

# Kelime Anlamları

Mehmet Fatih AMASYALI



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

## İçerik

- WORDNET
- Wordnet ile anlamsal benzerlik ölçümleri
- Wordnet yoksa / yeterli değilse
  - birlikte geçiş
  - sinirsel modeller



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Kemik

## Kelime benzerliđi

- Ses benzerliđi (edit-distance)
- Anlam benzerliđi
  - Kavramsal ağaçlarla (wordnet)
  - Kavramsal ağaç yoksa ?



Kemik

## Kelime Anlamları ve WORDNET\*

- WordNet sistemi kelimeleri, anlamlarını (tanımlarını), ilişkili olduđu kelimeleri ve ilişki türlerini, kelimelere erişim için çeşitli yazılımları içerir.
- Eş anlamlı kelimeler, eşküme (synset)ler içinde birlikte yer alırlar.
- İlişkiler eşküme arasında tanımlanmıştır.
- Elle oluşturulmuş. (yüksek kalite)
- 10 yıldan fazla emek. (yüksek iş gücü)
- **İNGİLİZCE** için.

[\*] [http://www.ise.gmu.edu/~carlotta/teaching/INFS-795-s06/readings/Similarity\\_in\\_WordNet.pdf](http://www.ise.gmu.edu/~carlotta/teaching/INFS-795-s06/readings/Similarity_in_WordNet.pdf)

[\*] <http://www.cs.utah.edu/~sidd/documents/msthe03ppt.pdf>



Kemik

## Rakamlarla Wordnet

Category	Unique Forms	# of Senses
Noun	117,097	145,104
Verb	11,488	24,890
Adjective	22,141	31,302
Adverb	4,601	5,720



Kemik

## WordNet Noun Relations

Relation	Definition	Example
Hypernym	From concepts to superordinates	<i>breakfast</i> → <i>meal</i>
Hyponym	From concepts to subtypes	<i>meal</i> → <i>lunch</i>
Has-Member	From groups to their members	<i>faculty</i> → <i>professor</i>
Member-Of	From members to their groups	<i>copilot</i> → <i>crew</i>
Has-Part	From wholes to parts	<i>table</i> → <i>leg</i>
Part-Of	From parts to wholes	<i>course</i> → <i>meal</i>
Antonym	Opposites	<i>leader</i> → <i>follower</i>

*Hyponym/Hypernym* (Is-A relationships)  
Bütün ilişkilerin %60'ı



Kemik

### Verb Relations in WordNet

Relation	Definition	Example
Hypernym	From events to superordinate events	<i>fly</i> → <i>travel</i>
Troponym	From events to their subtypes	<i>walk</i> → <i>stroll</i>
Entails	From events to the events they entail	<i>snore</i> → <i>sleep</i>
Antonym	Opposites	<i>increase</i> ↔ <i>decrease</i>

### Adjective & Adverb Relations in WordNet

Relation	Definition	Example
Antonym	Opposite	<i>heavy</i> ↔ <i>light</i>
Adverb	Opposite	<i>quickly</i> ↔ <i>slowly</i>



Kemik

## Wordnet'le Anlamsal Benzerlik Bulma

- İki temel yaklaşım
  - Bağ sayma
    - Taksonomi yeterli.
  - Ortak/ Müşterek bilgi (Mutual Information)
    - Taksonomi ve corpus kullanır.



Kemik

## Leacock & Chodorow (1998)

$$sim_{LC}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2L}$$

- **len(c1,c2)** iki synset arasındaki en kısa yolun uzunluğu. (*benzerlik değeriyle ters orantılı*)
- **L**, tüm taksonominin derinliği



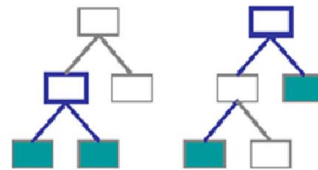
Kemik

## Wu & Palmer (1994)

$$sim_{Wu\&Palmer}(c_1, c_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3}$$

• **N1** ve **N2**, en yakın ortak üst synset'lerine iki synsetin IS-A bağlarıyla uzaklıkları (*benzerlik değeriyle ters orantılı*)

• **N3**, en yakın ortak üst synset'in kök synset'e IS-A bağlarıyla uzaklığı (*büyüklüğü ortak synset'in spesifikliğini gösterir*)



Kemik

- Bir kavramın bir korpus'taki olasılığı, korpus'ta geçme sayısı ile bulunur.

