



Kaba Küme Akış Çizgilerinin Olasılıksal Yapısı

Mert Bal^{1*}

^{1*} Yıldız Teknik Üniversitesi, Kimya Metalürji Fakültesi, Matematik Mühendisliği Bölümü, İstanbul, Türkiye, (ORCID: 0000-0001-6250-929X), mertbal@yildiz.edu.tr

(İlk Geliş Tarihi 28 Eylül 2020 ve Kabul Tarihi 11 Nisan 2021)

(DOI: 10.31590/ejosat.801014)

ATIF/REFERENCE: Bal, M. (2021). Kaba Küme Akış Çizgilerinin Olasılıksal Yapısı. *Avrupa Bilim ve Teknoloji Dergisi*, (23), 730-741.

Öz

Tanecikli Hesaplama yöntemlerinden biri olan Kaba Kümeler Teorisi 1980'li yılların başlarında Zdzislaw Pawlak tarafından, belirsiz ve muğlak bilgi ile uğraşmak için ortaya atılmıştır. Belirsizliğin en temel nedenlerinden biri, belirsizliğin ilgili olduğu alanın tüm değişkenlerini gözlemlemenin güçlüğüdür. Ayrıca, gözlemlenebilen değişkenlerin ait oldukları dünya deterministik olmasına rağmen, rastgele davranış gösterir. Kaba kümeler teorisi evrende her bir nesneden bilgi elde edebileceğimiz varsayımı üzerine kuruludur. Teorinin ortaya atılmasından günümüze kadar geçen süre içerisinde kaba kümeler teorisi üzerinde birçok çalışma yapılmıştır. Bunlardan biri de, 2000'li yılların başlarında kuramı ortaya atan Pawlak tarafından sunulan akış çizgileridir. Akış çizgeleri; verilerden akıl yürütmek ve akıllı veri analizi için veri tablolarından bilgi akışı dağılımını temsil etmek amacıyla grafiksel bir çerçeve sunmaktadır. Pawlak, akış çizgelerini Łukasiewicz tarafından önerilen olasılığı mantıksal terimlerle ifade etmeyi öneren kavramdan yola çıkarak açıklamıştır. Akış çizgeleri teorik bakış açısından, Łukasiewicz'in fikirlerinin bir genellemesi olarak görülebilir. Kaba küme teorisine dayalı akış çizgeleri, belirsiz ve eksik bilgi ile ilgilenen diğer yöntemler ile de örtüşmektedir. Bunlardan birisi de karmaşık alanlarda belirsizliği yönetmek için anlamsal bir modelleme aracı olarak kullanılan Bayes ağlarıdır. Kaba küme teorisi tarafından sunulan Bayes teoremine bakış, herhangi bir veri kümesinin toplam olasılık kuralı ve Bayes teoremini karşıladığını ortaya koymaktadır. Bayes teoremi, yeni bir kanıtın varlığında o ana kadar olan inançlarımızı nasıl değiştirmemiz gerektiğini açıklayan matematiksel bir kuraldır. Diğer bir deyişle, yeni bilgiler ile hali hazırda bulunan verilerin ve bilgilerin birleştirilmesini sağlar. Bu nedenle, akış çizgelerini Bayes ağlarının özel bir durumu olarak görebilmek mümkündür. Ayrıca, akış çizgeleri, karar tablolarından elde edilen kuralları Yönlendirilmiş Çevrimsiz Çizge (YDC) olarak düzenleyerek geleneksel kaba küme araştırmasını genişletir. Pawlak'ın akış çizgeleri, bilgi akışını görselleştirme yetenekleri nedeniyle hem pratik hem de teorik birçok araştırmacının ilgisini çekmiş ve birçok alanda başarı ile uygulanmıştır. Bu çalışmada, akış çizgelerinin temel kavramları ve özellikleri incelenmiş; akış çizgelerinin Bayes teoremi ve Bayes ağları ile ilişkisi gösterilmiştir. Ayrıca, akış çizgeleri ile ilgili geniş bir literatür araştırması yapılmış ve ilgili alanda yer alan uygulamalar ile teorik çalışmalara değinilmiştir. Çalışmanın son bölümünde, bir uygulama üzerinde, karar algoritması, sonlu bir "Eğer....O halde...." şeklinde karar kuralları kümesi olarak ifade edilmiştir. Ayrıca, bu karar kurallarının taşıdığı anlamlar Bayes teoremini sağlayan güç, kesinlik ve kapsama katsayıları ile ifade edilmiş ve değerlendirilmiştir. Burada her karar kurallarının olasılıksal özellikleri ortaya çıkardığı ve Bayes teoremi ile toplam olasılık kuralını karşıladığı görülmektedir.

Anahtar Kelimeler: Akış Çizgeleri, Kaba Kümeler, Bayes Teoremi, Bayes Ağları, Belirsizlik, Karar Kuralları, Karar Algoritmaları.

Probabilistic Structure of Rough Set Flow Graphs

Abstract

The Rough Set Theory, one of the granular computing methods, was put forward by Zdzislaw Pawlak in the early 1980s to deal with uncertain and vague information. One of the main reasons for uncertainty is that the uncertainty stems from the difficulty of observing all the variables of the domain it relates to. Also, although the world to which the observable variables belong is deterministic, it shows random behavior. The rough set theory is based on the assumption that we can obtain information from every single object in

