

Uçtan Uca Anahtar Sözcük Arama



Alican Gök ve Murat Saraçlar

Boğaziçi Üniversitesi, Elektrik Elektronik Bölüm Başkanlığı

alicangok@gmail.com, murat.saraclar@boun.edu.tr

1. Özet

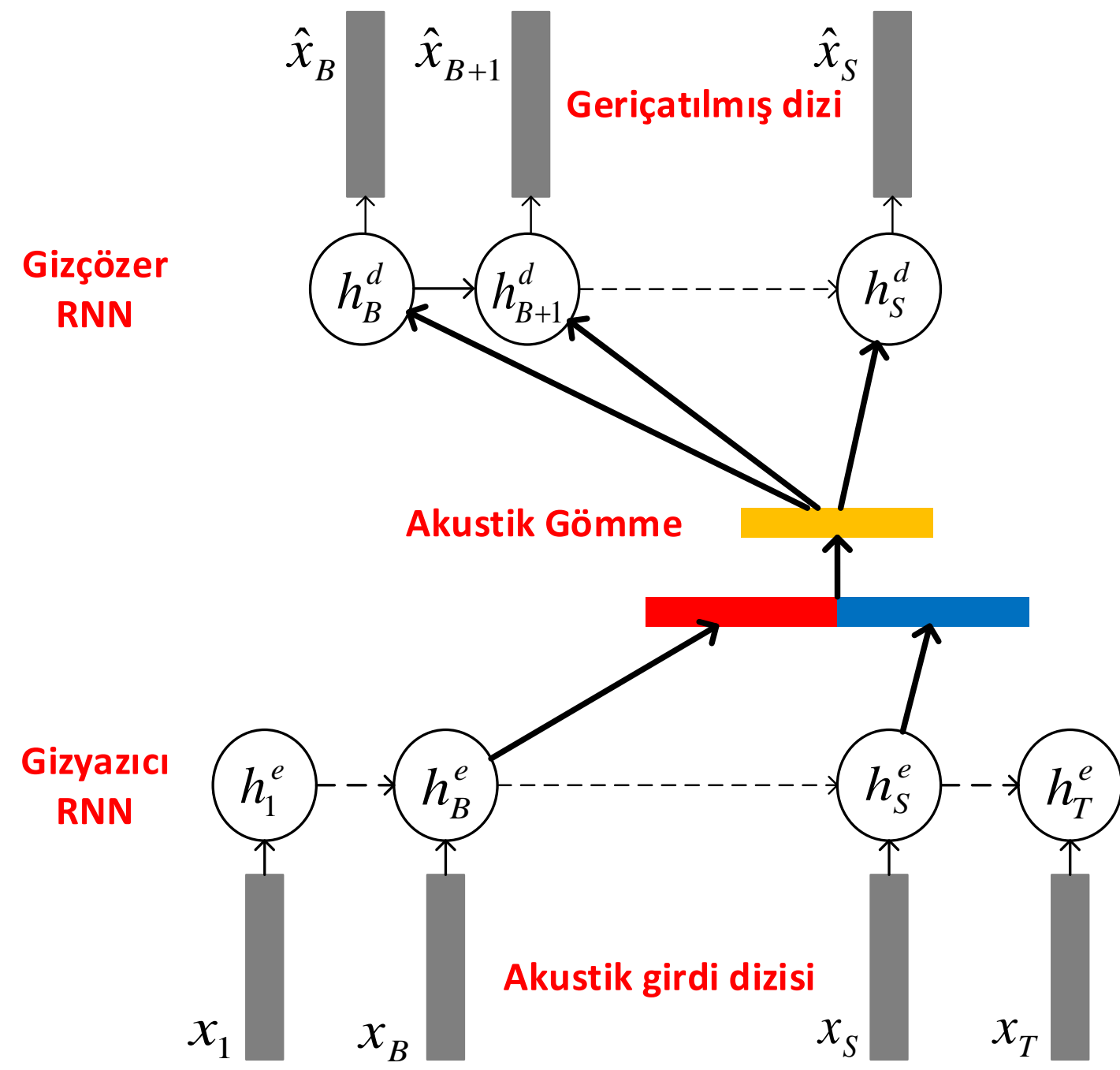
Bu çalışmada, klasik otomatik konuşma tanıma (OKT) tabanlı sistemlerin aksine, düşük gözetimle eğitilmiş bir **uçtan uca anahtar sözcük arama** sistemi tanıtılmaktadır. Bu sistem **üç alt-sistem**den oluşmaktadır. İlki, konuşmayı sabit boyutlu bir gömme kullanarak geriçatmak için eğitilmiş, **özyinelemeli sinir ağı (RNN) tabanlı bir akustik özgizyazıcı**; ikincisi ise karakter seviyesinde **evrişimli sinir ağı - özyinelemeli sinir ağı (CNN-RNN) tabanlı**, sorgu terimleri için sabit boyutlu bir gömme çıkartan bir **dil modelidir**. Bu iki gömme farklı gösterim uzaylarında olukları için, anahtar sözcüğün konuşmada yer alıp almadığını öngören **üçüncü bir ileri beslemeli sinir ağına** girdi olarak eklenmektedir. Klasik OKT-tabanlı sistemlere göre çok daha hızlı eğitilebilen bu uçtan uca sistem, özellikle dağarcık dışı terimlerin bulunmasında önemli derecede iyileşme sağlamaktadır.

