

Türkçe için Konuşma Tanıma ve Derin Öğrenmeyle Dil Modelleme

Ebru Arısoy

Elektrik-Elektronik Mühendisliği Bölümü
MEF Üniversitesi

3 Temmuz 2018



EBRU ARISOY

- 2002 Boğaziçi Üniversitesi EE (BS)
- 2004 Boğaziçi Üniversitesi EE (MS)
- 2009 Boğaziçi Üniversitesi EE (PhD)
- 2010 - 2013 IBM T.J. Watson Research Center
- 2014 - 2015 IBM Türkiye
- 2015 - MEF Üniversitesi

İçerik

- Part 1: Otomatik Konuşma Tanıma (Genel Bilgi)
- Part 2: Türkçe Konuşma Tanıma
- Part 3: Derin Öğrenme ile Dil Modelleme

Part I

Otomatik Konuşma Tanıma (Genel Bilgi)

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

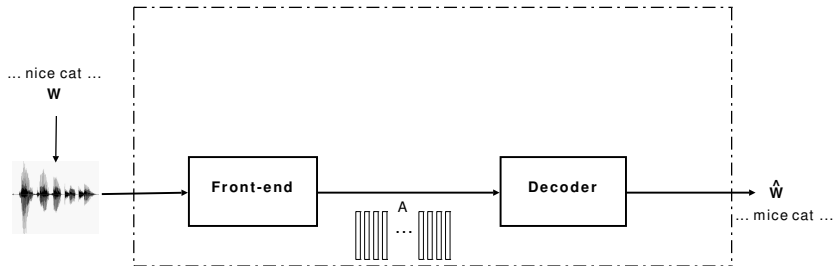
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

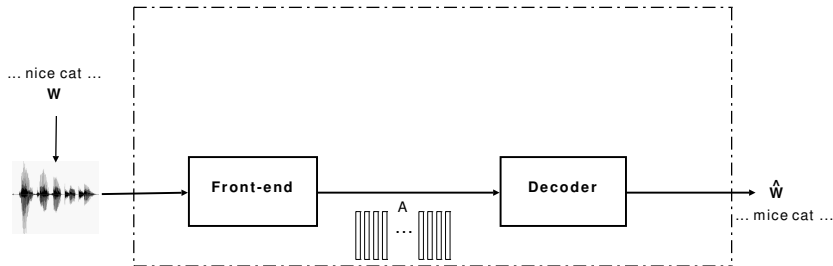
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

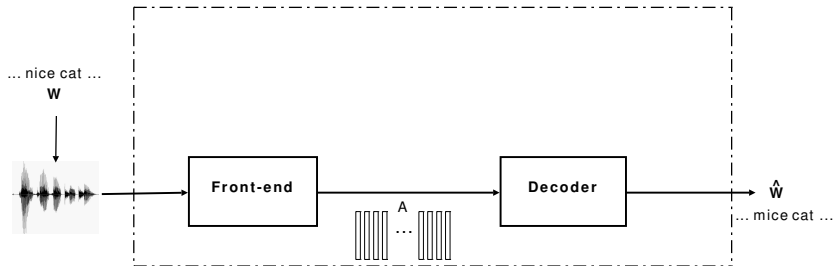
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

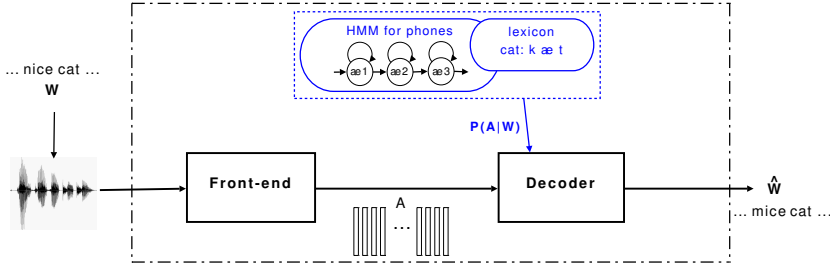
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

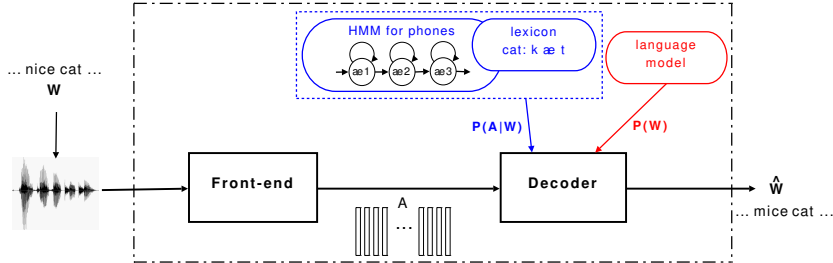
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

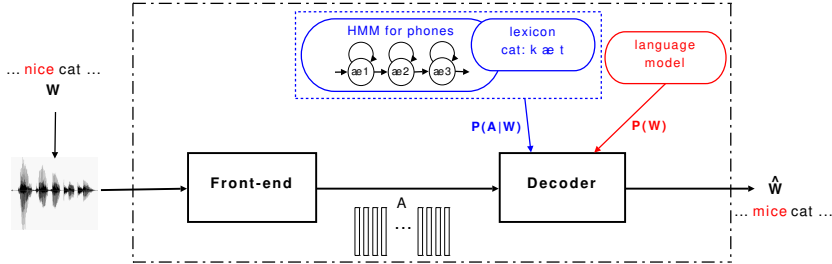
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

Otomatik Konuşma Tanıma

Amaç:

Bir ses sinyaline karşılık gelen metinleri otomatik olarak belirlemek.



$W = \{w_1 \dots \text{nice cat} \dots w_n\}$, $w_i \in \mathcal{V}$: sabit ve sınırlı dağarcık \rightarrow Dağarcık Dışı Kelime (DDK)

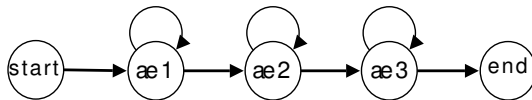
$A = \{a_1 a_2 \dots a_t\}$, $a_i \in \mathcal{A}$: sinyal işleme yöntemleri ile öznelilik çıkarımı

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|A) = \underset{W}{\operatorname{argmax}} \frac{P(A|W)P(W)}{P(A)} = \underset{W}{\operatorname{argmax}} P(A|W)P(W)$$

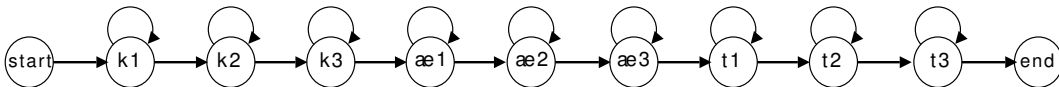
Akustik Model

Akustik Model:

- Akustik model olarak Saklı Markov Model (SMM) kullanılır.
- Ses kayıtları ve bunlara karşılık gelen referans metinleri kullanılarak eğitilir.



"æ" için 3-durumlu SMM



"cat: k æ t" kelimesi için SMM.

Akustik Model

SMM'deki **durum çıktı olasılıkları** için kullanılan modeller.

$P(A/Q)$ A: Akustik öznelik vektörü
Q: Sesçik/durum etiketleri

- Gauss Karışım Modeli (GKM)
 - $P(A/Q)$ hesabı GKM ile yapılır.
- Derin Sinir Ağları (DSA)
 - $P(Q/A)$ hesabı DSA ile yapılır.
 - $P(A/Q) \approx P(Q/A)/P(Q)$
 - SMM'de $P(A/Q)$ kullanılır.

Dil Modeli

İstatistiksel Dil Modeli

Bir dildeki her kelime dizilimine olasılık atar, $P(W)$.

n 'li Diziliimli Dil Modeli

$W = \{w_1 \dots w_N\} = \text{bu kedi çok sevimli}$

$$\begin{aligned} P(W) &= P(\text{bu kedi çok sevimli}) \\ &= P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{bu kedi})P(\text{sevimli}|\text{bu kedi çok}) \end{aligned}$$

Dil Modeli

İstatistiksel Dil Modeli

Bir dildeki her kelime dizilimine olasılık atar, $P(W)$.

n 'li Diziliimli Dil Modeli

$W = \{w_1 \dots w_N\} = \text{bu kedi çok sevimli}$

$$\begin{aligned} P(W) &= P(\text{bu kedi çok sevimli}) \\ &= P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{bu kedi})P(\text{sevimli}|\text{bu kedi çok}) \end{aligned}$$

Dil Modeli

İstatistiksel Dil Modeli

Bir dildeki her kelime dizilimine olasılık atar, $P(W)$.

n 'li Diziliimli Dil Modeli

$W = \{w_1 \dots w_N\} = \text{bu kedi çok sevimli}$

$$\begin{aligned} P(W) &= P(\text{bu kedi çok sevimli}) \\ &= P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{bu kedi})P(\text{sevimli}|\text{bu kedi çok}) \end{aligned}$$

Dil Modeli

İstatistiksel Dil Modeli

Bir dildeki her kelime dizilimine olasılık atar, $P(W)$.

n 'li Dizilimli Dil Modeli

$W = \{w_1 \dots w_N\} = \text{bu kedi çok sevimli}$

$$\begin{aligned} P(W) &= P(\text{bu kedi çok sevimli}) \\ &= P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{bu kedi})P(\underbrace{\text{sevimli}}_{\text{tahmin}} | \underbrace{\text{bu kedi çok}}_{\text{geçmiş}}) \end{aligned}$$

Dil Modeli

İstatistiksel Dil Modeli

Bir dildeki her kelime dizilimine olasılık atar, $P(W)$.

n 'li Diziliimli Dil Modeli

$W = \{w_1 \dots w_N\} = \text{bu kedi çok sevimli}$

$$\begin{aligned} P(W) &= P(\text{bu kedi çok sevimli}) \\ &= P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{bu kedi})P(\underbrace{\text{sevimli}}_{\text{tahmin}} | \underbrace{\text{bu kedi çok}}_{\text{geçmiş}}) \end{aligned}$$

Eğer $n = 2 \rightarrow$ (2'li dizilim)

$$P(W) \approx P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{kedi})P(\text{sevimli}|\text{çok})$$

