

Derin Öğrenme Deep Learning

Hazırlayan: M. Ali Akcayol
Gazi Üniversitesi
Bilgisayar Mühendisliği Bölümü

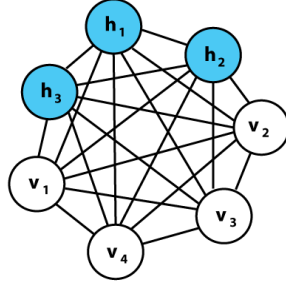
Bu dersin sunumları, "Simon Haykin, Neural Networks and Learning Machines, Prentice Hall, 2016." kitabı kullanılarak hazırlanmıştır.

İçerik

- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ RBM ve autoencoder
- ▶ RBM yapısı
- ▶ RBM eğitimi
- ▶ RBM hiper parametreleri
- ▶ RBM ile işbirlikçi filtreleme
- ▶ RBM uygulamaları

Boltzmann makinesi

- ▶ **Boltzmann makinesi** (**Boltzmann machine**), stokastik ve üretken (**generative**) yapay sinir ağıdır.
- ▶ Boltzmann makinesi üretken bir derin öğrenme modelidir.



- ▶ **Görünür (visible)** ve **gizli (hidden)** düğümlere sahiptir.
- ▶ Boltzmann makinesi giriş düğümleri ve çıkış düğümleri arasında bağlantıya sahiptir.

3

İçerik

- ▶ Boltzmann makinesi
- ▶ **Kısıtlanmış Boltzmann makinesi**
- ▶ RBM ve autoencoder
- ▶ RBM yapısı
- ▶ RBM eğitimi
- ▶ RBM hiper parametreleri
- ▶ RBM ile işbirlikçi filtreleme
- ▶ RBM uygulamaları

4

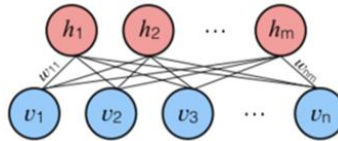
Kısıtlanmış Boltzmann makinesi

- ▶ **Kısıtlanmış Boltzmann Makinesi (Restricted Boltzmann Machine-RBM)**, yapay sinir ağıdır.
- ▶ **RBM yapısı**, feedforward ANN ve CNN yapılarından farklıdır.
- ▶ 1986 yılında **Harmonium** adıyla geliştirildi (Paul Smolensky).
- ▶ **2001**: **Contrastive divergence** kullanıldı (Geoffrey E. Hinton).
- ▶ **2007**: Büyük veri ile ilk **işbirlikçi filtreleme uygulaması** geliştirildi (Netflix movie rating).
- ▶ **2006-2010**: Çok sayıda RBM birlikte kullanılarak **Deep Belief Network** geliştirildi.
- ▶ **2010 sonrası**: Diğer deep learning modelleri daha yaygın hale geldi.
- ▶ Diğer ağların parametrelerinin başlangıç değerlerini belirlemek için kullanılmaktadır.

5

Kısıtlanmış Boltzmann makinesi

- ▶ **RBM**, iki katmanlı üretken kapasiteye sahip yapay sinir ağıdır.
- ▶ Boltzmann makinesinin giriş düğümleri arasında ve çıkış düğümleri arasında bağlantıya sahip olmayan (**kısıtlanmış, restricted**) alt türüdür.
- ▶ Birden çok RBM birleştirilerek kullanılabilir (**Deep Belief Network**).



6

Kısıtlanmış Boltzmann makinesi

- ▶ Görünür ve gizli katman arasında **tam bağıli bipartite graph yapısına sahiptir.**
- ▶ **Giriş katmanındaki her düğüm çıkış katmanındaki tüm düğümlerle bağlantıya sahiptir.**
- ▶ Giriş kümesindeki olasılık dağılımını öğrenebilir.
- ▶ Ağın eğitimi gradient-based **contrastive divergence algoritması kullanılarak yapılır.**
- ▶ **Boyut indirgeme, sınıflandırma, işbirlikçi filtreleme, özellik öğrenme** için farklı uygulamalarda başarılı bir şekilde kullanılmaktadır.