$$P(\text{concept}) = \text{freq}(\text{concept}) / \text{freq}(\text{root})$$

$\text{freq}(X) \rightarrow$  korpus'ta X synset'inin tepesinde bulunduğu tüm synset'lerin frekanslarının toplamı



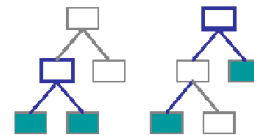
Kemik

## Jiang-Conrath (1997)- Lin (1998)

$$sim_{JC}(c_1, c_2) = \frac{2 \log(p(lcs(c_1, c_2)))}{\log(p(c_1)) + \log(p(c_2))}$$

$$sim_L(c_1, c_2) = \frac{2 * \log(p(lcs(c_1, c_2)))}{\log(p(c_1)) + \log(p(c_2))}$$

$lcs(c_1, c_2)$  en yakın ortak üst synset



- A, her iki kavramı da içeren en spesifik kavramı kullanır (iki kavramın beraber geçtiği doküman sayısına benzer)
- B, iki kavramdan herhangi birini içeren doküman sayısına benzer



Kemik

## Benzerlik Metotlarının Karşılaştırılması

- Bütün metotlar İngilizce 38 kelime çiftine uygulanmış.
- Bulunan benzerlik değerlerinin, insan yargılarıyla olan korelasyonları yandaki tabloda.

<i>Method</i>	<i>Type</i>	<i>Correlation</i>
Wu & Palmer 1994	Edge Counting	0.74
Li 2003	Edge Counting	0.82
Leacock & Chodorow 1998	Edge Counting	0.82
Resnik 1999	Info. Content	0.79
Lin 1998	Info. Content	0.82
Lord 2003	Info. Content	0.79
Jiang & Conrath 1998	Info. Content	0.83
Tversky 1977	Feature Based	0.73
Adapted Lesk 2002	Feature Based	0.37*
Rodriguez 2003	Hybrid	0.71



Kemik

## Wordnet vb. yoksa? Anlamsal Uzay

- Kavramların anlamsal yakınlıklarına göre yer aldıkları bir uzay
- Kavramların uzaydaki koordinatları, anlamsal yakınlıklarına göre belirlenir
- Anlamsal yakınlık nasıl ölçülür?



Kemik

## Birlikte Geçme Matrisleri

- Harris der ki:
  - Birlikte kullanılan kavramlar birbirlerine anlamsal olarak benzerler.
- Birlikte kullanım: aynı cümlede, aynı metinde, sabit bir kelime penceresinde



Kemik

## Kelimelerin Anlamsal Benzerlik Ölçümü Amasyalı, 2006

- İki kelimenin İnternet'te yer alan sayfaların kaçında yan yana kullanıldıkları bulunarak belirlenmiştir.
- Bunun için arama motoruna “kelime1 kelime2” ve “kelime2 kelime1” sorguları gönderilerek gelen sonuç sayfalarındaki sonuç sayıları toplanmıştır.



Kemik



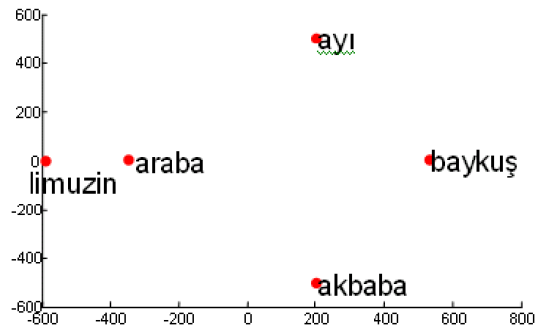
## Örnek benzerlik matrisi

	akbaba	ayı	baykuş	araba	limuzin
akbaba		0	20	1	0
ayı	0		33	4	0
baykuş	20	33		0	0
araba	1	4	0		38
limuzin	0	0	0	38	



