



T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

SOSYAL AĞ ANALİZİ İÇİN BAYES AĞLARININ KULLANIMI

Betül AKKOÇ

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Temmuz-2012
KONYA

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Betül AKKOÇ

19.07.2012

ÖZET

YÜKSEK LİSANS

SOSYAL AĞ ANALİZİ İÇİN BAYES AĞLARININ KULLANIMI

Betül AKKOÇ

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Ahmet ARSLAN

2012, 52 Sayfa

Jüri
Prof. Dr. Ahmet ARSLAN
Yrd. Doç. Dr. Ö. Kaan BAYKAN
Yrd. Doç. Dr. Murat CEYLAN

İnanç ağları olarak da bilinen bayes ağları belirsizlik hakkında bilgi sunmak için kullanılır. Sosyal ağlar kişilerin sosyal ilişkilerine göre birbirlerine bağlandığı yapılardır. Bu yapıda kişiler düğüm, ilişkiler bağlantı olarak ifade edilir.

Tez çalışmasında kişilerin etkinliklere katılım verileri kullanılarak sosyal ağ analizi için bayes ağı kurulmuştur. Bayes ağı ile bireyler arasındaki etkileşim modellenmiştir. K2, bayes ağlarının eğitimi için kullanılan bir algoritmadır. Tezde K2 algoritması kullanılarak ağ yapısı oluşturulmuştur. K2 algoritmasının performansı ağı oluşturan düğümlerin sırasına bağlıdır. Doğru bir ağ yapısı için bir düğümünün ebeveyninin kendinden önceki sırada verilmiş olması gerekmektedir. K2 algoritması için düğümlerin sırasının belirlenmesi amacıyla bir yöntem önerilmiştir. Ayrıca sıralamayı bulmak için genetik algoritma kullanılmıştır. Düğümler rassal, genetik algoritma ve önerilen yöntem kullanılarak sıralanmıştır. Her bir sıralama K2 algoritmasına girdi olarak verilmiş ve her biri için otomatik olarak ağ yapıları oluşturulmuştur. Bulunan ağ yapılarının skorları karşılaştırılmıştır. Önerilen yöntemin bulduğu sıralama genetik algoritmanın başlangıç popülasyonuna eklenerek hibrit bir yöntem ile sonuçlar iyileştirilmiştir.

Bayes ağı kurulduktan sonra, bayes ağı üzerinden sosyal ağ analizi yapılmıştır. Ağ içindeki bir bireyin bir etkinliğe katılma olasılığı başka bireylerin aynı etkinliğe katılma durumlarına göre analiz edilmiş ve analiz sonuçları sunulmuştur.

Anahtar Kelimeler: Bayes Ağlar, Bayes Teorisi, Genetik Algoritma, Makine Öğrenmesi, Sosyal Ağlar

ABSTRACT

MS THESIS

THE USE OF BAYESIAN NETWORK FOR SOCIAL NETWORK ANALYSIS

Betül AKKOÇ

The Graduate School of Natural and Applied Science of Selçuk University
The Degree of Master of Science in Computer Engineering

Advisor: Prof. Dr. Ahmet ARSLAN

2012, 52 Pages

Jury
Prof. Dr. Ahmet ARSLAN
Asst. Prof. Dr. Ö. Kaan BAYKAN
Asst. Prof. Dr. Murat CEYLAN

Bayesian networks that also known as belief networks are used to provide information about the uncertainty. Social networks are structures that connected to each other according to people social relations. In this structure, people are expressed as the nodes, and relationships are expressed as connections.

In the thesis, bayesian network is established for social network analysis using data of person's participation in activities. Interaction of between people is modelled with bayesian network. K2 is an algorithm used for the training of Bayesian networks. In this thesis, the network structure has been created using K2 algorithm. The performance of K2 algorithm depends on the order of the nodes. The parent of node must be given earlier than node in ranking for a right network structure. A method is proposed in order to determine the order of the nodes for K2 algorithm. Also, genetic algorithm is used to find ranking. Nodes are ordered random, using genetic algorithm and proposed method. All orders are given as input to K2 algorithm, then network structures are automatically formed for each of them. Scores of founded network structures are compared. Results are improved with a hybrid method that is obtained by ranking of proposed method addition to initial population of genetic algorithm.

After bayesian network has been established, social network analysis is done through bayesian network. Possibility of participate in an activity of an individual within the network has been analyzed according to situations of the other individuals that are participated in same activity, and the results of analyze have been presented.

Keywords: Bayesian Networks, Bayes Theory, Genetic Algorithm, Machine Learning, Social Networks

ÖNSÖZ

Bu çalışmamda bana yol gösteren danışman hocam Prof. Dr. Ahmet ARSLAN'a ve Selçuk Üniversitesi Mühendislik-Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü'nün tüm öğretim elemanlarına teşekkür ederim.

Çalışmalarım sırasında sağlamış oldukları öğrenim bursu için TUBİTAK-BİDEB 'e teşekkürlerimi sunarım.

Maddi ve manevi yönden beni her zaman destekleyen, üzerimde büyük hakları olan aileme teşekkürü bir borç bilirim.

Betül AKKOÇ
KONYA-2012

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **v**

ÖNSÖZ ..... **vi**

İÇİNDEKİLER ..... **vii**

SİMGELER VE KISALTMALAR ..... **ix**

1. GİRİŞ **1**

1.1. Tezin Amacı ve Önemi 1
1.2. Literatüre Katkısı 2
1.3. Tezin Organizasyonu 2

2. KAYNAK ARAŞTIRMASI **4**

3. BAYES TEOREMİ VE BAYES AĞLARI **6**

3.1. Olasılık ve Belirsizlik 6
3.2. Koşullu Olasılık 6
3.3. Bayes Teoremi 7
3.4. Bayes Ağları 8
3.5. BA ile Karar Verme 9
3.6. BA'ların İnteraktif El ile Oluşturulması 11
3.7. BA'da Öğrenme 14
3.7.1. Parametre öğrenme 15
3.7.2. Ağ yapısını öğrenme 18

4. GENETİK ALGORİTMA **21**

4.1. Uygunluk Fonksiyonu 22
4.2. Çaprazlama işlemi 23
4.3. Mutasyon İşlemi 23

5. SOSYAL AĞ ANALİZİNDE BAYES AĞLARININ KULLANIMI **25**

5.1. Sosyal Ağ 25
5.2. Sosyal Ağ Analizi 25
5.3. Sosyal Ağ Analizinde Bayes Ağlarının Kullanımı 26
5.4. Uygulama Ortamının Geliştirilmesi 27
5.1.1. Düğüm Ekleme/Silme ve Yer Değiştirme 27
5.1.2. Bağlantı Oluşturma 28
5.1.3. Olasılık Tablosu Oluşturma 29
5.1.4. Eğitim Yapma 30
5.1.5. Çıkarsama Yapma 31

6. YAPILAN ÇALIŞMA VE ALINAN SONUÇLAR 33

6.1. Önerilen Yöntem 33
6.2. Örnek Çalışma 38
6.3. Gerçek Veri Kümesi Uygulaması 40
6.4. SA'da Çıkarsama Yapma 45

7. SONUÇLAR VE ÖNERİLER 48

7.1. Sonuçlar 48
7.2. Öneriler 49

KAYNAKLAR 50

ÖZGEÇMİŞ.....52

SİMGELER VE KISALTMALAR

Simgeler

V : Veri tabanı
π_i : x_i düğümünün ebevenylerinin listesi
ϕ_i : V veri tabanı içindeki x_i'nin ebeveynlerinin tüm mümkün olan örneklerinin listesi
q_i : $| \phi_i |$
V_i : x_i özelliğinin tüm olası değerlerinin listesi
r_i : $| V_i |$
α_{ijk} : ϕ_i içindeki örneklerden π_i içindeki x_i'nin ebeveynlerinin j ile ve x_i'nin özelliklerinin k değeri ile örneklendiği V veri tabanı içindeki durum sayısı
N_{ij} : ϕ_i içindeki örneklerden π_i içindeki x_i'nin ebeveynlerinin j ile örneklendiği V veri tabanı içindeki durum sayısı
$f(i,\pi_i)$: π_i'nin x_i'nin ebeveyni olma skoru
P_c : Çaprazlama olasılığı
P_m : Mutasyon olasılığı

Kısaltmalar

ADA : Açgözlü Denklik Arama
BA : Bayes Ağı
GA : Genetik Algoritma
SA : Sosyal Ağ
SAA : Sosyal Ağ Analizi

1. GİRİŞ

Bayes teoremi Bayes tarafından (1763) bulunan koşullu olasılıkların hesaplanmasını sağlayan bir teoremdir. Bayes Ağı (BA) ilk olarak Judea Pearl (1985) tarafından sunulmuştur. BA grafiksel bir yapı üzerinde rasgele değişkenleri düğümler ve ilişkileri eğriler şeklinde ifade eder, düğümler arasındaki nedensel ilişkileri tablolar halinde tutar. Belirsizlik durumlarında bu grafik yapı ve tablolardan faydalanılarak çıkarsama yapılabilir.

Sosyal ağ (SA) "düğüm" olarak adlandırılan arkadaşlık, akrabalık, ortak ilgi, finansal değişim ya da inanç, bilgi veya prestij ilişkileri gibi bir veya daha fazla bağımlılıkla belirli türde bağlı sosyal yapıdır (Anonymous, 2010).

Sosyal ağ analizi (SAA), 1960 ve 1970'li yıllarda ağırlıklı olarak sosyal psikoloji içinde sosyolog ve araştırmacılar tarafından geliştirilmiştir. Matematik, istatistik ve bilgisayar ile işbirliği içinde daha da geliştirilmesiyle ekonomi, pazarlama ve endüstri mühendisliği gibi diğer disiplinlerde cazip bir araç olarak analiz tekniklerinin hızlı gelişmesine yol açmıştır (Scott, 2000).

1.1. Tezin Amacı ve Önemi

Bu çalışmada, kişilerin etkinliklere katılım verileri kullanılarak BA oluşturmak istenmektedir. Daha sonra oluşturulan bu BA kullanılarak analiz yapılması amaçlanmaktadır.

Etkinliğe katılma verileri bir gruba katılma, aynı sayfaya yorum yazma vb. olabilir. Eğer kullanıcı etkinliğe katılmışsa "1" değilse "0" değerini almaktadır. Etkinliklere katılma verileri sayesinde kişilerin birbirleriyle ortak hareket ettikleri, ortak beğenilerinin olduğunu veya da ortak beğenilerinin, etkileşimlerinin olmadığını söyleyebiliriz. Bu veriler ile BA kullanılarak bireyler arasındaki etkileşim modellenmek istenmektedir.

BA uzmanların bilgileriyle interaktif el ile kurulabildiği gibi veriden eğitilerek de oluşturulabilir. K2, BA'nın eğitimi için kullanılan bir algoritmadır. Tezde K2 algoritması kullanılarak ağ yapısı oluşturulmuştur.

K2 algoritması için düğümlerin sırasını öğrenmek amacıyla bir yöntem önerilmiştir. Ayrıca sıralamayı bulmak için genetik algoritma (GA) kullanılmıştır. Düğümler rassal, genetik algoritma ve önerilen yöntem kullanılarak sıralanmıştır. Her

bir sıralama K2 algoritmasına girdi olarak verilmiş ve her biri için ağ yapısı oluşturulmuştur. Bulunan ağ yapılarının skorları karşılaştırılmıştır. Ayrıca önerilen yöntem ve GA'nın beraber kullanıldığı hibrit bir yapı oluşturulmuştur.

BA eğitimi tamamlandıktan sonra SAA'nın nasıl yapılacağı anlatılmıştır. Analiz sonucunda SA içindeki bireylerin etkinliğe katılma durumları göz önüne alınarak başka bir bireyin etkinliğe katılma olasılığı hesaplanabilmektedir.

1.2. Literatüre Katkısı

K2 algoritmasının performansı ağı oluşturan düğümlerin sırasına bağlıdır. Algoritmada bir düğümün başka bir düğümün ebeveyni olabilmesi için sıralamada o düğümden önce gelmesi gerekmektedir. Her bir düğüm için kendinden önceki sırada verilmiş olan düğümlerle skorları hesaplanır. Bu skorlara göre düğümün ebeveynleri tespit edilir.

Tezde K2 algoritması için düğümlerin sırasının belirlenmesi amacıyla bir yöntem önerilmiştir. Ayrıca sıralamayı bulmak için GA kullanılmıştır. Düğümler rassal, genetik algoritma ve önerilen yöntem kullanılarak sıralanmıştır. Her bir sıralama K2 algoritmasına girdi olarak verilmiş ve her biri için ağ yapısı oluşturulmuştur. Bulunan ağ yapılarının skorları karşılaştırılmıştır.

GA ile elde edilen sonuçların başlangıç popülasyonuna bağlı olduğu gözlemlenmiştir. GA başlangıç popülasyonunun oluşturulmasında rassal sıralamaya ilave olarak önerilen yöntemin bulduğu sıralamada popülasyona eklenerek hibrit bir yöntem ile ağ skoru iyileştirilmeye çalışılmış ve gözlemlenen sonuçlar sunulmuştur.

Önerilen bu hibrit yöntemle kişilerin etkinliğe katılma verileri kullanılarak BA oluşturulmuştur. Kurulan bu BA üzerinden SAA'nın nasıl yapılacağı anlatılmış, gerçek veriler üzerinden analiz yapılmış ve analiz sonuçları sunulmuştur.

1.3. Tezin Organizasyonu

Bu tez çalışması 7 bölümden oluşmaktadır.

Birinci bölümde yapılan çalışma tanıtılmış, amacı ve önemi anlatılmış ve literatüre katkısından söz edilmiştir.

İkinci bölümde tez çalışmasının konusu olan BA ve SA ile ilgili yapılan önceki çalışmalardan bahsedilmiştir.

Tezin üçüncü bölümünde olasılık kavramı ve bayes teoremi anlatılmış, BA'lar hakkında bilgiler verilmiştir. K2 algoritması da bu bölümde anlatılmıştır.

Dördüncü bölümde GA hakkında bilgiler verilmiş ve tez kapsamında nasıl kullanılacağı anlatılmıştır.

Beşinci bölümde SA ve SAA terimi açıklanmıştır. BA'da SAA'nın nasıl yapılacağı anlatılmıştır. Bu bölümde ayrıca BA'nın kurulması için geliştirilmiş uygulama ortamından bahsedilmiştir.

Altıncı bölümde K2 algoritmasında sıralamayı belirlemek için bir yöntem önerilmiştir. Bu yöntemin diğer yöntemlerle karşılaştırılması ve sonuçları bu bölümde yer almaktadır. Ayrıca bu bölümde oluşan ağ üzerinden analiz yapılmış ve sonuçları sunulmuştur.

Son bölümde sonuçlar özet olarak verilmiş ve ileride yapılacak çalışmalar için önerilere yer verilmiştir.

2. KAYNAK ARAŞTIRMASI

Bayes ağları bir çok alanda kullanılmaktadır. Akademik çalışmalarda, biyoloji, finans, nedensel öğrenme, bilgisayar oyunları, bilgisayarla görme, bilgisayar donanımları, bilgisayar yazılımları, veri madenciliği, tıp, doğal dil işleme, planlama, psikoloji, güvenilirlik analizi, zaman çizelgelemesi, konuşma, taşıt kontrolü ve arıza teşhisi, hava tahmini gibi birçok farklı alanda BA'lara başvurulmaktadır (Neapolitan, 2003).

BA'lar interaktif el ile kurulabildiği gibi veriden eğitilerek de oluşturulabilir. BA yapısını veriden öğrenmek zorlu bir süreçtir. Chickering (1996) en fazla 2 derecelik kıstlamada bile BA yapısının öğrenilmesinin NP-zor olduğunu göstermiştir.

Verma ve Pearl (1991) ağ yapısını bulmak için sınırlama tabanlı yaklaşım önermişlerdir.

Ağ yapısını bulmak için Cooper ve Herskovit (1992) K2 algoritmasını önermişlerdir. Bu algoritmada düğümlerin sıralı bir şekilde veri kümesi içinde yer aldığı düşünülür. Algoritmanın performansı düğümlerin sırasına bağlıdır.

Larranaga ve ark. (1996) K2 algoritması için düğümlerin doğru sıralamasını bulmak için GA kullanarak en uygun ağı oluşturmaya çalışmışlardır.

SAA, öncelikle bireyler ve gruplar arasındaki ilişkilere analitik teknikler uygulamaya ve bireyler ve gruplar hakkında ek bilgi elde etmek için bu ilişkileri nasıl kullanacağını araştırmaya odaklanmıştır (Degenne ve ark. , 1999).

SAA farklı alanlarda kullanılmıştır. Krebs (2002) 11 Eylül 2001 saldırısını analiz etmek için bir çalışma gerçekleştirmiştir. Saldırıya katılanlar ve bu kişilerle bağlantıda bulunan kişiler arasındaki bir ağ haritası oluşturmuştur.

Ehrlich ve Carboni (2005) çalışmalarında iyi bir çalışma imkanı sağlamak amacıyla işçiler arasındaki etkin ilişkileri tanımlamak için SAA kullanmışlardır.

