

# BİÇİMBİLİMSEL BELİRSİZLİK GİDERME PROBLEMİNİN ÇOKLU SINIFLANDIRMA PROBLEMİNE DÖNÜŞTÜRÜLÜREK ÇÖZÜLMESİ



İSTANBUL TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR VE BİLİŞİM FAKÜLTESİ



OZAN ARKAN CAN

040090573

DANIŞMAN: YRD. DOÇ. DR. GÜLŞEN CEBİROĞLU ERYİĞİT

22 OCAK 2014

# İÇERİK

- Giriş
- İlgili Çalışmalar
- Kavramlar
- Veri
- Kullanılan Yöntem
- Testler & Sonuçlar
- Sonuç & Öneriler
- Soru & Cevap

# GİRİŞ

- ❖ Biçimbilimsel çözümleme bir sözcüğün en küçük anlamlı yapı birimlerinin (morfem) ve kökünün bulunması işlemidir.
- ❖ Çoğu dil için çözümleme işlemi sözcüklerin çözümlemesinin listelenerek, gerektiğinde sorgulanmasıyla gerçekleştirilmektedir.  
Örnek: İngilizce: “çiçek” -> tekil (flower) ve çoğul (flowers)
- ❖ Tekil ve çoğul hallerinin listelenmesi çözümleme işlemi için yeterlidir

- ❖ Türkçe karmaşık bir dil yapısına sahip, sondan eklemeli bir dildir.
- ❖ Teorik olarak bir kelimenin sonuna sonsuz sayıda ek gelebilmektedir.

Örnek: ***arabaları***

araba + İsim + Çoğul + 3.tekil kişi iyelik

araba + İsim + Çoğul + -i hali

araba + İsim + Çoğul + 3.çoğul kişi iyelik

araba + İsim + Tekil + 3.çoğul kişi iyelik

❖ Bir sözcük için birden fazla çözümleme olabilmektedir.

❖ Geçerli çözümleme kelimenin bulunduğu bağlama göre değişmektedir.

Bu durum bilgisayar tarafından çözümleme yapılırken belirsizliğe sebep olmaktadır.

❖ Bir sözcük için bağlama bağlı olarak doğru çözümlemenin belirlenmesi işlemi *biçimbilimsel belirsizlik giderme* olarak nitelendirilmektedir.

# İLGİLİ ÇALIŞMALAR

1. Kural Tabanlı Sistemler

2. İstatistiksel Yöntemler

3. Hibrit Sistemler

# 1. KURAL TABANLI YÖNTEMLER

- Elle belirlenmiş dil kuralları
- Bağlam bilgisi
- Amaç doğru etiketi seçmek ya da bir sözcüğe etiket atamak
- Karlsson, F., Voutilainen, A., Heikkilä, J., Anttila A. (1995)
- Brill, E. (1992)
- Oflazer, K. and Tür, G. (1997)

## 2. İSTATİSTİKSEL YÖNTEMLER

- ❖ Hakkani-Tür, D.Z., Oflazer, K., Tür, G. (2002)
  - Kökler ve Etiketler için iki ayrı 3-gram dil modeli
  - Saklı Markov Modelleri yardımıyla en olası etiket dizilimi
- ❖ Sak, H., Güngör, T., Saraçlar, M. (2007)
  - Morfolojik özellikler *perceptron* yardımıyla ağırlıklandırılıyor
  - Viterbi algoritmasıyla en olası çözümleme seçiliyor



## 2. İSTATİSTİKSEL YÖNTEMLER

❖ Ehsani, R., Alper, M., E., Eryiğit, G., Adalı, E. (2012)

- Etiketler gruplara ayrılmakta
- Her grup için Koşullu Rassal Alanlar yöntemine dayalı model

❖ Görgün, O., Yıldız, O. T. (2011)

- Dizi sınıflandırması yerine çoklu sınıflandırma problemi
- Her etiket bir sınıf
- WEKA ile SVM, Karar Ağaçları, Naive Bayes, KStar vs.

### 3. HİBRİT YÖNTEMLER

#### ❖ Yüret, D., Türe, F. (2006)

- Kuralları içeren karar listeleri
- Karar listelerinden elde edilen istatistiksel bilgiyle doğru çözümleme seçilmektedir

#### ❖ Kutlu, M, Çiçekli, İ. (2013)

- İstatistiksel model yardımıyla aday etiketler belirlenmekte
- Bu adaylar elle belirlenmiş kurallar yardımıyla sınırlandırmakta
- Kelime ve eklerin istatistiksel bilgilerinden edindikleri kuralları ve yine bu istatistiğe dayalı geliştirdikleri modeli yinelemeli bir şekilde kullanılmaktadır

# KAVRAMLAR

- ❖ **Morfem:** Dildeki en küçük anlamlı birimdir. “evdeki” sözcüğü için “ev”, “-de”, “-ki” birimleri dört ayrı morfemdir.
- ❖ **Kök:** Bir sözcüğün eklerinden arındırıldıktan sonra kalan anlamlı birimdir.
- ❖ **Çekim Kümesi:** Bir yapım ekiyle başlayıp diğer yapım ekine kadar olan morfemlerin tümüne çekim kümesi denir.

❖ **Etiket:** Sözcüğün biçimbilimsel çözümlemesi yapıldıktan sonra kökün dışında kalan morfemlerinin bütününe etiket denmektedir.