Kemik

## Benzerlik Matrisinden ÇBÖ ile elde edilen harita



Kemik

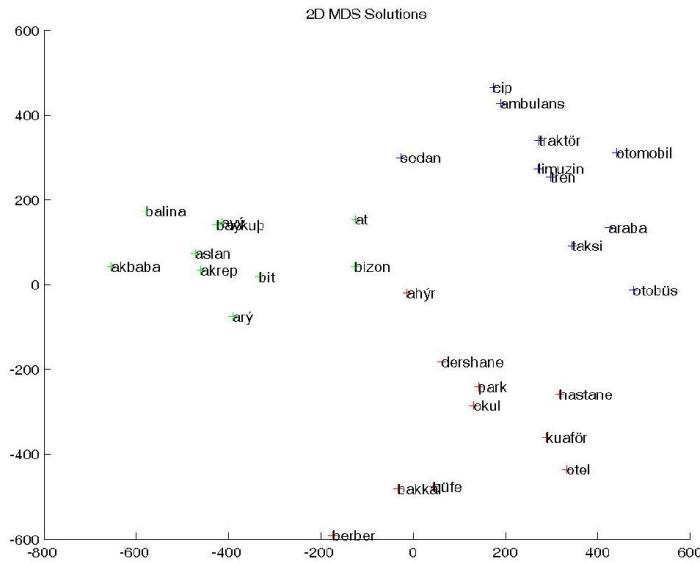
# Kelimeler

Veni kümesi -I	<i>Yer</i> berber - bakkal- kuaför- hastane- otel- okul- dershane- büfe- ahır- park	<i>Hayvan</i> akbaba- arı- baykuş- balina- aslan- at - bizon - akrep - ayı- bit	<i>Taşıt</i> otobüs- cip- taksi- ambulans- araba- sedan- tren- limuzin- otomobil- traktör			
Veni kümesi -II	<i>Ev eşyası</i> divan - televizyon- avize- merdiven- bilgisayar- buzdolabı- kapı- sandalye- lamba- koltuk	<i>Hayvan</i> akbaba- arı- baykuş- balina- aslan- at - bizon - akrep - ayı- bit	<i>Çiyecek</i> bikini- mayo- bere- süveter- etek- ayakkabı- eldiven- gömlek- şapka- gözlük			
Veni kümesi -III	<i>Çiyecek</i> tereyağı kaymak baharat peynir şeker tuz	<i>Çiyecek</i> ayakkabı eldiven şapka gömlek gözlük etek	<i>Ev eşyası</i> avize sandalye lamba koltuk buzdolabı televizyon	<i>Hayvan</i> aslan arı at akrep ayı akbaba	<i>Yer</i> otel dershane park berber hastane okul	<i>Taşıt</i> taksi limuzin ambulans cip otomobil araba

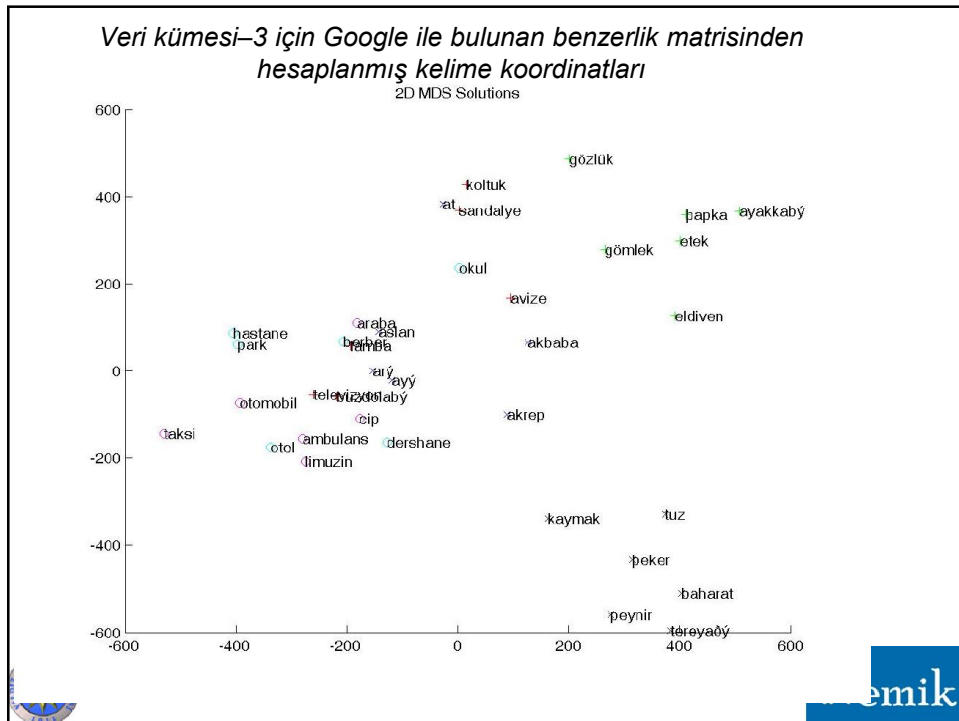
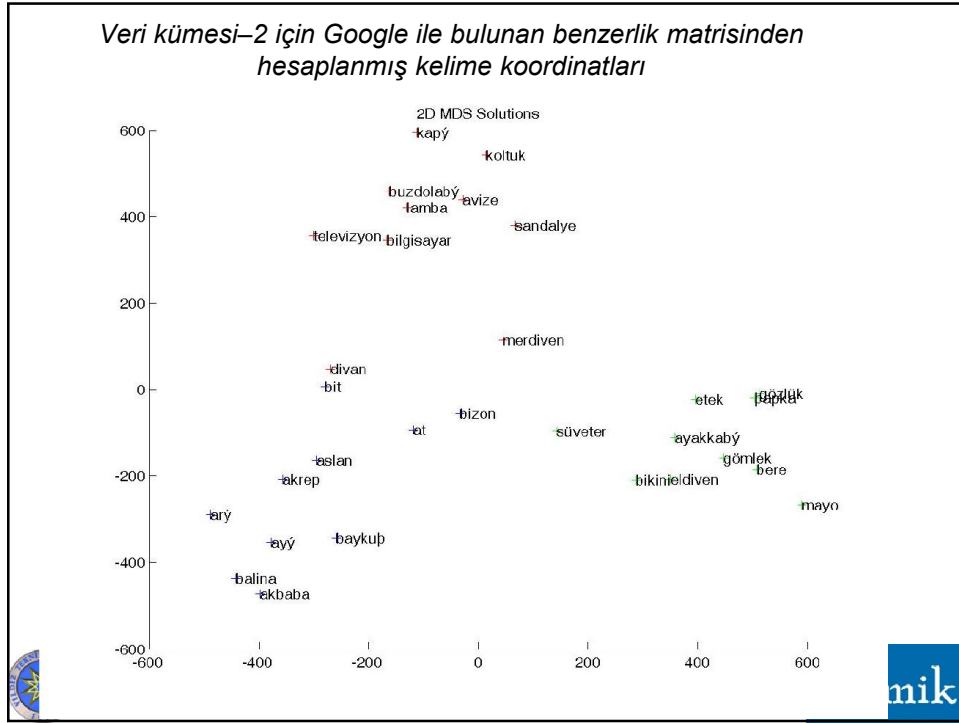