Kooelle ve ark. (2006) SAA için BA'ları bir araç olarak kullanmışlardır. Çalışmada geleneksel SAA teknikleri kullanarak elde edilmesi mümkün olmayan kavramları elde etmek için BA SAA'ya uygulanmıştır. BA kullanarak üç tip analiz; belirsizlik ile SA algoritmaları çoğaltma, düğümler için ağ arama ve ağ içinde yeni bağlantılar çıkarsama tartışılmıştır.

Lu ve ark. (2006) müşterilerin çevrimiçi alışveriş davranışları ve müşteri gereksinimleri arasındaki ilişkileri analiz etmek için BA kullanmışlardır.

Goldenberg'in (2007) tez çalışmasında insanlar arasında gözlemlenen etkileşimler temelinde sosyal ağların çeşitli modelleri sunulmuş ve seyrek ilişkisel veri için yüksek boyutlu olasılıksal modellerin etkili öğrenme sorunları ele alınmıştır.

Doğan'ın (2010) tez çalışmasında bilişim teknolojileri öğretmenlerinin yaygın olarak kullandıkları açık, yarı yapılandırılmış, çevrimiçi bir forum SAA yöntemiyle analiz edilmiş, öğretmenlerin mesleki gelişimleri için çevrimiçi iletişim desenleri tanınmış, SA ilişkileri ortaya koyulmuş ve bu sosyal ağın yapısal özellikleri incelenmiştir.

Park ve ark. (2010) BA'ları kullanıcılar arasındaki anlamsal ilişkileri tanımlamak için kullanmış ve bu anlamsal ilişkilerle mobil SA kurmuşlardır.

Bursa ve Ünalır (2010) profil yönetimi gerçekleştirilirken SA'ların profiller üzerine etkileri ve genişletilebilirlik etkilerinden bahsetmişlerdir.

3. BAYES TEOREMİ VE BAYES AĞLARI

3.1. Olasılık ve Belirsizlik

Olasılık, belirsizliği ifade etmek için kullanılan bir kavramdır. Olasılık kavramı, bir olayın gerçekleşebilmesinin "0" ve "1" arasında değişen matematiksel değeridir. "1" olasılığı olayın kesin gerçekleşeceğini, "0" olasılığı ise kesin gerçekleşmeyeceğini gösterir.

Günlük yaşamın çoğu durumlarında belirsizlik hakimdir. Örneğin yarın havanın yağmurlu olup olmayacağından emin olunamaz. Belirsizlik her zaman vardır.

Gerçek hayatta belirsizlikle karşılaşılan durumlarda makul bir muhakeme yapılır. Bir olayın gerçekleşip gerçekleşmeyeceği inancı başka bir olayın ortaya çıkıp çıkmamasına göre ayarlanır. Olasılık kavramları belirsizlik durumlarında makul fikir yürütmek için mantığa yardım eder (Bolstad, 2004).

3.2. Koşullu Olasılık

Bir olayın gerçekleşeceği bilgisi, başka bir olayın meydana gelme olasılığını nasıl etkileyeceğini tahmin etmek için koşullu olasılığa bakmak gerekir.

Bir A olayı gerçekleştiğinde B olayının koşullu olasılık denklemi aşağıda verilmiştir.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{3.1}$$

Koşullu olasılık için bağımsız olaylar; A ve B'nin bağımsız olaylar olduğu düşünüldüğünde;

$$P(B|A) = P(B) \tag{3.2}$$

şeklinde ifade edilir.

Bağımsız olaylar için $P(B \cap A) = P(B) \times P(A)$ olacaktır ve $P(A)$ faktörü iptal edilecektir. A ve B bağımsız olaylar olduğunda A hakkındaki bilgi B olasılığını etkilemez.

Çarpım Kuralı: Biçimsel olarak A ve B olaylarının rollerini değiştirdiğimizde, B ye göre A'nın koşullu olasılığı bulunmak istendiğinde denklem aşağıdaki şekildedir;

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3.3}$$

$$P(A \cap B) = P(B) \; x \; P(A|B) \tag{3.4}$$

Benzer şekilde;

$$P(A \cap \overline{B}) = P(\overline{B}) \; x \; P(A|\overline{B}) \tag{3.5}$$

şeklinde ifade edilir.

3.3. Bayes Teoremi

Bayes teoremi bir din adamı olan Thomas Bayes tarafından bulunmuştur. Bayes'in kendisine ait olasılık kuramı "Essay towards solving a problem in the doctrine of chances" (1763) adlı makalede yayınlanmıştır.

Bayes teoremi koşullu olasılıkları hesaplayan basit bir matematik formülüdür.

$$\text{Sonraki Olasılık} = \frac{\text{Koşullu Olasılık x Önceki Olasılık}}{\text{Marjinal Olasılık}}$$

P(A): A olayının bağımsız olasılığı (Öncül (prior) olasılık, marjinal olasılık)

P(B): B olayının bağımsız olasılığı

P(A|B): B olayından sonra A olayının meydana gelme olasılığı

P(B|A): A olayından sonra B olayının meydana gelme olasılığı

Koşullu olasılık denklemi;

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{3.6}$$

A olayının marjinal olasılığı tüm ayrık parçalarının toplamıdır.

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \tag{3.7}$$

Koşullu olasılık içinde bu tanım;

$$P(B|A) = \frac{P(A \cap B)}{P(A \cap B) + P(A \cap \bar{B})} \tag{3.8}$$

Bu ortak olasılıkların her birini bulmak için çarpma kuralı kullanılır.

Tek olay için bayes teoremi;

$$P(B|A) = \frac{P(A|B) \; x \; P(B)}{P(A|B) \; x \; P(B) + P(A|\bar{B}) \; x \; P(\bar{B})} \tag{3.9}$$

İki olay ve daha fazlası için bayes teoremi;

$$P(Bi|A) = \frac{P(A|B_i) \; x \; P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \; x \; P(B_j)} \tag{3.10}$$

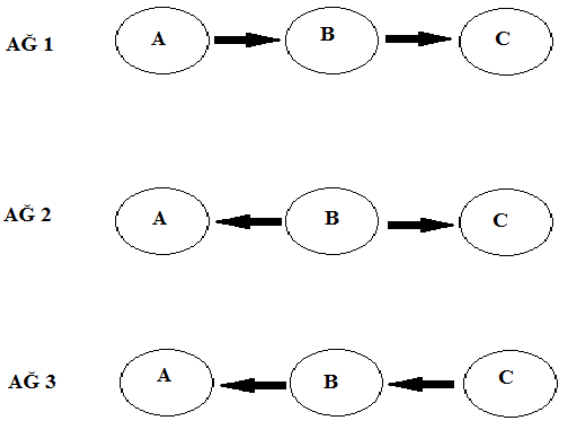
şeklinde ifade edilir.

3.4. Bayes Ağları

BA ilk olarak Judea Pearl (1985) tarafından "Bayesian Network: A model of self-activated memory for evidential reasoning" adlı makalesinde sunulmuştur. Son zamanlarda giderek popülerliği artmış ve birçok alanda kullanılır hale gelmiştir.

BA'lar ayrıca inanç ağları olarak da bilinir. Olasılıklı grafik modelleri ailesine aittir. Bu grafiksel yapılar belirsizlik etkisi hakkında bilgiyi sunmak için kullanılır. Her düğüm rasgele bir değişkeni temsil eder, düğümler arasındaki yollar ilgili değişkenler arasındaki olasılıksal bağımlılıkları ifade eder. Bu grafik içindeki koşullu bağımlılıklar genellikle bilinen istatistik ve hesaplama yöntemleri kullanılarak tahmin edilebilir. Bu nedenle BA'lar grafik teorisi, olasılık teorisi, bilgisayar bilimi ve istatistiği bir araya getirir (Ben-Gal, 2007).

Değişkenler düğümlerle ifade edilir ve her değişken için bir tablo bulunur. Değişkenler birbirleriyle etkileşimin yönüne göre oklarla (kenarlarla) birbirlerine bağlanır. Örneğin A düğümünden B düğümüne çizilen yol A'nın B'ye etkisini gösterir. Değişkenlerin aldıkları değerlere de durum adı verilir.



Şekil 3.1. BA'larda düğümlerin etkileşim grafiği

Şekil 3.1 de bulunan ağların koşullu olasılık dağılımları;

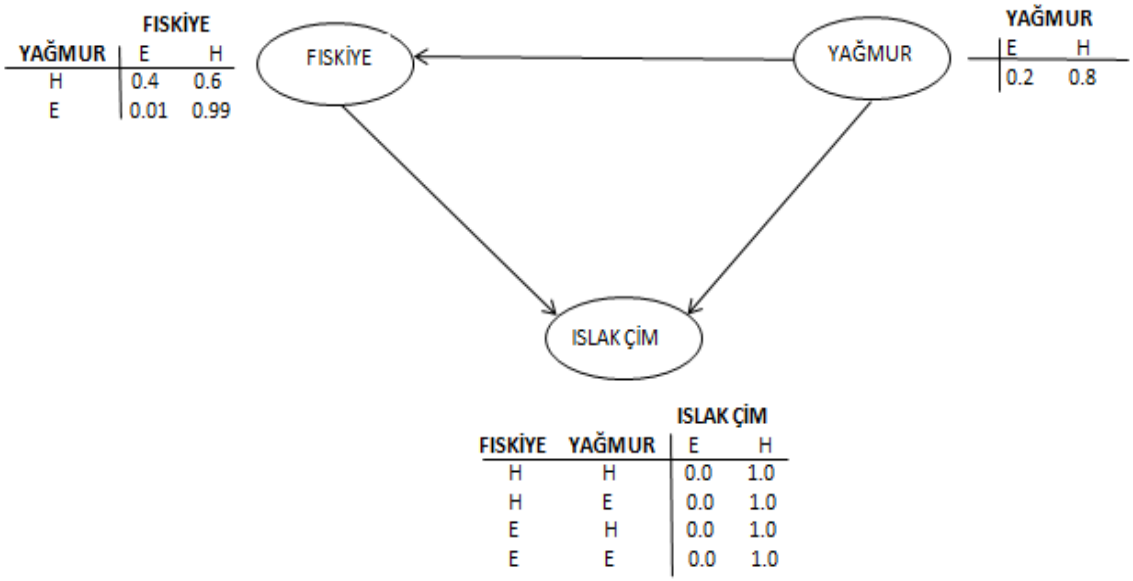
AĞ 1 $P(A,B,C) = P(A) \times P(B \mid A) \times P(C \mid B)$

AĞ 2 $P(A,B,C) = P(A \mid B) \times P(B) \times P(C \mid B)$

AĞ 3 $P(A,B,C) = P(A \mid B) \times P(B \mid C) \times P(C)$

3.5. BA ile Karar Verme

BA ile nasıl karar verildiğini anlayabilmek için literatürde sıkça karşılaşılan Pearl'in (1988) fiskiye problemi örnek olarak verilebilir. Örnekte çimlerin ıslak olmasını yağmurun yağması ve fiskiyenin açık olması etkilemektedir. Ayrıca fiskiye ile yağmur arasında da koşullu olasılık bulunmaktadır.



Şekil 3.2. BA fiskiye örneği

Örneğin BA'ya aktarılmış biçimi şekil 3.2 de görülmektedir. Düğümler yağmur, fiskiye, ıslak çim değişkenlerini tutmaktadır. Şekildeki tablolar içinde düğümlerin koşullu olasılıkları bulunmaktadır.

Bu örneğe göre çimler ıslak olduğunda yağmurun yağmış olma olasılığının bulunması aşağıda anlatılmıştır.

Ç: Islak Çim, F: Fıskiye, Y: Yağmur

$P(Ç,F,Y) = P(Ç|F,Y)x\ P(F|Y)x\ P(Y)$

$P(Y=E|Ç=E)$; Çimler ıslaksa yağmurun yağmış olma olasılığı

$P(Ç=E,Y=E)$; Çimlerin ıslak ve yağmurun yağmış olduğu durumlar

$P(Ç=E)$; Çimlerin ıslak olma olasılığı

$$\boldsymbol{P(Y=E|Ç=E) = \frac{P(Ç=E,Y=E)}{P(Ç=E)}}$$

$$\begin{aligned} P(Ç=E,Y=E) &= P(Ç=E,F=H,Y=E) + P(Ç=E,F=E,Y=E) \\ &= 0.8 * 0.99 * 0.2 + 0.99 * 0.01 * 0.2 \\ &= 0.1584 + 0.00198 \\ &= 0.16038 \end{aligned}$$

$$P(Ç=E)=P(Ç=E,F=H,Y=E) + P(Ç=E,F=E,Y=E) + P(Ç=E,F=H,Y=H) + P(Ç=E,F=E,Y=H)$$
$$=0.8 * 0.99 * 0.2 +0.99 * 0.01 * 0.2 + 0 * 0.6 * 0.8 + 0.9 * 0.4 * 0.8$$
$$= 0.1584 + 0.00198 + 0 + 0.288$$
$$= 0.44838$$

$$P(Y=E|Ç=E) = \frac{P(Ç=E,Y=E)}{P(Ç=E)} = \frac{0.16038}{0.44838} \approx 35.77\%$$

Bu sonuca göre çimler ıslaksa yağmurun yağmış olma olasılığı %35.77 dir.

3.6. BA'ların İnteraktif El ile Oluşturulması

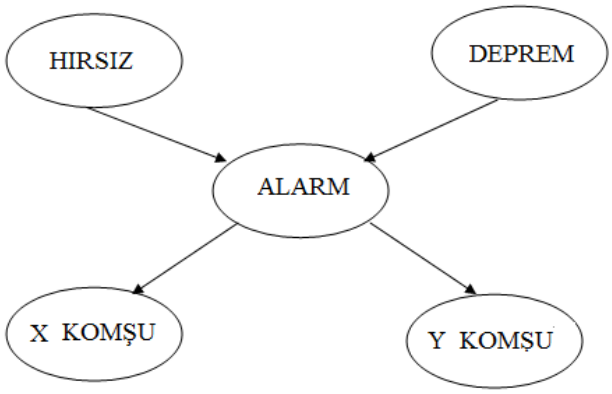
Koşullu olasılık teriminin ortak dağılım kuralı incelenecek olursa;

$$P(x_1, \ldots, x_n) = P(x_n|x_{n-1}, \ldots, x_1)\, P(x_{n-1}, \ldots, x_1)$$
$$P(x_1, \ldots, x_n) = P(x_n|x_{n-1}, \ldots, x_1)\, P(x_{n-1}|x_{n-2}, \ldots, x_1) \ldots P(x_2 | x_1)\, P(x_1)$$
$$= \prod_{i=1}^{n} P(x_i| \, ebeveyn(X_i))$$

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i| \, ebeveyn(X_i)) \qquad (3.11)$$

Denkleme göre BA'nın etki alanını doğru temsil etmesi için bir düğümün ebeveyninin kendinden önceki sırada verilmiş olması gerekmektedir. Doğru yapıya sahip bir BA oluşturulabilmesi için düğüm sırasının doğru olması lazımdır. Bundan dolayı her düğüm için ebeveynler seçmemiz gerekmektedir. x_i 'nin ebeveyni kendinden önceki sırada verilmiş düğümler x_i , . . . , x_{i-1} arasından kendisiyle direk etkisi olan düğümlerdir (Russell ve Norvig, 1995).

Ağ oluşumunda sıralamanın önemini anlayabilmek için literatürde sık rastlanılan Pearl'in (1988) BA alarm problemi örnek olarak verilebilir.



Şekil 3.3. BA alarm örneği

Örnekte deprem olması veya eve hırsız girmesi alarmı etkilemektedir. X ve Y ev sahibinin komşularıdır ve alarm durumunda ev sahibini aramaktadırlar.

Şekil 3.3 deki alarm örneğine bakacak olursak X komşu düğümü, Hırsız ve Deprem düğümlerinden etkilenir fakat bu düğümler X komşu düğümü ile doğrudan etkili değillerdir. Bunlar alarm düğümünün etkisiyle X komşu düğümünü etkilemektedir. Öte yandan X komşu ve Y komşu düğümleri arasında herhangi bir ilişki olup olmadığını da alarm düğümünün durumu belirler. Buna göre;

H: Hırsız, D: Deprem, A: Alarm, X: X Komşu, Y: Y Komşu

$$P(X \mid Y, A, D, H) = P(X \mid A)$$

Aynı işlemi tüm değişkenler için uyguladığımız zaman;

$$P(H, D, A, X, Y) = P(H)\, P(D)\, P(A|H,D)\, P(X|A)\, P(Y|A)$$

Pearl'in ağ kurulum algoritması;

1.Uygulama alanında tanımlanmış değişkenler kümesi $\{X_i\}$ seçilir.

2.Değişkenler için sıra belirlenir, $<X_1, \ldots, X_n>$

3.Solda değişkenler varken;

(a) Sonraki değişken X_i ağa eklenir.