7

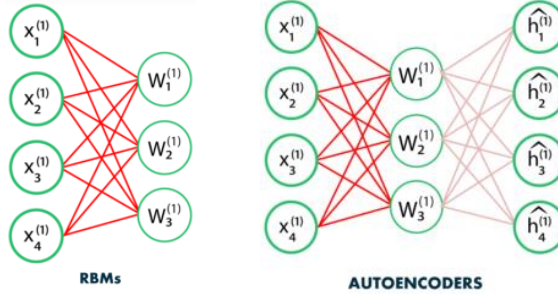
İçerik

- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ **RBM ve autoencoder**
- ▶ RBM yapısı
- ▶ RBM eğitimi
- ▶ RBM hiper parametreleri
- ▶ RBM ile işbirlikçi filtreleme
- ▶ RBM uygulamaları

8

RBM ve autoencoder

- ▶ **Autoencoder**, basit bir 3 katmanlı sinir ağıdır.
- ▶ **Gizli düğüm sayısı** genellikle **giriş-çıkış katmanlardaki düğüm sayısından azdır**.
- ▶ **Eğitimde** çıkışları girişlere yaklaştıracak **ağırlık değerleri hesaplanır**.
- ▶ **RBM'de, girişleri tekrar elde etmek için ileri ve geri geçiş yapılır**.



9

İçerik

- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ RBM ve autoencoder
- ▶ **RBM yapısı**
- ▶ RBM eğitimi
- ▶ RBM hiper parametreleri
- ▶ RBM ile işbirlikçi filtreleme
- ▶ RBM uygulamaları

10

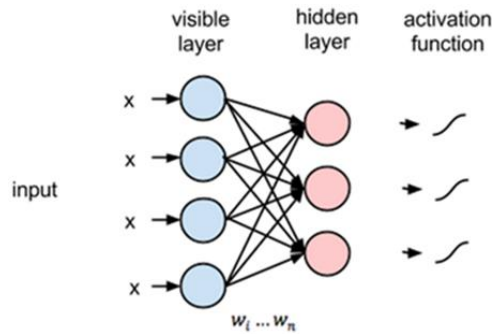
RBM yapısı

- ▶ RBM, **en basit yapay sinir ağı** modellerindedir.
- ▶ Giriş ve gizli katmandaki düğümlerin kendi arasında bağlantı yoktur.
- ▶ **İleri geçişte:**
 - ▶ Girişleri alır.
 - ▶ Ağırlıklarla çarpar.
 - ▶ Görünür katman bias değerlerini ekler.
 - ▶ Sigmoid aktivasyon fonksiyonu ile çıkışı hesaplar.
- ▶ **Geri geçişte:**
 - ▶ Gizli katman çıkışlarını giriş olarak alır.
 - ▶ Gizli katman bias değerlerini ekler.
 - ▶ Aynı ağırlıklarla çarpar.
 - ▶ Sigmoid aktivasyon fonksiyonu çıkışı ile girişleri yeniden elde eder.

11

RBM yapısı

- ▶ İteratif bir şekilde **gizli katman ve giriş katmanı değerleri hesaplanır.**



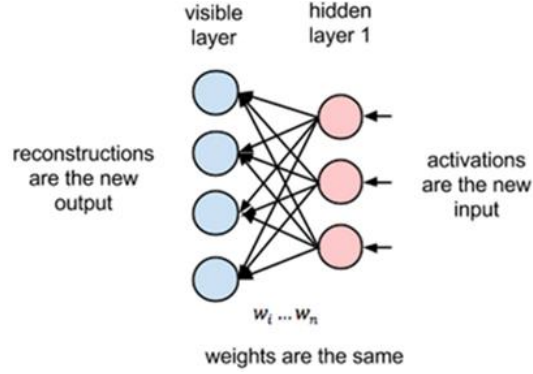
$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

$$\mathbf{h}^{(1)} = S(\mathbf{v}^{(0)T} \mathbf{W} + \mathbf{a})$$

12

RBM yapısı

- ▶ Orijinal **giriş vektörü ile yeniden elde edilen giriş vektörü arasında fark vardır.**



$$\mathbf{v}^{(1)} = S(\mathbf{h}^{(1)} W^T + \mathbf{a})$$

13

İçerik

- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ RBM ve autoencoder
- ▶ RBM yapısı
- ▶ **RBM eğitimi**
- ▶ RBM hiper parametreleri
- ▶ RBM ile işbirlikçi filtreleme
- ▶ RBM uygulamaları

