiştir.



Şekil 2: DMOZ – Açık Dizin Projesi



Şekil 3: Sistemin Akış Şeması

Açık Dizin Projesi internetteki en büyük ve en kapsamlı, insanlar tarafından düzenlenen dizindir. Dünyanın her tarafından katılımda bulunan geniş bir gönüllü editörler topluluğu tarafından inşa edilmiştir ve varlığı onlar tarafından sürdürülmektedir. Sistemde veritabanı olarak MySQL seçilmiştir. MySQL, altı milyondan fazla sistemde yüklü bulunan çoklu iş parçacıklı (multi-threaded), çok kullanıcı (multi-user), hızlı ve sağlam (robust) bir veritabanı yönetim sistemidir. Programlama dili olarak C# kullanılmıştır. Microsoft'un geliştirmiş olduğu yeni nesil bir programlama dilidir. Yapay sinir ağlarının tasarımı ve eğitimi için MATLAB programından faydalanılmıştır (Şekil 3).

Bu projede kullanılmak üzere beş kategori seçilmiştir: **Bilim, Kültür ve Sanat, Spor, Ekonomi ve İş Dünyası, Sağlık.**

ID	Kategori	URL
1	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.tubitak.gov.tr/
2	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.mam.gov.tr/
3	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.tudav.org/
4	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.mta.gov.tr/
5	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://Atk.org.tr/
6	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.tagem.gov.tr/
7	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.arastirma.yalova.gov.tr/
8	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.upav.org.tr/
9	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.tesav.org.tr/
10	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.mamagrubu.org/
11	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.tuik.gov.tr/
12	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.sav.org.tr/
13	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.tuba.gov.tr/
14	Top/World/Türkçe/Bilim/Araştırma_Kurumlan	http://www.yetam.hacettepe.edu.tr/

Kelime	Frekans
konu	1564
bilim	1428
matematik	1422
bilgi	1329
ders	1259
devam	1116
tarih	1113
yeni	1087
kullan	1029
program	1004
tarih	974
çalış	971
bölüm	965
fizik	911
alan	857

Şekil 4: Site Adresleri ve Kelime Haritası

İlk aşamada, bu kategoriler içerisindeki siteler içerikleri incelenmiş ve her kategoride en fazla geçen 50 kelime saptanmıştır. Toplamda beş kategoriye ait ikiyüzlü kelimelik bir kelime haritası oluşturulmuştur.

Bilim Kategorisi:

"araştırma", "bilim", "matematik", "bilgi", "ders", "formül", "kampüs", "yenilik", "icat", "program", "tarih", "çalış", "bölüm", "fizik", "makale", "bilgisayar", "sayfa", "soru", "haber", "ilgi", "zaman", "araştırma", "proje", "yorum", "genel", "sistem", "kimya", "üniversite", "yapı", "eğitim", "tez", "deha", "öğretmen", "öğrenci", "sayı", "öğrenci", "yayın", "dünya", "teori", "kitap", "gerek", "insan", "duyuru", "kitap", "akademi", "dergi", "lisans", "sınıf", "temel", "teknoloji"

Kültür ve Sanat Kategorisi:

"film", "sanat", "sinema", "sanatçı", "heykel", "haber", "özgün", "kitap", "fotoğraf", "konu", "yorum", "oyun", "insan", "forum", "mesaj", "sanat", "sayfa", "fotoğraf", "senaryo", "mimar", "bilgi", "sinema", "yazar", "resim", "ders", "video", "hayat", "biental", "dünya", "politika", "siyaset", "bilet", "müzik", "dans", "kültür", "pop", "tarih", "gezi", "şiir", "tiyatro", "roman", "link", "yönetmen", "tarih", "yazı", "festival", "eğitim", "konser", "ilgi", "şair"

Spor Kategorisi:

"takım", "spor", "oyun", "futbol", "favori", "galatasaray", "hafta", "lig", "oyun", "skor", "yorumcu", "sezon", "kurs", "faul", "yarış", "maç", "basketbol", "hakem", "fenerbahçe", "türkiye", "grup", "kupa", "faul", "saha", "kadro", "kulüp", "playoff", "federasyon", "kurul", "büyük", "kazan", "kaybet", "milli", "avrupa", "menejer", "şampiyon", "puan", "teknik", "taraf", "yedek", "haber", "oyuncu", "kart", "beşiktaş", "kamp", "derbi", "transfer", "avrupa", "yayın", "taraf"