Dil Modeli

İstatistiksel Dil Modeli

Bir dildeki her kelime dizilimine olasılık atar, $P(W)$.

n 'li Diziliimli Dil Modeli

$W = \{w_1 \dots w_N\} = \text{bu kedi çok sevimli}$

$$\begin{aligned}
 P(W) &= P(\text{bu kedi çok sevimli}) \\
 &= P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{bu kedi})P(\underbrace{\text{sevimli}}_{\text{tahmin}} | \underbrace{\text{bu kedi çok}}_{\text{geçmiş}})
 \end{aligned}$$

Eğer $n = 2 \rightarrow$ (2'li dizilim)

$$\begin{aligned}
 P(W) &\approx P(\text{bu})P(\text{kedi}|\text{bu})P(\text{çok}|\text{kedi})P(\underbrace{\text{sevimli} | \text{çok}}_{\text{geçmiş}}) \\
 &\quad \text{n-1 kelime}
 \end{aligned}$$

Dil Modeli

n 'li Diziliimli Dil Modeli

- Koşullu olasılıklar **en büyük olabilirlik kestirimi** ile **metin verisi** üzerinden öğrenilir.

$$P(w|h) = P(w_k | w_{k-n+1} \dots w_{k-1}) = \frac{C(w_{k-n+1} \dots w_{k-1} w_k)}{C(w_{k-n+1} \dots w_{k-1})}$$

3'lü dizilim:

$$P(\text{çok} | \text{bu kedi}) = \frac{C(\text{bu kedi çok})}{C(\text{bu kedi})}$$

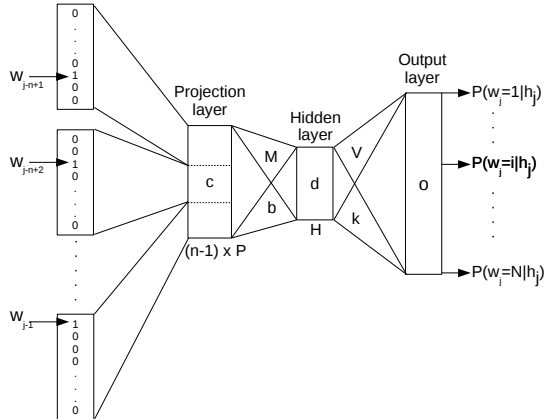
- Veri tabanında geçmeyen n 'li dizilimlere küçük olasılık dağıtımları yapılır.

Dil Modeli

Sinir Ağları Dil Modeli

- Koşullu olasılıklar **sinir ağları** kullanılarak hesaplanır.
- Kelimeler için **sürekli uzay** gösterimleri öğrenilir.

İleribesleme Sinir Ağı Dil Modeli



Dil Modeli

Öznitelik Tabanlı Dil Modeli

- Dilbilimsel bilginin dil modelinde kullanılmasını kolaylaştırır.
- Dilbilimsel bilgi öznitelik vektörüne dönüştürülüp dil modelinde kullanılır.
- En büyük entropi dil modeli:

$$P(w|h) = \frac{1}{Z(h)} e^{\sum_i \alpha_i \Phi_i(w,h)}$$

- Ayırıcı dil modeli:

$$F(W) = \sum_i \alpha_i \Phi_i(W)$$

Dil Modeli

Öznitelik Tabanlı Dil Modeli

- Dilbilimsel bilginin dil modelinde kullanılmasını kolaylaştırır.
- Dilbilimsel bilgi öznitelik vektörüne dönüştürülüp dil modelinde kullanılır.
- En büyük entropi dil modeli:

$$P(w|h) = \frac{1}{Z(h)} e^{\sum_i \alpha_i \phi_i(w,h)}$$

- Ayırıcı dil modeli:

$$F(W) = \sum_i \alpha_i \phi_i(W)$$

Part II

Türkçe Konuşma Tanıma

Türkçe'nin Dil Özellikleri

- ① Türkçe üretken biçimbilimsel yapıya sahip sondan eklemeli bir dildir.

bildiridekiler → bildir-i -de -ki -ler

ruhsatlandırılammamasındaki →

ruhsat -lan -dır -ıl -ama -ma -sı -nda -ki

- ② Türkçe serbest kelime dizilimine sahiptir.

Ben çocuğa kitabı verdim

Çocuğa kitabı ben verdim

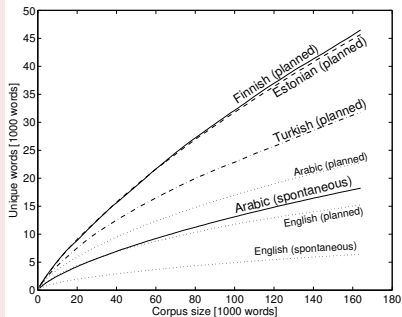
Ben kitabı çocuğa verdim

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

1 Sondan eklemeli (üretken biçimbilimsel yapı)

bildiridekiler → bildiri -de -ki -ler

*Şekil için Mathias Creutz'a teşekkür ederiz.

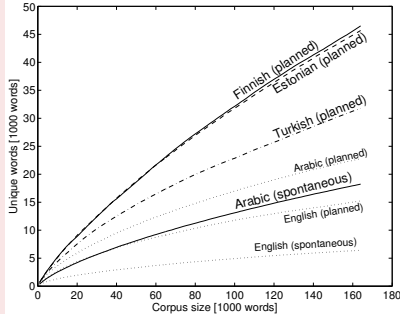


Sondan eklemeli diller: Türkçe, Fince, Estonyaca, Macarca

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

- ❶ Sondan eklemeli (üretken biçimbilimsel yapı) → yüksek DDK oranı
Her DDK 1.5 Kelime Hata Oranı (KHO) verir!

*Şekil için Mathias Creutz'a teşekkür ederiz.



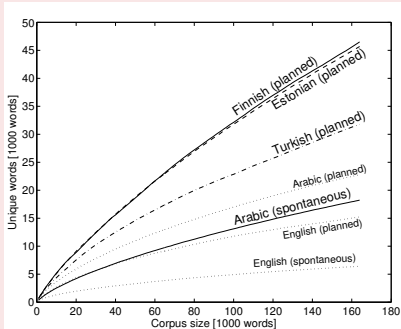
Sondan eklemeli diller: Türkçe, Fince, Estonyaca, Macarca

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

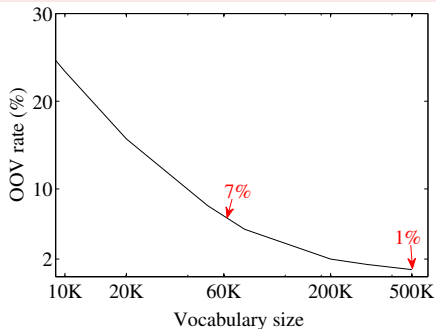
❶ Sondan eklemeli (üretken biçimbilimsel yapı) → yüksek DDK oranı

Her DDK **1.5 Kelime Hata Oranı (KHO)** verir!

*Şekil için Mathias Creutz'a teşekkür ederiz.



Türkçe: 500K → %1 DDK oranı



İngilizce: ≈60K → %1 DDK oranı

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

- 2 Serbest kelime dizilimi → Koşullu olasılıklar için gürbüz kestirimler yapılamaz

Ben çocuğa **kitabı** verdim

$$P(W) \approx P(ben)P(\text{çocuğa}|\text{ben}) \dots P(\text{verdim}|\text{kitabı})$$

Çocuğa kitabı **ben** verdim

Ben kitabı **çocuğa** verdim

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

- ② Serbest kelime dizilimi → Koşullu olasılıklar için gürbüz kestirimler yapılamaz

Ben çocuğa **kitabı** **verdim**

$$P(W) \approx P(ben)P(\text{çocuğa}|\text{ben}) \dots P(\text{verdim}|\text{kitabı})$$

Çocuğa kitabı **ben** **verdim**

$$P(W) \approx P(\text{çocuğa})P(\text{kitabı}|\text{çocuğa}) \dots P(\text{verdim}|\text{ben})$$

Ben kitabı **çocuğa** **verdim**

$$P(W) \approx P(ben)P(\text{kitabı}|\text{ben}) \dots P(\text{verdim}|\text{çocuğa})$$

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

Sondan eklemeli diller için önerilen çözüm:

Kelimeler yerine kelime-altı birimlerin konuşma tanımada kullanılması.

Kelime-altı birimler: Anlamalı kelime bölütleri

bildiridekiler → *bildir* *i* *-de* *-ki* *-ler*

• hatalı ses değişimi:

• hatalı ek dizilimi:

• anlamlı olmayan dizilim:

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

Sondan eklemeli diller için önerilen çözüm:

Kelimeler yerine kelime-altı birimlerin konuşma tanımada kullanılması.

Kelime-altı birimler: Anlamalı kelime bölütleri

bildiridekiler → bildir-i -de -ki -ler

Kelime-altı birimlerin konuşma tanıma açısından zorlukları

- hatalı ses değişimi: bildir-i-de-ki-ler
- hatalı ek dizilimi: bildir-i-de-ki-ler
- anlamlı olmayan dizilim: bildir-i-ki-de-ler

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

Sondan eklemeli diller için önerilen çözüm:

Kelimeler yerine kelime-altı birimlerin konuşma tanımada kullanılması.

Kelime-altı birimler: Anlamalı kelime bölütleri

bildiridekiler → bildir-i -de -ki -ler

Kelime-altı birimlerin konuşma tanıma açısından zorlukları

- hatalı ses değişimi: bildir-i-de-ki-lar
- hatalı ek dizilimi: bildir-i-ki-de-ler
- anlamlı olmayan dizilim: -ki-de-ler

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

Sondan eklemeli diller için önerilen çözüm:

Kelimeler yerine kelime-altı birimlerin konuşma tanımada kullanılması.

Kelime-altı birimler: Anlamalı kelime bölütleri

bildiridekiler → bildir-i -de -ki -ler

Kelime-altı birimlerin konuşma tanıma açısından zorlukları

- hatalı ses değişimi: bildir-i-de-ki-lar
- hatalı ek dizilimi: bildir-i-ki-de-ler
- anlamlı olmayan dizilim: -ki-de-ler

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

Sondan eklemeli diller için önerilen çözüm:

Kelimeler yerine kelime-altı birimlerin konuşma tanımada kullanılması.