(b)Ağ içine önceden eklenmiş aşağıdaki koşullu bağımsızlık özelliğini sağlayan düğümlerin alt kümesi Ebeveyn(X_i)'den X_i 'ye eğri çizilir.

$$P(X_i| X_1, \ldots, X_m) = P(X_i| Ebeveyn(X_i))$$

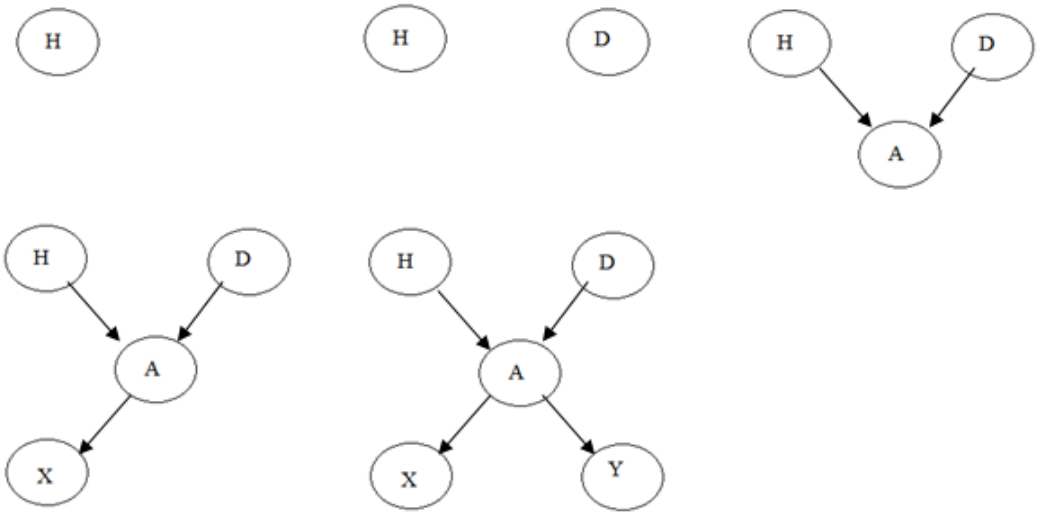
$X_1, \ldots, X_m$ değişkenlerinin hepsi X_i den önce eklenmiştir.

(c) X_i için parametreler tanımlanır.

Sıralama ağ yapısının oluşumunda etkilidir. Farklı sıralamalara göre oluşan ağ yapılarını inceleyecek olursak;

Sıra1: H, D, A, X, Y ye göre oluşmuş yapı şekil 3.4 de görülmektedir.

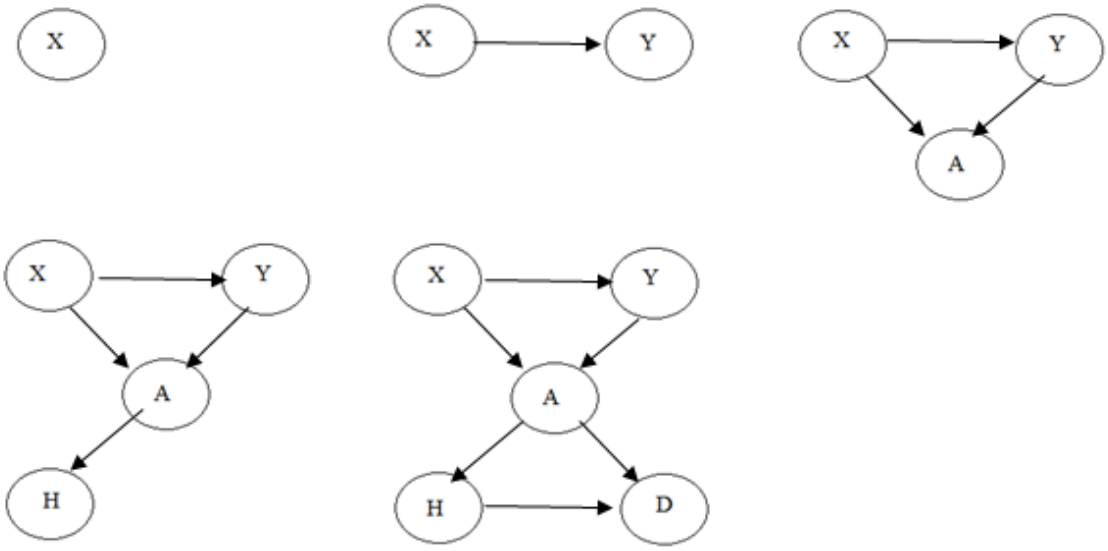
Ebeveyn(H)={}, Ebeveyn(D)={}, Ebeveyn(A)={H,D}, Ebeveyn(X)={A}, Ebeveyn(Y)={A}



Şekil 3.4. H, D, A, X, Y sırasına göre oluşturulmuş BA

Sıra 2: X, Y, A, H, D ye göre oluşmuş yapı şekil 3.5 de görülmektedir.

Ebeveyn(X)={}, Ebeveyn (Y)={X}, Ebeveyn(A)={X,Y}, Ebeveyn(H)={A}, Ebeveyn (D)={A,H}



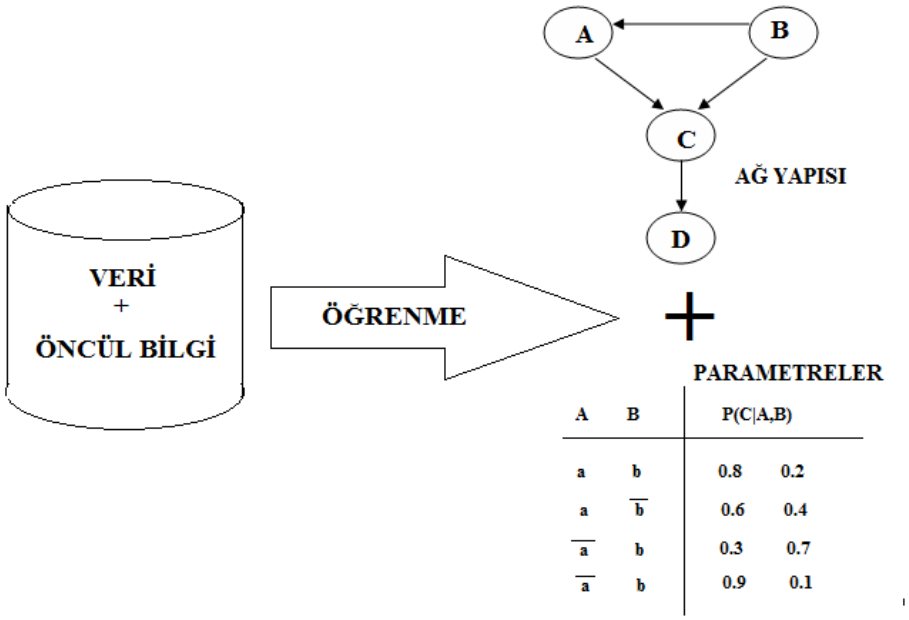
Şekil 3.5. X, Y, A, H, D sırasına göre oluşturulmuş BA

Pearl e göre X, Y, A, H, D sıralaması kötü bir sıralamadır çünkü X ve Y düğümleri, A düğümünden etkilenmesine rağmen ondan önce gelmişlerdir. H, D, A, X, Y sırası ise iyi bir sıralamadır çünkü bu sıralamada nedensel ilişkilere bağlı kalınmıştır.

3.7. BA'da Öğrenme

BA uzmanların bilgileriyle interaktif el ile kurulabildiği gibi veriden eğitilerek de oluşturulabilir. İnteraktif el ile ağları oluşturmanın bazı dezavantajları bulunmaktadır. Bilgi edinmek pahalı bir süreçtir. Ayrıca uzman bulmakta da sıkıntılar çekilebilir. Buna karşın öğrenme için sadece veri kullanılır. Ham veriler kullanılarak model oluşturulabilir. Gerçek dünyadaki birçok dağılım modellenebilir.

BA öğrenme sürecinde ağ yapısı ve parametreler öğrenilir (Şekil 3.6).



Şekil 3.6. BA'da öğrenme

3.7.1. Parametre öğrenme

Parametre öğrenme işlemi düğümlerin nedensellik ilişkilerine göre olasılık tablolarının oluşturulmasıdır. BA da parametre öğrenme için maksimum olabilirlik tahmini ve bayes tahmini yöntemleri kullanılır.

3.7.1.1. Maksimum olabilirlik tahmini

Maksimum olabilirlik tahmininde θ tahmin edici önerilir, verilen L (θ:V) verisinin olabilirliğini maksimize eden θ değeridir.

Binomsal olabilirlik için maksimum olabilirlik tahmini;

Binomsal olabilirlik iki olasılık değerine sahip durumlara denilir. Parayı havaya attığımızda yazımı, turamı geleceği olasılığı buna örnektir.

N_0= *X in 0 gelme olasılığı*

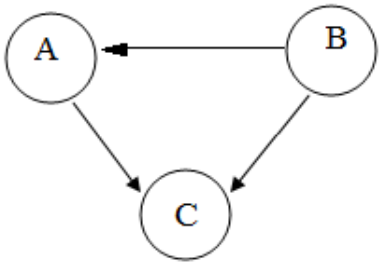
N_1=*X in 1 gelme olasılığı*

$$\hat{\theta} = \frac{N_0}{N_0 + N_1} \tag{3.12}$$

Maksimum olabilirlik tahmininde sorun küçük örneklere sahip olunduğunda ortaya çıkar. Örneğin parayı sadece iki kere havaya atarak elde edilen iki örneğe bakılarak tahmin doğru sonuçlar vermeyecektir.

Çok terimli durumlar için maksimum olabilirlik tahmini;

$$\hat{\theta} x_i \mid Ebeveyn_i = \frac{N(x_i,\ Ebeveyn_i)}{N(\ Ebeveyn_i)} \tag{3.13}$$



A	B	C
0	1	0
1	1	0
1	0	1
0	1	0
1	0	0
1	0	1
0	0	1
1	1	1
1	0	0
1	0	1

Şekil 3.7. BA parametre tahmin örneği

ÖRNEK;

Şekil 3.7 de gösterilen ağ yapısı ve verilen verilere göre olasılıkların maksimum olabilirlik tahmini ile hesaplanması;

$P(A=0) = 3/10$

$P(B=0) = 6/10$

$P(A=0|B=1) = 2/4$

$P(C=0|A=1,B=0) = 2/5$

Diğer parametrelerde aynı şekilde hesaplanabilir.

3.7.1.2. Bayes yaklaşımı

Bayes yaklaşımında tahmin için öncül değer kullanılır.

Binomsal olabilirlik için bayes yaklaşımı;

Laplace düzeltme ile maksimum bir değer değil ortalama bir değer bulunur.

$$\theta^* = \frac{N_0 + 1}{N_0 + N_1 + 2} \tag{3.14}$$

V örnek veri kümesini, m durum sayısını, α öncül değeri, Y yazı gelme durumunu, T tura gelme durumunu temsil ederken bir parayı havaya attığımızda yazı gelme olasılığı için bayes ile parametre dağılımında beta tahmini;

$$P(V_{m+1} = Y|V) = \frac{m_y + \alpha_y}{m + \alpha} \tag{3.15}$$

Örnek kümemizde 10 örnek bulunsun. Bunlardan ikisi yazı sekizi ise tura olsun;

Maksimum olabilirlik tahmini ile yazı gelme olasılığı 0.2 dir, bu orana bakılarak paranın hileli olduğunu söylemek doğru değildir. Elimizde 100,000 örnek olduğunda bunun 20,000 yazı 80,000 tura olduğu durumda maksimum olabilirlik tahmini ile yazı gelme olasılığı yine 0.2 dir. Fakat bu durumda paranın hileli olduğu yargısına ulaşabiliriz.

Bayes yaklaşımı için aynı örnekte ilk durumu göze alalım, $\alpha_y = \alpha_t = 100$ olduğunu varsayalım.

$$P(V_{m+1} = Y|V) = \frac{2+100}{10+100+100} \approx 0.5$$

İkinci durum için;

$$P(V_{m+1} = Y|V) = \frac{20,000+100}{100,000+100+100} \approx 0.2$$

Görüldüğü gibi bayes yaklaşımında öncül değerler kullanıldığı için bir yargıya ulaşabilmemiz için elimizde yeterli örnek olması gerekmektedir.

Çok terimli durumlar için bayes yaklaşımı;

$$\tilde{\theta} x_i \mid Ebeveyn_i = \frac{\alpha(x_i, Ebeveyn_i) + N(x_i, Ebeveyn_i)}{\alpha(Ebeveyn_i) + N(Ebeveyn_i)} \quad (3.16)$$

3.7.2. Ağ yapısını öğrenme

BA yapıları düğümlerin alan değişkenlerini ve düğümler arasındaki eğrilerin olasılık bağımlılıklarını gösteren yönlendirilmiş, çevrimsiz grafiklerdir (Pearl, 1988). BA yapısını veriden öğrenmek zorlu bir süreçtir. Chickering (1996) en fazla 2 derecelik kısıtlamada bile BA yapısının öğrenilmesinin NP-zor olduğunu göstermiştir.

Bayes ağlarının yapısını öğrenmek için temelde 2 yaklaşım vardır:

1. *Sınırlama Tabanlı Yaklaşım*: Bu yaklaşım Verma ve Pearl (1991) tarafından önerilmiştir, öğrenme değişkenler arasındaki koşullu bağımsızlık ilişkileri tespitiyle yapılır.
2. *Puanlama ve Arama Tabanlı Yaklaşım*: Bu süreçte her aday yapı veri kümesini ifade edebilme ölçüsüne göre puan alır. Veriler için en uygun BA yapısı aranır.

Puan tabanlı algoritmalar puanı en üst düzeye çıkarmak için çalışır. V veri kümesi ve A ağ yapısı varsayılarak skor tahmin edilmesi;

$$Skor\ (A,V)=P(A|V) \quad (3.17)$$

Bayes kanunları kullanarak skorun ifade edilişi;

$$Skor\ (A,V)=P(A|V)= \frac{P(V|A)P(V)}{P(V)} \quad (3.18)$$

Skor hesaplama için Cooper ve Herskovit (1992) tarafından K2 algoritması geliştirilmiştir. V veri kümesini ve B_a BA için ağ yapısını ifade ettiğinde, V veri kümesinin B_a yapısıyla skoru;

$$P(B_a \mid V) = P(B_a) \prod_{i=1}^{n} f(i,\pi_i) \tag{3.19}$$

$$f(i,\pi_i) = \prod_{j=1}^{qi} \frac{(r_i - 1)\,!}{(N_{ij} + r_i - 1)\,!} \prod_{k=1}^{ri} \alpha_{ijk}\,! \tag{3.20}$$

$P(B_a)$ bütün yapılar için eşit bir değere sahip olduğu varsayılır. Bu durumda en yüksek skora sahip yapı $P(B_a \mid V)$'yi maksimize eden yapıdır.

$$mak\,[P(B_a \mid V)] = c\ mak\,[\prod_{i=1}^{n} f(i,\pi_i)] \tag{3.21}$$

Ağın skorunu yükseltmek için K2 algoritması her düğüm için $f(i,\pi_i)$ skorunu maksimize etmeye çalışır.

3.7.2.1. K2 algoritması

Düğümlerin sıralandığı varsayılır ve her düğüm için maksimum alabileceği ebeveyn sayısı belirlenir. İlk düğümün ebeveyni yoktur. Düğümler ağa eklendiği anda ebeveyn listeleri boştur. Düğüm ağa eklendikten sonra kendinden önce gelen düğümlerin ebeveyn olarak kabul edildiğindeki skora bakılır ve skora göre ebeveyn ilişkisi kabul edilip edilmeyeceğine karar verilir. Her düğüm için ebeveyn ekleme skorda artış oldukça ve maksimum ebeveyn sayısına ulaşıncaya kadar devam eder.

Cooper ve Herskovit (1992) tarafından geliştirilen K2 algoritması;

1) **procedure** K2:
   {Giriş: n düğümlerin sıralanmış dizisi, u bir düğüm sahip olabileceği maksimum ebeveyn sayısı ve m durum içeren V veri tabanı}
   {Çıkış: Her düğüm için ebeveynlerinin çıktısı}
2) **for** $i := 1$ **to** n **do**
3) $\pi_i := \emptyset$;
4) $P_{eski} := f(i, \pi_i)$; { bu fonksiyon denklem 3.20 kullanılarak hesaplanır. }
5) OKToProceed := true;
6) **while** OKToProceed and $|\pi_i| < u$ do

{*f(i, π_i* U *{z})* ifadesini maksimize eden x_i'nin ebeveynleri dışında x_i düğümünden önce gelmiş düğüm *z* olsun}

7) $P_{yeni} := f(i, \pi_i \cup \{z\})$;

8) **if** $P_{yeni} > P_{eski}$ **then**

9) $P_{eski} := P_{yeni}$;

10) $\pi_i := \pi_i \cup \{z\}$

11) **else** OKToProceed := false;

12) **end** {while} ;

13) **write** ('Düğüm:', x_i ,' Düğümün ebeveynleri :' , π_i)

14) **end**{for};

15) **end**{K2}

Ayrıca algoritmayı hızlandırmak için $f(i,\pi_i)$ skoru yerine $log(f(i,\pi_i))$ skoru hesaplanır.

