



Yapay Sınır Ağları ile Web İçeriklerini Sınıflandırma

Yazarlar: Esra Nergis Güven, Hakan Onur ve Şeref Sağırođlu

Sunan : Esra Nergis Güven

İçerik

- Giriş
- Amaç ve Kapsam
- Sınıflandırma
- Geliştirilen Sistem (WeSaKa)
- Mimari Yapı
- YSA yapısı
- Kelime Haritası
- Eğitim ve Test Vektörleri
- WeSaKa arayüzü
- Test sonuçları

Giriş

- Elektronik ortamların gün geçtikçe yaygınlaşması ve kullanımının artmasıyla birlikte bilgi miktarında da hızlı bir artış (Miniwatts, 2006)
- Elektronik ortamlarda tutulan bilginin sınıflandırılması, doğru bilgiye hızla erişimin kolaylaştırması
- Wikipedia'da doküman sınıflandırma/kategorizasyon : "bir elektronik dokümanın içeriğinin bir veya daha çok kategoriye ayrılması işlemi"
- Bilgi erişim sistemleri temelde, kullanıcıların bilgi ihtiyaçlarını karşılaması muhtemel olan ilgili belgelerin tümüne erişir ve ilgili olmayanları da ayıklar
- Hızlı bilgi artışında otomatik metin kategorizasyonu önemli bir hedef

Amaç ve Kapsam

Amaç: Web sayfalarının belirlenen konulara göre otomatik olarak sınıflandırılabilmesi için çok katmanlı yapay sinir ağı modeli kullanılarak zeki bir sistem tasarlanması.

Kapsam: Mimari yapının tasarlanması, uygulama arayüzünün geliştirilmesi, özellik vektörünün seçilmesi, YSA yapısının oluşturulması, eğitim ve test vektörlerinin toplanması, eğitim ve test işlemleri yapılması, sonuçların değerlendirilmesi.

Doküman Sınıflandırma

- Sınıflandırma Yöntemleri:
 - Karar ağaçları (Moulinier ve Ganascia,1996),
 - Kural öğrenme (Apte, Dameran ve Weiss, 1994),
 - Sinir ağıları (Ng, Goh ve Low, 1997; Wiener, Pedersen ve Wiegend, 1995),
 - Lineer sınıflandırıcılar (Lewis, Schapire Callan ve Papka, 1996),
 - En yakın komşuluk algoritmaları (kNN) (Yang ve Pedersen, 1997),
 - Destek vektör makinaları (Joachims, 1997),
 - tf*idf değerleri (terim sıklığı* devrik belge sıklığı),
 - Kavram madenciliği (content mining),
 - Gizli anlam analizi (LSA-Latent Semantic Analysis),
 - Naive Bayes metodları (Lewis ve Ringuette, 1994; McCallum ve Nigam, 1998).
- Web tarayıcısı kullanan sınıflandırma sistemlerinin çoğunda işlemler insan desteğiyle (Shanks ve Williams, 2001)

Geliştirilen Sistem (WeSaKa)