Kemik

Veri kümesi-1 için Google ile bulunan benzerlik matrisinden hesaplanmış kelime koordinatları



Kemik



## Kelimeleri Sınıflandırma Başarıları

	v1	v1	v2	v2	v3	v3
	10 boyut	2 boyut	10 boyut	2 boyut	10 boyut	4 boyut
SVM	96,6	100	83,3	96,6	88,8	77,7
C4.5	83,3	83,3	90	90	77,7	75
RF	90	90	93,3	93,3	72,2	75
EM	66,6	96,6	66,6	96,6	66,6	83,3



Kemik

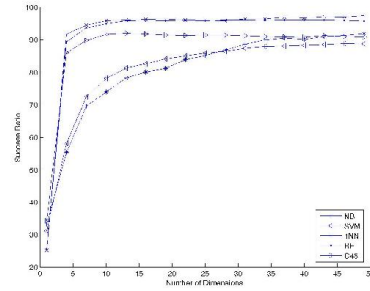
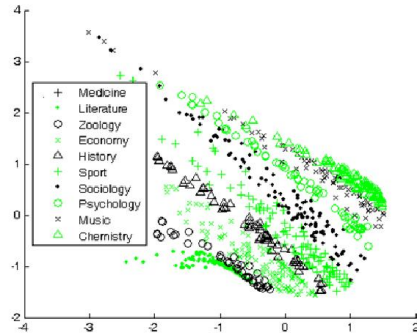
- En başarılı sınıflandırıcı SVM.
- Boyut azaltmanın genelde başarıyı yükseltiyor.
- SVM algoritması ile Veri kümesi-1 2 boyutta %100 başarılı.
- 3 veri kümesinde sınıflandırıcıların başarılarının ortalaması %87'dir.
- Ayrıca EM algoritmasıyla hiçbir sınıf bilgisi kullanmadan düşük boyutlarda ulaşılan başarı gerçekten iyi.
- Bununla birlikte yapılan denemeler küçük ölçeklidirler.
- Daha sağlıklı yorumlara ulaşabilmek için daha fazla sınıf ve kelime içeren veri kümeleriyle çalışmak gerekmektedir.



Kemik

## Teknik Terim sınıflandırma

- 967 teknik terimin koordinatları, 15 bin web sitesi üzerinde birlikte geçme matrisi



Kemik

## Benzerlikten (Cij) uzaklığa (Dij) geçişte formüller:

- $D_{ij} = 1 / (C_{ij} + 0.0001)$
- $D_{ij} = 2^{(1 - C_{ij})}$



Kemik

## Anlamsal Uzaylar Uygulama alanları

- Sorgu zenginleştirme (birbirine çok benzer kelimeleri sorgulara ekleme/değişirme)
- Metin sınıflandırma (Amasyalı, Beken, 2009)
- Kelime yörüngeleri (Amasyalı, Yener, Kaplan, 2012)
- ?