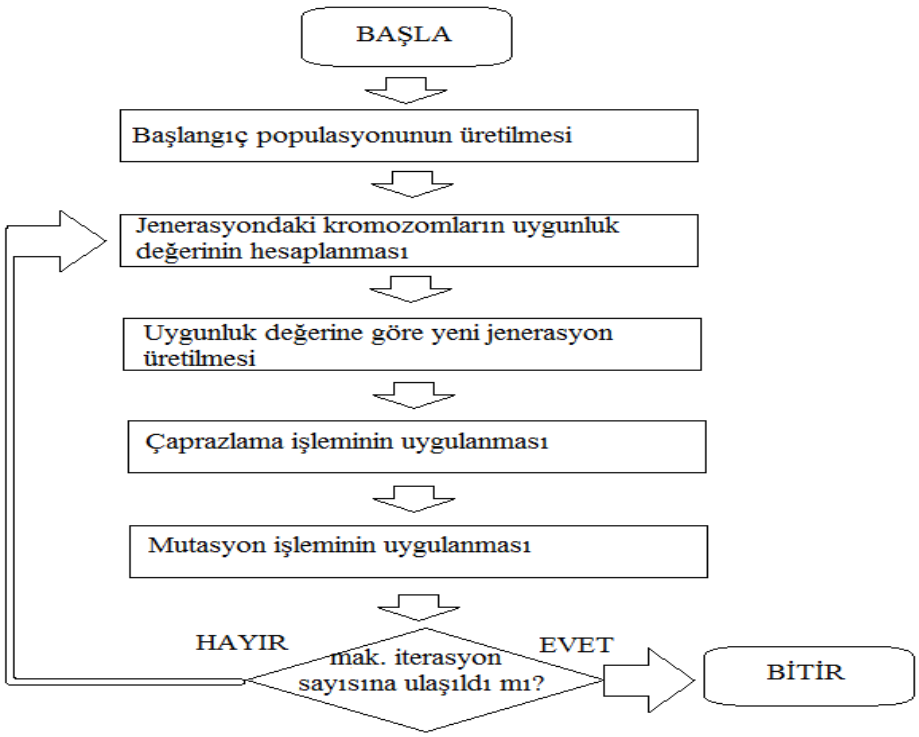
4. GENETİK ALGORİTMA

GA konusunda ilk çalışma Michigan Üniversitesi'nde John Holland tarafından 1975 yılında yapılmıştır (Holland, 1975). GA rastlantısal arama tekniklerini kullanarak çözüm bulmaya çalışan ve değişken kodlama esasına dayanan sezgisel bir arama tekniğidir (Goldberg, 1989). İteratif bir yöntemdir. Doğadaki evrimsel sürece benzer çalışmaktadır. GA en iyinin hayatta kalması ilkesine dayanır.

GA içinde parametreler genler olarak ifade edilmektedir. Genler bir araya gelerek kromozomları oluşturur. Kromozomlar problemin mümkün çözümlerini gösterir. Kromozomlar bir araya gelerek popülasyonu oluşturur. Popülasyondaki kromozom sayısı popülasyonun boyutunu ifade eder.

GA'da ilk olarak başlangıç popülasyonu üretilir. Başlangıç popülasyonu genellikle rassal üretilir. Daha sonra popülasyondaki kromozomların uygunluk değerleri uygunluk fonksiyonu kullanılarak hesaplanır. Bu uygunluk değerlerine göre yeni popülasyon üretilir. Ardından çaprazlama ve mutasyon işlemleri uygulanır. Bu işlemlerden sonra maksimum iterasyon sayısına ulaşılıp ulaşılmadığı kontrol edilir. Eğer ulaşıldıysa son popülasyondaki en iyi çözümler alınarak işlem tamamlanır, değilse yeniden uygunluk değerleri hesaplanarak aynı işlemler tekrarlanır.

GA'nın adımları aşağıdaki şekilde gösterilmiştir.



Şekil 4.1. GA adımları

Larranaga ve ark. (1996) yaptıkları çalışmada K2 algoritması için uygun sıralamayı bulmak amacıyla GA kullanmışlardır. Bu tez çalışmasında da önerilen yöntemin doğruluğunu karşılaştırmak ve önerilen yöntem ile hibrit bir yapı oluşturmak amacıyla K2 algoritmasında sıralamayı bulmak için GA kullanılmıştır.

GA ile sıralamayı bulmak için yapılan çalışmada genler düğümleri teşkil etmektedir. Kromozomlar düğümlerin sıralanması şeklinde kodlanmıştır. Bu şekilde sıralamalara permütasyon sıralama denir. GA düğümlerin sıralamasını bulmak için kullanılmıştır. Algoritma rassal bir başlangıç popülasyonu oluşturularak başlatılır. Bu oluşumdan sonra evrim başlar.

4.1. Uygunluk Fonksiyonu

GA da seçim işleminin yapılması için uygunluk fonksiyonu kullanılır. Bizim çalışmamızda amaç K2 algoritmasıyla BA oluşturabilmek için düğümlerin en uygun sıralamasını bulabilmektir. K2 algoritmasında skoru maksimize eden sıralama en uygun sıralamadır. GA içinde bu skor amaç fonksiyonu olarak kullanılmıştır.

$$uygunluk = log(\, P(B_a \mid V)) = \log(\, c \prod_{i=1}^{n} f(i,\pi_i)) \tag{4.1}$$

Amaç fonksiyonu içindeki c parametresi bütün yapılar için sabit olacağından göz ardı edilir;

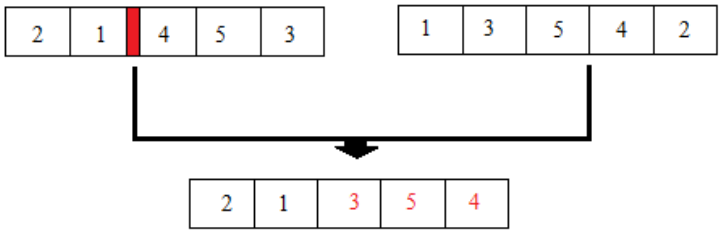
$$uygunluk = log(\, \prod_{i=1}^{n} f(i,\pi_i)\,) \tag{4.2}$$

GA da yeni popülasyonu oluşturacak bireyleri seçmek için farklı seçim yöntemleri vardır. Bu çalışmada elitist ve sıralı seçim kullanılmıştır. Elitist seçim yönteminde popülasyon içinde en iyi uygunluk değerine sahip kromozom korunur diğer kromozomlar uyum orantılı seçim yöntemlerinden biriyle seçilebilir. Uyum orantılı seçim yöntemi olarak sıralı seçim kullanılmıştır. Sıralı seçimde kromozomlar uygunluk değerlerine göre sıralanırlar. En az uygunluğu olan "1" en iyi uygunluğa sahip kromozom "n" değerini alacak şekilde sıralanır. Her kromozom sırası kadar seçim listesine eklenir.

4.2. Çaprazlama işlemi

GA içindeki kromozomlardan yeni kromozomlar oluşturmak için çaprazlama işlemi kullanılır. Bu işlem için 2 kromozom seçilir ve bu kromozomlardan seçilen çaprazlama yöntemine göre yeni kromozom oluşturulur.

Uygulamada tek noktalı çaprazlama yöntemi kullanılmıştır. Tek noktalı çaprazlama için ilk rassal bir x değeri seçilir. Bu değere göre ata kromozomun x noktasından önceki kısmı çocuk kromozoma aktarılır. Geriye kalan sıralama diğer ata kromozomun sıralamasına göre tamamlanır. Çaprazlamanın ne kadar sıklıkla yapılacağını belirtmek için çaprazlama olasılığı (P_c) kullanılır. Bu oran çok küçük olursa daha az birey yeniden üretilecektir. Çok yüksek olursa daha iyi bireyler üretmeden uygunluğu yüksek bireyler hızlı bir şekilde bozulur.



Şekil 4.2. Tek noktalı çaprazlama

4.3. Mutasyon İşlemi

Mutasyon, popülasyondaki bir kromozomun rassal seçilen geninin değerinin değiştirilmesidir. Yeni nesil üretimi devam ettikçe kromozomlar belli bir süre sonra tekrar etmeye başlayacaktır. Bunu engellemek ve çeşitliliği artırmak için mutasyon işlemi kullanılır. Mutasyon işleminin yapılacağı sıklığı mutasyon olasılığı (P_m) belirler. Bu olasılık çok büyük olursa popülasyon tamamen rassal bir hal alır. Çok küçük olursa arama uzayının tamamını araştırmaya engel olacaktır.

Uygulamada mutasyon işleminde kromozom üzerinde iki adet farklı rassal değer seçilir. Bu değerler birbirlerinin yerleriyle değiştirilir.



Şekil 4.3. Mutasyon İşlemi

Çaprazlama ve mutasyon işlemleri bittikten sonra maksimum iterasyon sayısına ulaşıldı mı kontrol edilir. Eğer ulaşıldıysa algoritma sonlandırılır, ulaşılmadıysa popülasyondaki kromozomların uygunluk değerleri hesaplanarak yukarıda anlatılan adımlar tekrar edilir.

5. SOSYAL AĞ ANALİZİNDE BAYES AĞLARININ KULLANIMI

5.1. Sosyal Ağ

Sözlük tanımıyla toplumsal veya sosyal ağ, bir veya daha fazla toplumsal ilişkiyle birbirine bağlanmış, dolayısıyla toplumsal bir bağ oluşturan bireyler (daha nadir durumlarda ortaklıklar ve roller) anlamına gelmektedir (Marshall, 1999) .

SA "düğüm" olarak adlandırılan arkadaşlık, akrabalık, ortak ilgi, finansal değişim ya da inanç, bilgi veya prestij ilişkileri gibi bir veya daha fazla bağımlılıkla belirli türde bağlı sosyal yapıdır (Anonymous, 2010).

SA çalışması ilk kez bireyler arasında çok merkezli mikro düzeyde bağlantıları öğrenmenin bir yolu olarak savaş sonrası sosyoloji ve antropoloji içinde ortaya çıkmıştır. O zamandan bu yana sosyal yapıların makro düzeyde niteliklerini karakterize eden bir araç olarak ağ fikrinin kapsam ve önemi genişlemiş, ve mikro ve makro düzeyleri bağlama anlamına gelmiştir (Turner, 2007).

Sosyalleşme internetin yaygın olarak kullanılması ile kullanım imkanı bulan ve Web 2.0 paradigmasının geliştirilmesi sonucunda uygulamaya geçirilmiş bir kavramdır. Web sayfaları içerisinde kendine ait bir uzay oluşturan ziyaretçiler, kendi aralarında ilişkiler tanımlayarak ve iletişim kurarak sosyalleşme kavramını gerçekleştirirler. SA'lar, bir topluluk içerisindeki ilişkilerin ziyaretçiler tarafından oluşturulması sonucunda oluşan ağlardır (Bursa ve Ünalır, 2010).

5.2. Sosyal Ağ Analizi

Sosyal ağ analizi (SAA), 1960 ve 1970'li yıllarda ağırlıklı olarak sosyal psikoloji içinde sosyolog ve araştırmacılar tarafından geliştirilmiştir. Matematik, istatistik ve bilgisayar ile işbirliği içinde daha da geliştirilmesiyle ekonomi, pazarlama ve endüstri mühendisliği gibi diğer disiplinlerde cazip bir araç olarak analiz tekniklerinin hızlı gelişmesine yol açmıştır (Scott, 2000).

SAA çağdaş sosyolojinin bir tekniği olarak ortaya çıkmıştır. SAA, düğümler ve bağlardan (kenarlar, bağlantılar veya ilişkiler olarak da adlandırılır) oluşan ağ teorisi terimi içinde sosyal ilişkileri incelemektedir. Düğümler ağ içindeki bireysel aktörlerdir ve bağlar aktörler arasındaki ilişkilerdir. Ortaya çıkan grafik tabanlı yapılar çok karmaşıktır. Düğümler arasındaki ilişkiler birçok türde olabilir. Akademik alanda

yapılan araştırmalar göstermiştir ki sosyal ağlar ailelerden milletlere kadar pek çok düzeyde kullanılmakta ve problemlerin çözümünün belirlenmesinde, kuruluşların çalışmasında ve bireylerin hedeflerine ulaşmada başarılı olma derecelerinde kritik rol oynamaktadır. En basit haliyle bir sosyal ağ çalışılan düğümler arasında arkadaşlık gibi belirli bağların bir haritasıdır. Bireyin bağlı olduğu düğümler o bireyin sosyal temasıdır (Anonymous, 2010).

5.3. Sosyal Ağ Analizinde Bayes Ağlarının Kullanımı

Tez çalışmasında SAA için kişilerin ortaklaşa katıldıkları etkinlik verileri kullanılmıştır. Bu temsilde eğer kişi bir etkinliğe katılmışsa "1" değerini katılmamışsa "0" değerini almaktadır.

Çizelge 5.1 de örnek olarak bireyler ve etkinliklere katılma verileri verilmiştir. Tablodan da anlaşılacağı üzere birey1 ve birey3 karşılaştırıldığında genel olarak ortak bir katılımın mevcut olduğunu ve bir etkileşimin olduğu söylenebilir. Buna karşın birey1 ve birey2 birbirinin katıldıkları etkinliklere katılmamış, bu durumda aralarında zıt bir etkileşimin olduğu söylenebilir. Çalışmada BA yardımıyla SAA yapılmıştır. BA yapısı bireylerin etkinliklere katılma verilerinden yararlanılarak öğrenilmiştir. Bu veriler ve K2 algoritması kullanılarak bireylerin aralarındaki etkileşim ağını bulmak ve bu ağdan yararlanılarak çıkarsamalar yapabilmek hedeflenmiştir.

Çizelge 5.1. Sosyal ağ içindeki bireylerin etkinliğe katılım verileri

	Birey1	Birey2	Birey3
Etkinlik1	**1**	**0**	**1**
Etkinlik2	**0**	**1**	**0**
Etkinlik3	**1**	**0**	**0**
Etkinlik4	**0**	**1**	**0**
Etkinlik5	**1**	**0**	**1**

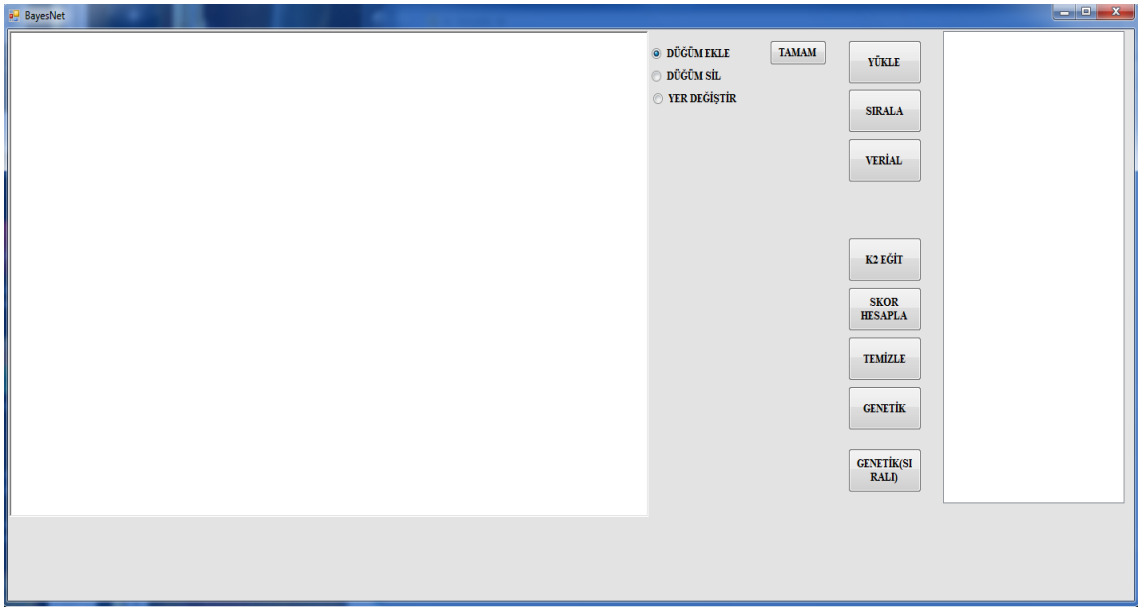
Etkinliklere katılım verilerini barındıran ağlar sayesinde iki kişinin birbirleriyle ortak hareket ettikleri, ortak beğenilerinin olduğu veya da tam tersi ortak beğenilerinin, etkileşimlerinin olmadığı söylenebilir. Bu türden ağlar özellikle gizli bağlantıların (ör: terör örgütü arasındaki haberleşme bağı vb.) belirlenmesine yardımcı olur. Ayrıca kişilerin etkileşim ağlarından elde edilecek bilgiler şirketlerde beraber çalışma verimini artırmak, öğrenciler arasında verimli çalışma ortamı oluşturmak gibi birçok amaç için kullanılabilir.