Web Sayfası Kategorizasyonunda YSA Kullanımı

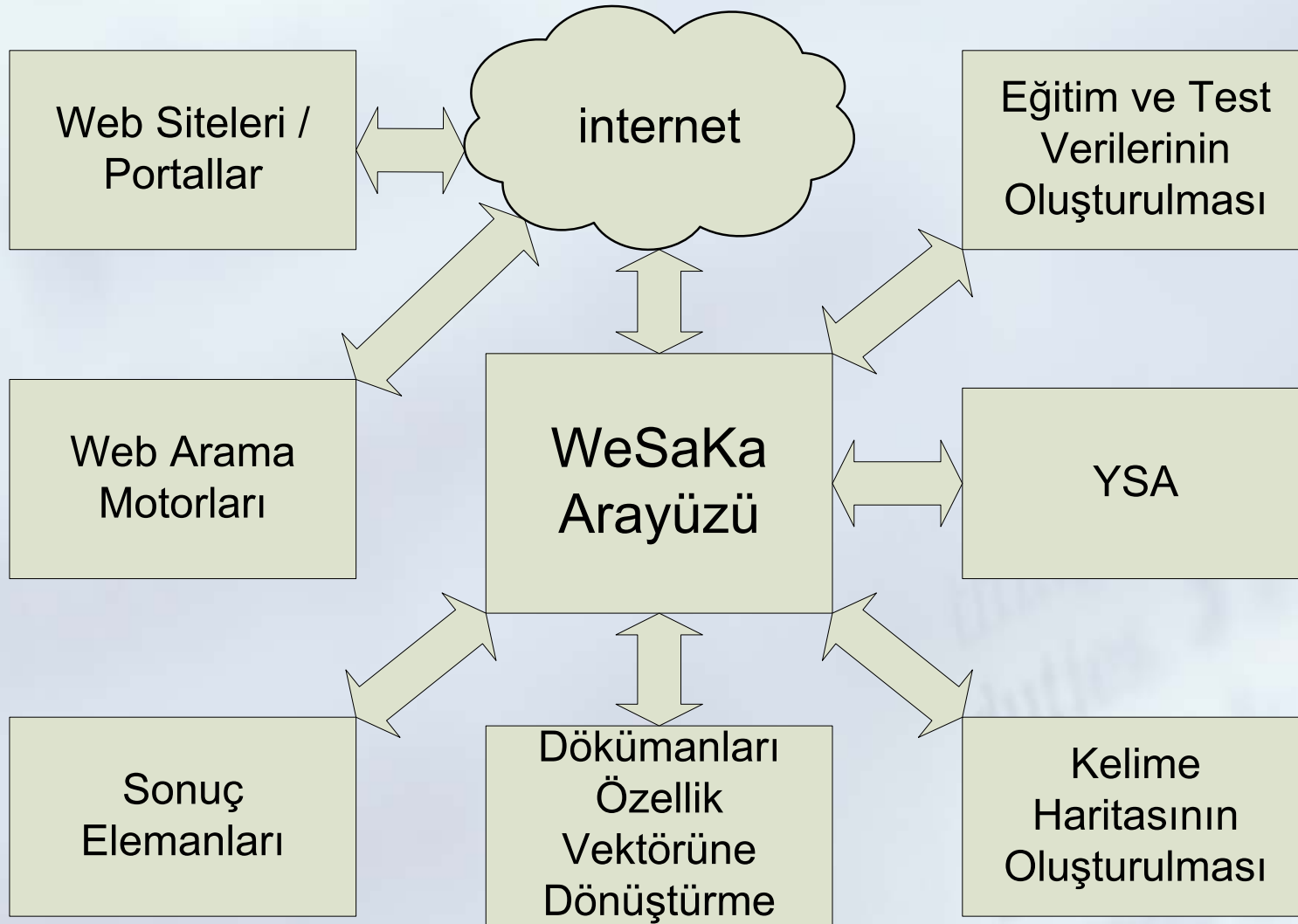
- Problemlere hızlı ve zeki çözüm
- Az veriyle genelleme
- Öğrenebilme yetisi
- Giriş ve çıkış verileri mevcut sistemlere genel bir model
- Karmaşık problemlere çözüm
- Farklı problemlere kolaylıkla uyarlanabilme

WeSaKa: YSA ile Web sayfası kategorizasyonu yapmak için geliştirilmiş bir sistemdir. Geliştirilmeye ve genişletilmeye açıktır.

Sınıflandırma:

- "spor",
- "ekonomi",
- "kültür"