* Sorumlu Yazar: mertbal@yildiz.edu.tr

the universe. Many studies have been conducted on the theory of rough sets in the period that passed since the theory was introduced. One of these is the flow charts presented by Pawlak, who introduced the theory in the early 2000s. Flow charts provide a graphical framework for reasoning from data and representing the distribution of information flow from data tables for intelligent data analysis. Pawlak explained the flow graphs from the concept proposed by Łukasiewicz that proposes to express probability in logical terms. Flow graphs can be viewed from a theoretical point of view as a generalization of the Łukasiewicz 'ideas. Flow graphs based on rough set theory overlap with other methods that deal with uncertain and incomplete information. One of these is Bayesian networks used as a semantic modeling tool to manage uncertainty in complex domains. The look at Bayes 'theorem presented by the rough set theory reveals that any data set meets the total probability rule and Bayes' theorem. Bayes' Theorem is a mathematical rule that explains how we should change our beliefs up to that point in the presence of new evidence. In other words, it ensures that new information is combined with already existing data and knowledge. Therefore, we can see flow graphs as a special case of Bayesian networks. In addition, flow graphs extend the traditional rough set research by organizing the rules derived from decision tables into Directed Acyclic Graphs (DAGs). Pawlak's flow graphs have attracted the attention of many practical and theoretical researchers due to their ability to visualize the flow of information and have been successfully applied in many areas. In this study, basic concepts and properties of flow charts are examined; the relationship of flow graphs with Bayes' theorem and Bayes networks is shown. In addition, a wide literature research on flow charts has been made and the applications and theoretical studies in the related field are mentioned. In the last section of the paper, the decision algorithm on an application is expressed as a finite set of decision rules in the form of "If...Then...". In addition, the meanings of these decision rules are expressed and evaluated with the strength, certainty and coverage coefficients that provide Bayes' theorem. It is seen here that each decision rules reveal probabilistic properties and meet Bayes' theorem and the total probability rule.

Keywords: Flow Graph, Rough Sets, Bayes Theorem, Bayesian Networks, Uncertainty, Decision Rules, Decision Algorithms.

1. Giriş

Kaba küme teorisi, muğlak, eksik veri analizi ve karmaşık alanlardaki belirsizliği modellemek için yeni bir matematiksel araç olarak Zdzislaw Pawlak (1982) tarafından tanıtılmıştır. Bu teorinin felsefesi, araştırma evreninin her nesnesiyle bazı bilgilerin ilişkilendirildiği varsayımına dayanır. Teori, karar desteği, mühendislik, bilgisayar bilimleri, bankacılık, tıp ve birçok alanda uygulama bulmuştur. Ayrıca, kaba küme teorisi, eksik ve belirsiz bilgiyle ilgilenen kanıt teorisi bulanık kümeler, Bayes ağları gibi birçok teori ile de örtüşmektedir. Örneğin, kaba küme teorisi, Bayes'in teoremine yeni bir bakış açısı sunar. Kaba küme teorisi tarafından sunulan Bayes teoremine bakış, herhangi bir veri kümesinin (karar tablosu) toplam olasılık kuralını ve Bayes teoremini karşıladığını ortaya koymaktadır. Bu özellikler, öznel ön bilgilere ve yeni kanıtlar mevcutsa revizyonuna başvurmadan nesnel verilerden sonuç çıkarmak için doğrudan kullanılabilir. Bu nedenle, Bayes'in teoremine ilişkin kaba küme görüşü, teoremin öznel "klasik" yorumunun aksine, oldukça nesneldir (Pawlak, 2002a).

Bayes ağları (Pearl,1988), karmaşık alanlarda belirsizlik ile baş etmek için anlamsal bir modelleme aracıdır. Bir Bayes ağı, Yönlendirilmiş Çevrimsiz Çizgeden (Directed Acyclic Graph-DAG) ve bu çizgeye karşılık gelen bir koşullu olasılık tabloları kümesinden oluşur. Yönlendirilmiş çevrimsiz bir çizge' de kodlanan olasılıksal koşullu bağımsızlıklar (Wong vd., 2000), koşullu olasılık tablolarının çarpımlarının tek bir ortak olasılık dağılımı olduğunu gösterir. Cooper (1990) çıkarımın karmaşıklığının NP-Zor olduğunu göstermesine rağmen, literatürde pratikte oldukça iyi sonuç veren birkaç yaklaşım geliştirilmiştir. Kaba kümeler ve Bayes ağlarının en önemli farklılıklarından birisi, kaba kümelerde Bayes ağlarından farklı olarak, söz konusu problem alanıyla ilgili hiçbir varsayımda

bulunulmaz. Bunun yerine, çıkarım süreci yalnızca örnek veriler tarafından yönetilir (Butz vd., 2005).

2002 yılında Pawlak kaba küme teorisinin önemli bir uzantısı olan ve verilerden akıl yürütmek için bir grafik çerçeve olarak matematiksel akış çizgelerini sunmuştur (Pawlak 2002b & Pawlak, 2003a). Akış çizgeleri, optimal akış analizi için Ford ve Fulkerson (1962) tarafından önerilen yöntemden farklıdır ve optimal akıştan ziyade, karar algoritmalarında bilgi akışı analizi için matematiksel bir araç olarak kullanılmaktadır.

Bir akış çizgesi, ağdaki akış dağılımını modeller ve verilen veri setinden bilgi akışının matematiksel modelini temsil eder. Akış çizgesinin dalları, karar kuralları olarak yorumlanırken, tüm akış çizgesi karar algoritmasının bir temsili olarak yorumlanabilir. Akış çizgesindeki bilgi akışı, Bayes'in kuralını sağlar, ancak kural olasılıksal bir anlama sahip değildir ve tamamen deterministiktir. Yani, Łukasiewicz'in öngördüğü olasılık ve doğruluk değerlerini kullanmak yerine, akış çizgelerinde deterministik akış analizi kullanılır. Bu özellik olasılıksal yapısına ve mantıksal yoruma atıfta bulunmadan, verilerden sonuç çıkarmak için kullanılabilir. Ayrıca, akış çizgesinin dalları arasındaki akış dağılım oranları olarak yorumlanabilmesini sağlar (Pawlak, 2003a & Pawlak,2003b). Teorik bakış açısından, akış çizgeleri, olasılığı mantıksal terimlerle ilk kez ifade etmeyi öneren Łukasiewicz 'in fikirlerinin bir genellemesi olarak görülebilir. Olasılığın önermesel fonksiyonların bir özelliği olduğunu ve [0,1] aralığına ait doğruluk değerleriyle değiştirilebileceğini iddia eder. Akış çizgesinde doğruluk değerleri ve dolayısıyla olasılıklar, bir akış çizgesinin dallarındaki akış yoğunluğu olarak yorumlanır (Pawlak, 2003a). Kaba küme teorisindeki akış çizgeleri, sezgisel ve açık biçimlendirme, basit hesaplama, paralel işleme ve Markov özelliği sergiler (Yao & Miao, 2019).