Kelime-altı birimler: Anlamalı kelime bölütleri

bildiridekiler → bildir-i -de -ki -ler

Kelime-altı birimlerin konuşma tanıma açısından zorlukları

- hatalı ses değişimi: bildir-i-de-ki-lar
- hatalı ek dizilimi: bildir-i-ki-de-ler
- anlamlı olmayan dizilim: -ki-de-ler

Türkçe'nin Konuşma Tanıma Açısından Zorlukları

Sondan eklemeli diller için önerilen çözüm:

Kelimeler yerine kelime-altı birimlerin konuşma tanımada kullanılması.

Kelime-altı birimler: Anlamalı kelime bölütleri

bildiridekiler → bildir-i -de -ki -ler

Kelime-altı birimlerin konuşma tanıma açısından zorlukları

- hatalı ses değişimi: bildir-i-de-ki-lar
- hatalı ek dizilimi: bildir-i-ki-de-ler
- anlamlı olmayan dizilim: -ki-de-ler

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve **anımlı** birimlere bölmek

harfler: bildiridekiler → b i l d i r i d e k i l e r
heceler: bildiridekiler → bil di ri de ki ler
kök ve ekler: bildiridekiler → de ki ler
kök ve kök sonrası: bildiridekiler → bildiri dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(r_i | \text{bil di})$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve **anımlı** birimlere bölmek

harfler: bildiridekiler → b i l d i r i d e k i l e r
heceler: bildiridekiler → b i l d i r i d e k i l e r
kök ve ekler: bildiridekiler → d e k i l e r
kök ve kök sonrası: bildiridekiler → bildiri dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(r_i | \text{bil di})$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve **anlamlı** birimlere bölmek

harfler: bildiridekiler → b i l d i r i d e k i l e r

heceler: bildiridekiler → bil di ri de ki ler

kök ve ekler: bildiridekiler → de ki ler

kök ve kök sonrası: bildiridekiler → bildiri dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(r|bil\ di)$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve anlamlı birimlere bölmek

harfler: bildiridekiler → b i l d i r i d e k i l e r

heceler: bildiridekiler → bil di ri de ki ler

kök ve ekler: bildiridekiler → bildiri de ki ler

kök ve kök sonrası: bildiridekiler → bildiri dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(r|bil\ di)$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve anlamlı birimlere bölmek

harfler:	bildiridekiler	→	b i l d i r i d e k i l e r
heceler:	bildiridekiler	→	bil di ri de ki ler
kök ve ekler:	bildiridekiler	→	<u>bildiri</u> de ki ler
kök ve kök sonrası:	bildiridekiler	→	<u>bildiri</u> dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(r|bil\ di)$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve anlamlı birimlere bölmek

harfler:	bildiridekiler	→	b i l d i r i d e k i l e r
heceler:	bildiridekiler	→	bil di ri de ki ler
kök ve ekler:	bildiridekiler	→	<u>bildir</u> i de ki ler
kök ve kök sonrası:	bildiridekiler	→	<u>bildir</u> i dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l →

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(r|bil\ di)$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve anlamlı birimlere bölmek

harfler:	bildiridekiler	→	b i l d i r i d e k i l e r
heceler:	bildiridekiler	→	b i l d i r i d e k i l e r
kök ve ekler:	bildiridekiler	→	<u>bildir</u> i d e k i l e r
kök ve kök sonrası:	bildiridekiler	→	<u>bildir</u> i d e k i l e r

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)

bil: b i l →

- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(\text{ri}|\text{bil di})$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

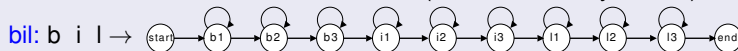
Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve **anlamlı** birimlere bölmek

harfler:	bildiridekiler	→	b i l d i r i d e k i l e r
heceler:	bildiridekiler	→	bil d i r i d e k i l e r
kök ve ekler:	bildiridekiler	→	<u>bildir</u> i d e k i l e r
kök ve kök sonrası:	bildiridekiler	→	<u>bildir</u> i d e k i l e r

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)



- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(\text{ri}|\text{bil di})$

Kelime-altı Birimler ile Konuşma Tanıma

Motivasyon:

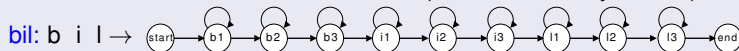
Konuşma Tanımadaki DDK problemini çözmek.

Yöntem:

- 1 kelimeleri daha kısa ve **anlamlı** birimlere bölmek

harfler:	bildiridekiler	→	b i l d i r i d e k i l e r
heceler:	bildiridekiler	→	bil di ri de ki ler
kök ve ekler:	bildiridekiler	→	<u>bildir</u> i de ki ler
kök ve kök sonrası:	bildiridekiler	→	<u>bildir</u> i dekiler

- 2 kelime-altı birimleri sözlükte kullanmak (Akustik model aynı kalır!)



- 3 Kelime-altı birimler üzerinden dil modeli eğitmek

3'lü dizilim: $P(\mathbf{ri}|\mathbf{bil\ di})$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki kelimeleri üretebilmeli.

$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$

- dil modeli olasılıklarının hesabı açısından anlamli birimler olmalı.

$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$

- fazladan (anlamsız) kelime üretmeye karşı dayanıklı olmalı.

$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından anlamli birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- fazladan (anlamsız) kelime üretmeye karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından **anlamli** birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- fazladan (anlamsız) kelime üretmeye karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından **anlamli** birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- fazladan (anlamsız) kelime üretmeye karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından **anlamli** birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- fazladan (anlamsız) kelime üretmeye karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından **anlamli** birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- **fazladan (anlamsız) kelime üretmeye** karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından **anlamli** birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- **fazladan (anlamsız) kelime üretmeye** karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Kelime-altı birimler

- bir dildeki **kelimeleri üretebilmeli**.

$$\mathcal{V}_{\text{harfler}} \ll \mathcal{V}_{\text{heceler}} < \mathcal{V}_{\text{kök ve ekler}} < \mathcal{V}_{\text{kök ve kök sonrası}} \ll \mathcal{V}_{\text{kelimeler}}$$

- dil modeli olasılıklarının hesabı açısından **anlamli** birimler olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

- **fazladan (anlamsız) kelime üretmeye** karşı dayanıklı olmalı.

$$\text{harfler} < \text{heceler} < \text{kök ve ekler} < \text{kök ve kök sonrası}$$

Anlamli Kelime-Altı Birimler

Dilbilgisel Birimler

Kural tabanlı **biçimbilimsel ayrıştırıcı** kullanılarak elde edilir.

- kök ve ekler, kök ve kök sonrası, ...

İstatistiksel Birimler

Metin verisinden **eğitmensiz öğrenme** ile elde edilir.

- Linguistica, Morfessor¹, Paramor, ...

kök ve ekler: bildiridekiler → bildiri -de -ki -ler
istatistiksel birimler: bildiridekiler → bildir -ide -kiler

- Kelimelerin sürekli uzay gösterimi ile bölütleme de yapılmaktadır².

¹Creutz M, Lagus K (2005) Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0. Publications in Computer and Information Science Report A81, HUT, Helsinki, Finland

²Üstün A, Can B. Unsupervised Morphological Segmentation Using Neural Word Embeddings. In Proceedings of SLSP 2016, Pilsen, Czech Republic, 2016

Anlamli Kelime-Altı Birimler

Dilbilgisel Birimler

Kural tabanlı **biçimbilimsel ayrıştırıcı** kullanılarak elde edilir.

- kök ve ekler, kök ve kök sonrası, ...

İstatistiksel Birimler

Metin verisinden **eğitmensiz öğrenme** ile elde edilir.

- Linguistica, **Morfessor**¹, Paramor, ...

kök ve ekler: bildiridekiler → bildiri -de -ki -ler
istatistiksel birimler: bildiridekiler → bildir -ide -kiler

- Kelimelerin sürekli uzay gösterimi ile bölütleme de yapılmaktadır².

¹Creutz M, Lagus K (2005) Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0. Publications in Computer and Information Science Report A81, HUT, Helsinki, Finland

²Üstün A, Can B. Unsupervised Morphological Segmentation Using Neural Word Embeddings. In Proceedings of SLSP 2016, Pilsen, Czech Republic, 2016

Türkçe Konuşma Tanıma Sistemi Geliştirilmesi

1 Veri tabanı oluşturulması

- Akustik veri: Ses kayıtları ve referans metinleri
- Metin verisi: Örün sitelerinden toplanmış metinler

2 Akustik model eğitimi

- SMM-GKM ve DSA ile eğitim
- Yöntemler dilden bağımsızdır
- Türkçe için sözlükte (genelde) sesler için harfler kullanılır.

3 Dil modeli eğitimi

- Türkçe için anlamlı kelime-altı birimlerinin araştırılması
- Ayırıcı dil modellerinin dilbilimsel ve istatistiksel özniteliklerle kullanılması
- Sinir Ağları (SA) dil modellerinin kullanılması

Türkçe Konuşma Tanıma Sistemi Geliştirilmesi

1 Veri tabanı oluşturulması

- Akustik veri: Ses kayıtları ve referans metinleri
- Metin verisi: Örün sitelerinden toplanmış metinler

2 Akustik model eğitimi

- SMM-GKM ve DSA ile eğitim
- Yöntemler dilden bağımsızdır
- Türkçe için sözlükte (genelde) sesler için harfler kullanılır.