Örnek: çiçekleri: çiçek+Noun+A3pl+Pnon+Acc

❖ **Makine Öğrenmesi [14]:** Bilgisayarların örnek veri ya da geçmiş deneyimi kullanarak bir ölçüte göre başarımlarını arttıracak biçimde programlanmasıdır.

- İstatistiki bilgi yardımıyla geliştirilen parametrik model
- Geçmiş veri ya da deneyimle bu modeli en iyilemek amaçtır
- Geliştirilen bu model, öngörücü ya da açıklayıcı olarak kullanılabilir.

- ❖ **Sınıflandırma:** Geliştirilen bir modele göre, bir girdi için modelin tasarlandığı problemdeki olası sınıflardan birinin atanmasına sınıflandırma denir.
- ❖ **Özellik ve Özellik Vektörü:** Bir nesneyi veya olayı sınıflandırmada kullanılacak o nesnenin veya olayın karakteristik değerlerinin her birine özellik denir [14]. Bu özellikler kullanılarak her bir örnek için bir özellik vektörü tanımlanabilir [22].

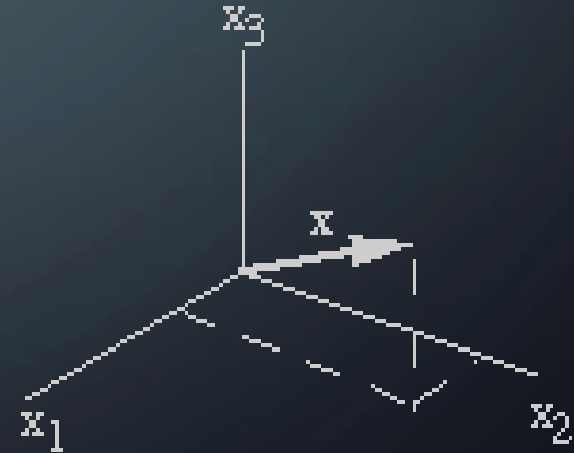
$X_1$  = Kelimenin cümlede geçtiği hali

$X_2$  = Kelimenin ana etiket sınıfı

...

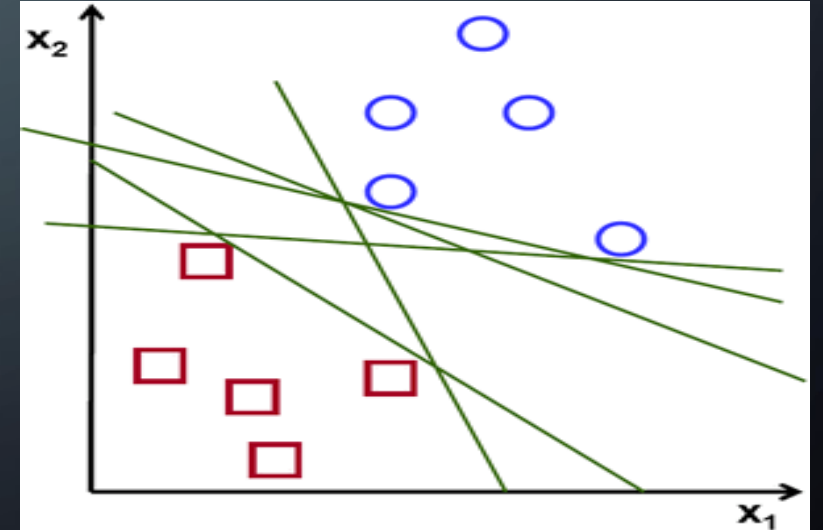
$X_d$  = Kelimenin son çekim kümesi

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$



## ❖ Karar Destek Makineleri:

- Örnek uzayının bir alt kümesini kullanarak, örnekleri bir çekirdek fonksiyonuyla ağırlıklandırarak doğrusal sınıflandırma yöntemidir.
- Ayırtaç tabanlı bir yöntemdir
- Ayırtıcı öğrenmek için  $p(x|C_i)$  sınıf olasılıklarını ya da  $P(C_i|x)$  sonsal olasılıklarını hesaplamak yerine  $P(C_i|x) = P(C_j|x)$  yapan sınır değerlerini belirlemek yeterlidir.
- Örneklerin bulunduğu sınıfları ayırt etmek için hiperdüzlem gerekir
- Birden fazla hiperdüzlem olabilmektedir.