14

RBM eğitimi

- ▶ Girişteki $\mathbf{v}^{(0)}$ ve $\mathbf{v}^{(1)}$ arasında hesaplanan fark, yeniden oluşturma (reconstruction) hatasını gösterir.
- ▶ Eğitim sürecinde hatanın minimize edilmesi gerekir.
- ▶ Her iterasyonda hatayı azaltacak şekilde ağırlıkların ayarlanması gerekir.
- ▶ **RBM'lerin eğitimi iki adımda gerçekleştirilir:**
 - ▶ Gibbs sampling
 - ▶ Contrastive divergence

15

RBM eğitimi

Gibbs sampling

- ▶ Eğitimde **ilk aşama Gibbs sampling** olarak adlandırılır.
- ▶ Girilen \mathbf{v} vektörü için \mathbf{h} vektörü hesaplanır.

$$p(\mathbf{h}^{(1)} \mid \mathbf{v}^{(0)}; W)$$

- ▶ Elde edilen \mathbf{h} vektörü için \mathbf{v} vektörü hesaplanır.

$$p(\mathbf{v}^{(1)} \mid \mathbf{h}^{(1)}; W)$$

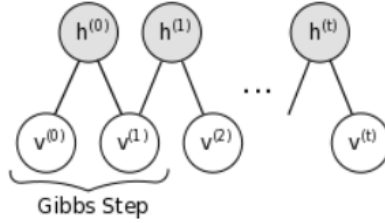
- ▶ İleri ve geri geçişte **aynı ağırlıklar kullanılır.**
- ▶ \mathbf{h} ve \mathbf{v} değerleri **k kez hesaplanır.**

16

RBM eğitimi

Gibbs sampling

- ▶ Toplam **k iterasyon** sonucunda **yeni bir \mathbf{v}_k vektörü** elde edilir.
- ▶ Elde edilen \mathbf{v}_k **vektörü**, giriş vektörü \mathbf{v}_0 **dan farklıdır.**



$$h^{(n+1)} \sim \text{sigm}(W^T v^{(n)} + c)$$
$$v^{(n+1)} \sim \text{sigm}(W h^{(n+1)} + b)$$

17

RBM eğitimi

Contrastive divergence

- ▶ **Ağırlıklar contrastive divergence** adımında **güncellenir.**
- ▶ Hesaplanan vektörler kullanılarak ağırlıklar güncellenir.
- ▶ \mathbf{v}_0 ile $p(\mathbf{h}_0|\mathbf{v}_0)$ ve \mathbf{v}_k ile $p(\mathbf{h}_k|\mathbf{v}_k)$ vektörlerinin **outer product'ları** ile ağırlık güncelleme matrisi elde edilir.
- ▶ **Ağırlık güncelleme matrisi ile** mevcut ağırlıklar toplanıp **yeni ağırlık değerleri hesaplanır.**

$$\Delta W = \mathbf{v}_0 \otimes p(\mathbf{h}_0|\mathbf{v}_0) - \mathbf{v}_k \otimes p(\mathbf{h}_k|\mathbf{v}_k)$$

$$W_{new} = W_{old} + \Delta W$$

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u} \mathbf{v}^T = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \\ u_4 v_1 & u_4 v_2 & u_4 v_3 \end{bmatrix}$$

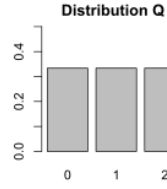
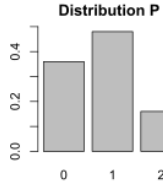
18

RBM eğitimi

KL divergence

- ▶ RBM, giriş değerleri ve tahmin edilen değerler arasındaki farkı **Kullback Leibler (KL) divergence** (göreceli entropi) ile hesaplar.
- ▶ KL-divergence, iki eğri altında çakışmayan alanı gösterir.