3 Dil modeli eğitimi

- Türkçe için anlamlı kelime-altı birimlerinin araştırılması
- Ayırıcı dil modellerinin dilbilimsel ve istatistiksel özniteliklerle kullanılması
- Sinir Ağları (SA) dil modellerinin kullanılması

Türkçe Veri tabanları

Akustik Veri Tabanı = Ses Kayıtları + Referans Metinleri

- Mikrofon ile Toplanan Kayıtlar:
 - GlobalPhone: Türkçe için ≈ 17 saat (ELRA ve Appen Butler Hill Pty Ltd.)
 - METU Microphone Speech: ≈ 10 saat (LDC - LDC2006S33)
 - Sabancı Üniversitesi'nde toplanan akustik veri: ≈ 30 saat³
- Telefon Üzerinden Toplanan Kayıtlar:
 - OrienTel Projesi: 1700 konuşmacı (ELRA)
 - IARPA Babel Turkish Language Pack: ≈ 213 saat (LDC - LDC2016S10)
 - SESTEK'in topladığı çağrı merkezi konuşmaları: ≈ 1000 saat⁴
- Haber Programı Kayıtları: ≈ 130 saat Türkçe programlar (LDC - LDC2012S06)

³Erdogan H, Büyük O, Oflazer K (2005) Incorporating language constraints in sub-word based speech recognition. In: Proceedings of ASRU, San Juan, PR, pp 98 –103

⁴Haznedaroğlu A, Arslan LM, Büyük O, Eden M (2010) Turkish LVCSR system for call center conversations. In: Proceedings of SIU, Diyarbakır, Turkey, pp 372 – 375

Türkçe Veri tabanları

MetinVeritabanı

- 184 M kelimelik gazete haberleri metinleri ⁵
- 81 M kelimelik haber ve e-kitap metinleri ⁶

⁵Sak H, Güngör T, Saraçlar M (2011) Resources for Turkish morphological processing. Language Resources and Evaluation 45(2):249–261

⁶Erdogan H, Büyük O, Oflazer K (2005) Incorporating language constraints in sub-word based speech recognition. In: Proceedings of ASRU, San Juan, PR, pp 98 –103

Türkçe için Akustik Model

- Akustik model olarak Saklı Markov Model (SMM) kullanılmıştır.
- SMM'deki durum çıktı olasılıkları için GKM ve DSA denemiştir.
- DSA Akustik Modeli
 - Derin sinir ağı mimarileri
 - Çok katmanlı ileribesleme
 - Yinelemeli ve türevleri
 - Zaman gecikmeli
 - Amaç işlevleri
 - Çapraz entropi
 - Dizisel ayırıcı eğitim
 - Örüsüz en büyük ortak bilgi

Türkçe için Akustik Model - Sonuçlar

Türkçe Haber Programları için GDSKT⁷

- Akustik Veri: 188 saat (eğitim), 3 saat (geliştirme), 3 saat (sınama)
- Metin Verisi: 182.3 M kelime (haber metinleri) + 1.8 M kelime (referans metinler)
- Dil Modeli: 200 K dağarcık, 3'lü dizilim kelime

Akustik Model	Sınama KHO (%)
SMM-GKM	18.30
Çok katmanlı ileribesleme (çapraz entropi)	14.24
Çok katmanlı ileribesleme (dizisel ayırıcı eğitim)	13.42
Zaman gecikmeli SA (çapraz entropi)	12.45
Zaman gecikmeli SA (örüsüz en büyük ortak bilgi)	10.92

⁷ Arısoy E, Saraçlar M (2018) Turkish Broadcast News Transcription Revisited. In: Proceedings of SIU 2018, Çeşme, Türkiye.

Türkçe için Akustik Model - Sonuçlar

Türkçe Haber Programları için GDSKT⁷

- Akustik Veri: 188 saat (eğitim), 3 saat (geliştirme), 3 saat (sınama)
- Metin Verisi: 182.3 M kelime (haber metinleri) + 1.8 M kelime (referans metinler)
- Dil Modeli: 200 K dağarcık, 3'lü dizilim kelime

Akustik Model	Sınama KHO (%)
SMM-GKM	18.30
Çok katmanlı ileribesleme (çapraz entropi)	14.24
Çok katmanlı ileribesleme (dizisel ayırıcı eğitim)	13.42
Zaman gecikmeli SA (çapraz entropi)	12.45
Zaman gecikmeli SA (örüsüz en büyük ortak bilgi)	10.92

⁷ Arısoy E, Saraçlar M (2018) Turkish Broadcast News Transcription Revisited. In: Proceedings of SIU 2018, Çeşme, Türkiye.

Türkçe için Dil Modeli – Kelime-altı Birimler

kelimeler:	sunulacak	bildiridekiler
istatistiksel birimler:	sunul -acak	bildir -ide -kiler
kök ve ekler	(yüzeysel): sun -ul -acak (sözlüksel ^B):	bildiri -de -ki -ler
kök ve kök sonrası	(yüzeysel): sun -ul-acak (sözlüksel ^B):	bildiri -de-ki-ler

/A/ yüzeysel gösterimde /a/ veya /e/ olabilir. /H/ yüzeysel gösterimde /ı/, /i/, /u/, veya /ü/ olabilir.

Sözlüksel Gösterim

- Farklı yüzeysel gösterimler için aynı sözlüksel gösterim → **Gürbüz** dil modeli kestirimleri
- Ses değişimlerinin doğruluğu → **Fazladan (anlamsız) kelimeleri üretme** hataları azalır

^BArısoy E, Sak H, Saraçlar M (2007) Language modeling for automatic Turkish broadcast news transcription. In: Proceedings of INTERSPEECH, Antwerp, Belgium, pp 2381 – 2384

Türkçe için Dil Modeli – Kelime-altı Birimler

kelimeler:	sunulacak	bildiridekiler
istatistiksel birimler:	sunul -acak	bildir -ide -kiler
kök ve ekler	(yüzeysel): sun -ul -acak	bildiri -de -ki -ler
	(sözlüksel ^B):	
kök ve kök sonrası	(yüzeysel): sun -ul-acak	bildiri -de-ki-ler
	(sözlüksel ^B):	

/A/ yüzeysel gösterimde /a/ veya /e/ olabilir. /H/ yüzeysel gösterimde /ı/, /i/, /u/, veya /ü/ olabilir.

Sözlüksel Gösterim

- Farklı yüzeysel gösterimler için aynı sözlüksel gösterim → **Gürbüz** dil modeli kestirimleri
- Ses değişimlerinin doğruluğu → **Fazladan (anlamsız) kelimeleri üretme** hataları azalır

^BArısoy E, Sak H, Saraçlar M (2007) Language modeling for automatic Turkish broadcast news transcription. In: Proceedings of INTERSPEECH, Antwerp, Belgium, pp 2381 – 2384

Türkçe için Dil Modeli – Kelime-altı Birimler

kelimeler:	sunulacak	bildiridekiler
istatistiksel birimler:	sunul -acak	bildir -ide -kiler
kök ve ekler	(yüzeysel): sun -ul -acak	bildiri -de -ki -ler
	(sözlüksel ⁸): sun -Hl -AcAk	bildiri -dA -ki -lAr
kök ve kök sonrası	(yüzeysel): sun -ul-acak	bildiri -de-ki-ler
	(sözlüksel ⁸): sun -Hl-AcAk	bildiri -dA-ki-lAr

/A/ yüzeysel gösterimde /a/ veya /e/ olabilir. /H/ yüzeysel gösterimde /h/, /i/, /u/, veya /ü/ olabilir.

Sözlüksel Gösterim

- Farklı yüzeysel gösterimler için aynı sözlüksel gösterim → **Gürbüz** dil modeli kestirimleri
- Ses değişimlerinin doğruluğu → **Fazladan (anlamsız) kelimeleri üretme** hataları azalır

⁸ Arısoy E, Sak H, Saraçlar M (2007) Language modeling for automatic Turkish broadcast news transcription. In: Proceedings of INTERSPEECH, Antwerp, Belgium, pp 2381 – 2384

Türkçe için Dil Modeli – Kelime-altı Birimler

kelimeler:	sunulacak	bildiridekiler
istatistiksel birimler:	sunul -acak	bildir -ide -kiler
kök ve ekler	(yüzeysel): sun -ul -acak	bildiri -de -ki -ler
	(sözlüksel ⁸): sun -Hl -AcAk	bildiri -dA -ki -lAr
kök ve kök sonrası	(yüzeysel): sun -ul-acak	bildiri -de-ki-ler
	(sözlüksel ⁸): sun -Hl-AcAk	bildiri -dA-ki-lAr

/A/ yüzeysel gösterimde /a/ veya /e/ olabilir. /H/ yüzeysel gösterimde /h/, /i/, /u/, veya /ü/ olabilir.

Sözlüksel Gösterim

- Farklı yüzeysel gösterimler için aynı sözlüksel gösterim → **Gürbüz** dil modeli kestirimleri
ev -lAr → ev -ler kitap -lAr → kitap -lar
- Ses değişimlerinin doğruluğu → **Fazladan (anlamsız) kelimeleri üretme** hataları azalır

⁸ Arısoy E, Sak H, Saraçlar M (2007) Language modeling for automatic Turkish broadcast news transcription. In: Proceedings of INTERSPEECH, Antwerp, Belgium, pp 2381 – 2384

Türkçe için Dil Modeli – Kelime-altı Birimler

kelimeler:	sunulacak	bildiridekiler
istatistiksel birimler:	sunul -acak	bildir -ide -kiler
kök ve ekler	(yüzeysel): sun -ul -acak	bildiri -de -ki -ler
	(sözlüksel ⁸): sun -Hl -AcAk	bildiri -dA -ki -lAr
kök ve kök sonrası	(yüzeysel): sun -ul-acak	bildiri -de-ki-ler
	(sözlüksel ⁸): sun -Hl-AcAk	bildiri -dA-ki-lAr

/A/ yüzeysel gösterimde /a/ veya /e/ olabilir. /H/ yüzeysel gösterimde /h/, /i/, /u/, veya /ü/ olabilir.

Sözlüksel Gösterim

- Farklı yüzeysel gösterimler için aynı sözlüksel gösterim → **Gürbüz** dil modeli kestirimleri

ev -lAr → ev -ler kitap -lAr → kitap -lar

- Ses değişimlerinin doğruluğu → **Fazladan (anlamsız) kelimeleri üretme** hataları azalır

ev -lAr ev -ler ✓
ev -lar ✗

⁸ Arısoy E, Sak H, Saraçlar M (2007) Language modeling for automatic Turkish broadcast news transcription. In: Proceedings of INTERSPEECH, Antwerp, Belgium, pp 2381 – 2384

Türkçe için Dil Modeli – Sonuçlar

Türkçe Haber Programları için GDSKT⁹

- Akustik Veri: 188 saat (eğitim), 3 saat (geliştirme), 3 saat (sınama)
- Metin Verisi: 182.3 M kelime (haber metinleri) + 1.8 M kelime (referans metinler)
- Akustik Model: SMM-GKM

Dil Modeli	Sınama KHO (%)
Kelime	23.1
Kök ve kök sonrası (yüzeysel)	21.9
Kök ve ekler (sözlüksel)	21.8
İstatistiksel birimler	21.7
Kök ve kök sonrası (sözlüksel)	21.3

⁹Sak H, Saraçlar M, Güngör T (2012) Morpholexical and discriminative language models for Turkish automatic speech recognition. IEEE Transactions on Audio, Speech, and Language Processing 20(8):2341–2351.