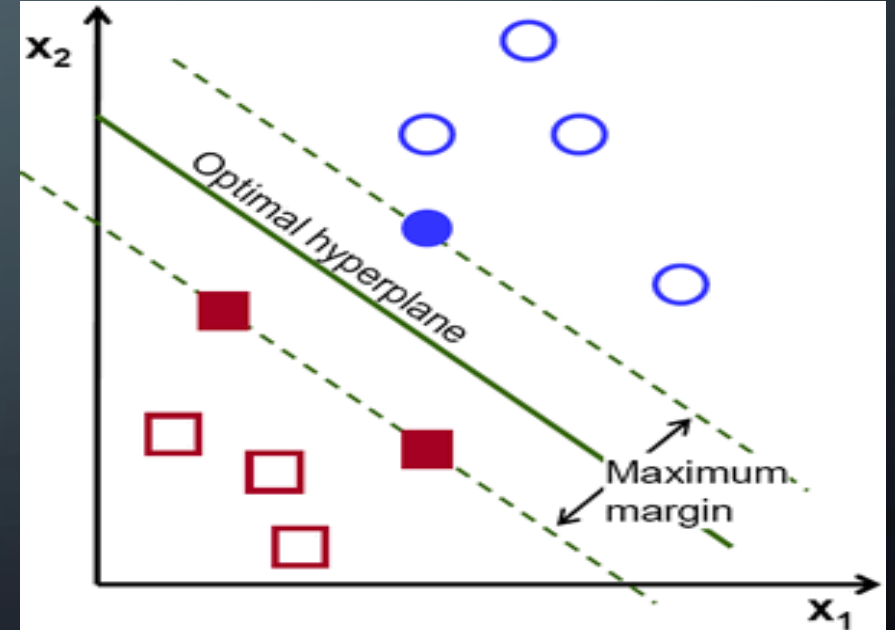


- En iyi ayırtedici düzlem tüm sınıflara en yakın uzaklıkta bulunan düzlemdir
- Problem doğrusal değilse, doğrusal olmayan bir model oturtmak yerine doğrusal olmayan taban işlevleriyle yeni bir uzaya gidip orada doğrusal bir model oluşturulabilir.

$\Phi: R^n \rightarrow R^m$  olmak üzere

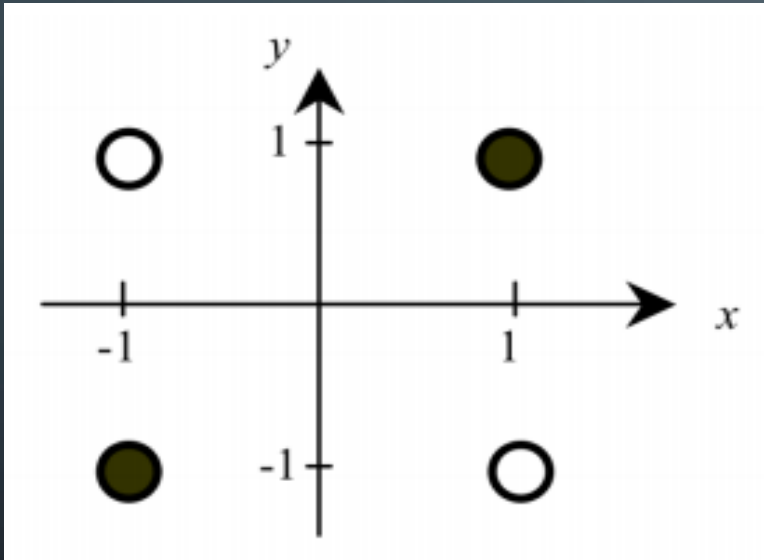
$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j)$$

- Çokterimli çekirdek
- Dairesel tabanlı (Gauss) işlev çekirdeği
- S biçimli çekirdek

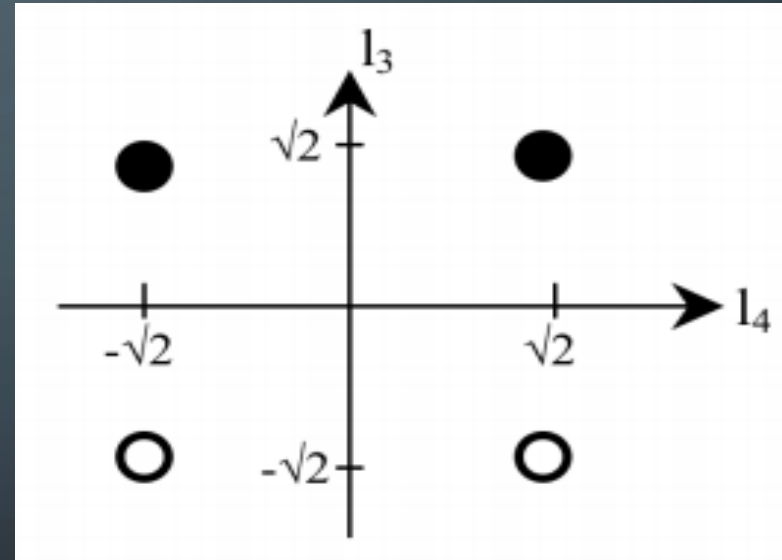


## ❖ XOR problemi

- $\Phi(x_1, x_2) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$



XOR probleminin koordinat sisteminde gösterimi



Çekirdek fonksiyonuyla farklı eksenlere taşınan girdiler



## ❖ *Entropi ve Enformasyon* [26]:

- Enformasyon, rassal bir olayın gerçekleşmesine ilişkin bilgi ölçütüdür.