Bir Karar Algoritması, sonlu bir "Eğer....O halde...." karar kuralları kümesidir. Her karar kuralı, Bayes'in teoremini

sağlayan güç, kesinlik ve kapsama olmak üzere üç temel katsayı ile ilişkilidir. Kapsama katsayısı belirli bir karar için nedenlerin koşullu olasılığını verir. Bu nedenle akış çizgeleri, Bayesci (Bayesian) çıkarım metodolojisine yeni bir bakış açısı sağlar. Ayrıca, akış çizgeleri Bayes ağlarının özel bir durumudur (Greco, vd., 2002) ve karar tablolarından elde edilen kuralları yönlendirilmiş çevrimsiz çizge olarak düzenleyerek geleneksel kaba küme araştırmasını genişletir (Butz vd., 2005 & Butz vd., 2006). Pawlak'ın akış çizgeleri, bilgi akışını görselleştirme yetenekleri nedeniyle hem pratik hem de teorik araştırmacıların ilgisini çekmiş ve bilgi keşfi, bilgi temsili, veri analizi ve veri madenciliği gibi birçok alanda başarıyla uygulanmıştır (Chitcharoen & Pattaraintakorn, 2010).

Bu çalışmada, akış çizgelerinin temel kavramları ve özellikleri incelenmiş ve Bayes ağları ile akış çizgelerinin ilişkisi gösterilmiştir. Ayrıca, akış çizgeleri ile ilgili geniş bir literatür araştırması yapılmış ve uygulamada ve teorik yapılan çalışmalara değinilmiştir. Son bölümde, bir uygulama üzerinde, karar algoritması, sonlu bir karar kuralları kümesi olarak ve Bayes'in teoremini sağlayan güç, kesinlik ve kapsama katsayıları ile ifade edilmiştir.

2. Literatür Özeti

Bu bölümde kaba kümeye dayalı akış çizgeleri ile ilgili hem teorik hem de uygulamaya yönelik literatürde yer alan çalışmalara yer verilecektir.

Pawlak (2002a & 2002b) çalışmasında, her karar algoritmasının, olasılık özelliklerini ortaya çıkardığı, özellikle toplam olasılık kuralı ve Bayes teoremini karşıladığını göstermiştir. Bu özellikler, Bayesci muhakeme ile doğası gereği ilişkilendirilen, önsel ve sonsal olasılıklara atıfta bulunmadan, verilerden sonuç çıkarmak için yeni bir yöntem sunar. Greco vd. (2002), karar kurallarının karşılıklı dışlanması kavramını gevşetmiş ve karar algoritmalarının özelliklerinin hala geçerli olduğunu göstermiştir. Ayrıca, karar kuralları kesinlik faktörünün bir derecesinde doğru olduğunda, toplam olasılık kuralının modus ponens ve modus tollens çıkarım kuralları ile nasıl ilişkili olduğunu ve gevşetilmiş koşullar altında, her karar algoritmasıyla bir akış grafiğinin ilişkilendirilebileceğini ve karar algoritmalarının yararlı bir yorumlanmasını sağladığını göstermişlerdir (Chitcharoen & Pattaraintakorn, 2010). Pawlak (2003b), akış çizgeleri ve olasılık arasındaki ilişkiyi ortaya çıkarmıştır. Ayrıca, modus ponens ve modus tollens kavramlarından, doğruluk değerlerini karşılık gelen olasılıklarla değiştirmiş ve kaba modus ponens ve kaba modus tollens elde etmiştir. Pawlak (2003c), Bayes teoremi ile kaba kümeler arasındaki bazı ilişkileri göstermiştir. Ayrıca bu çalışmada, her karar algoritması ile bir akış çizgesinin ilişkilendirilebileceğini göstermiştir. Kostek ve Czyzewski (2004a & 2004b), müzikal veritabanlarında bulunan meta verilerin akıllı analizine dayalı olarak müziğe erişimin sağlanmasına yönelik bir yöntem geliştirmişler ve akış çizgelerinden çıkarım kuralı türetme kavramından yararlanarak gelişmiş bir sorgu algoritması sunmuşlardır.

Pawlak (2005a) çalışmasında, akış çizgesinin hem kaba küme teorisi anlamında kümelerin yaklaşımlarını hesaplamak için biçimsel dil hem de veri yapısı için açıklama aracı olarak kullanılabileceği göstermiştir. Pawlak (2005b), bir akış çizgesindeki bilgi akışı dağılımına dayalı olarak veri madenciliği ve bilgi keşfi için yeni bir yaklaşım önermiş ve bir ağdaki akış dağılımını modellemiştir. Ayrıca, karar algoritmalarını basitleştirmenin yanında, akış çizgesinden ve koşullar ile karar kurallarının kararları arasındaki bağımlılık ile ilgili bir karar algoritması tanıtmıştır. 2006 yılında Butz vd., geleneksel kaba küme akış çizgesi çıkarım algoritmasının üstel zaman karmaşıklığına sahip olduğunu göstermiş ve geleneksel kaba küme akış çizgesinde çarpanlara ayırmayı kullanan yeni bir kaba küme akış çizgesi çıkarım algoritması önermiştir.