Türkçe için Dil Modeli – Sonuçlar

Türkçe Haber Programları için GDSKT⁹

- Akustik Veri: 188 saat (eğitim), 3 saat (geliştirme), 3 saat (sınama)
- Metin Verisi: 182.3 M kelime (haber metinleri) + 1.8 M kelime (referans metinler)
- Akustik Model: SMM-GKM

Dil Modeli	Sınama KHO (%)
Kelime	23.1
Kök ve kök sonrası (yüzeysel)	21.9
Kök ve ekler (sözlüksel)	21.8
İstatistiksel birimler	21.7
Kök ve kök sonrası (sözlüksel)	21.3

⁹Sak H, Saraçlar M, Güngör T (2012) Morpholexical and discriminative language models for Turkish automatic speech recognition. IEEE Transactions on Audio, Speech, and Language Processing 20(8):2341–2351.

Türkçe için Dil Modeli - Sonuçlar

Türkçe Haber Programları için GDSKT¹⁰

Akustik Model	Dil Modeli	Sınama KHO (%)
SMM-GKM	3'lü kelime	18.30
Zaman gecikmeli SA	3'lü kelime	10.92
Zaman gecikmeli SA	3'lü kök ve kök sonrası*	10.12

* Yüzeysel gösterim

¹⁰ Arısoy E, Saraçlar M (2018) Turkish Broadcast News Transcription Revisited. In: Proceedings of SIU 2018, Çeşme, Türkiye.

Türkçe için Dil Modeli - Sonuçlar

Türkçe Haber Programları için GDSKT¹⁰

Akustik Model	Dil Modeli	Sınama KHO (%)
SMM-GKM	3'lü kelime	18.30
Zaman gecikmeli SA	3'lü kelime	10.92
Zaman gecikmeli SA	3'lü kök ve kök sonrası*	10.12

* Yüzeysel gösterim

¹⁰ Arısoy E, Saraçlar M (2018) Turkish Broadcast News Transcription Revisited. In: Proceedings of SIU 2018, Çeşme, Türkiye.

Üretici Dil Modeli - Ayırıcı Dil Modeli

Üretici Dil Modeli: *n*-li dizilim klasik model

- Koşullu olasılıklar en büyük olabilirlik kestirimi ile hesaplanır.

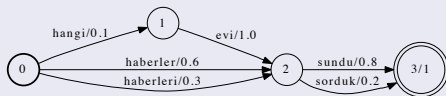
Ayırıcı Dil Modeli:

- Üretici modeli tamamlayıcı bir yöntemdir.
- Hata oranını en iyilemek için ses kayıtları ve onların referans metinleri ile eğitilir.
- **Avantajları:**
 - Model hem doğru hem de yanlış örneklerden öğrenir.
 - Farklı öznitelikler aynı modelde kolayca kullanılabilir.
 - biçimbilimsel, sözdizimsel, anlambilimsel,...

Ayırıcı Dil Modeli

haberleri sundu cümlesinin ses sinyalinin konuşma tanıma çıktısı

Örü Çıktısı



En iyi N hipotez

1. haberler sundu 0.48
2. haberleri sundu 0.24
3. haberler sorduk 0.12
- ⋮

- $\Phi(x, y)$: Öznitelik vektörü $\rightarrow x$: Konuşma y : Aday hipotez
 $\Phi_k(x, y) = \text{"haberleri sundu"}$ dizilimin y 'de kaç defa geçtiği
- $\bar{\alpha}$: Öznitelik parametreleri \rightarrow Ayırıcı eğitim ile öğrenilir.
- $F(x) = \operatorname{argmax}_{y \in \text{GEN}(x)} \Phi(x, y) \cdot \bar{\alpha} \rightarrow \text{GEN}(x)$: x 'e karşılık gelen en iyi N hipotez

Ayırıcı Dil Modeli

haberleri sundu cümlesinin ses sinyalinin konuşma tanıma çıktısı

Örü Çıktısı



En iyi N hipotez

1. haberler sundu 0.48
2. **haberleri sundu 0.24** (Doğru)
3. haberler sorduk 0.12
- ⋮

- $\Phi(x, y)$: Öznitelik vektörü $\rightarrow x$: Konuşma y : Aday hipotez
 $\Phi_k(x, y) = \text{"haberleri sundu"}$ dizilimin y 'de kaç defa geçtiği
- $\bar{\alpha}$: Öznitelik parametreleri \rightarrow Ayırıcı eğitim ile öğrenilir.
- $F(x) = \operatorname{argmax}_{y \in \text{GEN}(x)} \Phi(x, y) \cdot \bar{\alpha} \rightarrow \text{GEN}(x)$: x 'e karşılık gelen en iyi N hipotez

Türkçe Ayırıcı Dil Modeli ¹¹

Kullanılan öznitelikler:

- 1 Kelimelerin n 'li dizilimleri
- 2 Kelime-altı birimlerin n 'li dizilimleri
 - Dilbilimsel: kelimelerin biçimbilimsel analizinden elde edilir.
 - İstatistiksel: kelimelerin istatistiksel yöntemlerle bölütlenmesinden elde edilir.
- 3 Sözdizimsel öznitelikler: cümlelerin bağıllık analizinden elde edilir.

¹¹ Arısoy E, Saraçlar M, Roark B, Shafran I (2012) Discriminative language modeling with linguistic and statistically derived features. IEEE Transactions on Audio, Speech, and Language Processing 20(2):540–550

Part III

Derin Öğrenmeyle Dil Modelleme

Klasik ile Sinir Ağı Dil Modellerinin Karşılaştırması

Klasik n 'li Dizilimli Dil Modeli:

- Kelimeler **ayrık** birimlerdir.
- Veri tabanında **geçmeyen veya az geçen** dizilimler için gürbüz kestirimler yapılamaz.

$$P(\text{sevimli}|\text{bu kedi çok}) = \frac{C(\text{bu kedi çok sevimli})}{C(\text{bu kedi çok})} = 0.25$$

$$P(\text{sevimli}|\text{bu köpek çok}) = \frac{C(\text{bu köpek çok sevimli})}{C(\text{bu köpek çok})} = 0$$

Klasik ile Sinir Ağları Dil Modellerinin Karşılaştırması

Sinir Ağları (SA) Dil Modeli

- Kelimler için **sürekli uzay** gösterimleri öğrenilir.
 - Anlamsal ve dilbilimsel olarak benzer kelimeler sürekli uzayda yakın yerlerde bulunurlar.
- Olasılık kestirimleri **sürekli uzayda SA kullanılarak** yapılır.
- SA dil modeli ile daha **gülbüz** kestirimler yapılabilmektedir.
$$P(\text{sevimli}|\text{bu kedi çok}) \approx P(\text{sevimli}|\text{bu köpek çok})$$
- SA Mimarileri:
 - ileribesleme, yinelemeli, uzun kısa-süreli bellek

İleribesleme Sinir Ağı (SA) Dil Modeli^{12 13}

Olasılık kestirimleri **kelimelerin sürekli uzak gösterimleri** ile yapılır.

- Giriş Katmanı

- İzdüşüm Katmanı

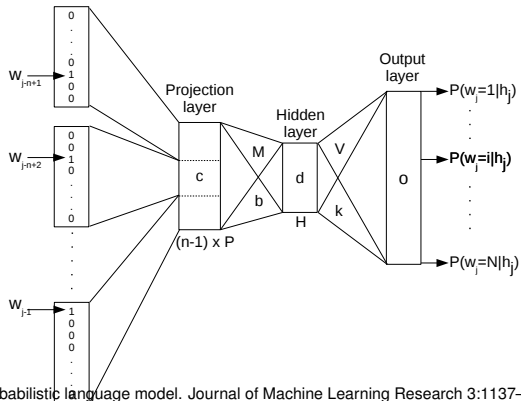
- Saklı Katman

$$d_j = \tanh \left(\sum_{l=1}^{(n-1) \times P} M_{jl} c_l + b_j \right)$$

- Çıktı Katmanı

$$o_i = \sum_{j=1}^H V_{ij} d_j + k_i$$

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$



¹²Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of Machine Learning Research 3:1137–1155

¹³Schwenk H (2007) Continuous space language models. Computer Speech and Language 21(3):492–518

İleribesleme Sinir Ağı (SA) Dil Modeli^{12 13}Olasılık kestirimleri **kelimelerin sürekli uzak gösterimleri** ile yapılır.

• Giriş Katmanı

• İzdüşüm Katmanı

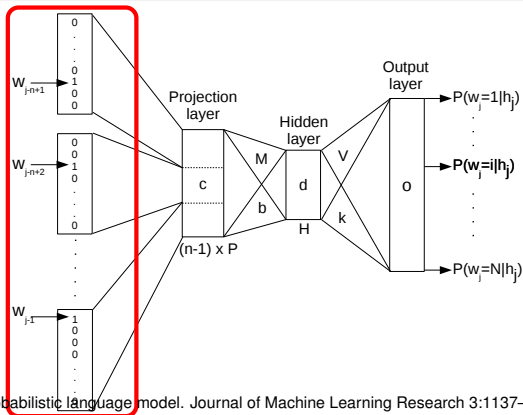
• Saklı Katman

$$d_j = \tanh \left(\sum_{l=1}^{(n-1) \times P} M_{jl} c_l + b_j \right)$$

• Çıktı Katmanı

$$o_i = \sum_{j=1}^H V_{ij} d_j + k_i$$

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$



¹²Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of Machine Learning Research 3:1137–1155

¹³Schwenk H (2007) Continuous space language models. Computer Speech and Language 21(3):492–518

İleribesleme Sinir Ağı (SA) Dil Modeli^{12 13}

Olasılık kestirimleri **kelimelerin sürekli uzak gösterimleri** ile yapılır.