WeSaKa Mimari Yapı



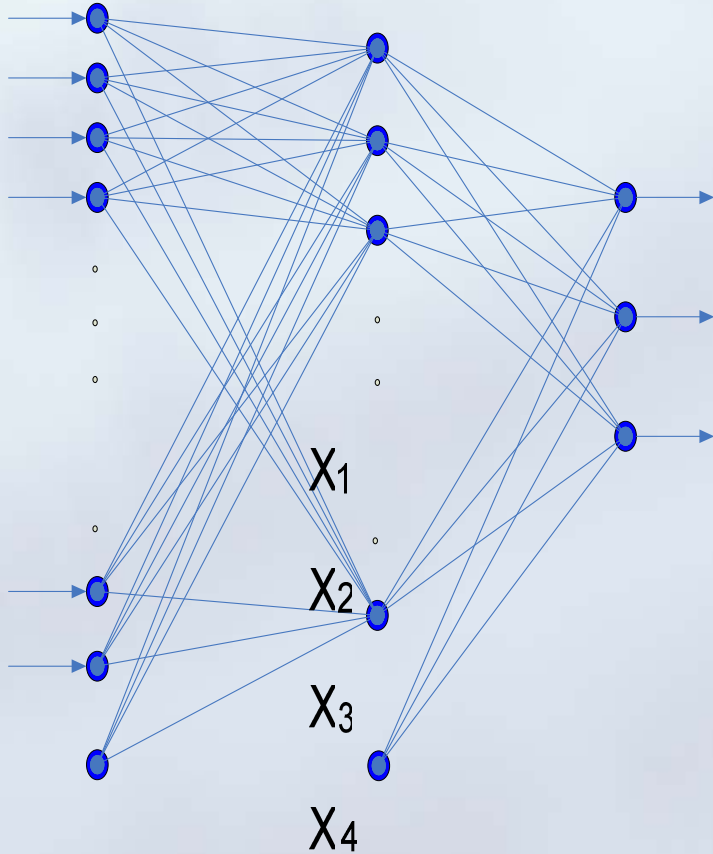
Yapı ve Uygulama

- Microsoft .NET 1.1 platformu
- VB.NET programlama dili
- YSA Alt motoru
- YSA öğrenme algoritması
- MATLAB entegrasyonu
- Web arayüz tasarımı

Kelime Haritası (özellik vektörü)

Kategori 1 (spor)		Kategori 2 (ekonomi)		Kategori 3 (kültür)	
KELİME	TEKRARLAMA	KELİME	TEKRARLAMA	KELİME	TEKRARLAMA
fark	24	devlet	43	aşk	15
fikstür	33	düşüş	10	bahar	5
forma	10	düzey	13	bale	9
futbol	34	eğilim	8	belgesel	5
galibiyet	8	ekonomi	143	film	17
gol	38	enflasyon	12	güzel	9
hakem	32	fırma	8	kitabı	11
kart	23	fiyat	15	klasik	5
lider	7	fuvar	2	konser	11
lig	119	ihale	13	koro	11
maç	57	ihraç	3	müzi	11
menajer	4	kalkınma	19	ödül	21
pozisyon	6	kamu	14	öykü	24
puan	63	petrol	13	resim	20
saha	16	taahhüt	5	sergi	12
spor	171	tutma	3	şarki	5
stad	9	ürün	28	şiir	14
şampiyon	25	yatırım	129	tiyatro	21
teşvik	5	yükseliş	4	tür	74
transfer	11	yüzde	124	yaz	23

YSA Bilgileri



- üç kategoriden 10'ar adet belge için toplamda 30 örnekten oluşan bir eğitim kümesi
- 60 Giriş, 3 Çıkış, 20'lik Gizli Katman
- 3 Çıkış: Spor, Kültür ve Ekonomi göstergeleri
- 60 Giriş: Özel kelimeler, dokümanda geçiş sayısı / dokümandaki toplam kelime sayısı
- Levenberg-Marquardt öğrenme algoritması kullanılarak eğitim
- Sigmoid transfer fonksiyonu
- Eğitim sırasında nöronlar arasındaki ağırlıklara ilk değer olarak $[-1, +1]$ arasında rastgele
- 350 tekrardan sonra eğitim tamamlanmış, toplam mutlak hata oranı $4.5e-10$

WeSaKa Arayüzü

URL : www.milliyet.com.tr

Sports : 0.0000
Art : 0.0000
Economy : 1.0000

Navigation Complete

Milliyet.com.tr
BASINDA GÜVEN

İnsankaynaklari.com
Alimoğulları Orman ürünleri Mobilya İR
İHRACAT SORUMLUSU

Sık Kullanılanlara Ekle | Açılış Sayfası Yap | Sitene Ekle | İletişim | Kurumsal | RSS

21 Ekim 2007 / Pazar

Bugünkü Milliyet | Cafe | Blog | Emlak | Arabam | İnsan Kaynakları | Cumartesi | Pazar | Ege | Seri İlanlar

Okur Yorum | Haber İndeksi | Arşiv | Haberci | Üyelik | Bülten | Kampanya Sözleşmesi | Hızlı Erişim

Arşiv Ara
Detaylı Arama ara

BABA BENİ
OKULA GÖNDER

EKONOMİ


TÜSİAD: YÜKSEK BÜYÜME, İSTİHDAM VE DEZENFLASYON DEVAM ETTİRİLİRSE...