Pattaraintakorn vd. (2006), sıralı tahmin için kaba kümeler kural öğrenimi, kuralların kaba çizge gösterimine dayanan bir çalışma ortaya koymuşlardır. Bu çalışmada, melanom verileri üzerinde bir vaka çalışması yapılmıştır. Sonuçlar, bu yenilikçi sistemin hem kuralları bulmak için hesaplama performansında hem de türetilen kuralların kullanılabilirliğinde kural öğrenmede bir gelişme sağladığını göstermiştir. Pawlak'ın akış çizgesi nicel bir çizge olduğundan, yani akış miktarını kullanarak düğümler arasındaki ilişkileri temsil ettiğinden, karar sistemlerinin özelliklerini tam olarak açıklayamaz. Sun vd. (2006) bu sorunu çözmek için akış çizgelerinin bir uzantısını geliştirmiştir. Bulanık akış çizgeleri kavramı, 2006 yılında Rolka ve Rolka tarafından tanıtılmıştır. Bu çalışmada, karar tablolarının bulanık özniteliklerle temsiline izin vermek için bulanık akış çizgeleri kavramı önerilmiştir (Chitcharoen & Pattaraintakorn, 2010). Karar ağaçları ve akış çizgeleri arasındaki ilişki Pawlak (2006) tarafından incelenmiştir. Bu çalışmada önerilen akış çizgesi, veri yapısında ilgili karar ağacından daha iyi kavrama sağlar ve doğrudan ağaçtan görülemeyen, karar ağaçlarının çok ilginç yeni özelliklerini ortaya çıkarır. Birçok şekilde kullanılabilirler ve özellikle verilerden verimli bir karar kuralları oluşturulmasını sağlarlar.

Suraj ve Pancercz (2007), zamansal bilgi sisteminin ardışık zaman pencerelerinde kural madenciliği için bir araç olarak akış çizgelerini kullanmıştır. Elde edilen kurallar, analiz edilen sistemin gelecekteki davranışını tahmin ederek, sistemdeki bileşenlerin gelecekteki değişikliklerini ve ardından gelenleri tahmin etmek için kullanılır. Chien ve Tsumoto (2007) çalışmasında, Çoklu küme (Multiset) Karar Tabloları kavramını kullanarak karar kuralları ve akış çizgeleri arasındaki ilişkiyi incelemişlerdir. Bu çalışma ile, çoklu küme karar tablosunun akış çizgelerinin minimal bir temsiline sağladığını ve tutarsız verilerden karar kurallarının öğrenilmesini kolaylaştırdığı gösterilmiştir. Kurallar, Pawlak tarafından sunulan Bayes faktörleriyle karakterize edilir (Chitcharoen & Pattaraintakorn, 2010).

Pal ve Chakraborty (2007) makalesinde, video gözetleme için, tanecikli hesaplama çerçevesinde yeni bir uyarlanabilir kural oluşturma yöntemi, kaba kural tabanı ve ayrıntılı akış

çizgesine dayalı olarak yeni bir kavram ve yüksek performanslı bir yöntem tanıtmışlardır. Bu çalışmada, akış çizgesi hem eğitim hem de doğrulama görevlerinde uygulanabilir kılmak için iki yeni özellik, yani beklenen akış çizgesi ve akış çizgeleri arasındaki karşılıklı bağımlılık tanımlanmıştır. Tüm bu teknikler, komşu tanecikler seviyesinde gerçekleştirilir. Denetimsiz video gözetleme için üretilen kaba akış çizgesi tabanlı uyarlanabilir tanecikli kural tabanlı sistem, çerçevelerdeki belirsizlikleri ve eksiklikleri yönetebilir, ilk manuel etkileşimler olmadan ortaya çıkan bilgilerdeki eksikliklerin üstesinden gelebilir ve üstün performans ve hesaplama süresinde kazanç sağlar.

Sun vd., (2008) tanecikli hesaplama kullanarak genişletilmiş bir akış çizgesi için hem basit hem de somut bir model geliştirmişler ve akış çizgeleri ile tanecikli hesaplama arasındaki ilişkiyi incelemişlerdir. Matusiewicz ve Pancercz (2008), durum tahmin problemlerinde akış çizgelerini ve bulanık ilişki denklemlerinin maksimum çarpım bileşimini birleştirmişler ve karar tablolarının bulanık özniteliklerle temsiline izin vermek için bulanık akış çizgeleri kavramını önermişlerdir.

Chitchareon ve Pattaraintakorn (2008a), akış çizgeleri ve birliktelik kuralları arasındaki ilişkiyi ortaya çıkarmak için iki yeni önerme sunmuşlar ve gerçek dünya verileri üzerinde deneyler gerçekleştirmişlerdir. Chitchareon ve Pattaraintakorn (2008b), çeşitli koşullar altında bulanık akış çizgelerinin kesinlik, kapsama ve güç katsayılarını hesaplamak için yeni bir yöntem geliştirmişlerdir. Ayrıca, karar algoritmasının karşılıklı dışlama özelliğinin gevşetilmesini önermişler ve kesinlik ve kapsama katsayılarını hesaplamışlardır. Pattaraintakorn (2009), Pawlak'ın matematiksel akış çizgesinde bir entropi ölçüsü tanıtmıştır. Bu çalışma ile özellikle, akış çizgelerindeki entropi ölçümlerinin, verilerden yeni bir akıl yürütme metodolojisine yol açtığı ve akış çizgeleri, entropi ve karar ağaçları arasındaki sıkı ilişkiler gösterilmiştir.

Chitchareon ve Pattaraintakorn (2010), çalışmada bir akış grafiğinin akışını ve matris biçiminde hem normalleştirilmiş hem de birleştirilmiş normalleştirilmiş akış çizgelerinin güç, kesinlik ve kapsama katsayılarını temsil etmek için yeni bir yöntem önermiştir. Matris formunun hesaplamayı basitleştirdiği ve zaman karmaşıklığını iyileştirdiği gösterilmiştir.