- Giriş Katmanı

- İzdüşüm Katmanı**

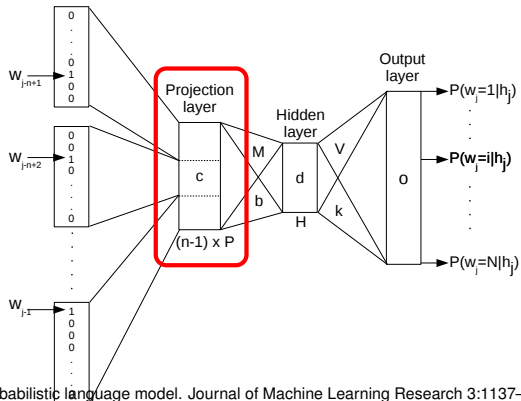
- Saklı Katman

$$d_j = \tanh \left(\sum_{l=1}^{(n-1) \times P} M_{jl} c_l + b_j \right)$$

- Çıktı Katmanı

$$o_i = \sum_{j=1}^H V_{ij} d_j + k_i$$

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$



¹²Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of Machine Learning Research 3:1137–1155

¹³Schwenk H (2007) Continuous space language models. Computer Speech and Language 21(3):492–518

İleribesleme Sinir Ağı (SA) Dil Modeli^{12 13}Olasılık kestirimleri **kelimelerin sürekli uzak gösterimleri** ile yapılır.

• Giriş Katmanı

• İzdüşüm Katmanı

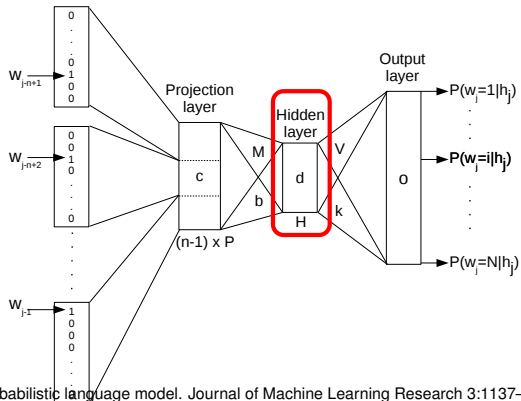
• **Saklı Katman**

$$d_j = \tanh \left(\sum_{l=1}^{(n-1) \times P} M_{jl} c_l + b_j \right)$$

• Çıktı Katmanı

$$o_i = \sum_{j=1}^H V_{ij} d_j + k_i$$

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$



¹²Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of Machine Learning Research 3:1137–1155

¹³Schwenk H (2007) Continuous space language models. Computer Speech and Language 21(3):492–518

İleribesleme Sinir Ağı (SA) Dil Modeli^{12 13}Olasılık kestirimleri **kelimelerin sürekli uzay gösterimleri** ile yapılır.

• Giriş Katmanı

• İzdüşüm Katmanı

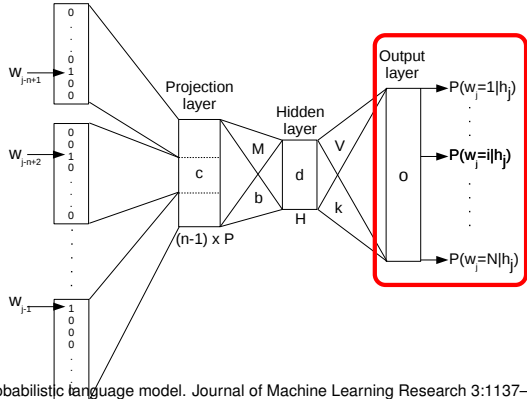
• Saklı Katman

$$d_j = \tanh \left(\sum_{l=1}^{(n-1) \times P} M_{jl} c_l + b_j \right)$$

• Çıktı Katmanı

$$o_i = \sum_{j=1}^H V_{ij} d_j + k_i$$

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$



¹²Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of Machine Learning Research 3:1137–1155

¹³Schwenk H (2007) Continuous space language models. Computer Speech and Language 21(3):492–518

İleribesleme Sinir Ağı (SA) Dil Modeli^{12 13}İleribesleme SA dil modeli **hata geri yayılım** algoritması ile eğitilir.

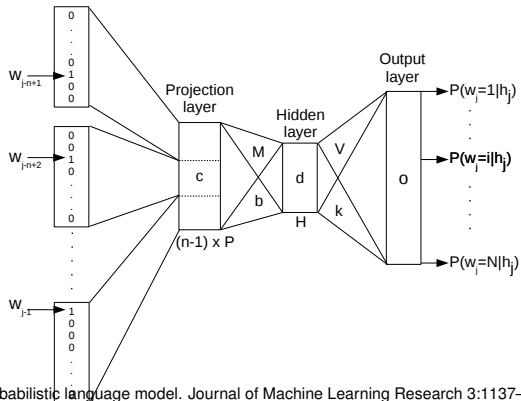
- Giriş Katmanı
- İzdüşüm Katmanı
- Saklı Katman

$$d_j = \tanh \left(\sum_{l=1}^{(n-1) \times P} M_{jl} c_l + b_j \right)$$

- Çıktı Katmanı

$$o_i = \sum_{j=1}^H V_{ij} d_j + k_i$$

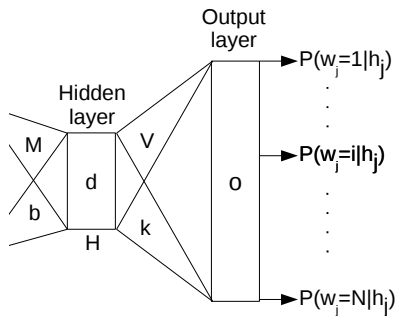
$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$



¹² Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. Journal of Machine Learning Research 3:1137–1155

¹³ Schwenk H (2007) Continuous space language models. Computer Speech and Language 21(3):492–518

İleribesleme Sinir Ağı (SA) Dil Modeli



SA Dil Modelindeki Zorluklar:

- Çıktı katmanındaki hesaplama karmaşıklığı ($H \times |V|$)

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{|V|} \exp(o_r)} = P(w_j = i | h_j)$$

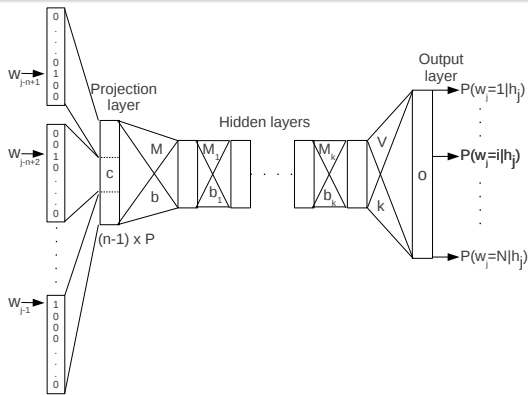
- $H \approx 300 - 500, |V| \approx 60K - 200K$

Önerilen Çözümler:

- Daha küçük çıkış dağarcığı kullanmak: $|V_o| \ll |V| \rightarrow |V_o| \approx 20K$
- Hiyerarşik bir çıkış katmanını tanımlamak: $P(w|h) = P(c(w)|h)P(w|h, c(w))$

Derin Sinir Ağı (DSA) Dil Modeli ¹⁴

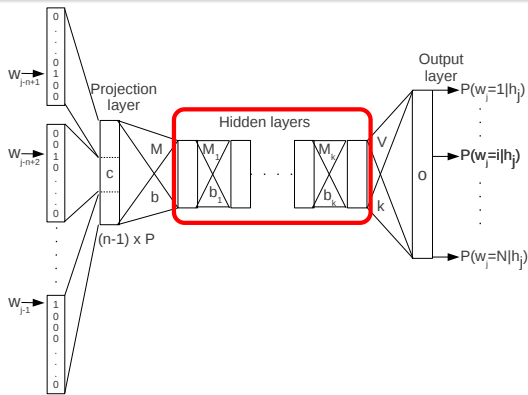
DSA birden fazla doğrusal olmayan katmana sahiptir.



¹⁴Arısoy E, Sainath T, Kingsbury B, Ramabhadran B (2012) Deep neural network language models. In: Proceedings of NAACL-HLT Workshop,

Derin Sinir Ağı (DSA) Dil Modeli ¹⁴

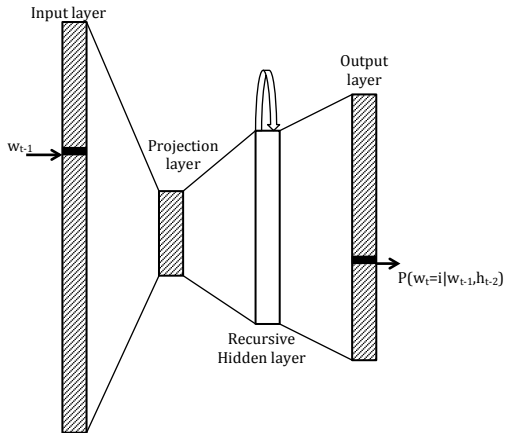
DSA birden fazla doğrusal olmayan katmana sahiptir.



¹⁴Arısoy E, Sainath T, Kingsbury B, Ramabhadran B (2012) Deep neural network language models. In: Proceedings of NAACL-HLT Workshop,

Yinelemeli Sinir Ağı (SA) Dil Modeli ¹⁵

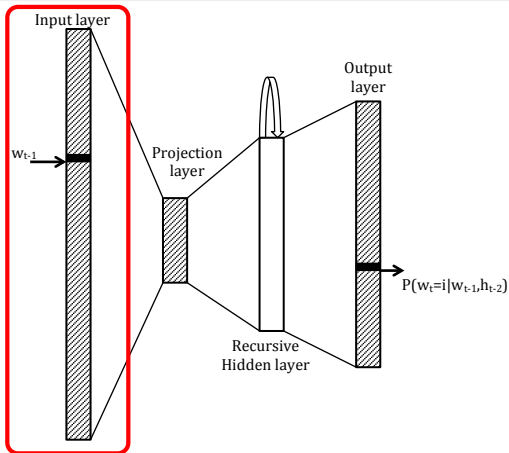
- Olasılık kestirimleri **yinelemeli SA** ile yapılır.
- Daha **uzun geçmiş kelime** kullanılarak koşullu olasılıklar hesaplanır.
- Zaman boyunca **geri yayılım** algoritması ile eğitim yapılır.