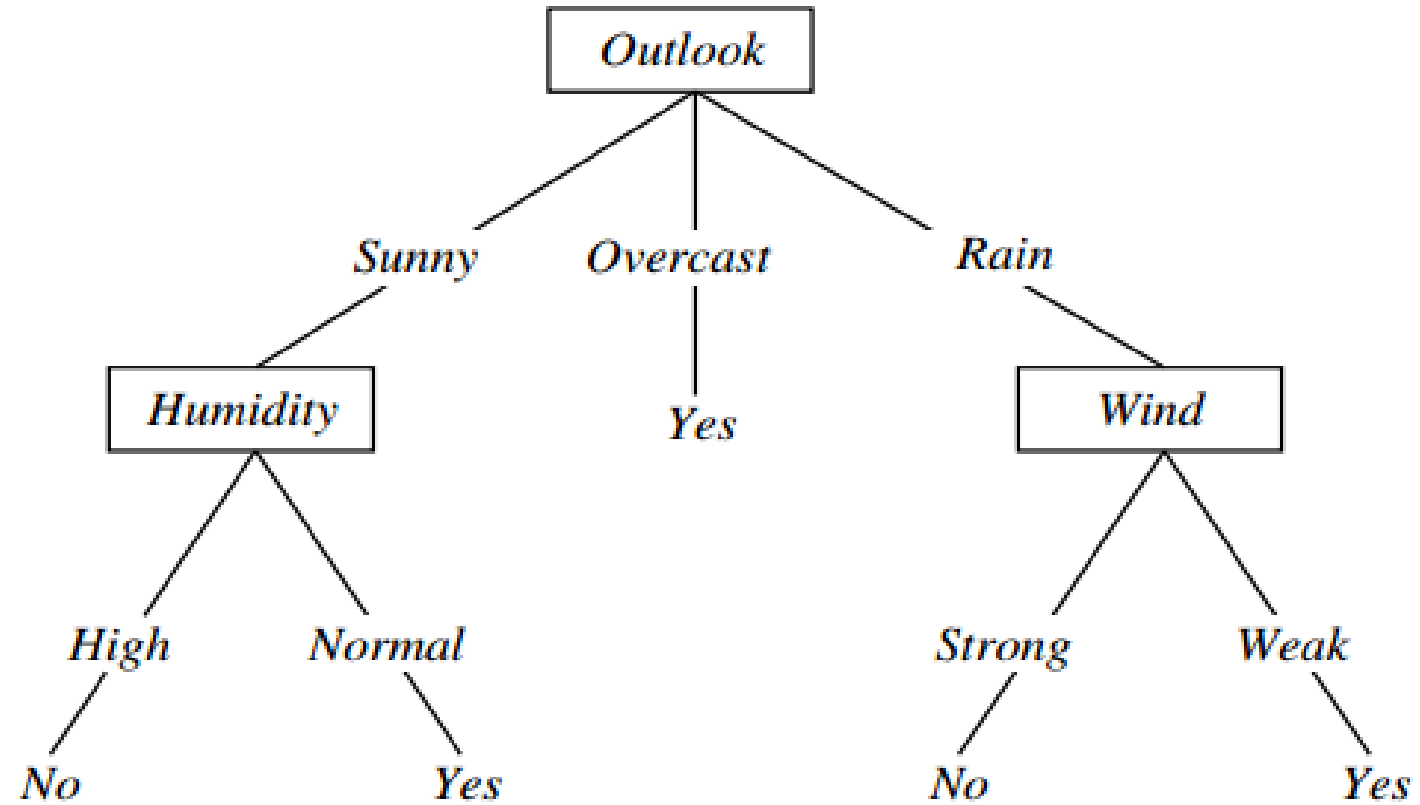
$$I(x) = \log \frac{1}{P(x)} = -\log P(x)$$

- Entropi, iletilen bir mesajın taşıdığı enformasyonun beklenen değeridir.

$$H(x) = E(I(X)) = \sum_{i=1}^n P(x_i)I(x_i) = -\sum_{i=1}^n P_i \log P_i$$

## ❖ *Karar Ağaçları* [26]:

- Böl ve yönet yaklaşımını gerçekleyen ağaç yapılı bir veri yapısıdır
- Karar ağaçları belirli özellik yardımıyla çok özellikli bir veriyi parçalara ayırır
- Her seferinde ağacın türüne göre en küçük entropiye dayalı bir yöntemle bu özellik seçilir.



Karar ağacı örneği

# VERİ

- Daha önceki çalışmalarda kullanılan veri Hakkani-Tür ve diğ. (2002)
- Eğitim ve Test (ORG ve STD)
- METU-Sabancı Ağaç Yapılı derlemi (MST)
- Eryiğit'in hazırlamış olduğu İTÜ onay test verisi (IVS)
- Eryiğit (2012) çalışmasında belirtilen yeni sürümleri
- Morfolojik çözümleyicinin değişen sürümlerindeki etiket farkları
- Noktalama işaretleri ve nümerik değerlerle ilgili düzeltmeler

# VERİ İSTATİSTİKLERİ

Veri	#Cümle	#Sözcük	#Belirsizliğe Sahip Sözcük
ORG	50674	837524	387857 (46.31%)
STD	42	862	357 (41.42%)
MST	5635	56424	26275 (46.57%)
IVS	300	3700	1942 (52.49%)

# KULLANILAN YÖNTEM

- ❖ Görgün ve Yıldız (2011)
- ❖ Dizi sınıflandırma problemi yerine çoklu sınıflandırma
  - Her etiket: *sınıf*
  - Bir kelimenin çözümlemesi için ortaya çıkan etiket kümesi : *problem*

**Kelime:** çocukları

**Çözümlemeler:**

çocuk+Noun+A3pl+Pnon+Acc  
çocuk+Noun+A3pl+P3sg+Nom  
çocuk+Noun+A3pl+P3pl+Nom  
çocuk+Noun+A3sg+P3pl+Nom