5.4. Uygulama Ortamının Geliştirilmesi

BA oluşturulması için geliştirilmiş çeşitli uygulama ortamları vardır. Netica, Hugin, Microsoft Belief Network bunlardan bazılarıdır. Bu bölümde tez kapsamında BA'lar için geliştirilmiş uygulama ortamı anlatılmıştır.

BA için geliştirilen uygulama ortamı ile BA'da interaktif el ile düğüm ekleme/silme, bağlantı ekleme, olasılık tablosu oluşturma işlemleri gerçekleştirilebilir. Bu işlemlere ek olarak eğitim yaparak da ağ kurulup, bu ağ ve veriden olasılık tabloları oluşturulabilir. Oluşan ağ yapısı ve olasılık tabloları kullanılarak çıkarsama işlemi yapılabilir.

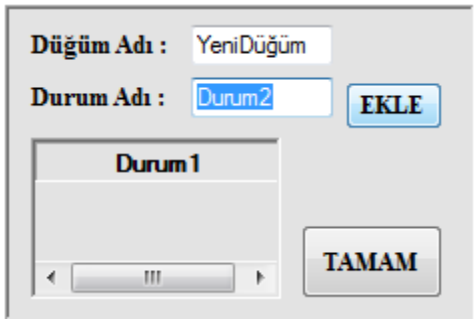
Uygulama başlatılmasıyla kullanıcıların karşısına BA'ları oluşturabilecekleri arayüz gelir (Şekil 5.1).



Şekil 5.1. Geliştirilen programın kullanıcı arayüzü

5.1.1. Düğüm Ekleme/Silme ve Yer Değiştirme

Düğüm eklemek için "Düğüm Ekle" seçeneği seçiliyken düğümün oluşturulması istenen bölgeye fareyle tıklanır. Kullanıcının düğüm ile ilgili bilgileri (düğüm adı ve durumları) girebileceği şekil 5.2 de gösterilen pencere açılır. Bu pencerede istenen değerler girilerek "Tamam" seçeneğinin seçilmesiyle düğümler oluşturulmuş olur.

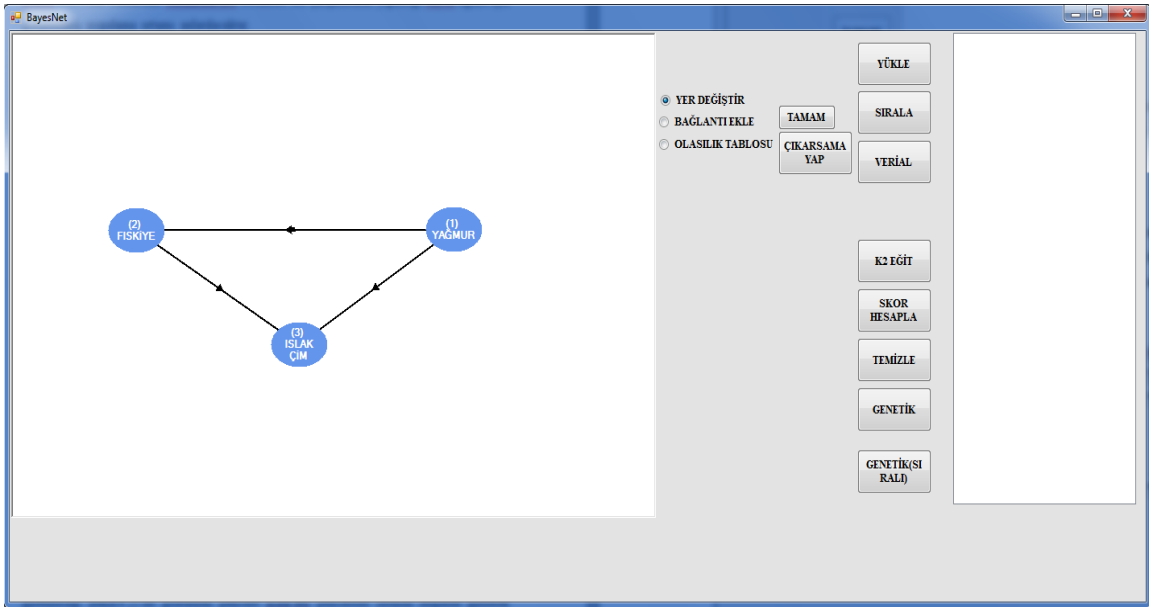


Şekil 5.2. Düğüm Oluşturma Penceresi

Oluşturulan düğümleri silmek için "Düğüm Sil" seçeneği seçiliyken düğüme tıklandığında düğümler silinir. "Yer Değiştir" seçeneği seçiliyken düğüm üzerine tıklanarak düğüm istenilen yere taşınabilir.

5.1.2. Bağlantı Oluşturma

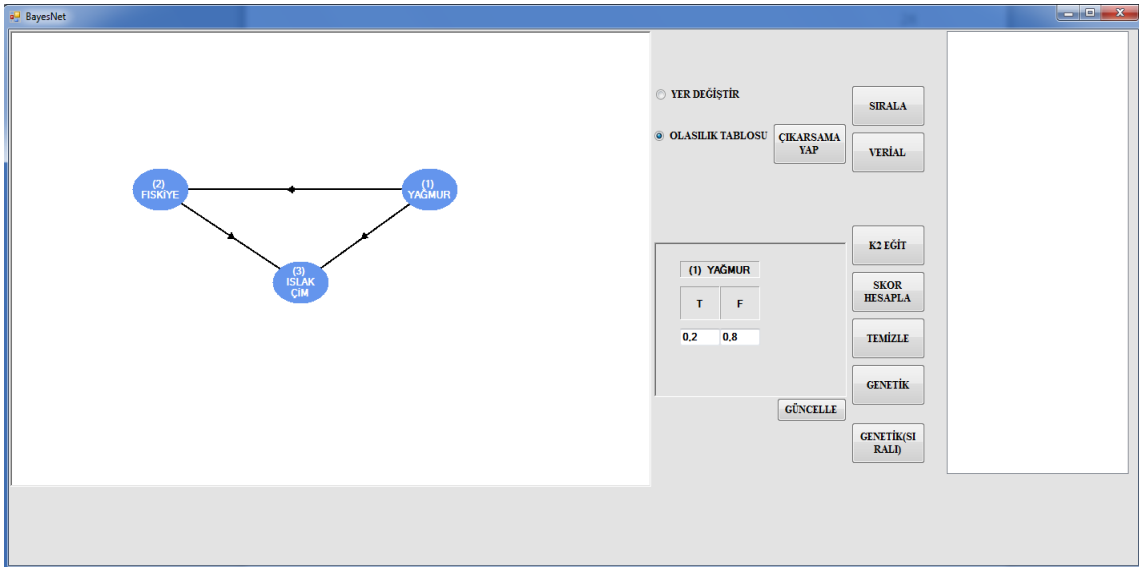
Düğümler oluşturulduktan sonra "Tamam" butonuna basılarak bağlantı oluşturma adımına geçilir. "Bağlantı Ekle " seçeneği seçiliyken sırasıyla ata düğüm ve çocuk düğüm olmak üzere iki düğüm üzerine tıklanarak şekil 5.3 de gösterildiği gibi aralarında bağlantı oluşturulur.



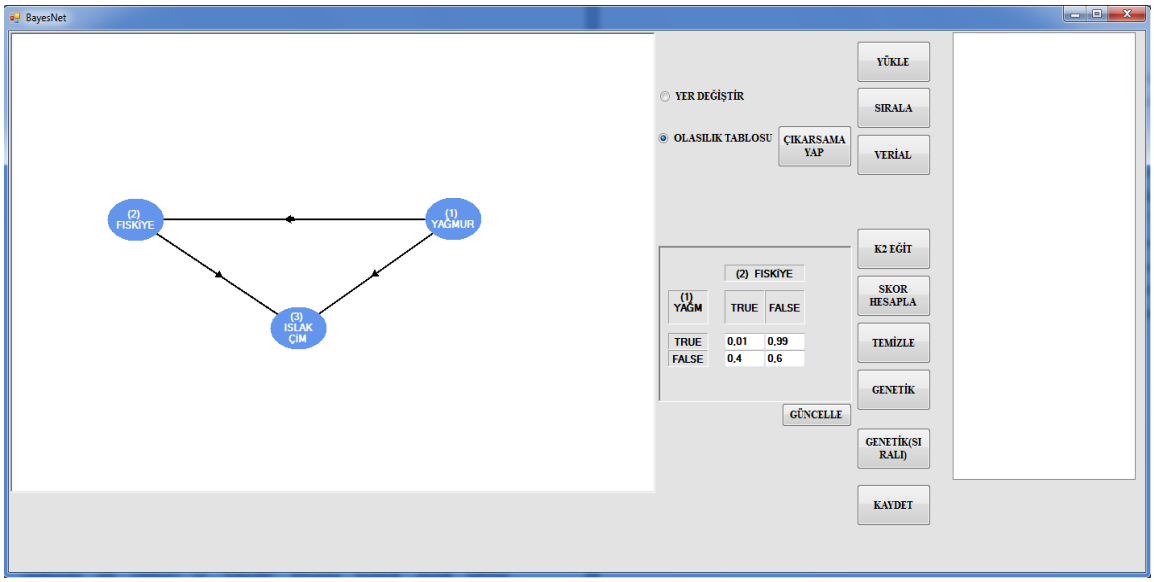
Şekil 5.3. Bağlantı ekleme işlemi

5.1.3. Olasılık Tablosu Oluşturma

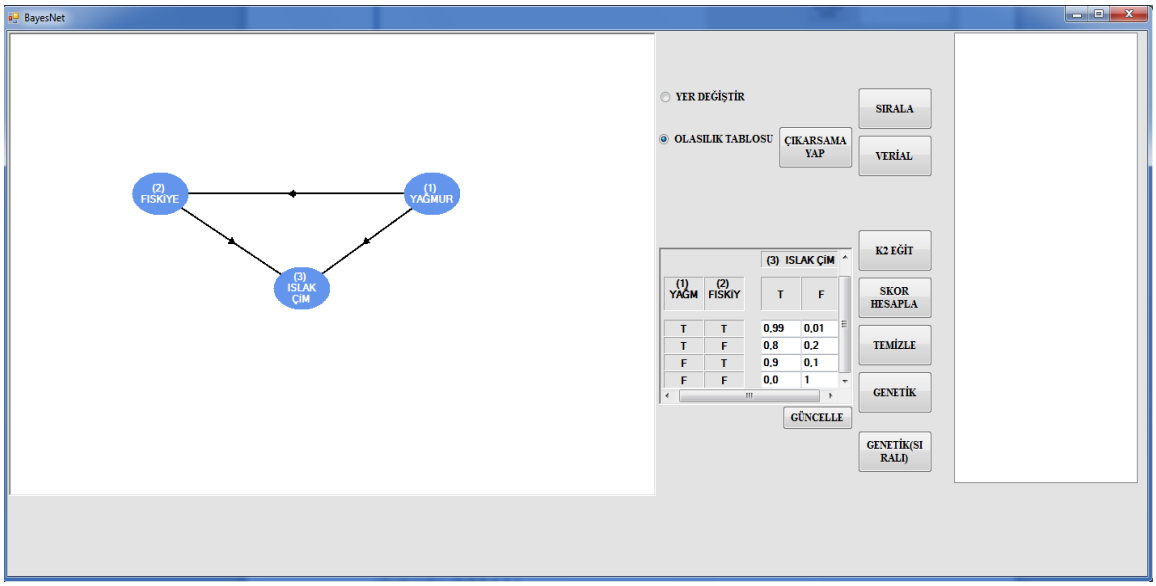
Bağlantı oluşturma işleminden sonra olasılık tablosu oluşturma işlemine geçilir. Bu işlem için "Olasılık Tablosu" seçeneği seçiliyken bir düğüm üzerine tıklanır. Seçilen düğümün bağlantı oluşturma adımında oluşturulmuş olan bağlantılarına göre bir tablo ekranı gelir.



Şekil 5.4. Yağmur düğümünün olasılık tablosu



Şekil 5.5. Fıskiye düğümünün olasılık tablosu



Şekil 5.6. Islak Çim düğümünün olasılık tablosu

Gelen tablo ekranında değerler her satır için satır değeri toplamı 1'e eşit olacak şekilde değerler girilir (satır değerlerinin toplamı 1'e eşit olmadıkça satır güncelleme işlemi yapılmasına izin verilmez) ve "Güncelle" butonuna basılarak olasılık tabloları oluşturulur.

5.1.4. Eğitim Yapma

BA uzmanların bilgileriyle interaktif el ile kurulabildiği gibi veriden eğitilerek de oluşturulabilir. Uygulamada BA'yı veriden eğiterek oluşturmak için K2 algoritması kullanılmıştır. K2 algoritmasının performansı düğümlerin sırasına bağlıdır. Bunun için K2 algoritmasına düğümler sıralı bir şekilde verilmelidir. Uygulama içinde bu sıralamayı bulmak için genetik algoritma ve 6. bölümde önerilen yöntem kullanılmaktadır. Ayrıca genetik algoritma ve önerilen yöntemin hibrit bir şekilde kullanıldığı bir yöntemle de sıralama bulunmaktadır.

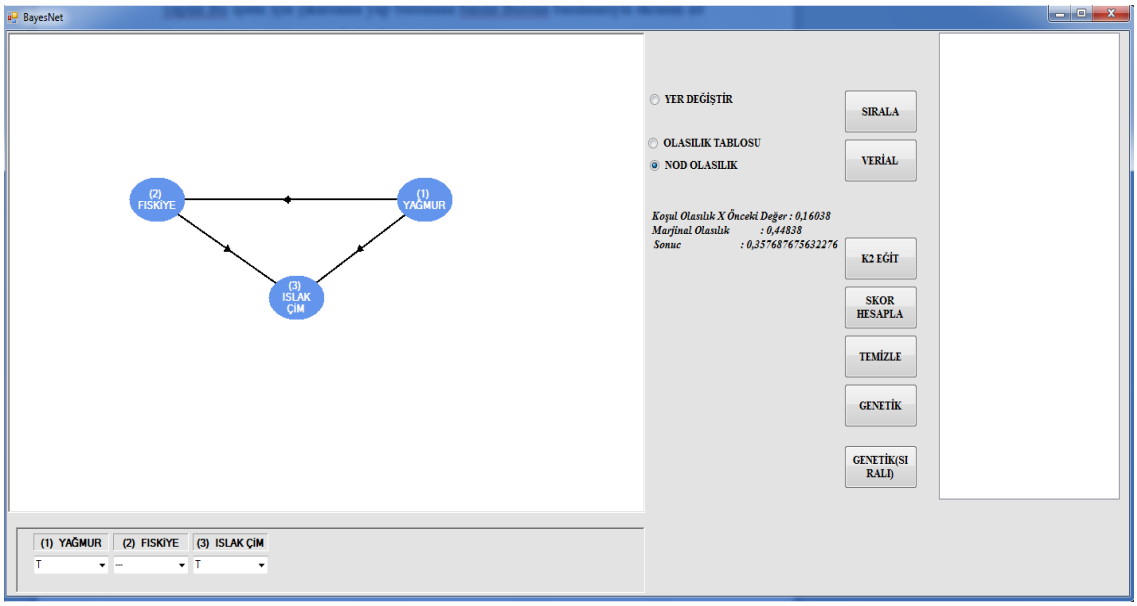
Ağ kurulumu tamamlandıktan sonra oluşan bu ağ ve veriler kullanılarak olasılık tabloları oluşturulmaktadır. Eğitim işleminde parametre öğrenme için maksimum olabilirlik tahmini yöntemi kullanılmaktadır.

Parametre öğrenme işleminde dikkat edilecek bir nokta eğitim kümesinde olmayan bir durumla karşılaşılmasıdır. Eğitim kümemizin 3 bireyli olduğunu düşünelim. Bu kümede birey3 hem birey1 hem de birey2 ile etkileşimde olduğunu düşünelim. Fakat birey1 ve birey2 hiçbir etkinliğe beraber katılmamış olsunlar. Bu durumda birey1=1, birey2=1 olduğunda birey3 için "0" ve "1" değerlerinin olasılıkları

hesaplanamaz. Örnekteki gibi eğitim kümesinde karşılaşılmayan bir durum olduğunda uygulamada olasılıklar eşit olarak dağıtılır. Yani birey1 ve birey2 "1" değerinde olduğunda birey3 için "0" ve "1" değerlerinin olasılıkları eşit olarak 0,5'tir. Buradaki varsayımımız eğitim kümesinde bulunmayan (hakkında bir bilgimiz olmayan) durumlara bağlı başka bir olayın meydana gelme olasılığı ve meydana gelmeme olasılığı eşittir.

5.1.5. Çıkarsama Yapma

BA oluşturulduktan sonra oluşan ağ yapısı üzerinden çıkarsama işlemi yapılır. Bu işlem için "Çıkarsama Yap" butonuna basılır. Butona basılmasıyla ekranın alt kısmına her bir düğüm için değerini seçebileceğimiz değerler listesi gelir.