¹⁵Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S (2010) Recurrent neural network based language model. In: Proceedings of INTERSPEECH, Japan, pp 1045–1048

Yinelemeli Sinir Ağı (SA) Dil Modeli ¹⁵

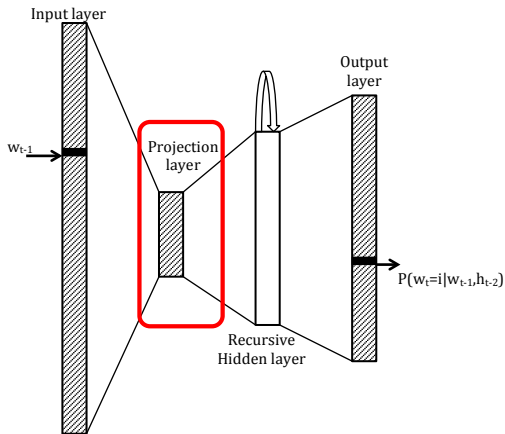
- Olasılık kestirimleri **yinelemeli SA** ile yapılır.
- Daha **uzun geçmiş kelime** kullanılarak koşullu olasılıklar hesaplanır.
- Zaman boyunca **geri yayılım** algoritması ile eğitim yapılır.



¹⁵Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S (2010) Recurrent neural network based language model. In: Proceedings of INTERSPEECH, Japan, pp 1045–1048

Yinelemeli Sinir Ağı (SA) Dil Modeli ¹⁵

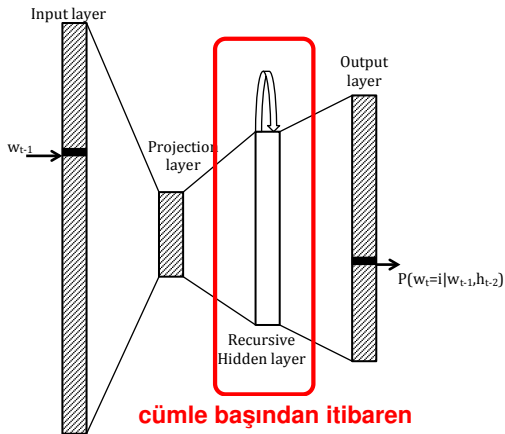
- Olasılık kestirimleri **yinelemeli SA** ile yapılır.
- Daha **uzun geçmiş kelime** kullanılarak koşullu olasılıklar hesaplanır.
- Zaman boyunca **geri yayılım** algoritması ile eğitim yapılır.



¹⁵Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S (2010) Recurrent neural network based language model. In: Proceedings of INTERSPEECH, Japan, pp 1045–1048

Yinelemeli Sinir Ağı (SA) Dil Modeli ¹⁵

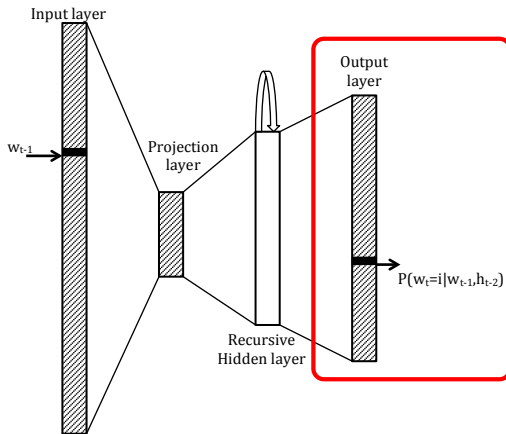
- Olasılık kestirimleri **yinelemeli SA** ile yapılır.
- Daha **uzun geçmiş kelime** kullanılarak koşullu olasılıklar hesaplanır.
- Zaman boyunca **geri yayılım** algoritması ile eğitim yapılır.



¹⁵Mikolov, T., Karafiat, M., Burget, J., Cernocký, J., Khudanpur, S. (2010). Recurrent neural network based language model. In: Proceedings of

Yinelemeli Sinir Ağı (SA) Dil Modeli ¹⁵

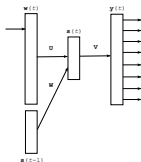
- Olasılık kestirimleri **yinelemeli SA** ile yapılır.
- Daha **uzun geçmiş kelime** kullanılarak koşullu olasılıklar hesaplanır.
- Zaman boyunca **geri yayılım** algoritması ile eğitim yapılır.



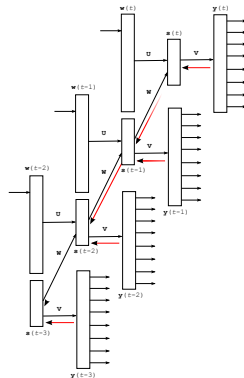
¹⁵Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S (2010) Recurrent neural network based language model. In: Proceedings of INTERSPEECH, Japan, pp 1045–1048

Yinelemeli Sinir Ağı (SA) Dil Modeli

Yinelemeli SA modeli zaman boyunca **hata geri yayılım** algoritması ile eğitilir.



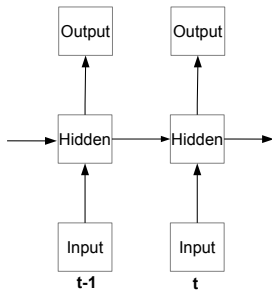
Şekiller Mikolov 2012'den alınmıştır ¹⁶.



¹⁶Mikolov T (2012) Statistical Language Models Based on Neural Networks. Ph. D. thesis, Brno University of Technology

Çift Yönlü Yinelemeli Sinir Ağı (SA) Dil Modeli

Zamanda açılmış **tek yönlü** yinelemeli SA

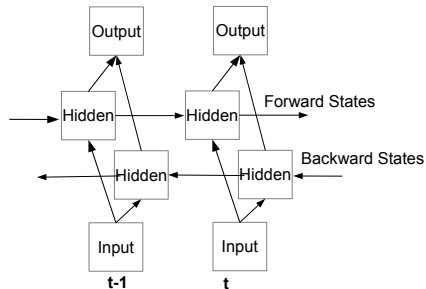


$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

$$y_t = W_{hy}h_t + b_y$$

$$p(w_t = i | w_{t-1}, h_{t-2}) = \frac{\exp(y_t^i)}{\sum_{j=1}^N \exp(y_t^j)}$$

Zamanda açılmış **çift yönlü** yinelemeli SA



$$h_t^F = \tanh(W_{xh}^F x_t + W_{hh}^F h_{t-1}^F + b_h^F)$$

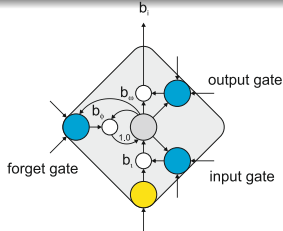
$$h_t^B = \tanh(W_{xh}^B x_t + W_{hh}^B h_{t+1}^B + b_h^B)$$

$$y_t = W_{hy}^F h_t^F + W_{hy}^B h_t^B + b_y$$

Uzun Kısa-Süreli Bellik Sinir Ağı (SA) Dil Modeli ¹⁷

Yinelemeli SA Dil Modelinin Kısıtları:

- Yinelemeli SA pratikte 5-10 zaman adımından bilgi kullanır.
- Gradyan zaman içerisinde üstel olarak azalır.
- **Uzun Kısa-Süreli Bellik SA:** Doğrusal olmayan birimler hafıza bloklarıyla değiştirilir. Zaman içindeki bağımlılıkları daha iyi modeller.



$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$$

$$h_t = o_t \tanh(c_t)$$

¹⁷Sundermeyer M, Schluter R, Ney H (2012) LSTM Neural Networks for Language Modeling. In: Proceedings of INTERSPEECH, Portland, OR, USA

Sinir Ağı Dil Modeli – Sonuçlar

İngilizce Haber Programları için GDSKT¹⁸

Dil Modeli	KHO (%)
4'lü kelime*	13.0
+ tek yönlü - yinelemeli**	12.7
+ çift yönlü - yinelemeli**	12.5
+ tek yönlü - uzun kısa-süreli bellek**	12.4
+ çift yönlü - uzun kısa-süreli bellek**	12.4

* eğitim verisi: 350 M kelime

** $P = 180$, $H = 300$, eğitim verisi: 12 M kelime

¹⁸ Arısoy E, Sethy A, Ramabhadran B, Chen S (2015) Bidirectional recurrent neural network language models for automatic speech recognition. In: Proceedings of ICASSP, Brisbane, Australia, pp 5421-5425

Sinir Ağı Dil Modeli – Sonuçlar

İngilizce Haber Programları için GDSKT¹⁸

Dil Modeli	KHO (%)
4'lü kelime*	13.0
+ tek yönlü - yinelemeli**	12.7
+ çift yönlü - yinelemeli**	12.5
+ tek yönlü - uzun kısa-süreli bellek**	12.4
+ çift yönlü - uzun kısa-süreli bellek**	12.4

* eğitim verisi: 350 M kelime

** $P = 180$, $H = 300$, eğitim verisi: 12 M kelime

¹⁸ Arısoy E, Sethy A, Ramabhadran B, Chen S (2015) Bidirectional recurrent neural network language models for automatic speech recognition. In: Proceedings of ICASSP, Brisbane, Australia, pp 5421-5425

Süreklî Uzak Gösterimleri

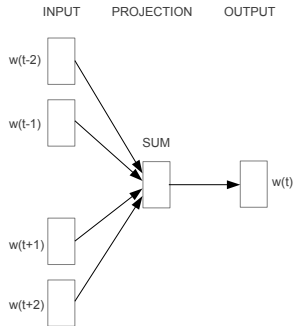
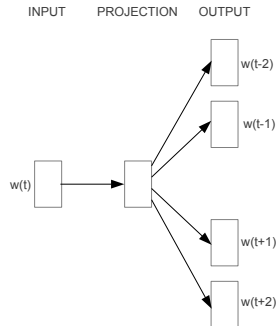
- **Süreklî uzak gösterimleri** doğal dil işleme uygulamaları için önemli bilgiler taşır ¹⁹.

$$\overrightarrow{France} - \overrightarrow{Paris} + \overrightarrow{Italy} = \overrightarrow{Roma}$$

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

¹⁹Mikolov T, et al. (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of NIPS 2013, pages 3111–3119

Sürekli Uzay Gösterimleri

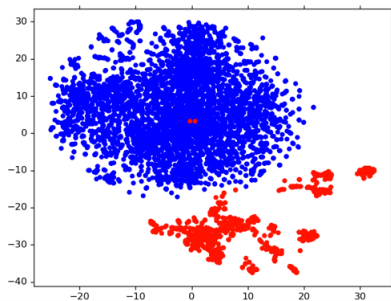
**CBOW****Skip-gram**

Şekiller Mikolov 2013'den alınmıştır ²⁰.