2. Anahtar Sözcük Arama Nedir?

Anahtar Sözcük Arama (ASA) metin olarak verilen bir sorgu teriminin konuşma verisi içerisinde aranması işlemidir.

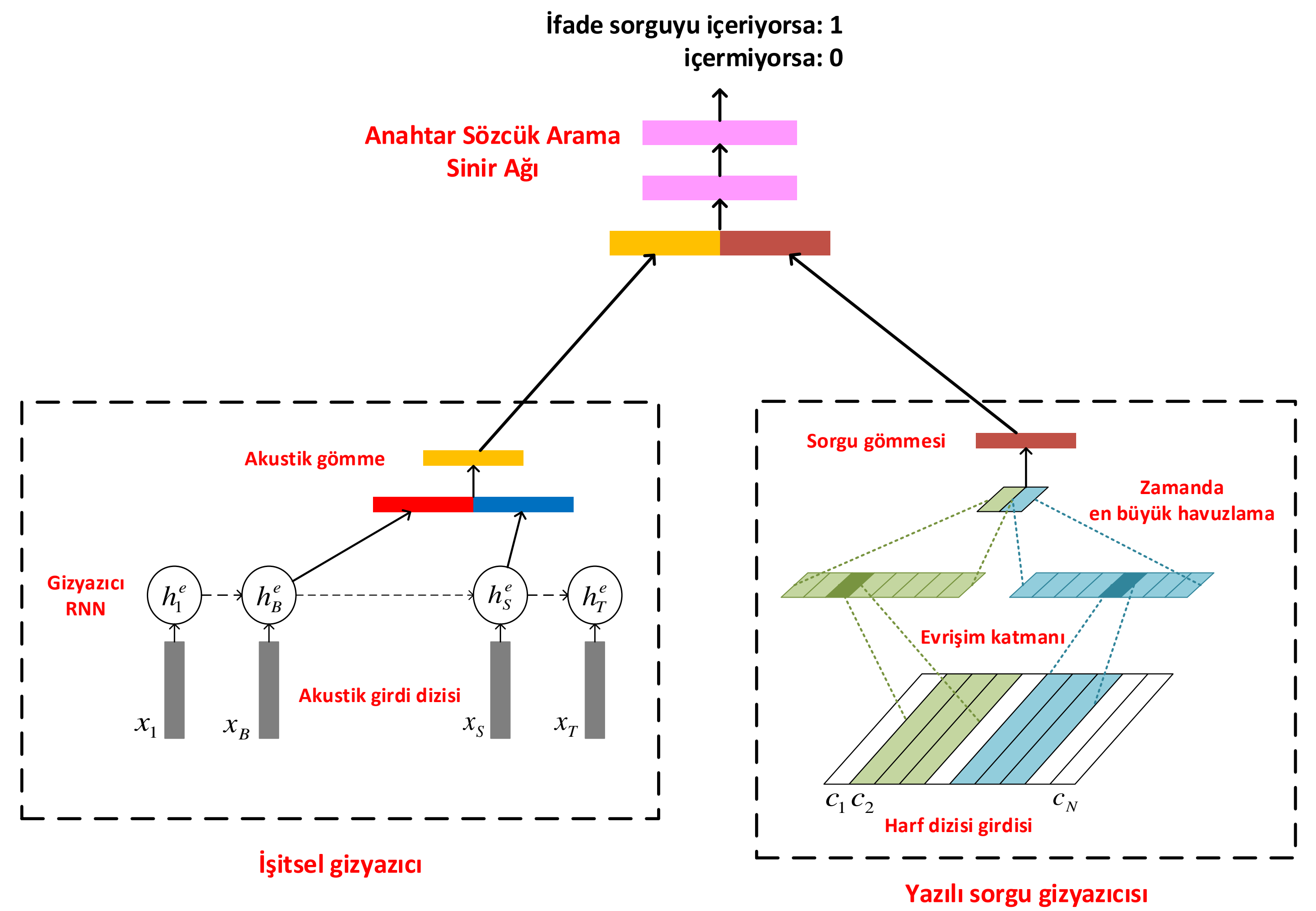
3. Ses Girdisinin Modellenmesi

Bu ağın amacı, anahtar sözcüklerin taratılacağı, farklı uzunluklarda olabilen ses dizilerini betimleyen sabit boyutlu gömmeler elde etmektir. Gözetimsiz eğitilen bu özgizyazıcıda girdi olarak ses dizilerinin posteriorgramı ya da çok dilli darboğaz öznitelikleri kullanılabilir. Önerdiğimiz modelin temeli, [1] numaralı çalışmaya dayanmaktadır.