Şekil 5.7. Çıkarsama işlemi

Burada çimlerin ıslak olması durumunda yağmurun yağma olasılığını bulmak isteyelim. Yağmur değişkeninin "T" değeri ve Islak Çim değişkeninin "T" değeri seçilir (Şekil 5.7). Daha sonra "Yağmur" düğümü üzerine fare ile tıklanır. Böylece çıkarsama sonuçları elde edilir.

Çıkarsama sonucu:

$$P(Y=T|Ç=T) = \frac{P(Ç=T,Y=T)}{P(Ç=T)} = \frac{0.16038}{0.44838} \approx 35.77\%$$

Yaklaşık %35.77 olasılıkla “çimler ıslaksa sebebi yağmur yağmasıdır” çıkarsaması yapılır.

6. YAPILAN ÇALIŞMA VE ALINAN SONUÇLAR

Tez kapsamında SA'daki bireyler arasındaki etkileşimi barındıran BA'nın oluşturulması ve bu ağ kullanılarak çıkarım yapılması hedeflenmektedir. BA'nın eğitiminde SA verileri ve K2 algoritması kullanılmıştır. K2 algoritmasının performansı ağı oluşturacak bireylerin temsil edildiği düğümlerin sırasına bağlıdır.

Tezin bu bölümünde ağı oluşturan düğümlerin sırasını bulabilmek için bir yöntem önerilmiştir. Gerçek veriler ile rassal sıralama, GA ile bulunan sıralama ve önerilen yöntemin bulduğu sıralama kullanılarak K2 algoritmasıyla ağ oluşturulmuştur. Elde edilen ağ skorları karşılaştırılmıştır. Ayrıca skoru iyileştirmek için GA ve önerilen yöntem hibrit bir şekilde kullanılarak sıralama bulunmuştur.

6.1. Önerilen Yöntem

K2 algoritması için doğru sıralamanın bulunması amacıyla bir yöntem önerilmiştir. Bu yöntem için K2 skor hesaplama formülü kullanılmıştır.

K2 skor hesaplama formülü;

$$f(i,\pi_i) = \prod_{j=1}^{qi} \frac{(r_i - 1)\,!}{(N_{ij} + r_i - 1)\,!} \prod_{k=1}^{ri} \alpha_{ijk}\,! \qquad (6.1.)$$

K2 algoritmasında bu formülle ilk olarak düğümün kendi skoru hesaplanır. Buna $P_{eskiskor}$ denilir. Daha sonra bu düğümden önce sırada gelmiş düğümlerin her biri ile formül kullanılarak skor hesaplanır. Bu skorların en büyüğü $P_{yeniskor}$ olarak alınır. Eğer yeni skor eski skordan büyükse ($P_{yeniskor} > P_{eskiskor}$) eski skor değeri yeni skor değerini alır ($P_{eskiskor} = P_{yeniskor}$), düğüm ebeveyn olarak kabul edilir ve ebeveynden düğüme bir ok çizilir. Bu işlem eski skor yeni skordan küçük oldukça ya da maksimum ebeveyn düğüm sayısına ulaşıncaya kadar devam eder. K2 algoritması bu formülü sıralanmış düğümlere uygular ve formülden elde edilen skorlara göre düğümlerin ebeveynleri belirlenir.

Önerilen yöntemde bu formül kullanılarak düğümlerin sırasının bulunması amaçlanmaktadır. Her bir düğümün ağ içindeki diğer tüm düğümlerle bu formül uygulanarak skoru bulunur. K2 algoritmasında bu formül bir düğüm için sıralamaya

göre sadece kendinden önce gelmiş düğümlere uygulanmaktadır. K2 algoritmasından farklı olarak önerilen yöntemde bir düğüm için formül ağ içindeki bütün düğümlere uygulanır. Formül uygulanmasından sonra her bir düğüm için bir skor elde edilir. Bu skorlar içinde en büyük skora sahip düğüm sıralamada ilk sıradadır. Sıralama düğümlerin skorları büyükten küçüğe sıralayarak yapılır. Ayrıca K2 algoritmasındaki gibi burada da maksimum ebeveyn sayısı belirlenmelidir.

Çizelge 6.1. 3 düğümlü örnek veri kümesi

Durumlar	Düğüm1	Düğüm2	Düğüm3
1	1	0	0
2	1	1	1
3	0	0	1
4	1	1	1
5	0	0	0
6	0	1	1
7	1	1	1
8	0	0	0
9	1	1	1
10	0	0	0

3 adet düğümün olduğu 10 durum içeren bir veri kümesinin olduğunu varsayalım (Çizelge 6.1). Burada düğümler bireyleri, durumlar ise etkinlikleri temsil etmektedir. r1,r2,r3 değeri her düğüm için değişkenler {0,1} den oluştuğu için 2 ye eşittir. Maksimum ebeveyn sayısı 2 olarak belirlenmiştir.

Düğüm1 için skor hesaplama;

1. $\pi_1 = Ø$

2. $P_{eskiskor} := f(1, Ø) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$

Formüldeki gerekli değerler hesaplanır. İlk olarak düğüm için $P_{eskiskor}$ hesaplanır. Bu ilk (ebeveynsiz) düğüm hesaplamada q_i=0 dır. Bu yüzden j formülden yok edilerek ilk hesaplama yapılır.

- α_{11}=5 (x1=0)
- α_{12}=5 (x1=1)
- $N_1 = \alpha_{11} + \alpha_{12} = 10$

Bunun sonucu ;

$$P_{eskiskor} := f(1, \emptyset) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}! = (1/11!) * 5! * 5!$$

$$= 1/2772$$

3. Daha sonra her bir düğüm ile $P_{yeniskor}$ hesaplanır. K2 algoritmasından farklı olarak her bir düğüm için skor hesaplanır. K2 algoritmasında düğümlerin sıralı geldiği varsayıldığı için sadece kendinden önceki düğümler ile skor hesaplaması yapılmaktadır. Önerilen yöntemde ise veri kümesindeki her bir düğüm ile skor hesaplaması yapılmaktadır.

Bu adımda $f(1,\pi_1 \cup \{x_2\})$ ve $f(1,\pi_1 \cup \{x_3\})$ değerleri hesaplanır.($\pi_1 = \emptyset$)

- $f(1,\pi_1 \cup \{x_2\}) = f(1,\{x_2\}) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$
  - $\phi_1 = V$ içindeki $\{x_2\}$ nin değerleri =$((x_2=0), (x_2=1))$
  - q1=$|\phi_1|$=2
  - α_{111}=4 (x2=0 ve x1=0)
  - α_{112}=1 (x2=0 ve x1=1)
  - α_{121}=1 (x2=1 ve x1=0)
  - α_{122}=4 (x2=1 ve x1=1)
  - $N_{11} = \alpha_{111} + \alpha_{112} = 5$
  - $N_{12} = \alpha_{121} + \alpha_{122} = 5$
  - $f(1,\{x_2\}) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$

    $= (1/6!)* 4!* 1!* (1/6!)* 1!* 4! = 1/900$
- $f(1,\pi_1 \cup \{x_3\}) = f(1,\{x_3\}) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$
  - $\phi_1 = V$ içindeki $\{x_3\}$ 'ün değerleri =$((x_3=0), (x_3=1))$
  - q1=$|\phi_1|$=2
  - α_{111}=3 (x3=0 ve x1=0)
  - α_{112}=1 (x3=0 ve x1=1)
  - α_{121}=2 (x3=1 ve x1=0)

- α_{122}=4 (x3=1 ve x1=1)
- $N_{11} = \alpha_{111} + \alpha_{112}$ =4
- $N_{12} = \alpha_{121} + \alpha_{122}$ =6
- $f(1,\{x_3\}) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$

  = (1/5!)* 3!* 1 !* (1/7!)* 2! * 4 ! = *1/2100*

4. $P_{yeniskor}$ = mak(f(1,$\pi_1 \cup \{x_2\}$)=1/900, f(1,$\pi_1 \cup \{x_3\}$)=1/2100)=1/900

$P_{yeniskor}$=1/900 > $P_{eskiskor}$ =1/2772

$P_{yeniskor}$ değeri $P_{eskiskor}$ değerinden büyük olduğu için x2 değeri ebeveyn listesine eklenir (π1={x2}) ve $P_{eskiskor}$=$P_{yeniskor}$ olarak işleme devam edilir.

5. Sonraki adımda maksimum ebeveyn sayısına ulaşılıp ulaşılmadığı kontrol edilir. Maksimum ebeveyn sayısına ulaşılmadıysa işleme devam edilir. Düğüm1'in ebeveyn sayısı maksimum ebeveyn sayına ulaşmadığı için işleme devam edilir. Düğüm1'in 2. ebeveyninin olup olmadığını bulmak için hesaplamalar yapılır.

- $f(1,\pi_1 \cup \{x_3\}) = f(1,\{x_2, x_3\}) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$
  - $\phi_1 = V$ içindeki {x2,x3} ün değerleri =((x3=0,x2=0), (x3=0,x2=1), (x3=1,x2=0), (x3=1,x2=1))
  - q1=| ϕ_1 |=4
  - α_{111}=3 (x3=0 ,x2=0, x1=0)
  - α_{112}=1 (x3=0 ,x2=0, x1=1)
  - α_{121}=0 (x3=0 ,x2=1, x1=0)
  - α_{122}=0 (x3=0 ,x2=1, x1=1)
  - α_{131}=1 (x3=1 ,x2=0, x1=0)
  - α_{132}=0 (x3=1 ,x2=0, x1=1)
  - α_{141}=1 (x3=1 ,x2=1, x1=0)
  - α_{142}=4 (x3=1 ,x2=1, x1=1)
  - $N_{11} = \alpha_{111} + \alpha_{112}$ =4

- $N_{12} = \alpha_{121} + \alpha_{122} = 0$
- $N_{13} = \alpha_{131} + \alpha_{132} = 1$
- $N_{14} = \alpha_{141} + \alpha_{142} = 5$
- $f(1,\{x_2, x_3\}) = \prod_{j=1}^{q1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r1} \alpha_{1jk}!$

$= (1/5!)* \alpha_{111}!* \alpha_{112}!* (1/1!)* \alpha_{121}! * \alpha_{122}! * (1/2!)* \alpha_{131}!* \alpha_{132}!* (1/6!)* \alpha_{141}! * \alpha_{142}!$

$= (1/5!)* 3! * 1! * (1/1!)* 0! * 0! * (1/2!)* 1! * 0! * (1/6!)* 1! * 4! = 1/1200$

6. $P_{yeniskor}=1/1200 < P_{eskiskor} =1/900$, yeni skor değeri eski skordan büyük olmadığı için işlem sonlandırılır X3 ebeveyn listesine eklenmez.

$P_{skor}(1)$, $P_{eskiskor}$ değerini alır, ($P_{skor}(1)= P_{eskiskor}$)

$P_{skor}(1) =1/900$

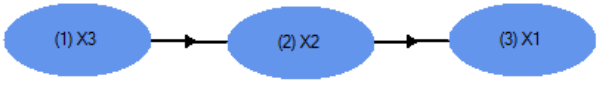
Aynı işlemler diğer düğümler içinde uygulanır ve skorları hesaplanır.

- $P_{skor}(1) =1/900$
- $P_{skor}(2) =1/210$
- $P_{skor}(3) =1/180$

Düğümler skorları hesaplandıktan sonra birbirleriyle karşılaştırılırlar. En büyük skora sahip düğüm sıralamanın en başına geçirilir. Sıralama skor değerleri büyükten küçüğe doğru olacak şekilde yapılır.

Bu örnek için bulunan sıralama ={ Düğüm3, Düğüm 2, Düğüm 1}

Anlatılan bu yöntem ile sıralama bulunduktan sonra bu sıralamaya göre veri kümesi sıralanır. Sıralanmış veriye K2 algoritması uygulanır ve ağ yapısı oluşturulur. Bu sıralamaya göre oluşturulmuş ağ şekil 6.1 de gösterilmiştir.

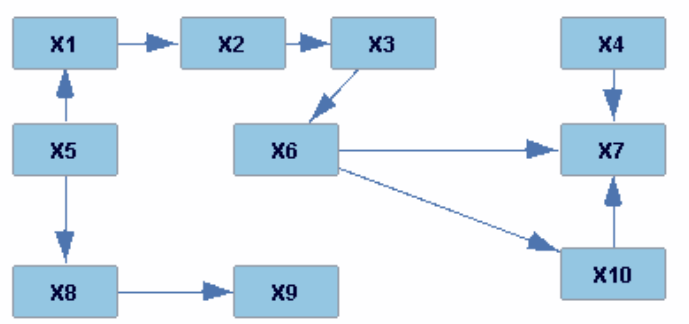


Şekil 6.1. Önerilen sıralamayla bulunan BA

Hesaplama işlemleri yapılırken K2 algoritmasını hızlandırmak için $f(i,\pi_i)$ skoru yerine $log(f(i,\pi_i))$ skoru hesaplanmaktadır. Önerilen yöntem içinde hızın artırılması amacıyla $f(i,\pi_i)$ skoru yerine $log(f(i,\pi_i))$ skoru kullanılmıştır.

6.2. Örnek Çalışma

Önerilen yöntemin sonuçlarını değerlendirmek için ilk olarak Tetrad IV paket programı tarafından üretilmiş veriler kullanılmıştır. Tetrad IV, istatistiksel ve nedensel modeller için geliştirilmiş ücretsiz Java tabanlı bir programdır (Ramsey ve ark., 2004). Tetrad IV programı ile şekil 6.2 de gösterilen 10 düğüm içeren küçük bir ağ oluşturulmuştur.



Şekil 6.2. Örnek BA yapısı

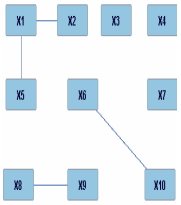
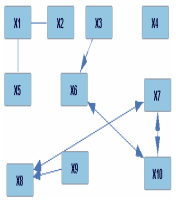
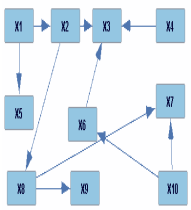
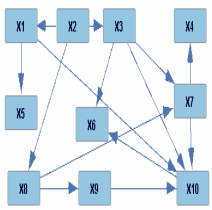
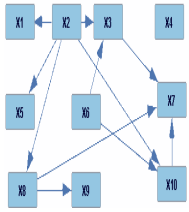
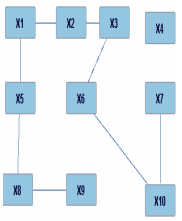
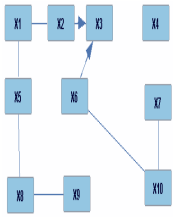
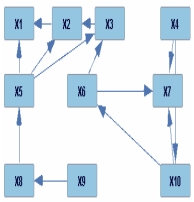
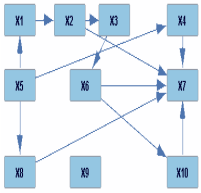
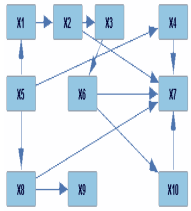
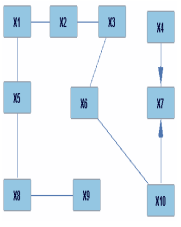
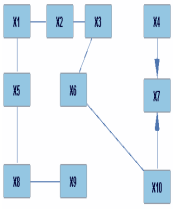
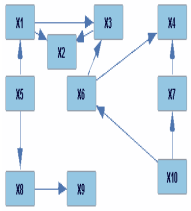
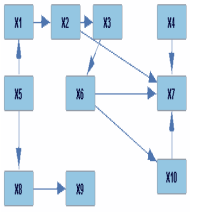
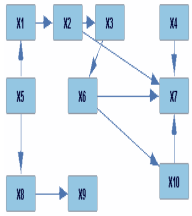
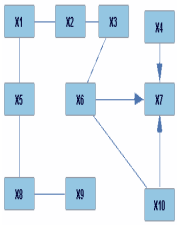
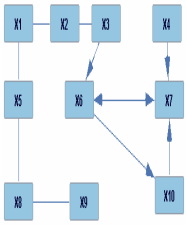
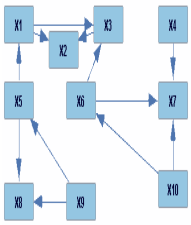
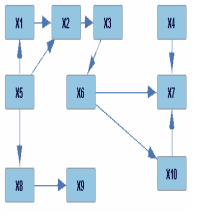
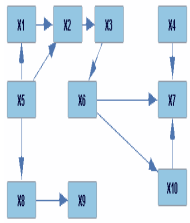
Tetrad IV programı ile kurulmuş şekil 6.2 de gösterilen ağ yapısına bağlı 100, 500, 1000, 2000, 3000 kayıt içeren veri kümeleri Tetrad IV programı ile oluşturulmuştur. Program, verileri ağ yapısına bağlı ve rassal olarak oluşturmaktadır. Oluşturulan bu veri kümeleri algoritmalara giriş verisi olarak verilerek sonuç olarak başlangıçta kurulan ağ yapısının bulunması amaçlanmaktadır.

