

Yapay Zeka Teknikleri ve Yapay Sinir Ağları Kullanılarak Web Sayfalarının Sınıflandırılması

Yrd. Doç. Dr. Aybars UĞUR

Ege Üniversitesi
Bilgisayar Mühendisliği Bölümü
aybars.ugur@ege.edu.tr

Ahmet Cumhuri KINACI

Ege Üniversitesi
Bilgisayar Mühendisliği Bölümü
cumhur.kinaci@ege.edu.tr

ÖZET

Bu makalede Yapay Zeka ve Yapay Sinir Ağları kısaca tanıtılmış ve yapay zekanın internet üzerindeki uygulama alanları üzerinde durulmuştur. Yapay sinir ağları içeren programları geliştirmek için kullanılan API'ler hakkında bilgiler verilmiştir. Bu teknikler kullanılarak web sayfalarını sınıflandırmayı amaçlayan bir araç geliştirilmiştir.

ABSTRACT

In this paper artificial intelligence and artificial neural networks introduced briefly and application areas of artificial intelligence on internet are mentioned. Information about APIs for developing applications including artificial neural networks, is given. Using these techniques, a tool for classifying web pages is developed.

Anahtar Kelimeler: Yapay zeka, Web'de Zeka, Yapay sinir ağları

1. GİRİŞ

Zeki yazılımlar geliştirmeye yönelik çalışmalar günümüzde önem kazanmıştır. Zeki olmayan yazılımlar, kullanıcılarını gereksiz veya tekrarlı işleri yapmak durumunda bırakarak verim kaybına yol açmaktadırlar. Kullanıcılar, aynı türde yazılımların zeki olanlarını tercih etmektedirler. D. Lenat ve E. Feigenbaum'un tanımlarına göre zeka "Karmaşık bir problemi çözmek için gerekli bilgileri toplayıp birleştirme kabiliyetidir" veya "Karmaşık bir problemi, çözüm arama alanını daraltarak kısa yoldan çözebilme kabiliyetidir". Zeka'nın sözlük anlamı ise, insanın düşünme, akıl yürütme, nesnel gerçekleri algılama, kavrama, yargılama, sonuç çıkarma yeteneklerinin tümüdür. Ayrıca; soyutlama, öğrenme ve yeni durumlara uyma gibi yetenekler de zeka kapsamı içindedir. Yapay zeka ise, bu özelliklere sahip organik olmayan sistemlerdeki zekadır. Luger ve Stubblefield, 1993 yılında yapay zekayı "zeki davranışların otomasyonu ile ilgili bilgisayar bilimleri dalıdır" şeklinde tanımlamıştır. [1][2][3][4]

Yazılımların öğrenilmesi, tecrübelerine dayanarak davranışlarını değiştirebilmesini sağlamaktadır. Öğrenen yazılımlar ve robotlar gibi bu yazılımları içeren donanım destekli sistemler günlük yaşamda giderek yaygınlaşmaktadır. Yapay sinir ağları,

makinelere öğrenmesini sağlayan en güçlü yöntemlerdendir. Bu çalışmada, web üzerinde yapay sinir ağı yardımı ile öğrenen bir uygulama geliştirilmiştir.

2. YAPAY SİNİR AĞLARI

Yapay Sinir Ağları (Artificial Neural Networks) veya kısaca Sinir Ağları (YSA) insan beyninden esinlenerek geliştirilmiş, ağırlıklı bağlantılar aracılığı ile birbirine bağlanan işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. En önemli özelliği, deneyimlerden (tecrübe) yararlanarak öğrenmesidir. Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilmişlerdir. Yapay sinir ağları, öğrenmenin yanı sıra bilgiler arasında ilişkiler oluşturma yeteneğine de sahiptir.

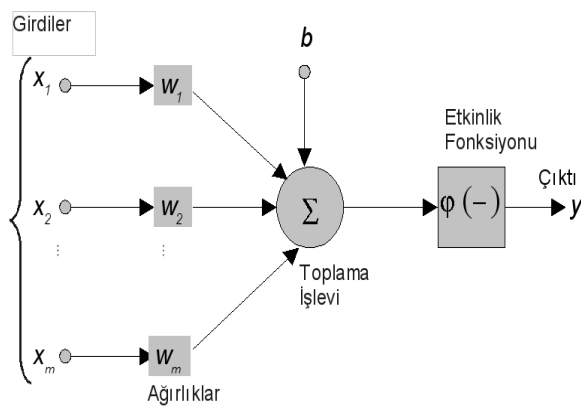
Yapay sinir ağlarının temel işlevleri şu şekilde belirtilebilir :

- Öngörü (Prediction) veya tahminleme : İleriki satışlar, hava tahminleri, at yarışları, çevresel risk, ...
- Sınıflandırma (Classification) ve Kümeleme (Clustering) : Müşteri profilleri, tıbbi teşhis, ses ve şekil tanıma, hücre tipleri ...
- Kontrol (Control) : Erken uyarı için uçaklarda ses ve titreşim düzeyleri, ...