**Kelime:** kitapları

**Çözümlemeler:**

kitap+Noun+A3pl+Pnon+Acc  
kitap+Noun+A3pl+P3sg+Nom  
kitap+Noun+A3pl+P3pl+Nom  
kitap+Noun+A3sg+P3pl+Nom

Problem #1

Sınıf 0: Noun+A3pl+Pnon+Acc  
Sınıf 1: Noun+A3pl+P3sg+Nom  
Sınıf 2: Noun+A3pl+P3pl+Nom  
Sınıf 3: Noun+A3sg+P3pl+Nom

❖ Veri cümle cümle işleniyor

❖ Özellik vektörü çıkarılırken:

- Belirsizliği giderilmeye çalışılan sözcük  $w_i$  için
- $w_{i-1}$  ve  $w_{i-2}$  kelimelerinin çözümlemelerindeki tüm morfemler
- Var-Yok şeklinde ikilileştirilerek kullanılmıştır
- Eğer  $w_{i-1}$  ya da  $w_{i-2}$  belirsizse tüm çözümlemelerindeki morfemler alınmıştır

❖ Sınıflandırıcı

- Her problem için ayrı ayrı sınıflandırıcı eğitilmiştir
- Karar Destek Makineleri, Naive Bayes, Karar Ağaçları, KStar, K En Yakın Komşu

- ❖ Görgün ve Yıldız (2011) çalışmasındaki yöntem yeniden gerçekleştirilen
- ❖ Ele alınmayan iki durum görülmüştür

**1. Kök belirsizliği :** Kök kısmı hariç aynı etikete sahip çözümler

**Örnek Kelime:** sorunun

Sınıf 0: soru+Noun+A3sg+Pnon+Gen

Sınıf 1: soru+Noun+A3sg+P2sg+Gen

Sınıf 2: sorun+Noun+A3sg+P2sg+Nom

Sınıf 0: sorun+Noun+A3sg+Pnon+Gen

## 2. Test verisinde olup da eğitim verisinde olmayan problemler :

Eğitilmiş model bulunmuyor

Eğitim & Test Dosyaları	#Eğitimde Görülmüş Problemler	#Kök Belirsizliğine Sahip Problemler	#Eğitimde Görülmemiş Problemler
ORG - STD	336 (94,1%)	19 (5,3%)	2 (0,6%)
ORG - MST	25424 (96,8%)	375 (1,4%)	476 (1,8%)
ORG - IVS	1860 (95,8%)	43 (2,2%)	39 (2%)
MST - IVS	1791 (88,9%)	43 (2,1%)	180 (8,9%)
ORG + MST - IVS	1865 (96%)	43 (2,2%)	34 (1,8%)



Bu durumlara uygulanan çözüm yöntemleri

- ❖ En az sayıda morfem içeren etiket in seçilmesi
- ❖ En uzun gövdeye sahip çözümlemenin seçilmesi\*
- ❖ Karakter uzunluğu en fazla olan etiket in seçilmesi

## ❖ Yapılan geliřtirmeler:

- Özellik vektörü çıkarımında bir önceki ve iki önceki kelimenin tüm morfemleri yerine farklı yaklaşımlar
  - -2, +2 aralığındaki kelimelerin çözümlemelerindeki morfemler
  - Tüm çekim kümelerinin ya da sadece son çekim kümesinin kullanılması
  - Kök bilgisinin kullanılması
  - Özelliklerin sırasının deęiřtirilmesi
- Sınıflandırıcıda farklı çekirdek türleri
- $w_{i-1}$ ,  $w_i$ ,  $w_{i+1}$  kelimelerinin tüm çekim kümelerindeki morfemler ve kök bilgilerinin çok terimli çekirdek kullanılmasıyla oluşturulan model en başarılı sonucu göstermiştir

Geliştirilen yeni modellere ek olarak, yeni yaklaşımlar da denendi

### 1. Yaklaşım:

- ✓ Önceki yaklaşımda, özellik vektörü çıkarımı sırasında kullanılan  $w$  kelimesinde belirsizlik varsa tüm çözümlerler kullanılıyordu. Bunun yerine eğer belirsizliği giderildiyse seçilen çözümleme kullanıldı
- ✓ Yinelemeli bir şekilde cümlelerin belirsizliği giderildi
- ✓ Dizi sınıflandırmaya benzer bir yaklaşım

## 2. Yaklaşım: Hasim ve diğ. (2007) aracıyla kaskat bir yapı

- ✓ Test edilecek veri önce bu araçtan geçirildi
- ✓ Özellik vektörü belirlenirken belirsizliği giderilecek kelime dışındaki kelimeler için bu aracın çıktıları kullanıldı
- ✓ Yeni belirlenen çözümleme eskisinin yerine konuldu

### 3. Yaklaşım:

- ✓ Bu aşamaya kadar bahsedilen yaklaşımlarda Karar Destek Makineleri kullanılmaktadır. Bu sınıflandırıcı yerine Karar Ağaçları kullanan modeller denenmiştir
- ✓ Özellik vektörü çıkarımında kök bilgisi kullanılırsa vektörün boyutu aşırı arttığı için eğitim ve test tamamlanamamakta (+20000 - 60000)
- ✓ Karar Destek Makineleri kullanan modellerden daha başarısızdır