²⁰Mikolov T, et al. (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of NIPS 2013, pages 3111–3119

Sürekli Uzay Gösterimleri

Türkçe kök ve kök sonrası sürekli uzay vektörlerinin iki boyutta gösterimi (TSNE)



+in

+ların

+lerinin

+sinin

+ın

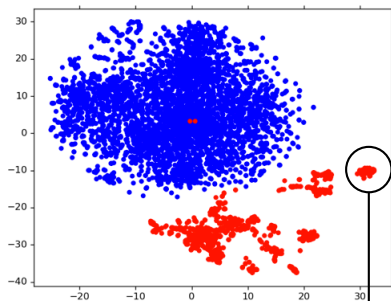
+lerin

+larının

+sının

Sürekli Uzay Gösterimleri

Türkçe kök ve kök sonrası sürekli uzay vektörlerinin iki boyutta gösterimi (TSNE)



+in	+ların	+lerinin	+sinin
+ın	+lerin	+larının	+sının

Türkçe Yinelemeli Sinir Ağı Dil Modeli

Türkçe Haber Programları için GDSKT²¹

Akustik Model	Dil Modeli	Sinama KHO (%)
Zaman gecikmeli SA	3'lü kelime	10.92
Zaman gecikmeli SA	3'lü kök ve kök sonrası* + Uzun Kısa-Sürekli Bellek**	10.12 9.83

* Yüzeysel gösterim

** $P = 180$, $H = 300$, eğitim verisi: 1.3 M kelime

²¹ Arısoy E, Saraçlar M (2018) Turkish Broadcast News Transcription Revisited. In: Proceedings of SIU 2018, Çeşme, Türkiye

Türkçe Yinelemeli Sinir Ağı Dil Modeli

Türkçe Haber Programları için GDSKT²¹

Akustik Model	Dil Modeli	Sinama KHO (%)
Zaman gecikmeli SA	3'lü kelime	10.92
Zaman gecikmeli SA	3'lü kök ve kök sonrası* + Uzun Kısa-Süreli Bellek**	10.12 9.83

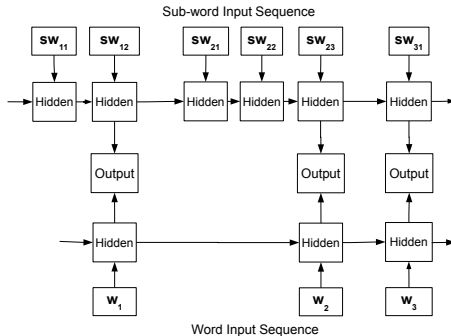
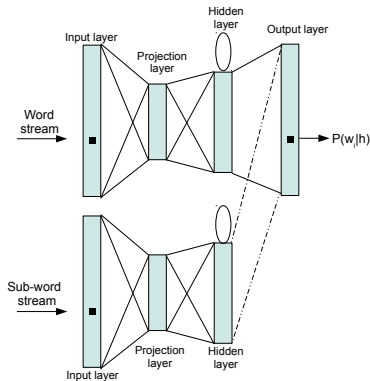
* Yüzeysel gösterim

** $P = 180$, $H = 300$, eğitim verisi: 1.3 M kelime

²¹ Arısoy E, Saraçlar M (2018) Turkish Broadcast News Transcription Revisited. In: Proceedings of SIU 2018, Çeşme, Türkiye

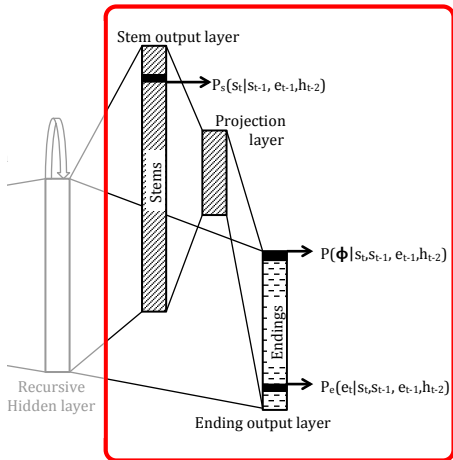
Türkçe Sinir Ağı Dil Modeli Mimarileri

Kelime ve Kelime-altı Birimlerin Yinelemeli Sinir Ağı Dil Modelinde Beraber Kullanılması ²²



²² Arısoy E, Saraçlar M (2015) Multi-Stream Long Short-Term Memory Neural Network Language Model. In: Proceedings of Interspeech, Dresden, Germany

Türkçe Sinir Ağı Dil Modeli Mimarileri



Kök Çıktı Katmanı

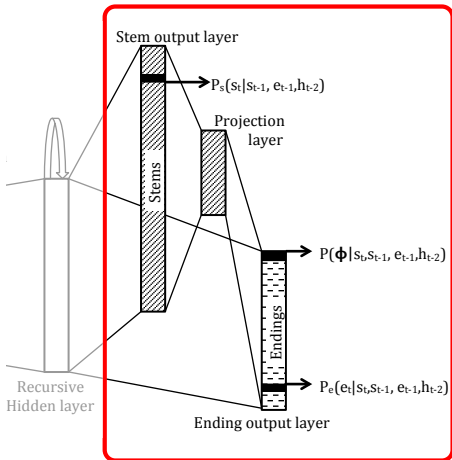
- $P_s(s_t | s_{t-1}, e_{t-1}, h_{t-2})$

Kök Sonrası Çıktı Katmanı

- $P_\phi = P(\phi | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P_e = P_e(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2}) =$

$$\begin{cases} P_\phi & \text{ek yoksa,} \\ (1 - P_\phi) \times P_e & \text{ek varsa.} \end{cases}$$

Türkçe Sinir Ağı Dil Modeli Mimarileri



Kök Çıktı Katmanı

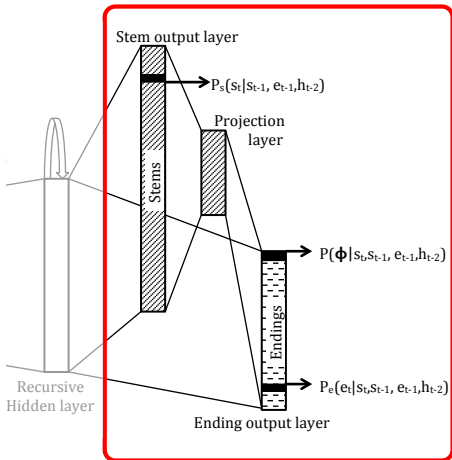
- $P_s(s_t | s_{t-1}, e_{t-1}, h_{t-2})$

Kök Sonrası Çıktı Katmanı

- $P_\phi = P(\phi | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P_e = P_e(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2}) =$

$$\begin{cases} P_\phi & \text{ek yoksa,} \\ (1 - P_\phi) \times P_e & \text{ek varsa.} \end{cases}$$

Türkçe Sinir Ağı Dil Modeli Mimarileri



Kök Çıktı Katmanı

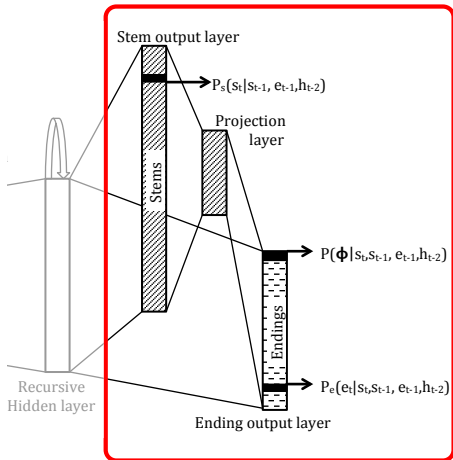
- $P_s(s_t | s_{t-1}, e_{t-1}, h_{t-2})$

Kök Sonrası Çıktı Katmanı

- $P_\phi = P(\phi | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P_e = P_e(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2}) =$

$$\begin{cases} P_\phi & \text{ek yoksa,} \\ (1 - P_\phi) \times P_e & \text{ek varsa.} \end{cases}$$

Türkçe Sinir Ağı Dil Modeli Mimarileri



Kök Çıktı Katmanı

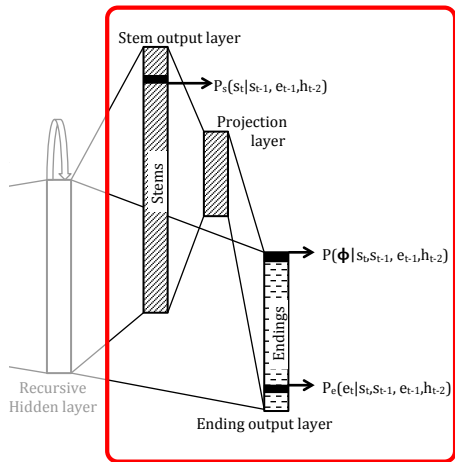
- $P_s(s_t | s_{t-1}, e_{t-1}, h_{t-2})$

Kök Sonrası Çıktı Katmanı

- $P_\phi = P(\phi | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P_e = P_e(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2})$
- $P(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2}) =$

$$\begin{cases} P_\phi & \text{ek yoksa,} \\ (1 - P_\phi) \times P_e & \text{ek varsa.} \end{cases}$$

Türkçe Sinir Ağı Dil Modeli Mimarileri



Dil Modelinde Kullanılan Olasılıklar

- $$P(s_t, e_t | s_{t-1}, e_{t-1}, h_{t-2}) =$$

$$P_s(s_t | s_{t-1}, e_{t-1}, h_{t-2})$$

$$\times P(e_t | s_t, s_{t-1}, e_{t-1}, h_{t-2})$$

TEŞEKKÜRLER