Kemik

## Anlamsal Uzayların Metin Sınıflandırmada Kullanımı

- 5 farklı haber sınıfına (ekonomi, magazin, sağlık, siyasi, spor) ait 230'ar metin
- Her sınıftan 150'şer haber metni eğitim, 80'er adedi test
- PCKimmo ile kelime kökleri
- farklı gövde sayısı yaklaşık 4500



Kemik

## Anlamsal Uzayların Metin Sınıflandırmada Kullanımı

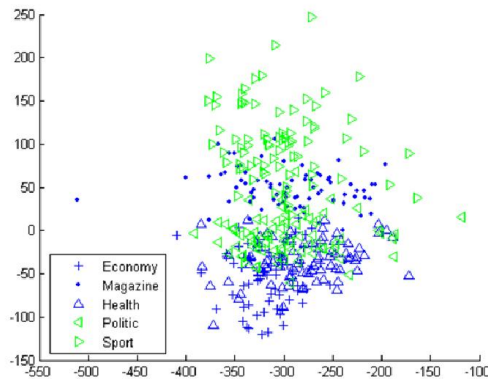
- 4500 kelimenin sayısal karşılıklarını elde etmek için 15.000 web sitesinden oluşan bir külliyat kullanılmış.
- Kelimelerin birbirlerine anlamsal yakınlık matrisi bu külliyatta birlikte geçtikleri doküman sayıları
- uzaklık=1/yakınlık
- MDS (ÇBÖ)
- MDS'de ilk 100 ve ilk 10 boyut alınmış



Kemik

## Anlamsal Uzayların Metin Sınıflandırmada Kullanımı

- Metinler, içerdikleri kelimelerin koordinatlarının ortalamaları
- Metinler kelimelerle aynı boyutlu
- 1 ve 2. boyutlar



nik

# Anlamsal Uzayların Metin Sınıflandırmada Kullanımı

Klasik terim\*doküman matrisi yaklaşımı

Sınıflandırma Algoritması	Metinlerin Boyut Sayısı	Başarı yüzdesi (f)	Başarı yüzdesi (tfidf)
Klasik Naive Bayes	4500	87.25	
Diskrit Naive Bayes	4500	85.75	<b>89.25</b>
En Yakın Komşu	4500	34.25	43.5
Destek Vektör Makineleri	4500	87	86.5
Rastgele Ormanlar (100 ağaçlı)	4500	X	X
Lineer Regresyonla Sınıflandırma	4500	X	X
Pace Regresyonla Sınıflandırma	4500	X	X
C4.5	4500	74.75	23.5
Lineer Regresyonla Sınıflandırma	100	81.75	84.5
Pace Regresyonla Sınıflandırma	100	81.5	83.25
En Yakın Komşu	100	71.25	76.25
Destek Vektör Makineleri	100	80.25	87.75
Rastgele Ormanlar (100 ağaçlı)	100	85	84.5

Anlamsal uzay yaklaşımı

Sınıflandırma Algoritması	100 Boyutlu Metinler	10 Boyutlu Metinler
Lineer Regresyonla Sınıflandırma	<b>93.25</b>	89
Pace Regresyonla Sınıflandırma	92.75	88.5
En Yakın Komşu	70	79
Destek Vektör Makineleri	90	89.75
Rastgele Ormanlar (100 ağaçlı)	90.75	87.75



Kemik

# Anlamsal Uzayların Metin Sınıflandırmada Kullanımı

- Daha yüksek başarı, Metin temsilinde daha az hafıza
- Benzerlik matrisi bulurken Google neden kullanılmadı?
- Yöntem dilden bağımsız mı?
- Eşsesli kelimelerin durumu?
- Metinlerin koordinatlarının bulunmasında başka bir yol ? **Benzerliği, kelimelerin kaç dokümanda değil de kaç sınıfta birlikte geçtiklerine bakarak bulmak**



Kemik



## Kelime Yörüngeleri

- *İnsanların düşünce süreçleri birbirine ne kadar benzemektedir?*
- *İnsanların düşünce süreçlerini gözlemek, direkt ölçmek mümkün değilse de onu, ürünlerinin bazılarıyla (konuşma ve yazılarıyla) dolaylı olarak gözlemek mümkündür.*



Kemik

## Kelime Yörüngeleri

- Yazıları, onları oluşturan kelimelerin zamana göre sıralanmış hali olarak düşünelim.
- Elimizde kelimelerin koordinatları olsa, bu koordinatları yazıdaki sıralarıyla birleştirerek bir yazıyı  $X$  boyutlu bir uzayda bir yörünge olarak ifade edebiliriz.