Akış çizgeleri, video gözetleme sistemlerinde özellikle de, sınırlı algılamaları nedeniyle insan operatörler tarafından ele alınması sorunlu olan dağıtılmış çoklu kamera sistemlerinde kullanılabilir. Bu tür sistemlerde otomatik video analizinin

3. Akış Çizgeleri ve Bayes Ağları

Bu bölümde Bayes teoremi ile Kaba kümeler teorisi arasında bir ilişkinin varlığı gösterilecektir. Herhangi bir karar algoritması, Bayes teoremini klasik Bayes metodolojisi ile birleştirilen önsel veya sonsal olasılıkları kullanmadan sağlar. Bunun yanında, her karar algoritması ile bir akış çizgesinin

uygulanması gerekir. Bu analizin önemli bir parçası, tek bir kamera içinde ve kameraların görüş alanları arasında nesneyi takip etmektir. Ayrıca, akış çizgeleri belirli bir nesnenin gelecekteki hareketini tahmin etmek için de kullanılabilir. Lisowski ve Czyzewski (2015) çalışmada, gözetlenen alandaki nesnelerin yollarına dayalı akış çizgesi oluşturma, akış çizgesi temelinde bir olasılık ağacı inşa etmenin bir yöntemini ve akış çizgesini kamera ağının değişen topolojisine uyarlamak için bir yöntem sunmuşlardır. Kaur vd. (2016), akış çizgeleri kavramını, algısal nesnelerin ayrık kümeler arasındaki yakınlığı belirlemede çok önemli bir rol oynadıkları, algısal sistemlere uygulamış ve İçerik Tabanlı Görüntü Erişimi (Content Based Image Retrieval-CBIR) gerçekleştirmek için dijital görüntüleri göre tanımlanan bir algısal sistem tarafından oluşturulan akış çizgelerinin pratik bir uygulamasını sunmuşlardır. Liou vd. (2016) çalışmalarında, havayolu müşterilerinden oluşan geniş bir örneklemeden, hizmet kalitesine ilişkin müşteri tutumlarını belirlemek, hizmet kalitesini iyileştirmek ve böylece müşterilerin ihtiyaçlarını karşılamak için akış çizgeleri yaklaşımıyla bir dizi kural türetmişler ve akış çizgelerinin tahmin yeteneğini değerlendirmişlerdir. Ayrıca, çalışmalarında Kaba Küme Teorisi modeli ve akış çizgelerinin kombinasyonunun belirlenmesine ve bir iyileştirme stratejisinin geliştirilmesine yardımcı olduğunu göstermişlerdir. Takács ve Csajbók (2017) çalışmada, kaba küme akış çizgeleri ile istatistiksel olasılık tabloları arasındaki bağlantıyı göstermişlerdir. Lewicki ve Eberbach (2018) çalışmalarında, çok amaçlı optimizasyon probleminde yaklaşık kümeleri bulmak için karınca kolonilerinin değiştirilmiş bir kolektif davranış stratejisinin kullanımını sunmuşlardır. Ayrıca, dağıtık bilişsel ortamlarda yapay karıncaların iletişim stratejisini kontrol etmek için kaba küme akış çizgelerinin kullanımına dayalı yeni bir yaklaşım önermişlerdir. Yao ve Miao (2019) çalışmalarında, akış çizgelerinin doğrudan öznitelik-değer tabloları biçiminde verilerden çıkarılmasına ve bu tür tablolardaki değişkenler arasındaki nedenselliğin tanımlanmasına odaklanmışlar; ayrıca nedenselliğin varlığı ile değişkenler arası akış çizgesi arasında ilişki kurmuşlardır. Söz konusu çalışmada Yao ve Miao (2019) denklik sınıfları ve doğrudan veri tablolarından türetilen bölümler kullanılmak suretiyle Markov özelliğine sahip değişkenlerin, akış çizgelerine entegre edilebilecek yapıları oluşturmak amacıyla kurulabileceğine vurgu yapmışlardır. Daha sonra, bu yapılarla dayanarak veri tablolarında gizlenen nedensellik, Pearl (2009) tarafından önerilen kriterler aracılığıyla belirlenebilir.

birleştirilmesi mümkündür. Bayes teorisi çizgedeki bilgi akışı ile yakından ilişkilidir. Ayrıca, her akış çizgesi, Öklidyen (Euclidean) uzayı ile ifade edilebilir. Burada, Öklidyen uzayının noktaları karar algoritması tarafından belirlenen kararları temsil ederken noktalar arasındaki mesafe ise, karar algoritmasındaki kararlar arasındaki mesafeyi gösterir (Pawlak, 2003c). Koşul ve karar özelliklerinin birbirinden ayırt edildiği bilgi tablosu karar tablosu olarak adlandırılır. Karar tablosundan

“Eğer.....O halde.....” biçiminde karar kuralları şeklinde bazı örüntüler çıkarılabilir (Greco vd., 2004).

Karar algoritması, “Eğer.....O halde.....” karar kurallarının sonlu bir kümesidir. Karar kuralları örüntüleri üç katsayı ile ifade edilir. Bu katsayılar, bir kuralın gücü, kesinliği ve kapsama faktörleridir. Bahsi geçen katsayılar veriden hesaplanabilir veya öznel bir değerlendirme ile belirlenebilir. Güç, kesinlik ve kapsama katsayıları Bayes teoremini sağlar. Bayes çıkarım metodolojisi önsel olasılıkları güncellemeden oluşur ve bunları da sonsal olasılıklar yoluyla gerçekleştirir. Bunun anlamı da, verinin mevcut olması durumunda güncellenmiş bilgiyi ifade etmesidir. Güç, kesinlik ve kapsama faktörleri, sınır hattındaki olasılıklar veya gerçeklik derecesi olarak yorumlanabilir. Ayrıca bunlar karar algoritmaları ile birleştirilmiş olarak akış çizgelerindeki bir deterministik akış dağılımı olarak yorumlanabilir.