$$D_{KL}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$



x	0	1	2
Distribution P(x)	0.36	0.48	0.16
Distribution Q(x)	0.333	0.333	0.333

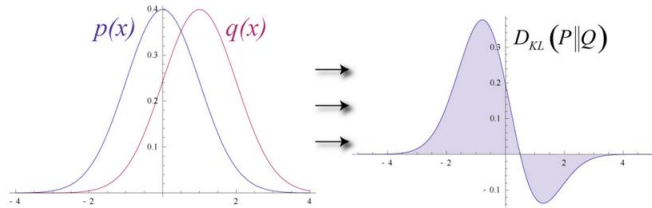
$$\begin{aligned} D_{KL}(P \parallel Q) &= \sum_{x \in \mathcal{X}} P(x) \ln \left(\frac{P(x)}{Q(x)} \right) \\ &= 0.36 \ln \left(\frac{0.36}{0.333} \right) + 0.48 \ln \left(\frac{0.48}{0.333} \right) + 0.16 \ln \left(\frac{0.16}{0.333} \right) \\ &= 0.0852996 \end{aligned}$$

19

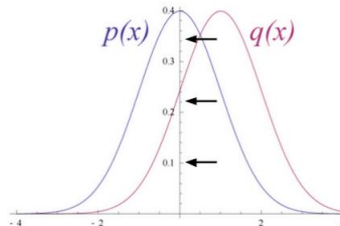
RBM eğitimi

KL divergence

- ▶ RBM eğitim algoritması KL-divergence değerini minimize eden ağırlıkları hesaplar (p orijinal giriş, q tahmin edilen değer).



- ▶ Eğitim sürecinde q vektörü p vektörüne yaklaştırılır.



20

İçerik

- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ RBM ve autoencoder
- ▶ RBM yapısı
- ▶ RBM eğitimi
- ▶ **RBM hiper parametreleri**
- ▶ RBM ile işbirlikçi filtreleme
- ▶ RBM uygulamaları

21

RBM hiper parametreleri

Öğrenme oranı

- ▶ Öğrenme oranı ile ağırlıkların değişim hızı belirlenir.
- ▶ Farklı öğrenme hızında sonuca ulaşma süresi farklıdır.
- ▶ Dinamik öğrenme oranı hata miktarına göre adım boyutunu belirler.
- ▶ **Öğrenme oranı** genellikle $[0, 1]$ arasında hata oranına göre **adaptif seçilir.**

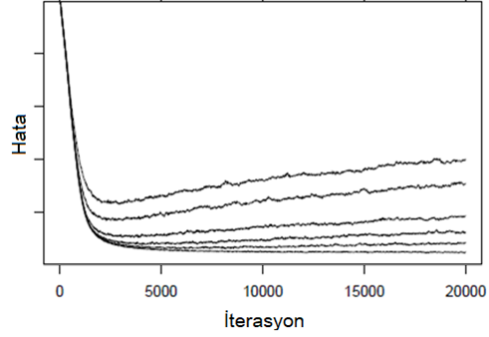
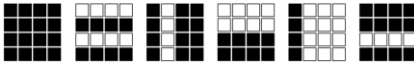
22

RBM hiper parametreleri

k örnekleme sayısı

- ▶ RBM'de farklı k örnekleme değeri için öğrenme hızı farklıdır.
- ▶ Bazı k değerleri için öğrenme istenen oranda gerçekleşmeyebilir.
- ▶ BAS (Bars-and-Stripes) veriseti için iterasyon sayısına göre sonuç değişir (k=1, 2, 5, 10, 20, 100 yukarıdan aşağıya).
- ▶ k'nın düşük değerleri için iterasyon arttıkça sonuç daha kötü hale gelmektedir.