#### 4. Yaklaşım: Problemlerin birleştirilmesi

Önceki yaklaşımlarda karşılaşılan sorunlar:

- Problemlerin bir kısmı için eğitim setinde çok az örnek bulunmaktadır
  - Bazı problemler diğer bir problemin içerisinde bulunabilmektedir
  - Sınıflar kendi içlerinde benzer morfemlere sahip olabilmektedir
- ✓ Bu tarz problemler birleştirilerek daha fazla bilgi kazanımı sağlanmaya çalışıldı

- ✓ Problemlerin içerdiği sınıflardaki morfemlerden küme yapısında bir vektör oluşturuldu
- ✓ Problemlerin içerdiği ortak morfem sayısı, problem vektörlerinin uzunluklarına bölünerek, problemlerin benzerliğine dair bir değer elde edildi
- ✓ Belli bir eşik değerin üstünde benzerlik gösteren problemler birleştirildi (0,75 veya 0,80)

#### Problem 1:

Sınıf 0: Noun+A3sg+Pnon+Nom

Sınıf 1: Noun+A3sg+P3sg+Nom

Vektör: [Noun, A3sg, **P3sg**, Pnon, Nom]

$$\text{Problem 1'in benzerlik oranı} = \frac{4}{5} = 0.8$$

$$\text{Problem 2'nin benzerlik oranı} = \frac{4}{5} = 0.8$$

#### Problem 2:

Sınıf 0: Noun+Prop+A3sg+Pnon+Nom

Sınıf 1: Noun+A3sg+Pnon+Nom

Vektör: [Noun, A3sg, **Prop**, Pnon, Nom]

Sınıf 0: Noun+A3sg+Pnon+Nom

Sınıf 1: Noun+A3sg+P3sg+Nom

Sınıf 2: Noun+Prop+A3sg+Pnon+Nom

# DEĞERLENDİRME METRİĞİ

$$\checkmark \text{Başarım} = \frac{\text{Doğru etikete sahip sözcük sayısı}}{\text{Sözcük sayısı}}$$

$$\checkmark \text{Başarım} = \frac{\text{Doğru etikete sahip belirsizliğe sahip sözcük sayısı}}{\text{Belirsizliğe sahip sözcük sayısı}}$$

$$\checkmark \text{Başarım} = \frac{\text{Doğru etikete sahip kök belirsizliğine sahip sözcük sayısı}}{\text{Kök belirsizliğine sahip sözcük sayısı}}$$

$$\checkmark \text{Başarım} = \frac{\text{Doğru etikete sahip eğitimde görülmemiş problem sayısı}}{\text{Eğitimde görülmemiş problem sayısı}}$$



# TEST SONUÇLARI

❖ Eğitilmiş model ile çözülemeyen problemler için geliştirilen yöntemlerin karşılaştırması

Sırasıyla :

- ✓ Bütün sözcükler üzerindeki başarım
- ✓ Belirsizliğe sahip kelimeler üzerindeki başarım
- ✓ Test setinde olup eğitim setinde de olan problemlerdeki başarım
- ✓ Kök belirsizliğine sahip problemlerdeki başarım
- ✓ Test setinde olup da eğitim setinde olmayan problemlerdeki başarım gösterilmiştir.

## Yeni problem türleri için geliştirilen sistemlerin değerlendirmesi

Model	ORG - STD	ORG - MST	ORG - IVS	MST - IVS	ORG+MST - IVS
[11] Gorgun & Yıldız	93.04   83,19   88,4   0   0	86,74   74,62   77.12   0   0	87,43   77,81   81.24   0   0	86,97   76,93   83.42   0   0	87,51   77,96   81.18   0   0
+ En az sayıda morfem	94,55   86,83   88,4   13/19   0/2	87,47   76,19   77.12   111/375   301/476	88,54   79,92   81.24   11/43   30/39	89,03   80,84   83.42   11/43   65/108	88,57   79,97   81.18   11/43   28/34
+ En uzun kök	94,2   85,99   88.4   10/19   0/2	87,41   76,07   77.12   111/375   269/476	88,92   80,64   81.24   20/43   35/39	89,43   81,62   83.42   20/43   71/108	88,86   80,54   81.18   20/43   30/34
+ En uzun etiket	93,74   84,87   88.4   6/19   0/2	87,02   75,22   77.12   87/375   69/476	87,73   78,37   81.24   7/43   4/39	87,51   77,96   83.42   7/43   13/108	87,76   78,42   81.18   7/43   2/34