3.1. Akış Çizgeleri

Bir akış çizgesi, yönlü, çevrimsiz, sonlu bir $G = (N, \mathfrak{B}, \varphi)$ çizgesi'dir. Burada, N düğümlerin kümesi $\mathfrak{B} \subseteq N \times N$ yönlü dalların kümesi, $\varphi: \mathfrak{B} \rightarrow R^+$ bir akış fonksiyonu ve R^+ pozitif gerçel sayılar kümesidir. $(x, y) \in \mathfrak{B}$ ise o halde x , y 'nin giriş ve y 'de x 'in çıkışıdır. Eğer, $x \in N$ ise bu durumda $I(x)$, x 'in tüm girişlerinin ve $O(x)$ 'de x 'in tüm çıkışlarının kümesidir G çizgesinin giriş ve çıkışı aşağıdaki gibi tanımlanır.

$$I(G) = \{x \in N : I(x) = \emptyset\} \quad (1)$$

$$O(G) = \{x \in N : O(x) = \emptyset\} \quad (2)$$

G çizgesinin giriş ve çıkışları G 'nin dış düğümleridir, diğer düğümler ise G 'nin iç düğümleri olarak adlandırılır. Eğer, $(x, y) \in \mathfrak{B}$ ise o halde $\varphi(x, y)$ x 'den y 'ye bir akış yoludur. Varsayalım ki, her $(x, y) \in \mathfrak{B}$ için $\varphi(x, y) \neq 0$ olsun. G akış çizgesinin her x düğümü ile giriş ve çıkış akışını birleştirilelim bu durumda aşağıdaki denklemler elde edilir.

$$\varphi_+(x) = \sum_{y \in I(x)} \varphi(y, x) \quad (3)$$

$$\varphi_-(x) = \sum_{y \in O(x)} \varphi(x, y) \quad (4)$$

Benzer şekilde, tüm G akış çizgesi için aşağıdaki gibi giriş ve çıkış akışı tanımlanır.

$$\varphi_+(G) = \sum_{x \in I(G)} \varphi_-(x) \quad (5)$$

$$\varphi_-(G) = \sum_{x \in O(G)} \varphi_+(x) \quad (6)$$

Herhangi x içsel düğümü için, $\varphi_+(x) = \varphi_-(x) = \varphi(x)$ 'dir ve burada $\varphi(x)$, x düğümünün akış yoludur. Açık bir şekilde, $\varphi_+(G) = \varphi_-(G) = \varphi(G)$ olur ve burada $\varphi(G)$, G çizgesinin akış yoludur. Yukarıdaki formüller, akış korunum denklemleri olarak ele alınabilir. Bir normalize edilmiş akış çizgesi bir yönlü, çevrimsiz, sonlu bir $G = \{N, \mathfrak{B}, \sigma\}$ çizgesi'dir. Burada, N düğümler kümesi, $\mathfrak{B} \subseteq N \times N$ yönlü dalların bir kümesi ve $\sigma: \mathfrak{B} \rightarrow \langle 0,1 \rangle$, (x, y) 'nin normalize edilmiş akışıdır ve

$$\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)} \quad (7)$$

(7) denkleminin (x, y) 'nin gücüdür ve 0 ile 1 arasında değerler alacaktır ($0 \leq \sigma(x, y) \leq 1$). Dalın gücü, dal boyunca toplam akışın basitçe yüzdesini ifade etmektedir. G akış çizgesinin her x düğümü ile onun normalize edilmiş giriş ve çıkış akışı birleştirilebilir. Bu durum aşağıdaki gibi tanımlanır (Pawlak, 2004a & Pawlak, 2004b).

$$\sigma_+(x) = \frac{\varphi_+(x)}{\varphi(G)} = \sum_{y \in I(x)} \sigma(y, x) \quad (8)$$

$$\sigma_-(x) = \frac{\varphi_-(x)}{\varphi(G)} = \sum_{y \in O(x)} \sigma(x, y) \quad (9)$$

Açıkça görülebilir ki, her x içsel düğümü için $\sigma_+(x) = \sigma_-(x) = \sigma(x)$ 'dir. Burada, $\sigma(x)$ normalize edilmiş akış yoludur. Ayrıca,

$$\sigma_+(G) = \frac{\varphi_+(G)}{\varphi(G)} = \sum_{x \in I(G)} \sigma_-(x) \quad (10)$$

$$\sigma_-(G) = \frac{\varphi_-(G)}{\varphi(G)} = \sum_{x \in O(G)} \sigma_+(x) \quad (11)$$

(10) ve (11) denklemlerinden, $\sigma_+(G) = \sigma_-(G) = \sigma(G) = 1$ olacağı açıkça görülmektedir.

3.2. Kesinlik ve Kapsama Faktörleri

Bir G akış çizgesinin her (x, y) dalı ile kesinlik ve kapsama faktörleri birleştirilebilir. (x, y) 'nin kesinlik ve

kapsama faktörleri aşağıdaki (12) ve (13) denklemleri ile tanımlanır.

$$\text{cer}(x, y) = \frac{\sigma(x, y)}{\sigma(x)} \quad (12)$$

$$\text{cov}(x, y) = \frac{\sigma(x, y)}{\sigma(y)} \quad (13)$$

Burada sırasıyla, $\sigma(x) \neq 0$ ve $\sigma(y) \neq 0$ olacaktır. Yukarıda verilen tanımlardan aşağıdaki sonuçlar elde edilir.