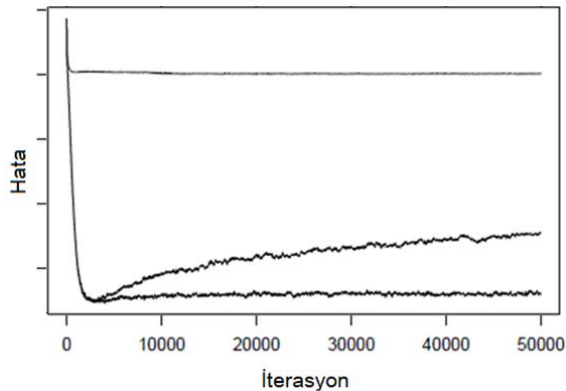
BAS veriseti



RBM hiper parametreleri

λ weight-decay değeri

- ▶ Weight-decay parametresi çok küçük veya çok büyük seçildiğinde istenen sonuç elde edilemeyebilir.
- ▶ BAS veriseti için iterasyon sayısına göre sonuç aşağıdaki gibidir ($\lambda=0.05, 0.00005, 0.0005$ yukarıdan aşağıya).



İçerik

- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ RBM ve autoencoder
- ▶ RBM yapısı
- ▶ RBM eğitimi
- ▶ RBM hiper parametreleri
- ▶ **RBM ile işbirlikçi filtreleme**
- ▶ RBM uygulamaları

25

RBM ile işbirlikçi filtreleme

- ▶ Kişilerin filmleri 1-5 arasında puanladığını düşünelim.
- ▶ **Klasik faktör analizinde her film bir grup faktör ile ilişkilendirilebilir.**
- ▶ "Harry Potter" ve "Fast and Furious" filmleri "fantasy" ve "action" ile ilişkilendirilebilir.
- ▶ **RBM ile** filmlerin aralarındaki **ilişki faktörleri tespit edilebilir.**
- ▶ RBM, kullanıcılardan alınan **rating değerleri kullanılarak eğitilir.**
- ▶ **RBM** eğitim sürecinde kullanıcıların **rating değerleri ile filmler arasındaki gizli faktörleri öğrenir.**

26

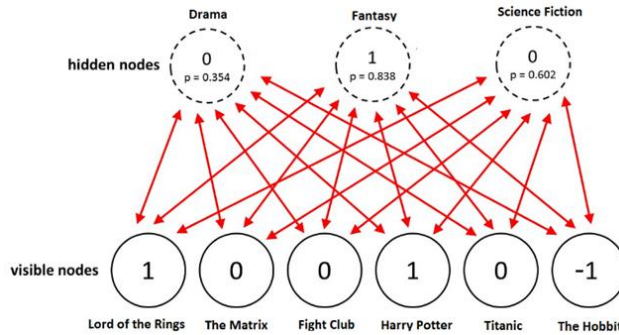
RBM ile işbirlikçi filtreleme

- ▶ Kullanıcıların film tercihlerine dayalı olarak **işbirlikçi filtreleme ile gizli faktörler belirlenmiş olur.**
- ▶ İşbirlikçi filtrelemede kullanıcıların **filmi beğenmeleri "1" ve beğenmemeleri "0"** ile gösterilebilir.
- ▶ **Oylanmayan filmler** ise **"-1"** ile gösterilebilir.
- ▶ Her kullanıcı için **binary rating değerleri giriş vektörünü oluşturur.**
- ▶ **Her gizli katman düğümü gizli faktörleri gösterir.**

27

RBM ile işbirlikçi filtreleme

- ▶ RBM ile **eğitim sürecinde üç gizli faktör kullanılabilir.**

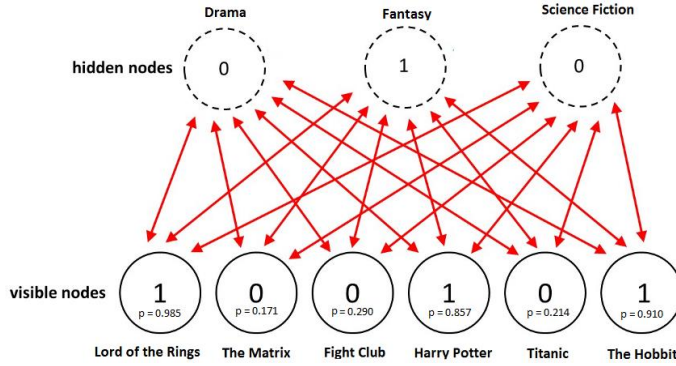


- ▶ "Lord of the Rings" ve "Harry Potter" filmleri "fantasy" türü ile ilişkilidir.
- ▶ Örnekteki kullanıcı "The Hobbit" dışındakileri oylamıştır.
- ▶ Örnekte, girilen rating değerlerine sahip kullanıcının **"Fantasy"** türü filmleri daha çok beğendiği görülmektedir.
- ▶ **Tüm kullanıcılar için ağ eğitilir.**