# Geliştirilen yöntemlerin değerlendirilmesi

Model	ORG - STD		ORG - MST		ORG - IVS		MST - IVS		ORG+MST - IVS	
[11] Görgün & Yıldız	93,04	83,19	86,74	74,62	87,43	77,81	86,97	76,93	87,51	77,96
$w_{i-2}, w_{i-1}$ Son ÇK	93,16	83,47	86,71	74,57	87,49	77,91	86,89	76,78	87,62	78,17
$w_{i-2}, w_{i-1}$ Son ÇK + $r_{i-2}, r_{i-1}$	93,16	83,47	86,86	74,88	87,3	77,55	86,97	76,93	87,43	77,81
$w_{i-2}, w_{i-1}, w_i$ Tüm ÇK	93,27	83,75	86,74	74,61	87,46	77,86	86,92	76,83	87,43	77,81
$w_{i-2}, w_{i-1}, w_i$ Tüm ÇK + $r_{i-2}, r_{i-1}, r_i$	92,69	82,35	86,41	73,91	86,92	76,83	87,08	77,14	86,7	76,42
$w_{i-1}, w_i, w_{i+1}$ Tüm ÇK	93,27	83,75	87,19	75,58	87,51	77,96	87,41	77,75	87,76	78,42
$w_{i-2}, w_{i-1}, w_{i+1}$ Tüm ÇK	93,27	83,75	87,11	75,42	87,51	77,96	87,3	77,55	87,65	78,22
$w_{i-2}, w_{i-1}, w_i, w_{i+1}$ Tüm ÇK	90,84	77,87	86,29	73,67	86,95	76,88	86,65	76,31	87,3	77,55
$w_{i-1}, w_i, w_{i+1}$ Tüm ÇK + $r_{i-1}, r_i, r_{i+1}$	92,11	80,95	86,71	74,55	87,43	77,81	87,38	77,7	87,7	78,32
$r_{i-1} + w_{i-1}, w_i, w_{i+1}$ Tüm ÇK	93,39	84,03	87,24	75,69	87,51	77,96	87,65	78,22	87,92	78,73
$r_i, w_i, w_{i-1}, r_{i-1}, r_{i+1}, w_{i+1}$ + En uzun kök	95,48	89,08	88,41	78,21	89,14	81,05	90,22	83,11	89,27	81,31

❖ Özellik vektörü çıkarımında bir çok model denenmiştir

❖ En başarılı olan model:

- $w_{i-1}$ ,  $w_i$ ,  $w_{i+1}$  kelimelerinin tüm çekim kümelerindeki morfemlerinin kullanılması
- Kök bilgisinin eklenmesi
- Sıralama:
  - ✓  $r_i$
  - ✓  $w_i$
  - ✓  $w_{i-1}$
  - ✓  $r_{i-1}$
  - ✓  $r_{i+1}$
  - ✓  $w_{i+1}$
- Çok terimli çekirdek

## Geliştirilen sistemin daha önceki araçlarla karşılaştırması

Model	ORG - STD		ORG - MST		ORG - IVS		Eğitim Süresi (ms)	Test Süresi(ms)
[8] Hasim ve diğ.	96,29	91,28	88,74	78,93	89,81	82,35	1989	2   30   4
[9] Yüret ve Türe	95, 82	90,19	87,84	77,01	88,57	79,98	840	10   700   53
[7] Tür ve diğ.	94,43	86,55	85,84	72,69	87,08	77,14	94	1   4   1
[11] Görgün ve Yıldız	93,04	83,19	86,74	74,62	87,43	77,81	311	7   210   12
$r_i, w_i, w_{i-1}, r_{i-1}, r_{i+1}, w_{i+1} + \text{En uzun kök}$	95,48	89,08	88,41	78,21	89,14	81,05	329	7   365   14

- Sistemin başarısı en başarılı araç olan [8] başarımına  $\sim\%1$  yakınlıkta
- STD test verisi dışında [9] aracından daha başarılı
- Veriyi parçalara bölüp eğittiği için eğitim süresi [7] aracı dışında diğer yaklaşımlardan hızlı
- Test aşamasında ise problemler tek tek arandığı için [9] aracı dışında yavaştır

## ❖ Sınanan farklı yaklaşımların değerlendirilmesi

Model	ORG - STD		ORG - MST		ORG - IVS	
Hasim [8]	96,29	91,28	88,74	78,93	89,81	82,35
Görgün [11]	93,04	83,19	86,74	74,62	87,43	77,81
$r_i, w_i, w_{i-1}, r_{i-1}, r_{i+1}, w_{i+1}$ + En uzun kök	95,48	89,08	88,41	78,21	89,14	81,05
Yinelemeli Yöntem: Karar destek makineleri	94,66	87,11	87,67	76,62	89,16	81,1
Hasim + $r_i, w_i, w_{i-1}, r_{i-1},$ $r_{i+1}, w_{i+1}$ + En uzun kök (KDM)	95,71	89,64	88,75	78,95	89,7	82,13
$w_{i-2}, w_{i-1}$ Tüm ÇK + En uzun kök: Karar ağacı	91,18	78,71	85	70,88	85,3	73,74
$w_{i-1}, w_i, w_{i+1}$ Tüm ÇK + En uzun kök: Karar ağacı	88,28	71,71	84,04	68,82	84,57	72,35
Yinelemeli Yöntem: Karar ağacı	89,79	75,35	85	70,88	85,3	73,74
Hasim + $w_{i-2}, w_{i-1}$ + En uzun kök (Karar Ağacı)	90,49	77,03	85,39	71,73	85,92	75,46
Birleştirilmiş Problemler (0.75)	95,24	88,52	87,18	75,58	88,16	79,2
Birleştirilmiş Problemler (0.8)	95,01	87,96	87,02	75,10	87,95	78,62

# SONUÇ VE ÖNERİLER

- ✓ Temel alınan sistemden daha başarılı
- ✓ Var olan sistemlere yakın başarımda
- ✓ Morfoloji açısından zengin dillere uyarlanabilir
- ✓ Problemlere daha yakında bakılıyor
- ✓ Üretken Sınıflandırıcı + Ayırt Edici Sınıflandırıcı kullanılabilir