$$\sum_{y \in O(x)} \text{cer}(x, y) = 1 \quad (14)$$

$$\sum_{x \in I(y)} \text{cov}(x, y) = 1 \quad (15)$$

$$\sigma(x) = \sum_{y \in O(x)} \text{cer}(x, y) \sigma(x) = \sum_{y \in O(x)} \sigma(x, y) \quad (16)$$

$$\sigma(y) = \sum_{x \in I(y)} \text{cov}(x, y) \sigma(y) = \sum_{x \in I(y)} \sigma(x, y) \quad (17)$$

$$\text{cer}(x, y) = \frac{\text{cov}(x, y) \sigma(y)}{\sigma(x)} \quad (18)$$

$$\text{cov}(x, y) = \frac{\text{cer}(x, y) \sigma(x)}{\sigma(y)} \quad (19)$$

Yukarıdaki özellikler olasılıksal bir niteliğe sahiptir. Örneğin, (16) ve (17) denklemleri toplam olasılık kuralı yapısına sahiptir ve (18) ve (19) formülleri ise Bayes kurallarıdır. Burada ki yaklaşımda bu özellikler deterministik bir yol ile tanımlanır ve ağdaki dallar boyunca akış dağılımını tanımlar.

3.3. Yollar ve Bağlantılar

x 'den y 'ye bir yönlü yol G çizgesinde $x \neq y$ olmak üzere her i için $(1 \leq i \leq n-1)$ $x_1 = x$, $x_n = y$ ve $(x_i, x_{i+1}) \in \mathfrak{B}$ gibi x_1, x_2, \dots, x_n ' e düğümler dizisidir. x 'den y 'ye bir yol $[x \dots y]$ ile gösterilir. $[x_1, x_2, \dots, x_n]$ yolunun gücü, kesinliği ve kapsamı sırasıyla aşağıdaki gibi tanımlanır.

$$\text{cer}[x_1, x_2, \dots, x_n] = \prod_{i=1}^{n-1} \text{cer}(x_i, x_{i+1}) \quad (20)$$

$$\text{cov}[x_1, x_2, \dots, x_n] = \prod_{i=1}^{n-1} \text{cov}(x_i, x_{i+1}) \quad (21)$$

$$\sigma[x \dots y] = \sigma(x) \text{cer}[x \dots y] = \sigma(y) \text{cov}[x \dots y] \quad (22)$$

G çizgesinde, x 'den y 'ye tüm yollar kümesi $\langle x, y \rangle$ ile gösterilir. Bu da G 'de x 'den y 'ye bir bağlantı olarak adlandırılır. Diğer bir ifadeyle, $\langle x, y \rangle$ bağlantısı x ve y düğümleri ile belirlenen G 'nin bir alt çizgesidir. Her $\langle x, y \rangle$ bağlantısı için güç, kesinlik ve kapsama faktörleri aşağıdaki gibi gösterilir.

$\langle x, y \rangle$ bağlantısının kesinliği,

$$\text{cer}\langle x, y \rangle = \sum_{[x \dots y] \in \langle x, y \rangle} \text{cer}[x \dots y] \quad (23)$$

$\langle x, y \rangle$ bağlantısının kapsamı,

$$\text{cov}\langle x, y \rangle = \sum_{[x \dots y] \in \langle x, y \rangle} \text{cov}[x \dots y] \quad (24)$$

$\langle x, y \rangle$ bağlantısının gücü,

$$\sigma\langle x, y \rangle = \sum_{[x \dots y] \in \langle x, y \rangle} \sigma[x \dots y] = \sigma(x) \text{cer}\langle x, y \rangle = \sigma(y) \text{cov}\langle x, y \rangle \quad (25)$$

$[x \dots y]$ bir yol olsun. Burada, x ve y sırasıyla G çizgesinin giriş ve çıkışıdır. Böyle bir yol tam olarak adlandırılır. x 'den y 'ye tüm yolların kümesi G çizgesinde x 'den y 'ye tam bağlantı olarak adlandırılacaktır. x ve y sırasıyla, G 'nin giriş ve çıkışı olsun. Eğer G 'deki her $\langle x, y \rangle$ tam bağlantısı için, $\sigma(x, y) = \sigma\langle x, y \rangle$ ile yer değiştirirsek $\text{cer}(x, y) = \text{cer}\langle x, y \rangle$ ve $\text{cov}(x, y) = \text{cov}\langle x, y \rangle$ olur. O halde, $\sigma(G) = \sigma(G')$ gibi yeni bir G' akış çizgesi elde ederiz. Yeni akış çizgesi, birleştirilmiş akış çizgesi olarak adlandırılır (Pawlak, 2004a & Pawlak, 2004b). Verilen bir akış çizgesi için birleştirilmiş akış çizgesi onun giriş ve çıkışları arasındaki ilişkiyi temsil eder.

4. Karar Algoritmaları

Bir karar kuralı, $\Phi \rightarrow \Psi$ biçiminde ifade edilir ve "Eğer Φ ise o halde Ψ " şeklinde okunur. Burada Φ ve Ψ kuralın koşul ve kararı olarak adlandırılan mantıksal formüllerdir. Bu mantıksal formüller çizgenin düğüm kümesini oluşturur. Mantıksal formüller, önermesel fonksiyonlar olarak ifade edilir ve her kenar (Φ, Ψ) karar kuralı olarak görülebilir. $|\Phi|$, U

evreninden tüm nesnel kümesini gösterir ve bunlar Φ özelliğine sahiptir. Eğer, $\Phi \rightarrow \Psi$ bir karar kuralı ise o zaman $\text{supp}(\Phi, \Psi) = \text{card}(\Phi \wedge \Psi)$ karar kuralının desteği olarak adlandırılır ve karar kuralı, bu kuralın gücü ($\sigma(\Phi, \Psi)$), kesinliği ($\text{cer}(\Phi, \Psi)$) ve kapsama ($\text{cov}(\Phi, \Psi)$) faktörleri ile karakterize edilir. Her $[\Phi_1, \Phi_2, \dots, \Phi_n]$ yolu, $\Phi_1 \rightarrow \Phi_2, \Phi_2 \rightarrow \Phi_3, \dots, \Phi_{n-1} \rightarrow \Phi_n$ karar kuralları dizisini oluşturur ve karar kurallarının bu dizisi tek bir karar kuralı $\Phi_1 \Phi_2 \dots \Phi_{n-1} \rightarrow \Phi_n$ olarak yorumlanır ve kısaca $\Phi^* \rightarrow \Phi_n$ olarak yazılır. Burada, $\Phi^* = \Phi_1 \wedge \Phi_2 \wedge \dots \wedge \Phi_{n-1}$ dir ve aşağıdaki gibi karakterize edilir.