28

RBM ile işbirlikçi filtreleme

- ▶ RBM eğitildikten sonra **kullanıcının önceden görmediği filmler için rating değeri tahmin edebilir.**
- ▶ **Yeni bir kullanıcının mevcut rating değeri girilir.**
- ▶ Giriş değerleri Bernoulli dağılımı ile 0 veya 1 olarak elde edilir.
- ▶ Örnekte, "Lord of the Rings" ve "Harry Potter" filmlerini beğenen kullanıcının "The Hobbit" filmini beğeneceği tahmin edilmiştir.



29

İçerik

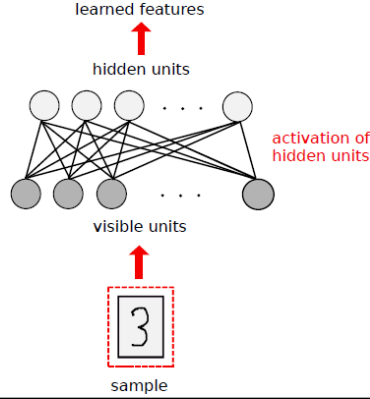
- ▶ Boltzmann makinesi
- ▶ Kısıtlanmış Boltzmann makinesi
- ▶ RBM ve autoencoder
- ▶ RBM yapısı
- ▶ RBM eğitimi
- ▶ RBM hiper parametreleri
- ▶ RBM ile işbirlikçi filtreleme
- ▶ **RBM uygulamaları**

30

RBM uygulamaları

Özellik çıkarımı

- ▶ RBM, örnek girişler için özellik çıkarımı amacıyla kullanılabilir.
- ▶ **Gizli katman nöron'larının aktif/pasif durumları girişin özellik haritasını gösterir.**
- ▶ Çıkarılan özellikler başka bir sınıflandırıcıya giriş olarak verilebilir.



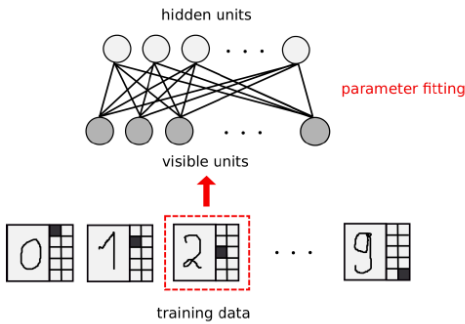
31

RBM uygulamaları

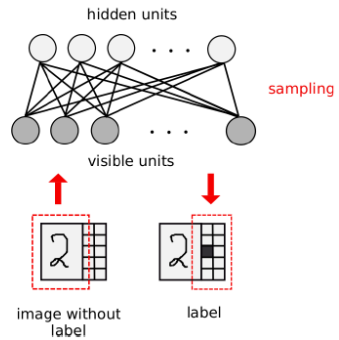
Sınıflandırma

- ▶ RBM, eğitim veri kümesi kullanılarak göre eğitilir.
- ▶ Yeni bir giriş verildiğinde görünür katmandan elde edilen değerlere göre **yeni girişi sınıflandırabilir.**

learning with labels



classification



32

RBM uygulamaları

- ▶ RBM günümüzde aşağıdaki başlıca alanlarda uygulanmaktadır:
 - ▶ Boyut indirme
 - ▶ Sınıflandırma
 - ▶ İşbirlikçi filtreleme
 - ▶ Karakter algılama
 - ▶ Özellik çıkarımı

33

Ödev

- ▶ RBM'lerin tavsiye sistemlerinde uygulamasını içeren SCI/E dergilerinde yayınlanmış bir makale hakkında ödev hazırlayınız.

34