Ekonomi ve İş Dünyası:

"para", "hizmet", "ürün", "bilgi", "borsa", "şirket", "firma", "döviz", "kalite", "fiyat", "haber", "hisse", "fuat", "işçi", "proje", "kurum", "ekonomi", "dolar", "üretim", "kur", "servis", "genel", "grup", "özel", "makine", "yapı", "model", "kaynak", "cihaz", "çalışmak", "teknik", "kamu", "malzeme", "işveren", "teknoloji", "eğitim", "maaş", "istanbul", "sanayi", "zaman", "uygun", "ücret", "çalışan", "patron", "pazarlama", "reklam", "satış", "müdür", "euro", "yönetim"

Sağlık Kategorisi:

"tedavi", "sağlık", "ilaç", "tıp", "hastalık", "bebek", "estetik", "doktor", "hasta", "cerrahi", "kan", "konu", "aile", "hizmet", "hekim", "çocuk", "taburcu", "acil", "merkezi", "uzman", "bakım", "taburcu", "dahili", "hemşire", "lazer", "yatarak", "serum", "ameliyat", "reçete", "kanser", "hastane", "genetik", "aile", "organ", "meme", "gebe", "cilt", "zaman", "diyet", "özel", "alan", "genel", "program", "merkez", "veteriner", "beslenme", "kongre", "durum", "kurul", "kalp"

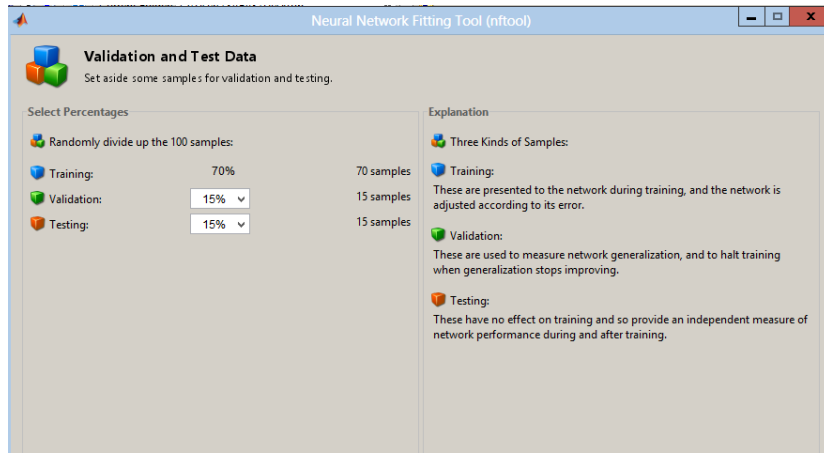
Örnek	Giriş	Çıkış
	0,002617801 0,039267016 0,002617801	
	0,007853403 0 0,028795812 0 0,007853403 0	
	0,060209424 0 0,002617801 0 0 0,010471204 0	
	0,005235602 0 0,007853403 0,007853403 0	
	0,04973822 0,031413613 0 0 0,002617801	
	0,002617801 0,005235602 0,002617801 0,007853403	
	0,007853403 0 0 0 0,005235602 0,002617801	
	0,002617801 0 0,002617801 0,002617801	
	0,002617801 0 0,007853403 0 0 0,010471204	
	0,005235602 0 0,002617801 0 0 0,007853403	
	0,028795812 0 0,007853403 0 0,002617801	
	0,002617801 0,002617801 0 0 0 0,005235602	
	0,002617801 0 0,007853403 0 0 0,002617801 0	
	0 0 0,002617801 0 0 0 0,010471204 0	
	0,002617801 0 0 0 0 0,002617801 0,007853403	
	0,007853403 0,002617801 0 0 0 0,007853403 0 0	
	0,028795812 0 0 0 0 0,007853403 0 0 0	
1	0,015706806 0,015706806 0 0,002617801 0 0 0	10000
	0 0,020942408 0 0 0 0,002617801 0,002617801 0 0	
	0,005235602 0,002617801 0,010471204 0 0 0 0 0	
	0 0,002617801 0,002617801 0,002617801	
	0,002617801 0,007853403 0,028795812 0,007853403	
	0 0 0,007853403 0,010471204 0 0,005235602	
	0,031413613 0,007853403 0,005235602 0,002617801	
	0,002617801 0 0 0 0 0 0,002617801 0 0	
	0,007853403 0,002617801 0,005235602 0,002617801	
	0,002617801 0 0,023560209 0,007853403 0 0	
	0,013089005 0 0,007853403 0 0 0,013089005 0	