# SORU & CEVAP





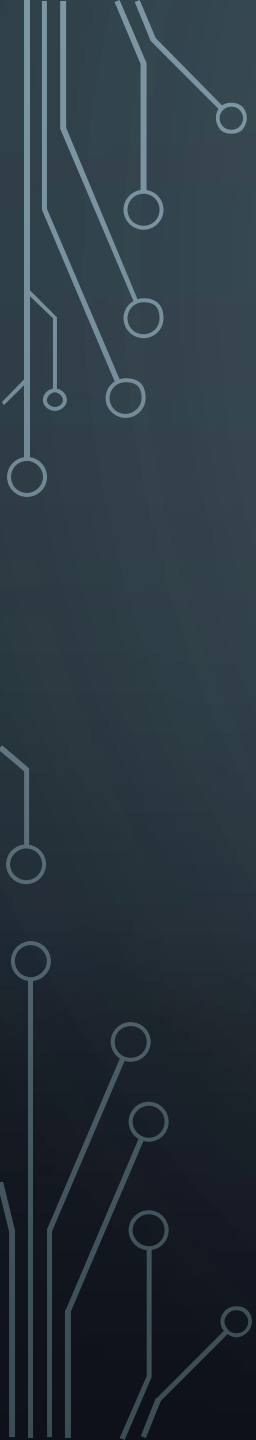
# KAYNAKLAR

[1] Gülşen Eryiğit. 2013. Biçimbilimsel Çözümleme . Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 6:33–55.

[2] Koskenniemi, K. (1983). Two-level morphology: a general computational model for word-form recognition and production. Department of General Linguistics. Helsinki: University of Helsinki .

[3] Oflazer, K.: Two-level description of Turkish morphology. Lit. Linguist. Comput. 9(2), 137–148 (1994)

[4] Karlsson, F., Voutilainen, A., Heikkilä, J., Anttila A.: Constraint Grammar-A Language-Independent System for Parsing Unrestricted Text (1995)



[5] Brill, E.: A Simple Rule-Based Part-of-Speech Tagger. Proceedings of Third Conference on Applied Natural Language Processing, Trento, Italy (1992)

[6] Oflazer, K. and TÜR, G. (1997). Morphological disambiguation by voting constraints. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL97, EACL97), Madrid, Spain.

[7] Hakkani-Tür, D.Z., Oflazer, K., TÜR, G.: Statistical morphological disambiguation for agglutinative languages. Comput. Humanit.36(4), 381–410 (2002)

[8] Sak, H., Güngör, T., Saraçlar, M.: Morphological disambiguation of Turkish text with perceptron algorithm. In: Gelbukh, A. (ed.) CCLING 2007, LNCS 4394, pp. 107–118 (2007)

[9] Yüret, D., Türe, F.: Learning morphological disambiguation rules for Turkish. In: Proceedings of HLT-NAACL (2006)





[10] Razieh Ehsani, Muzaffer Ege Alper, Gülşen Eryiğit, Eşref Adalı. 2012. Disambiguating Main POS tags for Turkish. In Proceedings of the 24th Conference on Computational Linguistics and Speech Processing (ROCLING 2012), Chung-Li, Taiwan, 21-22 September 2012.

[11] Gorgun, O, O. T. Yildiz, "A Novel Approach to Morphological Disambiguation for Turkish", International Conference on Computer and Information Sciences (ISCIS), pp. 77-83, London, UK, 2011.

[12] Kutlu, M, Çiçekli, İ. 2013 A Hybrid Morphological Disambiguation System for Turkish. Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 1230-1236, Nagoya, Japan.

[13] Gülşen Eryiğit. The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish. In Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, May 2012.

[14] E. Alpaydın, *Yapay Öğrenme*, Boğaziçi Üniversitesi Yay. 2011.



[15] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

[16] Oflazer, K., Hakkani-Tür, D. Z., Tür, G.: Design for a Turkish treebank. In: Proceedings of the Workshop on Linguistically Interpreted Corpora (1999)

[17] Gülşen Eryiğit. ITU Validation Set for Metu-Sabancı Turkish Treebank, Unpublished material, March 2007.

[18] Gülşen Eryiğit, “Doğal Dilde İstatistiksel Yöntemler”, 2011.

[19] McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit.«

[20] Git: Versiyon kontrol aracı. 2013

<http://git-scm.com/>

[21] GitHub: Versiyon kontrol için çevrimiçi depolama alanı. 2013

<https://github.com/>

[22] Perry Cook, “Human-Computer Interface Technology”, 2008

<https://www.cs.princeton.edu/courses/archive/fall08/cos436/>

[23] Duncan Fyfe Gillies, “Probabilistic Inference”, 2013

<http://www.doc.ic.ac.uk/~dfg/ProbabilisticInference/IDAPILecture18.pdf>

[24] OpenCV, “Support Vector Machines, 2013

[http://docs.opencv.org/doc/tutorials/ml/introduction\\_to\\_svm/introduction\\_to\\_svm.html](http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html)

[25] Eclipse IDE, 2013

<http://www.eclipse.org/>

[26] Umut Orhan, “Makine Öğrenmesi”, 2013

<http://bmb.cu.edu.tr/uorhan/DersNotu/Ders03.pdf>