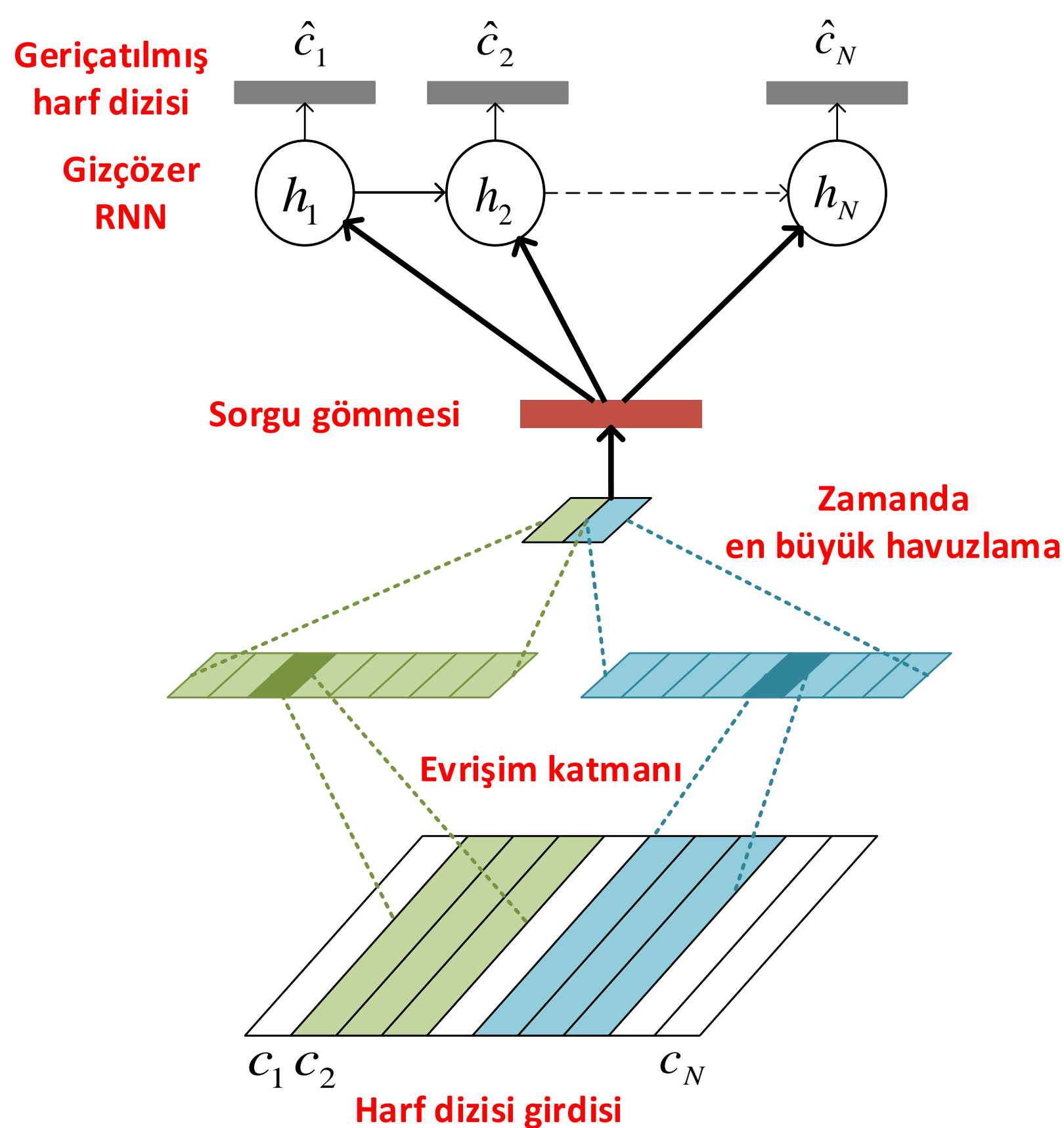
5. Anahtar Sözcük Arama Ağı

Anahtar sözcüğün taranacağı ses ile yazılı sözcüğün gömmeleri bu nihai ağda birleştirilir. İleri beslemeli bu ağ, olumlu ve olumsuz örneklerle gözetimli bir biçimde eğitilir.