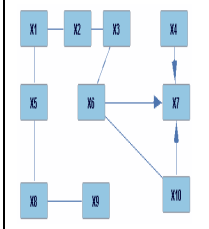
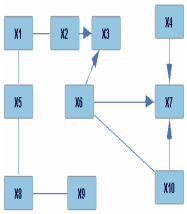
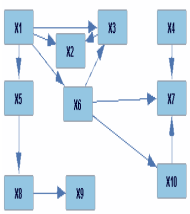
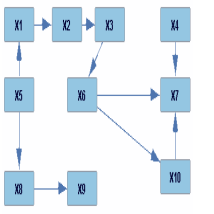
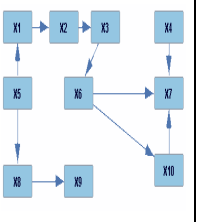
Bu veriler ile ADA (Açgözlü Denklik Arama) , PC algoritmalarının ve düğümlerin rassal, genetik algoritma ve önerilen yöntem kullanılarak sıralanmasıyla K2 algoritmasının buldukları ağ yapıları çizelge 6.2 de sunulmuştur.

ADA algoritması iki fazlı açgözlü arama algoritmasıdır. Algoritma hiçbir kenarı olmayan yönlendirilmiş çevrimsiz grafik içeren sınıf ile başlar. Sonra ilk fazda yerel bir maksimuma ulaşıncaya kadar sadece ileri yönde aç gözlü arama yapar. İkinci fazda birinci fazın yerel maksimumundan başlayarak ikinci aç gözlü arama yapar, fakat bu kez arama geri doğrultuda gerçekleştirilir. Algoritma ikinci fazda yerel maksimuma ulaşıldığında sona erer (Chickering ve Meek, 2002).

PC algoritması Spirtes ve Glaymour (1990) tarafından sunulmuş bir sınırlama arama algoritmasıdır. Değişken çiftlerinin diğer değişken kümeleri üzerinde bağımsızlık şartını sınayarak çalışır ve ilk fazda bir yönsüz yakınlık grafiği verir. Kenarlar belirli kurallara göre ikinci fazda yönlendirilir.

Çizelge 6.2. Örnek BA için ADA, PC, rassal sıralı K2, GA ile sıralanmış K2 ve önerilen yöntemle sıralanmış K2 algoritmalarının bulduğu ağ yapıları

Kayıt	ADA	PC	Rassal Sıralı K2	GA ile sıralanmış K2	Önerilen yöntemle sıralanmış K2
100					
500					
1000					
2000					

3000					

Çizelge 6.2 deki rassal sıralı K2 algoritmasında her kayıt için 10 farklı rassal sıralama üretilip bunlar içindeki en iyi skora sahip ağ yapısı tabloda gösterilmiştir

GA ile sıralanmış K2 algoritması için çaprazlama olasılığı 0.5, mutasyon olasılığı 0.1 seçilmiştir. Popülasyon boyutu 10 ve iterasyon sayısı 100 olarak ayarlanmıştır. GA sezgisel bir algoritma olduğu için algoritma her çalışmasında değişik sonuçlar üretebilir. Bu yüzden düğümlerin sıralamasını bulmak amacıyla GA için 10 rassal başlangıç popülasyonu üretilmiştir. Her bir başlangıç popülasyonuna GA uygulanmıştır. GA uygulanması sonucunda elde edilen en iyi uygunluk değerini veren sıralamaya göre oluşturulan ağ yapısı tabloya eklenmiştir.

Sonuçlardan da görüleceği gibi veri kümesinde 10 düğümlü ağı örnekleyen veri sayısı arttıkça her algoritma ağ yapısına daha çok yaklaşmaktadır. 3000 örnek ile örneklendiği zaman GA ile sıralanmış K2 ve önerilen yöntemle sıralanmış K2 algoritmalarının doğru ağ yapısını elde ettiği görülmektedir. Fakat rassal sıralamalı K2 algoritması 3000 örnekte de yanlış bağlantılar bulmuştur. ADA ve PC algoritmaları ise 3000 örnekte bağlantıları doğru bulmuş fakat yönleri tam bulamamışlardır.

6.3. Gerçek Veri Kümesi Uygulaması

Algoritmanın çalışmasını test etmek için gerçek veri kümesi kullanılmıştır. Veri kümesi (LabData) J. Kubica tarafından (2003) derlenmiştir. İçeriğinde Carnegie Mellon Üniversitesi Auton Laboratuvarı için üyeler arasındaki ortak bağlantıları bulundurur. 115 birey ve 94 bağlantıdan oluşmaktadır.

Öncelikle veri kümesi uygulamamızda kullanabilmek için uygun formata getirilmiştir. Bu yüzden bireyler sütun değişkenleri, bağlantılarsa satır değişkenleri olacak şekilde tablo oluşturulur. Eğer bir birey ilgili bağlantıda yer alıyorsa “1” değilse

"0" değeri verilir. K2 algoritması için maksimum ebeveyn sayısı 10 olarak belirlenmiştir.

İlk olarak rassal 30 sıralama yapılıp bu sıralamalara göre K2 algoritmasıyla ağ yapısı bulunup skor hesaplanmıştır.

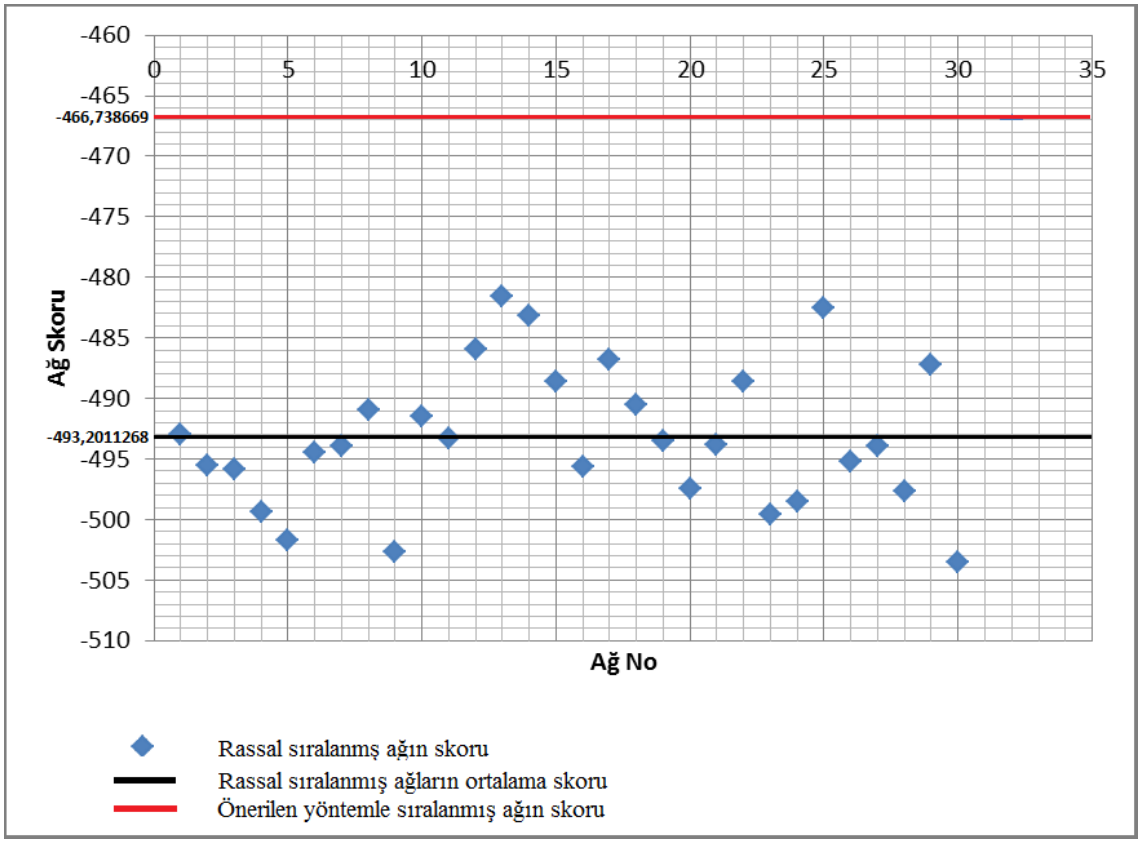
Çizelge 6.3. Rassal sıralı verilerin ağ skorları

No	Ağ Skoru	No	Ağ Skoru
1	-492.98925	16	-495.6093831
2	-495.5117107	17	-486.763685
3	-495.8991248	18	-490.5663736
4	-499.3172128	19	-493.5272166
5	-501.7277863	20	-497.4878308
6	-494.4367981	21	-493.858743
7	-493.9520584	22	-488.6430417
8	-490.8973927	23	-499.5511872
9	-502.6523101	24	-498.5002394
10	-491.4516869	25	-482.4997984
11	-493.2447153	26	-495.2483798
12	-485.9509313	27	-493.9493799
13	-481.5703819	28	-497.6477981
14	-483.2166745	29	-487.1887406
15	-488.6313	30	-503.542674

Çizelge 6.3 de gösterilen 30 adet rassal sıralı ağ skorunun ortalaması alınmıştır. Daha sonra önerilen yöntem kullanılarak sıralama bulunmuştur. Bu sıralamaya göre K2 algoritmasıyla ağ oluşturulmuş ve ağın skoru hesaplanmıştır. Önerilen yöntemle sıralanmış ağ skoru ile rassal sıralanmış ağ skorlarının karşılaştırılması aşağıda verilmiştir.

Rassal sıralı K2 skor (Ortalama) : -493.2011268

Önerilen yöntemle sıralanmış K2 Skor : -466.738669



Şekil 6.3. Rassal sıralanmış ağların ve önerilen yöntemle sıralanmış ağın skorları

Sonra düğümler GA ile sıralanmış ve GA ile sıralanmış ağların skorları bulunmuştur. GA için boyut=10 ve iterasyon=100 olarak belirlenmiştir. Popülasyondaki her birey ağı oluşturacak düğümlerin sıralamasını ifade etmektedir. 30 adet rassal üretilmiş farklı başlangıç popülasyonuyla GA uygulanmış ve bu çalışmaların sonucunda sıralamalar bulunmuştur. Bu sıralamalar kullanılarak K2 algoritmasıyla ağ yapıları oluşturulmuş ve ağların skoru bulunmuştur. Bulunan sonuçlar çizelge 6.4 de gösterilmiştir.

Çizelge 6.4. GA ile sıralanmış verilerin ağ skorları

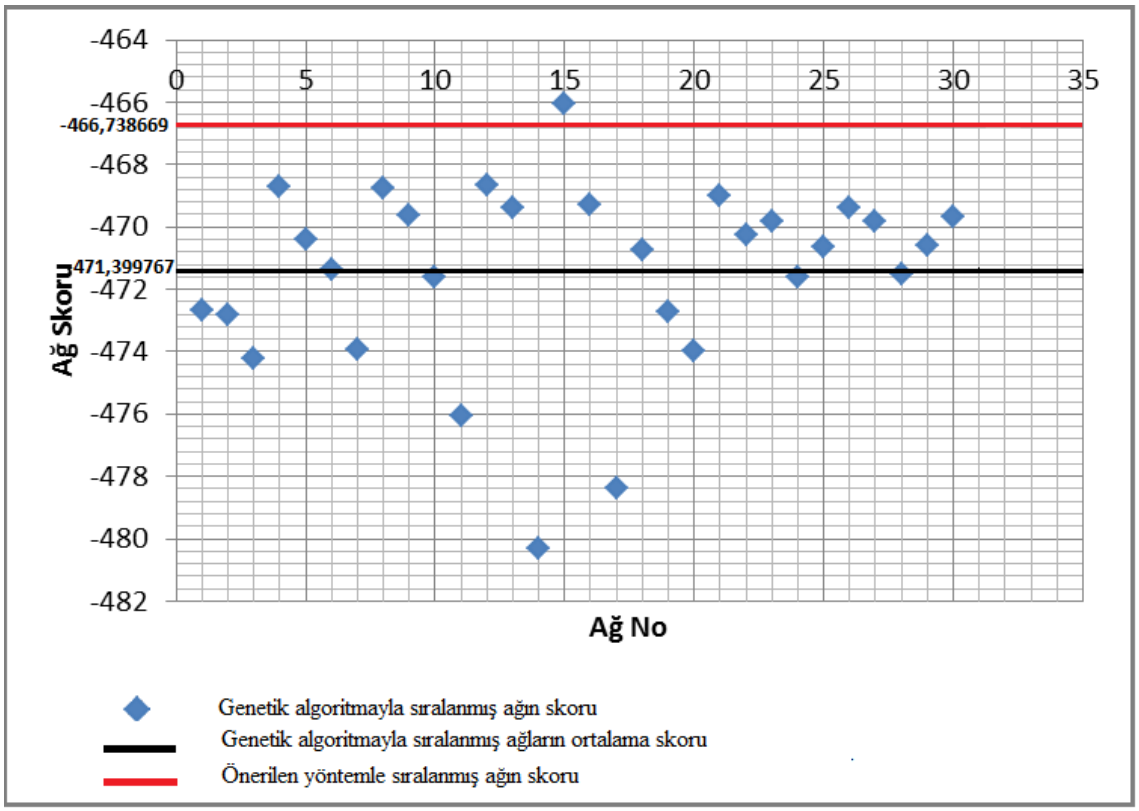
No	Ağ Skoru	No	Ağ Skoru
1	-472.6798417	16	-469.282169
2	-472.8192043	17	-478.3637472
3	-474.227131	18	-470.7258379
4	-468.730246	19	-472.7081865
5	-470.4012574	20	-473.9972401

6	-471.3886848	21	-468.9758408
7	-473.9454922	22	-470.2514915
8	-468.7744803	23	-469.8427824
9	-469.6238609	24	-471.6276212
10	-471.5998228	25	-470.6347563
11	-476.0451565	26	-469.3723689
12	-468.6425874	27	-469.8094984
13	-469.3984805	28	-471.5259133
14	-480.2913106	29	-470.5829754
15	-466.0284532	30	-469.696567

Çizelge 6.4 de 30 adet rassal üretilmiş farklı başlangıç popülasyonuyla GA uygulanarak bulunan sıralamalar ile oluşturulan ağ skorlarının ortalaması alınmıştır. Önerilen sıralama yöntemiyle bulunan ağ skoru ile karşılaştırılması aşağıda verilmiştir.

GA ile sıralı K2 skor (Ortalama) : -471.399767

Önerilen yöntemle sıralanmış K2 Skor : -466.738669



Şekil 6.4. Genetik algoritmayla sıralanmış ağların ve önerilen yöntemle sıralanmış ağın skorları

Şekil 6.4 de GA ile yapılan sıralamalarla oluşan ağların skorları, bu ağların ortalama skorları ve önerilen yöntemle sıralanmış ağın skoru gösterilmiştir. Önerilen yöntemle bulunan ağın skoru GA ile sıralanmış ağların ortalama skorundan daha yüksektir. Görüleceği gibi GA ile 30 farklı başlangıç popülasyonuyla bulunan sıralamalardan sadece bir tanesinin önerilen yöntemle bulunan skoru geçtiği görülmektedir. Geri kalan 29 tanesinin skoru önerilen yöntemle bulunan skordan düşüktür.

Sonuçların iyileştirilmesi amacıyla GA ile önerilen yöntem hibrit bir şekilde kullanılmıştır. Burada GA başlangıç popülasyonu tamamen rassal olarak üretilmemiştir. Başlangıç popülasyonuna önerilen yöntemle bulunmuş sıralama eklenmiş popülasyonun diğer bireyleri ise rassal üretilmiştir. Bu hibrit yöntem 30 farklı başlangıç popülasyonuyla çalıştırılmış ve sonuçları çizelge 6.5 de verilmiştir.

Çizelge 6.5. Hibrit yöntemle sıralanmış verilerin ağ skorları

No	Ağ Skoru	No	Ağ Skoru
1	-465.4227503	16	-465.1217687

2	-464.2092392	17	-465.60775
3	-464.9550706	18	-465.5202098
4	-464.5882382	19	-465.8025084
5	-465.0570286	20	-465.5202098
6	-464.8672506	21	-464.5937097
7	-465.158685	22	-465.6033871
8	-465.1587375	23	-465.2372882
9	-464.9502739	24	-465.5250064
10	-465.6033871	25	-465.1217687
11	-464.8712859	26	-465.895445
12	-464.5289695	27	-465.2560985
13	-464.973351	28	-464.973351
14	-465.2239764	29	-464.5845522
15	-466.4188248	30	-465.1720233

Çizelge 6.5 de gösterilen 30 adet hibrit yöntemle bulunan sıralamaların ağ skorlarının ortalaması alınmıştır. Önerilen sıralama yöntemiyle bulunan ağ skoru ile karşılaştırılması aşağıda verilmiştir.