Enflasyon hedefinde biraz gecikme çok önemli değil

TÜSİAD'ın Başbakan Erdoğan'a sunduğu raporda, maliye politikasında sıkılaştırmaya devam önerilirken, para politikasında, "Bütçe disiplini ve dezenflasyon patikası veri iken, gevşetileme imkânları dikkate alınmalıdır" denildi

EKONOMİ SERVİSİ

Türk Sanayicileri ve İşadamları Derneği'nin (TÜSİAD), Başbakan Erdoğan'a sunduğu raporda, makroekonomik istikrarın sağlanması sayesinde enflasyonla mücadeleyi devam ettirebilmenin, geçmiş dönemde olduğu gibi gelecek dönemde de birbirini destekleyecek bir süreç olduğu konusunda bazı çekinceler bulunduğu, ancak esas önemli olanın, dezenflasyon sürecinin devam etmesi olduğu kaydedildi. Belgede, "yüzde 4 enflasyon hedefine planlanan tarihten biraz daha geç ulaşılmasının, yüksek büyümenin, istihdam yaratmanın ve dezenflasyon sürecinin devam



Name	Category	IncidenceCou
fark	1	24
fiktür	1	33
forma	1	10
futbol	1	34
galibiyet	1	8
gol	1	38
hakem	1	32
kart	1	23
lider	1	7
lig	1	119
maç	1	57
menajer	1	4
pozisyon	1	6
puan	1	63
saha	1	16
spor	1	171
stad	1	9
şampiyon	1	25
teşvik	1	5
transfer	1	11
devlet	2	43
düşüş	2	10
düzey	2	13
eğilim	2	8
ekonomi	2	143
enflasyon	2	12
firma	2	8
fiyat	2	15
fuat	2	2

Sport

Result_Sport Result_Econo Result_Art aşk

Save ANNT Sport Generate ANN Test Values

Generate ANN Input Values

YSA'nın Test Edilmesi

- 4 Haber sitesi ile test
 - <http://www.hurriyet.com.tr>
 - <http://www.milliyet.com.tr>
 - <http://www.zaman.com.tr>
 - <http://www.ntvmsnbc.com>

Test Sonuçları

<i>Web Sitesi</i>	<i>Kategori</i>	<i>Ölçülen Değerler</i>		
		<i>Spor Değeri</i>	<i>Ekonomi Değeri</i>	<i>Kültür Değeri</i>
http://www.hurriyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.hurriyet.com.tr	Spor	0.9998	0.0000	0.1191
http://www.hurriyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.hurriyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.hurriyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.milliyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.milliyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.milliyet.com.tr	Spor	1.0000	0.0000	0.0000
http://www.milliyet.com.tr	Spor	0.9995	0.0032	0.0000
http://www.milliyet.com.tr	Spor	0.9754	0.0000	0.0026
http://www.zaman.com.tr	Spor	0.9982	0.0000	0.5260
http://www.zaman.com.tr	Spor	0.2371	0.0000	0.9742
http://www.zaman.com.tr	Spor	0.3084	0.0000	0.9936
http://www.zaman.com.tr	Spor	1.0000	0.0000	0.0000
http://www.zaman.com.tr	Spor	0.9984	0.0000	0.3168
http://www.ntvmsnbc.com	Spor	1.0000	0.0000	0.0120
http://www.ntvmsnbc.com	Spor	1.0000	0.0000	0.0000
http://www.ntvmsnbc.com	Spor	0.7847	0.0000	0.9980
http://www.ntvmsnbc.com	Spor	1.0000	0.0000	0.0000