Ayrıca, Veri Birleştirme (Data Association), Kavramsallaştırma (Data Conceptualization) ve Filtreleme (Data Filtering) için de kullanılabilir.

Yapay sinir ağlarının endüstriyel uygulamalar, finans uygulamaları, askeri ve savunma uygulamaları, tıp ve sağlık uygulamaları, mühendislik uygulamaları, robotbilim, görüntü işleme, örüntü tanıma dışında iletişim sanayi, eğlence amaçlı tahmin gibi özel uygulama alanları da bulunmaktadır.

Bir yapay sinir ağı hücresi temel olarak girdilerden, ağırlıklardan, toplama işlevinden ve çıktıdan oluşur (Şekil 1).



Şekil 1. Yapay Sinir Hücresi

Genelde, verilen bir girdi setine karşılık çıktı değerleri verilerek belirtilen öğrenme kuralına göre ağırlık değerleri otomatik olarak değiştirilmektedir. Eğitim verisinin tamamlanmasından sonra eğitilmiş olan ağ, ağırlık değerlerinin son durumuna göre, verilen herhangi bir veri setinin sonucunu tahminleyebilmektedir.

Yapay sinir ağı bir dizi sinir hücresinin ileri sürümlü ve geri beslemeli bağlantı şekilleri ile birbirine bağlanması ile oluşur.

Günümüzde, belirli amaçlarla ve değişik alanlarda kullanılmaya uygun birçok yapay sinir ağı modeli (Perceptron, Adaline, MLP, LVQ, Hopfield, Recurrent, SOM, ART ve PCA gibi) geliştirilmiştir. Öğrenme çeşitlerinden öğreticili öğrenme, öğreticisiz öğrenme, destekleyicili öğrenme ve karma stratejiler kullanılmaktadır.

3. WEB'DE ZEKA VE YAPAY ZEKA'NIN İNTERNET ÜZERİNDEKİ UYGULAMA ALANLARI

Web'de Zeka (Web Intelligence), yapay zeka yöntemlerinin web tabanlı sistemlere uyarlanması ile ilgilidir. Yeni nesil web tabanlı ürünler, sistemler, servisler ve etkinliklerde yapay zeka ve ileri bilgi teknolojilerinin kullanılması üzerine çalışmalar yapılan bir bilimsel araştırma geliştirme alanıdır.

Bu konuyla ilgili araştırmaları organize etmek ve yönlendirmek üzere "The Web Intelligence Consortium" (WIC) kurulmuştur. WIC tarafından belirlenen 9 ana konu bulunmaktadır:

1. Web Bilgi Erişimi (Web Information Retrieval)
2. Web Madenciliği ve Çiftçiliği (Web Mining and Farming)
3. Sosyal Ağ (Social Network)
4. "Grid" Hesaplama (Grid Computing)
5. Bilgi Ağı ve Yönetimi (Knowledge Network and Management)
6. Web Etmenleri (Web Agents)

7. Multi-modal Etkileşim (Multi-modal Interaction)
8. Her yerde olan Hesaplama (Ubiquitous Computing)
9. Web'de Zeka Sistemleri Uygulamaları (WI Systems Applications)

Bu başlıklar kendi içlerinde aşağıda bir kısmı verilen bir çok uygulama alanını barındırmaktadır :

- Web Bilgi Erişimi : Otomatik Web içeriği kataloglama ve indeksleme, kavramsal bilgi çıkarımı, çoklu ortam bilgilerine erişim.
- Web Madenciliği ve Çiftçiliği : Veri madenciliği ve bilgi keşfi, çoklu ortam veri madenciliği, kullanıcı profili öğrenme.
- Bilgi Ağı ve Yönetimi : Web sayfası otomatik yaratma ve güncelleme, bilgi çevrimi, çok boyutlu web veri tabanları.
- Web Etmenleri: E-posta filtreleme, e-posta yarı-otomatik cevaplama, bilgi filtreleme, küresel bilgi toplama.
- Web'de Zeka Sistemleri Uygulamaları : Web tabanlı dağıtık bilgi sistemleri, elektronik veri takası, karar destek sistemleri.