Hibrit yöntemle sıralanmış K2 skor (Ortalama) : -465.1840715
Önerilen yöntemle sıralanmış K2 Skor : -466.738669

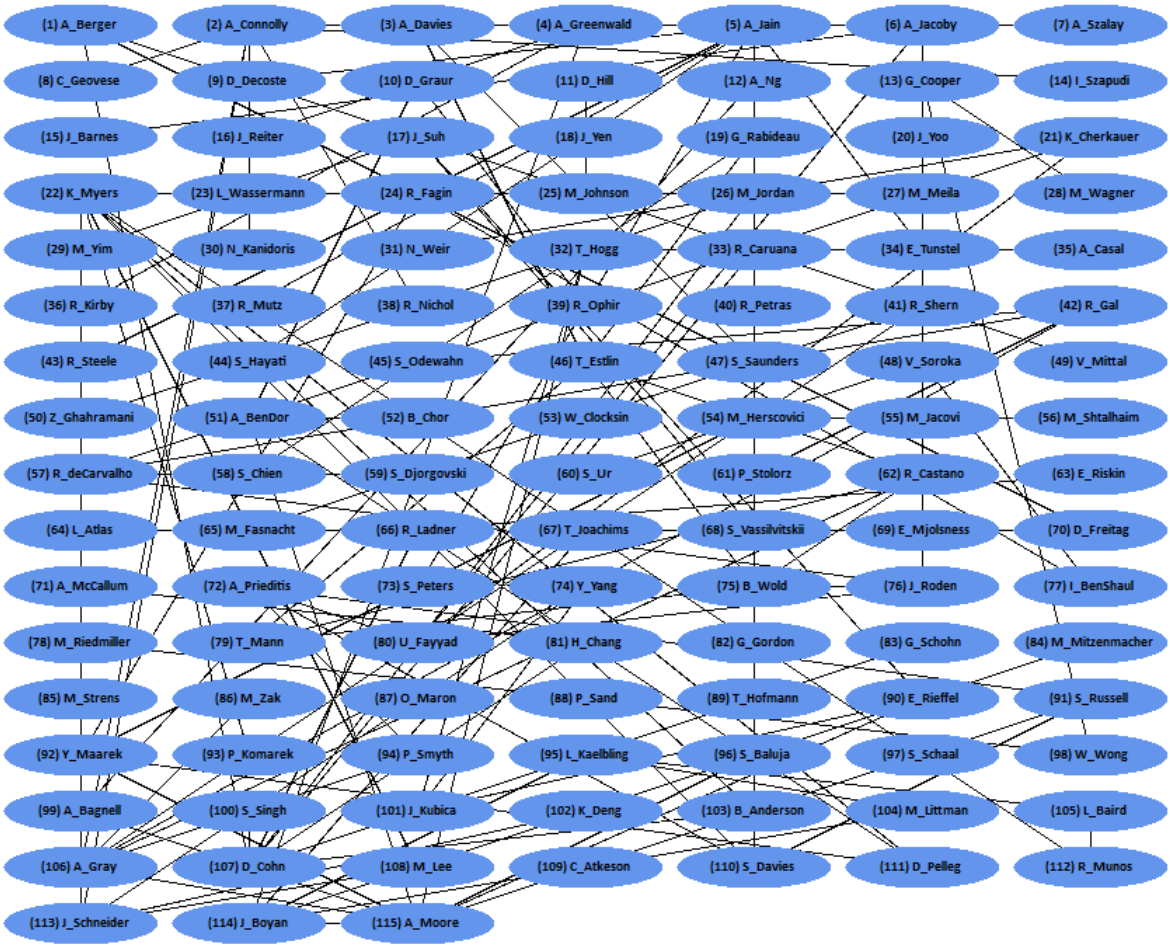
6.4. SA'da Çıkarsama Yapma

BA ile çıkarsama yapılırken tek bir yön takip edilmez. Başka bir deyişle bir düğüm için olasılık bulunurken sadece etkilendiği düğümlere değil ayrıca etkilediği düğümlerin durumuna da bakılır. B bireyi A bireyinden etkileniyor olsun. A bireyi bir etkinliğe katılmışsa B bireyinin bu etkinliğe katılma olasılığını etkileyecektir, aynı şekilde B bireyi bir etkinliğe katılmışsa A bireyinden etkilenip bu etkinliğe katılma olasılığı da olduğu için A bireyinin bu etkinliğe katılma olasılığını etkileyecektir.

Bireylerin etkinliğe katılma olasılıkları doğrudan veya dolaylı olarak aralarında etkileşim bağlantısı bulunan bireylerin durumlarından etkilenmektedir.

Tez kapsamında SAA yapmak amacıyla bireyler arasındaki ortak bağlantıları bulunduran gerçek veri kümesi (LabData) kullanılarak BA eğitimle oluşturulmuştur.

Veri kümesinde 115 birey bulunmaktadır. BA içinde bu bireyleri düğümler teşkil etmektedir. Düğümler arasındaki bağlantılar ise bu bireylerin etkilendiği veya etkilediği bireyleri göstermektedir. BA'nın eğitimi için K2 algoritması kullanılmıştır. Ağ içindeki düğümlerin sıralaması 6. bölümde önerilen hibrit yöntem ile yapılmıştır. K2 algoritmasında bir düğümün ebeveyninin kendinden önce gelen bir düğüm olması gerekir. Oluşan BA da iki düğüm arasında bir bağlantı varsa ön sırada ağa eklenen düğüm etkileyen, diğer düğüm etkilenen düğümdür.



Şekil 6.5. LabData veri kümesinden oluşturulan BA

Şekil 6.5 te gerçek veri kümesi ve önerilen hibrit yöntem kullanılarak oluşturulmuş BA gösterilmiştir. Bu BA üzerinden SAA yapılabilmektedir. Ağ üzerindeki bir bireyin etkinliğe katılma olasılığını başka bireylerin aynı etkinliğe

katılma durumlarına göre analiz edilebilmektedir. Uygulama üzerinden analiz yapılması istenen bireylerin değerleri seçilerek analiz gerçekleştirilebilmektedir.

Oluşan BA içinde M_Herskovici bir etkinliğe katılmış ve bu etkinliğe D_Pelleg katılmamışsa Y_Maarek düğümünün bu etkinliğe katılma olasılığı analiz edilmiştir. Y_Maarek bireyini temsil eden düğüm M_Herscovici bireyini temsil eden düğümden ve R_Fagin bireyini temsil eden düğümden doğrudan etkilenmektedir. A_Moore ve D_Pelleg düğümleri de Y_Maarek düğümünden etkilenmektedir.

BA üzerinden çıkarsama işlemi yapılırken D_Pelleg için D, M_Herskovici için M, Y_Maarek için Y kısaltmaları kullanılmıştır.

$$P(Y=1|D=0,M=1) = \frac{P(D=0,M=1,Y=1)}{P(D=0,M=1)} \approx 76.21\%$$

Geliştirilen uygulama yazılımında bahsedilen değerler seçilmiş ve çıkarsama işlemi yapılmıştır. Bu çıkarsamanın sonucunda M_Herskovici bir etkinliğe katılmış ve D_Pelleg bu etkinliğe katılmamışsa , %76,21 olasılıkla Y_Maarek'in de bu etkinliğe katılacağı söylenir.

Uygulama üzerinden analiz yapılmak istenen bireyler için değerler seçilmiş ve analiz işlemleri gerçekleştirilmiştir. Analiz sonuçları aşağıdaki çizelge 6.6 da sunulmuştur.

Çizelge 6.6. Analiz sonuçları

Bilinen Durumlar	Analiz Edilen Durum	Analiz Sonucu
M_Herskovici=1, D_Pelleg=0	Y_Maarek=1	%76,21
J_Suh=0, Y_Yang=0, J_Kubica=1	E_Rieffel=1	%40,16
M_Riedmiller=0, W_Wong=1	G_Cooper=0	%50,27
A_Davies=0, R_Castano=1, B_Wold=0	T_Mann=0	%33,57
M_Jordan=0, R_LAnder=0 ,B_Anderson=0, DChon=1	A_Berger=0	%85,55

7. SONUÇLAR VE ÖNERİLER

7.1. Sonuçlar

BA'lar belirsizlik hakkında bilgi sunmak için kullanılır. SA'lar kişilerin sosyal ilişkilerine göre birbirlerine bağlandığı yapılardır. Bu yapıda kişiler düğüm, ilişkiler bağlantı olarak ifade edilir.

Bu çalışmada kişilerin etkinliklere katılım verileri kullanılarak SAA için BA oluşturulmuştur. Bu veriler ile BA kullanılarak bireyler arasındaki etkileşim modellenmek istenmektedir.

Tez kapsamında BA yapısı veriden öğrenerek kurulmuştur. Eğitim için K2 algoritması kullanılmıştır. K2 algoritmasının performansı düğümlerin sırasına bağlıdır. Ağ içindeki düğümlerin sıralamasını bulmak için bir yöntem önerilmiştir.

Önerilen yöntemin performansını test etmek için ilk olarak Tetrad IV programı kullanılarak 10 düğüm içeren küçük bir ağ oluşturulmuştur. Bu ağı temsil edecek veriler 100, 500, 1000, 2000, 3000 durum içerecek şekilde program yardımıyla rassal üretilmiştir. Eğitim için üretilen bu veri kümeleri girdi değeri olarak verilip ağ bulunması istenmiştir. ADA, PC algoritmaları ve düğümlerin rassal, GA ve önerilen yöntem ile sıralanarak K2 algoritması ile BA yapıları oluşturulmuştur. Uygulama sonucunda önerilen yöntem ve GA ile önerilen yöntemle bulunan sıralama kullanan K2 algoritmaları 3000 durumda doğru ağ yapısını bulmuştur.

SAA için gerçek veri kümesi kullanılarak BA oluşturulmuştur. Bu veri kümesi ile düğümlerin rassal, GA ve önerilen yöntemle sıralanarak K2 algoritması ile BA yapıları oluşturulmuştur. Bulunan ağ yapılarının skorları karşılaştırılmıştır. Karşılaştırma sonucunda en iyi skoru önerilen yöntemle sıralanmış K2 algoritmasının bulduğu gözlemlenmiştir. Ayrıca sonuçları iyileştirmek için GA ve önerilen yöntem beraber kullanılmıştır. GA başlangıç popülasyonuna önerilen yöntem ile bulunan sıralama eklenerek hibrit bir yapı oluşturulmuş ve sonuçlar iyileştirilmiştir.

Önerilen bu hibrit yöntemle kişilerin etkinliğe katılma verileri kullanılarak BA oluşturulmuştur. Kurulan bu BA üzerinden SAA yapılmıştır. Ağ üzerindeki bir bireyin etkinliğe katılma olasılığı başka bireylerin aynı etkinliğe katılma durumlarına göre analiz edilmiş ve analiz sonuçları sunulmuştur.

7.2. Öneriler

Çalışmada kullanılan veri kümelerinin içindeki değerlerin tam olduğu düşünülmüştür. Ama gerçek hayatta karşılaşılan veriler genelde eksik, hatalı ve tutarsız olabilmektedir. Gelecek çalışmalarda kayıp ve eksik veriler göz önüne alınarak SAA için BA oluşturulması düşünülmektedir.

KAYNAKLAR

Anonymous, 2010, Kaynak: http://en.wikipedia.org/w/index.php?oldid=404640258 [Ziyaret Tarihi: 25 Ekim 2011]

Bayes, T., 1763, An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of Royal Society of Londan*, 370-418.

Ben-Gal, I., 2007, Encyclopedia of Statistics in Quality and Reliability, Ruggeri, F., Faltin, F., Kenett, R. (eds), *John Wiley & Sons*.

Bolstad, W. M., 2004, Introduction to Bayesian Statistics, *John Wiley & Sons,* 354p.

Bursa, O. ve Ünalır, M. O., 2010, Sosyal Ağlar ve Profil Yönetimine Etkileri, *Akademik Bilişim 2010*, Muğla Üniversitesi, 681-686.

Chickering, D. M., 1996, Learning Bayesian networks is NP-Complete, *Learning from Data: Artificial Intelligence and Statistics V*, 121-130.

Chickering, D. M., and Meek, C., 2002, Finding Optimal Bayesian Networks, *In Proceedings of the Eighteenth Conference on Uncertainty in Artifical Intelligence*, 94–102.

Cooper, G. F. and Heskovitz, E., 1992, A Bayesian method for the induction of probabilistic networks from data, *Machine Learning*, 309-347.

Degenne, A. and Forse, M., 1999, Introducing Social Networks, *Sage Publications*, London.

Doğan Akbay, H., 2010, Çevrimiçi öğrenme topluluklarının sosyal ağ analizi: Bir öğretmen forumu örneği, Yüksek Lisans Tezi, *Anadolu üniversitesi Sosyal Bilimler Enstitüsü,* Eskişehir.

Ehrlich, K. and Carboni, I., 2005, Inside Social Network Analysis, *IBM Technical Report 05-10*.

Goldberg, D. E., 1989, Genetic Algorithms in Search, Optimization and Machine Learning, *Kluwer Academic Publishers*, Boston.

Goldenberg, A., 2007, Scalable Graphical Models for Social Networks, PhD thesis, *Carnegie Mellon University,* Pittsburgh.

Holland, J. H., 1975, Adaptation in Natural and Artificial Systems, *University of Michigan Press*.

Krebs, V. E., 2002, Uncloaking Terrorist Networks [online], Kaynak: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/941/863 [Ziyaret Tarihi: 25 Ekim 2011]

Koelle, D., Pfautz, J., Farry, M., Cox, Z., Catto, G., Campolongo, J., 2006, Applications of Bayesian Belief Networks in Social Network Analysis , *In Proceedings of 4th Bayesian Modeling Applications Workshop at the 22nd Annual Conference on Uncertainty in AI: UAI '06* , Cambridge, Massachusetts.

Kubica, J., Moore, A., Cohn, D., Schneider., J., 2003, Finding Underlying Connections: A Fast Graph-Based Method for Link Analysis and Collaboration Queries, *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003),* Washington DC, 392-399.

Larranaga, P., Kuijpers, C. M. H., Murga, R.H., Yurramendi, Y., 1996, Learning Bayesian network structures by searching for the best ordering with genetic algorithms, *IEEE Transactions on System*, 26(4), 487-493.

Lu,Z., Lu,J., Bai, C. and Zhang, G., 2006, Customer online shopping behaviours analysis using Bayesian networks, *AI 2006: Advances in Artificial Intelligence*, Springer-Verlag, Berlin, Heidelberg, 1293–1297.

Marshall, G., 1999, Sosyoloji Sözlüğü, Çev.: O. Akınhay, ve D. Kömürcü, *Bilim ve Sanat Yayınları*, İstanbul.

Neapolitan, R. E., 2003, Learning Bayesian Networks, *Prentice – Hall*.

Park, H. S. and Cho, S. B., 2010, Building Mobile Social Network with Semantic Relation Using Bayesian Network-based Life-log Mining, *2010 IEEE Second International Conference on Social Computing*, 401-406.

Pearl, J., 1985 , Bayesian networks: A model of self-activated memory for evidential Reasoning, *In Proceedings, Cognitive Science Society*, 329–334.

Pearl, J., 1988, Probabilistic reasoning in intelligent systems: Networks of plausible inference, *Morgan Kaufmann Publishers*, San Mateo.

Ramsey, J. ve ark., 2004, Tetrad IV paket programı 4.3.10-5 sürümü [online], Kaynak: http://www.phil.cmu.edu/tetrad [Ziyaret Tarihi: 25 Ekim 2011]

Russel, S. and Norvig, P., 1995, Artificial Intelligence: A Modern Approach, *Prentice-Hall.*

Scott, J., 2000, Social Network Analysis A Handbook, *Sage.*

Spirtes, P. and Glymour, C., 1990, An Algorithm for Fast Recovery of Sparse Causal Graphs, *Department of Philosophy*.

Turner, B. S., 2006, The Cambridge Dictionary of Sociology, *Cambridge University Press.*

Verma, T. S. and Pearl, J., 1991, Equivalence and Synthesis of Causal Models, *Uncertainty in Artificial Intelligence 6,* 255-268.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:**	Betül AKKOÇ
Uyruğu	**:**	T.C.
Doğum Yeri ve Tarihi	**:**	Gebze/ 02.02.1989
Telefon	**:**	-
Faks	**:**	-
e-mail	**:**	betulakkoc@selcuk.edu.tr

EĞİTİM

Derece		**Adı, İlçe, İl**	**Bitirme Yılı**
Lise	:	Özel Envar Lisesi, Karatay, Konya	**2005**
Üniversite	:	Selçuk Üniversitesi Bilgisayar Mühendisliği, Selçulu, Konya	**2010**
Yüksek Lisans	:	Selçuk Üniversitesi	**2012**
Doktora	:	-	

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**
2011	Selçuk Üniversitesi Bilgisayar Mühendisliği Bölümü	Arş.Gör.
2009-2010	Selçuk Üniversitesi Bilgisayar Mühendisliği Bölümü	Öğrenci Asistan
2008-2009	Selçuk Üniversitesi Bilgisayar Mühendisliği Bölümü	Öğrenci Asistan

UZMANLIK ALANI

Bayes Ağları, Makine Öğrenmesi, Yapay Zeka

YABANCI DİLLER

İngilizce