Kemik

## Kelime Yörüngeleri

- Aynı kişinin farklı yazılarının yörüngeleri birbirine benzer midir?
- Farklı kişilerin yörüngelerini birbirinden ayırmak mümkün müdür?
- Cevaplar için önce yörüngeleri oluşturmak gerekir. Bunun için ise önce kelime koordinatlarını bulmak gerekir.



Kemik

## Kelime Koordinatları

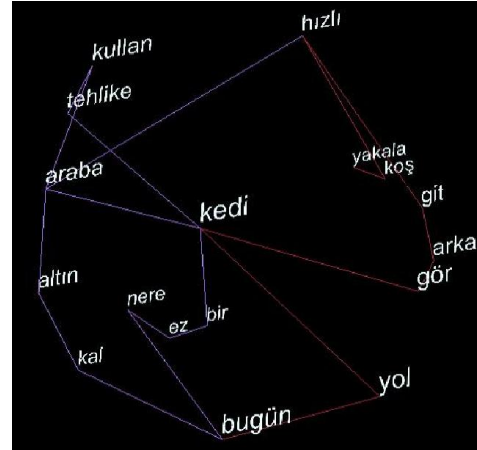
- Birlikte geçiş matrisi



Kemik

## Kelime Yörüngeleri

- Bugün yolda kedi gördüm. Arkasından gittim. Hızlı koşuyordu. Yakalayamadım.
- Hızlı araba kullanmak tehlikelidir. Kediler arabaların altında kalabilirler. Bugün neredeyse eziliyordu bir kedi.



Kemik

## Yörüngelerin Özellikleri

- 2 temel özellik:
- Kavramlar arası mesafeler,
- Kavramlar arası açılar.
- $(n, n+1, n+2, n+3)$  arka arkaya gelen 4 koordinat olmak üzere  $(n, n+1)$ ,  $(n, n+2)$ ,  $(n, n+3)$  arası mesafelerin 10'luk histogram değerleri, frekansları ve  $(n, n+1, n+2)$ ,  $(n, n+2, n+3)$  arası açıların (PI cinsinden) 10'luk histogram değerleri, frekansları



Kemik

## Yörüngelerin Özellikleri

- $d$  uzunluğundaki bir yörünge de  $d-1$  adet  $(n, n+1)$  arası mesafe ölçülmektedir. Bu  $d-1$  ölçümün eşit aralıklı 10 parçalık histogramı çıkarılmaktadır. Bu histogramın 10 adet değeri ve her değer için bir frekans bulunmaktadır. Dolayısıyla yörünge için  $(n, n+1)$  arası mesafelerini ifade eden 20 adet özellik çıkarılmaktadır. Bu işlem  $(n, n+2)$ ,  $(n, n+3)$  arası mesafelere ve  $(n, n+1, n+2)$ ,  $(n, n+2, n+3)$  arası açılara da uygulandığında bir metni ifade eden yörüngeye ait 100 özellik bulunmuş olmaktadır



Kemik

## DeneySEL Sonuçlar-1

2 yazara ait 35'er yazı,  
Bir yazı kime ait?  
10'lu çapraz geçerleme sonuçları

Algoritma	Sınıflandırma Başarısı (%)
C4.5	97.15
Naive Bayes	97.15
En yakın komşu	95.71
Destek Vektör Makineleri	98.57



Kemik

## Deneysel Sonular-2

Her yazara ait 35'er yazı,  
Bir yazı kime ait?  
10'lu apraz geerleme sonuları

Algoritma	3 yazar	4 yazar	8 yazar
C4.5	91.43	90.71	53.21
Naive Bayes	95.24	92.14	62.5
En yakın komşu	80.95	62.14	50
Destek Vektör Makineleri	94.29	85.71	68.57
Zero 0	33.33	25	12.5



Kemik

## Sonuç ve Tartışma

- 2 kişiye ait 35'er yazıdan oluşan veri kümesinde bir metnin yazarını tanıma başarısı %98
- Yazar sayısı arttıka başarının düşüyor (8 yazar için %68)
- Metinde kullanılan kavramlardan bağımsız olması (sadece yörünge nin özelliklerinin kullanılıyor olması) aynı kişinin farklı konularda yazdıklarından elde edilen yörüngeler benzer