4. YAPAY SINIR AĞLARI İÇİN HAZIR ARAYÜZLER

İnternet üzerinde yapay sinir ağları için bir çok yazılım kütüphanesi ve simülasyon ortamı bulunabilmektedir. Bunların bazıları sadece kütüphane kısmından oluşurken bazıları yapay sinir ağı tasarlamayı sağlayan araçları da içerir. Bu tür araçlar özellikle konuyu ileri düzeyde bilmeyenler için yazılımlara yapay sinir ağı eklemeyi kolaylaştırır.

Belli başlı yapay sinir ağı arayüzleri aşağıda belirtilmektedir:

Yazılım Kütüphanesi :

1. Joone : Açılımı Java Object Oriented Neural Engine olan kütüphane Java ortamında yapay sinir ağlarının kullanılması için geliştirilmiştir. Ayrıca görsel olarak yapay sinir ağının oluşturulması ve eğitilmesini sağlayan bir araca sahiptir.
2. Neural Network Toolbox for MATLAB: MATLAB için yapay sinir ağı aracıdır.
3. Annie : Açık kaynak kodlu bir kütüphane olup C++ dili için geliştirilmiştir. MLP, RBF ve Hopfield modellerini desteklemektedir.
4. FANN : ANSI C ile gerçekleştirilmiş bir yapay sinir ağı kütüphanesidir. Benzerlerinin çoğundan daha hızlı çalışmaktadır.

Simülasyon :

1. Genesis : Karmaşık sinir ağı simülasyon ortamıdır.
2. Stuttgart Neural Network Simulator : C de yazılmış ve X tabanlı grafik arabirimi olan

yapay sinir ağı uygulama simülasyon ortamıdır.

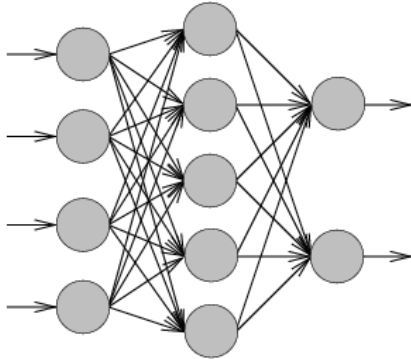
Yazılım ve eklentiler :

1. NeuroXL : MS Excel için yapay sinir ağı eklentisidir. Tahminleme, sınıflandırma ve özellikle finansal tahminleme için geliştirilmiştir.
2. NeuroSolutions : Bir çok sinir ağı ve öğrenme algoritmasını destekleyen ticari bir yazılımdır.
3. NeuralWorks Professional II/PLUS : Kapsamlı sinir ağı geliştirme ortamı sunan ticari bir yazılımdır.

5. YAPAY SİNİR AĞLARI KULLANILARAK WEB SAYFALARININ SINIFLANDIRILMASI

Web sayfalarını sınıflandıran bir uygulama gerçekleştirilmiştir. Bu örnek uygulamada amaç yapay sinir ağı yöntemiyle web sayfalarının kategorilere ayrılabilmesini sağlamaktır. Bunun için kullanılan yapay sinir ağı modeli çok katmanlı algılayıcıdır. Uygulama Java ortamında hazırlanmıştır. Yapay sinir ağlarını programda kullanmak için JOONE kütüphanesinden yararlanılmıştır.

Çok katmanlı algılayıcılar (Şekil 2) bir çok problemi çözmeye kullanılmaktadır. Bir girdi katmanı, en az bir tane gizli katman ve bir çıktı katmanına sahiptir. Girdi katmanından gelen sinyaller ileri doğrultulu olarak, katman katman ilerler. Genelde geri yayılım (Backpropagation) algoritması ile öğrenmektedir.



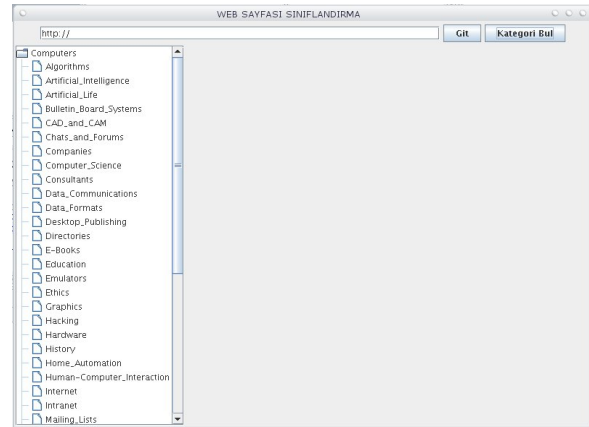
Şekil 2. Tek gizli katmana sahip bir çok katmanlı algılayıcı modeli.