Şekil 5: İlk Nörona Ait Giriş ve Çıkış Değerleri

Şekil 5’de ilk nörona ait giriş ve çıkış değerleri verilmiştir. İlk nöron <http://www.tubitak.gov.tr> adresine aittir. Bu şekilde rakamlar:

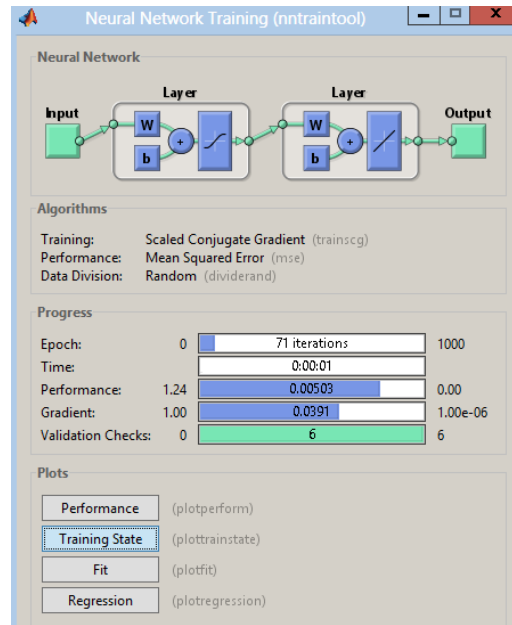
1. Kelime haritasındaki ilk gruptaki (bilim) ilk kelime olan “araştırma” kelimesininin toplam kelime sayısına olan oranını ifade etmektedir.
2. Kelime haritasındaki ikinci kelime olan “bilim” kelimesininin toplam kelime sayısına olan oranını ifade etmektedir.
3. Kelime haritasındaki üçüncü kelime olan “matematik” kelimesininin toplam kelime sayısına olan oranını ifade etmektedir.
4. Buradaki değerler beklenen ve çıkış değerleridir. Beş kategori sırasıyla ifade edilmiştir. Sayfanın dahil olduğu kategori “1” ile gösterilip, diğer kategoriler “0” ile gösterilmektedir. “1,0,0,0,0” değeri sayfanın ilk kategori yani “bilim” kategorisinde olduğunu gösterir.

Öngörülen 100 adet verinin %70’i eğitim, %15’i doğrulama, %15’i de test amaçlı kullanılmıştır.



Şekil 6: YSA'da Kullanılan Ölçüm Değerleri

Yapılan 71 iterasyondan sonra eğitim tamamlanmış, sonrasındaki performans ölçümleri Şekil 7'deki gibi gözlemlenmiştir.



Şekil 7: YSA'da Kullanılan Ölçüm Değerleri -2

7. Sonuç

Kelime haritası oluşturulurken yapılan tarama miktarı, test sonucu alınan sayfadaki metnin aslında içerik olarak farklı kategoriye ait olması, sayfanın bir bölümünde bulunan bilgilendirme niteliğindeki metinlerin yanıltıcı olması, eğitim kümesi oluşturulurken yapılan taramanın geniş kapsamlı olmaması vb. gibi. Ancak bu sebeplerin kolaylıkla ortadan kaldırılması ve başarı oranı düşük olan sayfalarda da bu oranın yükseltilmesi mümkündür. Konuların birbirine yakınlığı da sonuçlarda önemli rol oynamıştır. Spor konuları ekonomi ve kültür konularından uzak olduğu için oldukça yüksek başarı sağlanabilirken, sınıflandırmada ekonomi ve kültür haberleri genelde iç içe olduğundan hata oranının arttığı gözlemlenmiştir.