Kemik

## Olası kullanım alanları

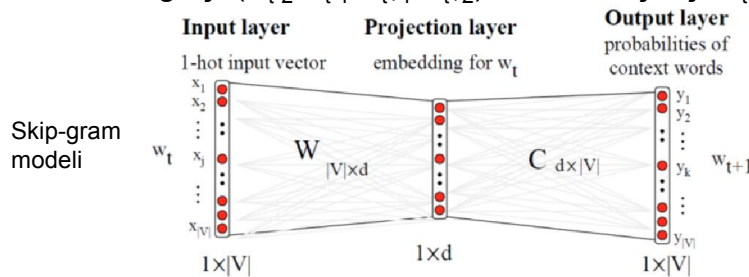
- Psikolojik hastalıkların tespiti
- Psikolojik hastalıkların düşünce süreçleri üzerindeki etkilerinin araştırılması
- Cinsiyet, yaş, eğitim farklılıklarının düşünce süreçleri üzerindeki etkilerinin araştırılması



Kemik

## Sinirsel Modeller

- Word2vec\*, fasttext\*\*
- Tek gizli katmana sahip bir yapay sinir ağı
- $C=2$  için bağlam =  $(w_{t-2}, w_{t-1}, w_t, w_{t+1}, w_{t+2})$
- Skip-gram: giriş:  $w_t$ , çıkış:  $(w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2})$  in herbiri
- C-bow: giriş:  $(w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2})$  in herbiri, çıkış:  $w_t$



\* 2013, T. Mikolov ve ark. "Distributed representations of words and phrases and their compositionality"

\*\* 2017, P. Bojanowski ve ark. "Enriching word vectors with subword information"

Kemik

## Sinirsel Modeller

- $W$  matrisi kelime vektörleri
- Başlangıçta rasgele atanırlar. Eğitim bittiğinde kullanıma hazırlar.
- Amaç: bir kelimenin ( $w$ ) genelde etrafında bulunan kelimelerle ( $c$ ) benzer vektörlere, diğerleriyle ( $w_i$ ) farklı vektörlere sahip olması



<https://medium.com/@Aj.Cheng/word2vec-3b2cc79d674>  
<https://web.stanford.edu/~jurafsky/slp3/slides/vector2.pdf>

Kemik

## Eğitim kümesi oluşturmak

- Text:  $w_1 w_2 w_3 w_4 w_5 w_6 w_7 \dots$
- Skip gram için eğitim örnekleri (giriş, çıkış)  $C=2$  için:
  - benzer olmasını istediğimiz ikililer:  $(w_3, w_1)$ ,  $(w_3, w_2)$ ,  $(w_3, w_4)$ ,  $(w_3, w_5)$
- Sonra  $w_4$  ile devam ...
- C-bow için ikilileri tersine çevir



Kemik

## Fasttext

- Birlikte bulunmaya ek olarak kelime içi n-gram benzerliklerini (sub-word similarity) de dikkate alır.
- Kelimeleri n-gram larına parçalar ve her bir n-gram içinde vektörler bulur.
- Bu sayede sözlükte olmayan (test örneklerinde çok karşılaşılan bir durum) kelimelerin vektörlerine bulur.
- Türkçe gibi ek alan diller için avantajlı (aynı kelimenin farklı ekler almış halleri, yanlış yazılmış halleri birbirine yakın koordinatlara sahip olur)
- Ama daha yavaş