Test Sonuçları

Web Sitesi	Kategori	Ölçülen Değerler		
		Spor Değeri	Ekonomi Değeri	Kültür Değeri
http://www.hurriyet.com.tr	Ekonomi	0.0000	0.9781	0.0027
http://www.hurriyet.com.tr	Ekonomi	0.0000	1.0000	0.0000
http://www.hurriyet.com.tr	Ekonomi	0.0000	0.9986	0.0001
http://www.hurriyet.com.tr	Ekonomi	0.0000	0.9835	0.0018
http://www.hurriyet.com.tr	Ekonomi	0.0000	0.9995	0.0259
http://www.milliyet.com.tr	Ekonomi	0.0000	1.0000	0.0001
http://www.milliyet.com.tr	Ekonomi	0.0000	1.0000	0.0000
http://www.milliyet.com.tr	Ekonomi	0.0000	0.8289	0.8330
http://www.milliyet.com.tr	Ekonomi	0.0000	0.9845	0.0022
http://www.zaman.com.tr	Ekonomi	0.0000	0.8757	0.0447
http://www.zaman.com.tr	Ekonomi	0.0000	0.0383	0.9242
http://www.zaman.com.tr	Ekonomi	0.0000	1.0000	0.0000
http://www.zaman.com.tr	Ekonomi	0.0000	0.9752	0.0012
http://www.zaman.com.tr	Ekonomi	0.0000	1.0000	0.0023
http://www.ntvmsnbc.com	Ekonomi	0.0022	0.1447	0.0028
http://www.ntvmsnbc.com	Ekonomi	0.0001	0.9997	0.0000
http://www.ntvmsnbc.com	Ekonomi	0.0000	1.0000	0.0000
http://www.ntvmsnbc.com	Ekonomi	0.0000	1.0000	0.0000
http://www.ntvmsnbc.com	Ekonomi	0.0012	0.8720	0.0134
http://www.ntvmsnbc.com	Ekonomi	0.0095	0.9642	0.0000

Test Sonuçları

Web Sitesi	Kategori	Ölçülen Değerler		
		Spor Değeri	Ekonomi Değeri	Kültür Değeri
http://www.hurriyet.com.tr	Kültür Sanat	0.0072	0.0000	0.9994
http://www.hurriyet.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.hurriyet.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.hurriyet.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.hurriyet.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.milliyet.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.milliyet.com.tr	Kültür Sanat	0.0001	0.0000	1.0000
http://www.milliyet.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.milliyet.com.tr	Kültür Sanat	0.0206	0.0000	0.9941
http://www.zaman.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.zaman.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.zaman.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.zaman.com.tr	Kültür Sanat	0.0001	0.0000	1.0000
http://www.zaman.com.tr	Kültür Sanat	0.0000	0.0000	1.0000
http://www.ntvmsnbc.com	Kültür Sanat	0.0041	0.0000	1.0000
http://www.ntvmsnbc.com	Kültür Sanat	0.0000	0.0000	1.0000
http://www.ntvmsnbc.com	Kültür Sanat	0.0441	0.0000	0.9843
http://www.ntvmsnbc.com	Kültür Sanat	0.0000	0.0000	1.0000
http://www.ntvmsnbc.com	Kültür Sanat	0.0001	0.0000	0.9993
http://www.ntvmsnbc.com	Kültür Sanat	0.0000	0.0000	1.0000

Sonuçlar ve Kazanımlar

Geliştirilen esnek yapı sayesinde; uygulama sırasında belirlenen kategorilerin sayısında ve tanımlamasında herhangi bir sınır yoktur, arttırılabilir.

Fakat yeni kategoriler tanımlandıkça;

- Özellik vektörü yenilenmeli,
- Yeni eğitim ve test vektörleri oluşturulmalı,
- YSA yapısının giriş sayısı özellik vektörüne güncellenmeli,
- Çıkış sayısı kategori sayısına göre değiştirilmeli
- Eğitim yenilenmelidir.

Sonuçlar ve Kazanımlar

Kategori içeriklerinin birbirine yakınlığı sonuçlardaki başarı oranını etkilemekte

Kategoriler	Spor	Ekonomi	Kültür Sanat
Spor	0,915075	0,000160	0,197115
Ekonomi	0,000650	0,882145	0,092720
Kültür Sanat	0,003815	0,000000	0,998855

Sonuçlar ve Kazanımlar

WeSaKa sistemi geliştirilmiş ve başarıyla sunulmuştur.

- Başarı
- Zaman
- Otomatik
- Doğru bilgi
- Güvenlik
- Kullanılabilirlik
- Esneklik
- Uyarlanabilirlik

Sonuçlar ve Kazanımlar

Karşılaşılan güçlükler

- Esnek yapı geliştirilmesinin zorluğu,
- Sınıf değişiminde eğitim gerekliliği
- Online işlem yapma zorluğu
- Bilgilendirme metinlerinin yanıtıcı olması
- Özellik vektörü çıkartılması

Gereğinden fazla olması sistemin öğrenmesini zorlaştırmakta,

Gereğinden az olması YSA sonuçlarındaki hata oranında artmasına sebep olmakta.

TEŞEKKÜR EDERİM.
eng@gazi.edu.tr