Bu çalışmanın geliştirilmesinde karşılaşılan güçlükler farklı ortamları biraraya getiren bir arayüz geliştirilmesi, kelime haritası oluşturulurken dikkat edilmesi gereken hususların incelenmesi ve kategori için belirleyici olacak olan optimum kelimelerin bulunması, oluşturulan kelime haritası için eğitim setinin oluşturulması, bu eğitim setinin uygulanacağı YSA yapısının belirlenmesi ve en uygun yapıyı bulmak için denemeler yapılması olarak söylenebilir. Bu sırada karşılaşılan en büyük güçlük eğitim aşamasının uzun zaman alması olmuştur. Fakat yapılan çalışmalardan sonra en iyi sonucu veren katman ve nöron sayıları, aktivasyon fonksiyonu ve öğrenme algoritması tercih edilmiştir.

Kaynaklar

- Andaç Demirtaş, *Sosyal Sınıflandırma, Kişiler arası Beklentiler ve Kendini Doğrulayan Kehanet*
- Ali Dönmez (1992), Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Araştırma Dergisi, (14), 131-46, *Bilişsel Sosyal Şemalar*
- Çavdur, F. 2005. Uludağ Üniversitesi, Yüksek Lisans Tezi. (yayınlanmamış), 153 S. *Arama Motorları Kullanıcı Oturumlarındaki Konu Değişikliklerinin Tespit ve Tahmin Yöntemleri*
- Çetin Elmas, 2003. Seçkin Yayıncılık, Ankara. 192S. *Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama)*
- D. Dubois, J. Lang, H. Prade, Part 2 : Logical Approaches", *Fuzzy Sets and Systems*, 40, 203-244 1991, *Fuzzy Sets in Approximate Reasoning*
- E. Öztemel Otomasyon Dergisi, Sayı:126, *Yapay Zekâ Ne Kadar Yapaydır?*
- Haykin, S. 1994. Macmillan College Publishing Company, USA, 696P, *Neural Networks*
- HARP Akademileri Komutanlığı Yayınlarından, HARP Akademileri Basımevi, İstanbul, Mart-1996 .HARP Akademileri Komutanlığı Yayınlarından, Adı Geçen Eser, S. 49-51, *Uzman Sistemler ve Yapay Zekâ*
- H.H. Önder, Uluslararası Eğitim Teknolojileri Sempozyumu Bildirileri, Sakarya, Türkiye, 2001, *Yapay Zeka Programlama Teknikleri Ve Bilgisayar Destekli Eğitim*
- J.F. Baldwin, Int. J.Man-Machine St., 11, 465-480, 1979, *Fuzzy Logic and Fuzzy Reasoning*
- Lynch, Clifford A. (1991). Ed. Filomena Simora. New Providence, NJ: Bowker Visions of electronic libraries in: *The Bowker Annual: Library and Book Trade Almanac (75-82) 36th ed.*
- Merrill, Deane et al. (1995 April). "The University of California CD-ROM Information System, 8(4): 51-52, *Communications of the ACM*
- M. J. Berry and G. Lino®, New York, NY, USA: John Wiley & Sons, Inc., 1997, *Data Mining Techniques: For Marketing, Sales, and Customer Support.*
- N. Taşkın, B. Kandemir, 3th International Computer&Instructional Technologies Symposium, October 07-09 2009, Trabzon, TURKEY, pp.409-415. *The Effects of Web Based Virtual Geometry Laboratory on Sixth Grade Elementary School Students' Academic Achievements And Educational Attainments*
- Önder Efe ,Okyay Kaynak. 2000. Boğaziçi Üniversitesi Yayını,İstanbul,141 S, *Yapay Sinir Ağları ve Uygulamaları*
- S. Serap. Kurbanoglu, (1995). Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi 12(1/2): 167-186. *Elektronik uzayda suç ve ceza*
- Yaşar Tonta, (1992). Susan Stone and Michael Buckland. Medford, NJ: Learned information, *Indexing in hypertext databases in: Studies in Multimedia: State-of-the-Art Solutions in Multimedia and Hypertext Ed*