Kemik

## Hazır fasttext\* vektörleri ile ilk satırdaki kelimelere en yakın 19 kelime

"cimbom"	"terlik"	"virüs"	"elma"	"aslan"	"koş"	"bırak"
"cimbombom"	"terliksi"	"virüsün"	"elma,armut"	"aslaner"	"koşo"	"bıraktın"
"lik--"	"terlikler"	"virüslü"	"armut"	"aslana"	"koşoy"	"bırakın"
"penaltılarla"	"terlikleri"	"virüsle"	",armut"	"aslani"	"koşaç"	"bırakmam"
"#beşiktaş"	"noterlik"	"virüse"	"şeftali"	"aslankeser"	"koşa"	"bırakmam"
"rövanşta"	"kasiyerlik"	"virüstür"	"meyveş"	"aslan"	"koşağaç"	"bırak"
"lik/"	"önük"	"virüsü"	"meyva"	"aslanın"	"koşovi"	"bıraksın"
"penaltisonuç"	"yeterlik"	"virüsler"	"elmaağacı"	"aslan yürek"	"koşun"	"bıraktım"
"beşiktaşlı"	"dülgerlik"	"virüsün"	"kayıs"	"aslanbaş"	"koşamaz"	"bırakayım"
"cimbilli"	"dizlik"	"virüsten"	"meyvelik"	"aslanhan"	"koşalı"	"bırakırsan"
"gülesin"	"göynek"	"virüsteki"	"ayva"	"özaslan"	"koşacı"	"bıra"
"galatasaray"	"yerlik"	"virüsleri"	"üzüm"	"aslanbay"	"koşaca"	"bırakırsam"
"iük"	"ayakkabıcılık"	"virüslere"	"elmakaya"	"aslanduz"	"koşköpür"	"bırakmayın"
"eledim"	"tonmaysterlik"	"mimivirüs"	"buğday,şeker"	"aslanpençesi"	"koşay"	"bıraktık"
","fenerbahçe"	"poşu"	"virütik"	"badem"	"aslan tuğ"	"koşan"	"bırakmalı"
"haftabjk"	"köynek"	"virüslerde"	"arpa,buğday"	"aslanla"	"koşrabat"	"bırakalım"
"beşiktaş"	"yakıcılık"	"virüsleri#grup"	"armutu"	"aslanları"	"beliy"	"bırakalı"
"rövanşında"	"rençberlik"	"poliovirüs"	"şeftalinin"	"sanaslan"	"koşanlar"	"bırakmadım"
"cim"	"terli"	"virüslerin"	"elmax"	"aslanlar"	"koşunca"	"bıraksam"
"kupadaki"	"üzerlik"	"virüslerle"	"armudu"	"yavuzaslan"	"koşkin"	"bırakma"



\*Fasttext vektörleri 416 bin kelime içermektedir.  
Ngram benzerliği de kullanılır.

Kemik



42 bin haber metniyle eğitilen vektörler\*,  
ngram benzerliği yok  
ilk satırdaki kelimelere en yakın 19 kelime

"cimbom"	"istanbul"	"virüs"	"elma"	"aslan"	"ölmek"	"bırak"
"kırmızılılar"	"sergi"	"virüsü"	"meyve"	"aslanın"	"yaşamak"	"bırakın"
"beşikta"	"gerçekleşecek"	"virüsün"	"meyveler"	"lakaplı"	"istemiyorum"	"dedikleri"
"wesley"	"sahipliğinde"	"bulaşan"	"lif"	"cimbom"	"asker"	"kardeşim"
"didier"	"katılımıyla"	"hiv"	"sebze"	"kırmızılı"	"erbakan"	"üzeri"
"galatasaraya"	"fashion"	"hastalık"	"limon"	"kırmızılılar"	"chavez"	"git"
"transfere"	"konser"	"yoluyla"	"çürük"	"serhat"	"lutfen"	"dediğini"
"kırmızılı"	"buluşacak"	"grip"	"salata"	"devler"	"bekleyen"	"hayattan"
"kırmızılıların"	"sarayında"	"aşı"	"pirinç"	"didier"	"kanserin"	"yahu"
"eboue"	"inci"	"zararlı"	"buğday"	"muhammet"	"ameliyata"	"genişletilmiş"
"cim"	"sergisi"	"yazılımı"	"zeytinyağı"	"toure"	"anın"	"memnunum"
"sarıkırmızılı"	"cnr"	"bilgisayar"	"patates"	"drogbayı"	"rio"	"aşkına"
"devler"	"seçkin"	"solunum"	"çiğ"	"lokomotiv"	"aileme"	"temsalcimiz"
"aysal"	"global"	"enfeksiyonları"	"muz"	"payını"	"ölümü"	"etkisine"
"ezel"	"modern"	"kullanıcının"	"ürünlerde"	"ulaş"	"istemediğini"	"özetle"
"terimin"	"festival"	"belirtilen"	"fındık"	"oya"	"ağabeyim"	"dedim"
"ligindeki"	"louis"	"ağrısına"	"sarımsak"	"bulut"	"koydular"	"açıklamalarda"
"deplasmanında"	"dünyaca"	"enfeksiyon"	"portakal"	"yitirdi"	"özbeğenin"	"gidin"
"futbolcunun"	"konseri"	"bulaşıcı"	"hapı"	"ceren"	"yasak"	"yazdı"
"beraberlik"	"tarihleri"	"saldırganlar"	"ihracatı"	"ekrem"	"söylemiş"	"oynuyor"



\*Vektörler 19 bin kelime içermektedir.

Kemik