Uygulamanın geliştirilmesinde öncelikli olarak kategorilere ayrılmış web sayfalarının bulunduğu bir veri seti üretmek için dmoz.org[5] sitesinin verileri kullanılmıştır. Bu site her kategori için mevcut olan editörleriyle, uygun web sayfalarının kategoriye eklenmesiyle büyüyen bir veri kümesine sahiptir. Benzerleri içinde bu sitenin verilerinin uygulama için kullanılmasının nedeni tüm verilerini XML

dosyalarında tutması ve dolayısıyla rahat işlenebilir olmasıdır.

Veri setinin çok büyük olmaması ve yapay sinir ağının eğitim süresinin uzamaması için sadece 'Computer' başlığı altındaki 44 ana başlık değerlendirmeye alınmıştır. Girdi verilerinin oluşturulmasında dmoz.org verileri basit bir doğal dil işleme işleminden geçirilerek uygun kelimelerin ayrıştırılması sağlanmış ve bu şekilde sadece isim ve sıfat olan kelimelerin girdi verisinde yer alması mümkün olmuştur. Doğal dil işleme için Stanford Parser[6] kütüphanesi kullanılmıştır. Ayrılan tüm kelimelerden bir indeks ve her bir internet sayfası için girdi verisi olarak bu indeksin boyutunda bir dizi oluşturulmuştur. Bu dizide, sayfada bulunan kelimeler 1.0 ile, olmayanlar ise -1.0 ile ifade edilmiştir.

Kullanılan yapay sinir ağının girdi katmanı oluşturulan kelime indeksi boyutunda yani 1983 elemana sahiptir. Gizli katmanda 50 ve çıktı katmanında da 44 eleman kullanılmıştır. Uygulamanın ekran görüntüsü Şekil 3'te gösterilmektedir.



Şekil 3. Örnek uygulamanın ekran görüntüsü

6. SONUÇ

İnternet kullanıcılarının aradığı sayfaların ve bilgilerin bulunmasında, zeki bir arama makinesi tahminlemeler yaparak işlem süresini kısaltmakta ve gereksiz ayrıntılarla uğraşılmasını önlemektedir. Web sitelerinde kullanılan zeki sanal yardımcıları, kişileri uygun şekilde yönlendirerek amaçlarına en kısa sürede ulaşmalarını sağlamaktadır. Sohbet robotları karşılıklı konuşma imkanı sunmakta, bilgi verme, yönlendirme ve eğlence amaçlı olarak kullanılmaktadır. Belirli kalitede ürünlerin en uygun fiyatla satıcıdan veya sanal marketten alınması için zeki etmenler kullanılmaktadır. Kişilik özelliklerini belirleyerek kullanıcılara, sayfalarını ve bilgilerini en uygun şekilde sunan siteler yapılmaktadır.

Yapay sinir ağlarının yazılımlarda kullanılması sırasında girdi verilerinin oluşturulması önemli bir problem oluşturur. Bir diğer önemli problem ise

uygun sinir ağı modelinin seçilmesi işlemidir. Öncelikle bu iki temel sorunun çözülmesi için geliştirilecek metotlar ve araçlar, yazılımlara bu tekniğin daha kolay eklenebilmesini sağlayacaktır. Geliştirilen örnek uygulama ile yapay zeka tekniklerinin internet üzerinde zeki yazılımlar oluşturmak için kullanılabilir olduğu görülmüştür. Gelecekte yapay zeka tekniklerinin gelişmesi ve internet üzerinde bütünleşik çalışabilen zeki yazılımlar ile bir çok alanda çok daha başarılı sonuçlar alınacaktır.

KAYNAKLAR

- [1] Luger, G.F., 2002, Artificial Intelligence : Structures and Strategies for ComplexProblem Solving, 4th edition, Addison-Wesley.
- [2] Nilsson, N. J., 1998, Artificial Intelligence :A New Synthesis, Morgan Kaufmann Publishers.
- [3] Russell, S., J., and Norvig, P., 1995, Artificial Intelligence : A Modern Approach, Prentice-Hall.
- [4] Uğur, A., Kınacı, A.C., “İnternet Üzerinde Yapay Zeka ve Yapay Sinir Ağları”, COMPOTEK 2005 Bilişim Seminerleri Programı, İzmir, Türkiye, Kasım 2005.
- [5] <http://dmoz.org> - Open Directory Project
- [6] <http://www-nlp.stanford.edu/software/lex-parser.shtml> - Stanford Parser