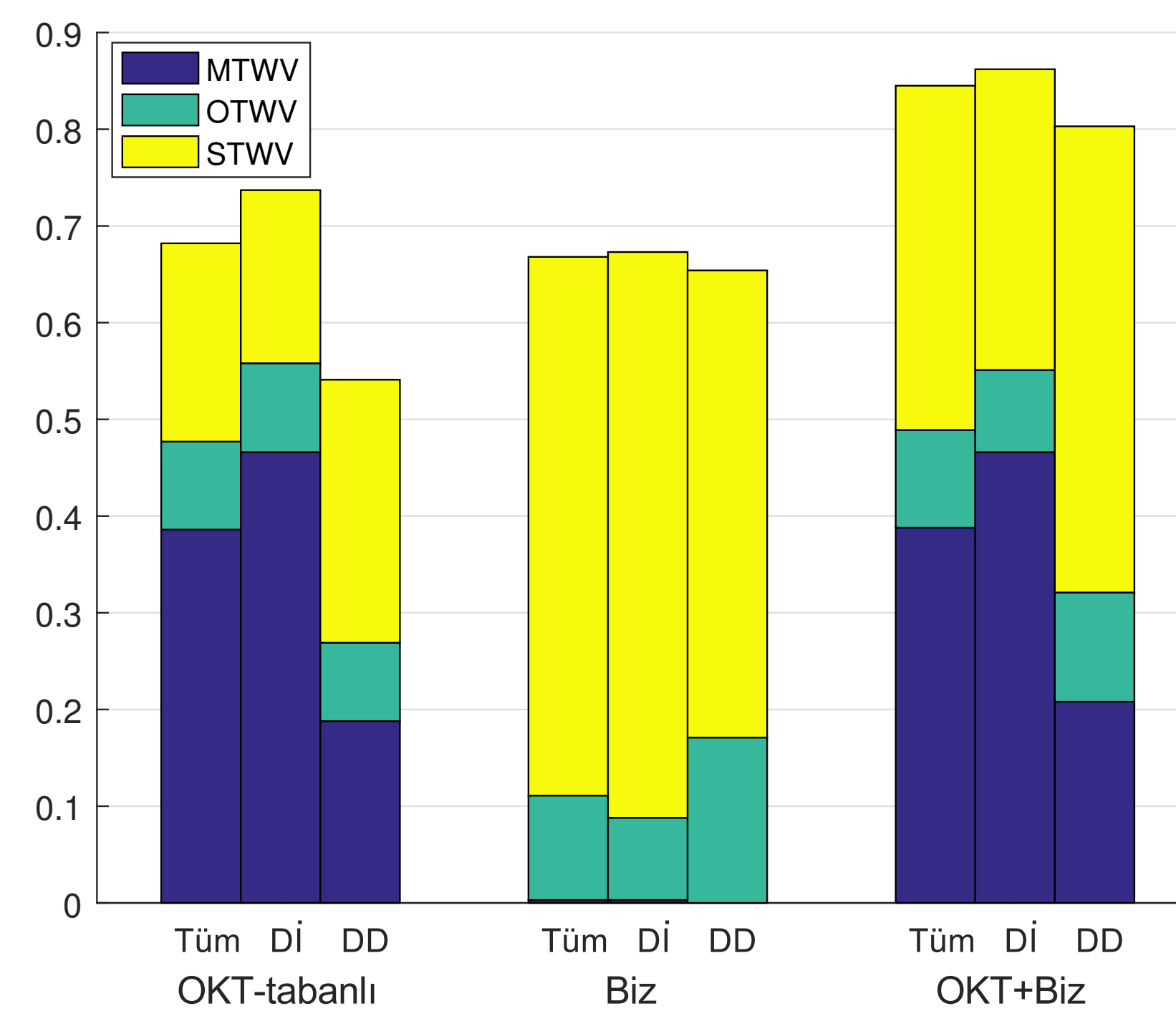
4. Metin Sorgunun Modellenmesi

Benzer şekilde, farklı uzunluklarda olabilen anahtar sözcükleri sabit boyutlu bir uzayda göstermek için aşağıdaki karakter-tabanlı dil modeli kullanılmıştır. Bu ağ, hedef dile ait herhangi bir grup yazı dizisi ile gözetimsiz olarak eğitilir.



6. Sonuçlar ve Tartışma

- Türkçe dili için, 10'ar saatlik eğitim ve sınav verileri ile eğitilmiş ve sınanmıştır.
- Eğitilmesi günler süren OKT-tabanlı daha karmaşık sistemlerin aksine birkaç saatte eğitilebilen bu sistem, bilgimiz dahilinde gerçekçi ölçevler ile sınanmış, uçtan uca yapay sinir ağı tabanlı ilk büyük dağarcıklı ASA çalışmasıdır.
- Sistemimiz klasik OKT-tabanlı sistemlerle birleştirildiğinde [2], sistemimizin ASA başarımını özellikle dağarcık dışı (DD) terimlerde önemli derecede artırdığı görülmüştür.
 - MTWV: Tüm anahtar sözcüklere aynı eşik değeri ile elde edilen tespit başarımı,
 - OTWV: Farklı sözcükler için en iyi eşik değeri ile elde edilen tespit başarımı,
 - STWV: Sahte alarmlara maliyet uygulanmaksızın elde edilen tespit başarımı.



7. Kaynakça

- [1] K. Audhkhasi, A. Rosenberg, A. Sethy, B. Ramabhadran, and B. Kingsbury. End-to-end ASR-free keyword search from speech. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1351–1359, Dec 2017.
- [2] B. Gündoğdu, B. Yusuf, and M. Saraçlar. Fusion of event detection systems: An application to keyword search. In *2018 26th Signal Processing and Communication Application Conference (SIU)*, May 2018.