$$\text{cer}(\Phi^*, \Phi_n) = \text{cer}[\Phi_1, \Phi_2, \dots, \Phi_n] \quad (26)$$

$$\text{cov}(\Phi^*, \Phi_n) = \text{cov}[\Phi_1, \Phi_2, \dots, \Phi_n] \quad (27)$$

$$\sigma(\Phi^*, \Phi_n) = \sigma(\Phi_1) \text{cer}[\Phi_1, \Phi_2, \dots, \Phi_n] = \sigma(\Phi_n) \text{cov}[\Phi_1, \Phi_2, \dots, \Phi_n] \quad (28)$$

Burada, $\sigma(\Phi)$ formülün doğruluk değeri ve $\sigma(\Phi, \Psi)$, $\Phi \rightarrow \Psi$ karar kuralının gücüdür. Benzer şekilde, her $\langle \Phi, \Psi \rangle$ bağlantısı tek karar kuralı $\Phi \rightarrow \Psi$ olarak yorumlanabilir ve aşağıdaki (29), (30) ve (31) bağıntıları ile gösterilir.

$$\text{cer}(\Phi, \Psi) = \text{cer}\langle \Phi, \Psi \rangle \quad (29)$$

$$\text{cov}(\Phi, \Psi) = \text{cov}\langle \Phi, \Psi \rangle \quad (30)$$

$$\sigma(\Phi, \Psi) = \sigma(\Phi) \text{cer}\langle \Phi, \Psi \rangle = \sigma(\Psi) \text{cov}\langle \Phi, \Psi \rangle \quad (31)$$

$[\Phi_1, \Phi_2, \dots, \Phi_n]$ bir yol olsun. Şöyle ki, Φ_1 bir giriş ve Φ_n G akış çizgesinin bir çıkışıdır. Böyle bir yol, uygun $\langle \Phi_1, \Phi_n \rangle$ bağlantısı tam olarak adlandırılır. Tüm karar kuralları $\Phi_{i_1} \Phi_{i_2} \dots \Phi_{i_{n-1}} \rightarrow \Phi_{i_n}$ tüm tam yollar $\Phi_{i_1} \Phi_{i_2} \dots \Phi_{i_n}$ ile ilişkilidir ve akış çizgesi tarafından çıkarılan karar algoritması olarak adlandırılır.

Bir karar kuralının gücü, kesinlik ve kapsama faktörleri aşağıdaki gibi tanımlanır.

Karar kuralının gücü,

$$\sigma(\Phi, \Psi) = \frac{\text{supp}(\Phi, \Psi)}{\text{card}(U)} \quad (32)$$

Karar kuralının kesinlik faktörü,

$$\text{cer}(\Phi, \Psi) = \frac{\text{supp}(\Phi, \Psi)}{\text{card}(\Phi)} \quad (33)$$

Karar kuralının kapsama faktörü,

$$\text{cov}(\Phi, \Psi) = \frac{\text{supp}(\Phi, \Psi)}{\text{card}(\Psi)} \quad (34)$$

Karar tablolarından “Eğer.....O halde.....” şeklinde karar kuralları örüntülerinin çıkarılabileceğine yukarıda değinilmişti. Örneğin, hasta örneklerinden tıbbi bilgi toplayan bir veri tablosunda belirtiler kümesi $S = \{S_1, S_2, \dots, S_n\}$ ’yi koşul özellikleri ve hastalıklar kümesi $D = \{d_1, d_2, \dots, d_m\}$ ’yi de karar özellikleri olarak ele alalım. Karar tablosunda karar kurallarını “Eğer $S_{i1}, S_{i2}, \dots, S_{in}$ belirtileri görülürse o halde d_j hastalığı vardır” biçiminde karar kurallarını çıkarabiliriz. Burada $S_{i1}, S_{i2}, \dots, S_{in} \in S$ ve $d_j \in D$ ’dir. Bir karar tablosundan çıkarılan her karar kuralı ile geleneksel olarak üç katsayı (güç, kesinlik ve kapsama) ilişkilidir. Örneğin, karar kuralı “Eğer $S_{i1}, S_{i2}, \dots, S_{in}$ belirtileri görülürse o halde d_j hastalığı vardır” şeklinde ise bu aşağıdaki şekilde karakterize edilebilir (Greco vd.,2004):

- $S_{i1}, S_{i2}, \dots, S_{in}$ belirtilerine sahip olan hastalar ve d_j hastalığı örnekteki tüm hastaların %15’ini oluşturur. Bu durumda kuralın gücü %15’dir.
- $S_{i1}, S_{i2}, \dots, S_{in}$ belirtilerine sahip olan hastaların %91’i ayrıca d_j hastalığına sahiptir. bu durumda kuralın kesinlik faktörü %91’dir.
- d_j hastalığına sahip hastaların %52’si ayrıca $S_{i1}, S_{i2}, \dots, S_{in}$ belirtilerine sahiptir. Bu durumda kuralın kapsama faktörü %52’dir.

Güç, kesinlik ve kapsama katsayıları, uzun zamandır veri tabanlarında ve makine öğreniminde kullanılmaktadır. Eğer $\text{cer}(\Phi, \Psi) = 1$ ise, $\Phi \rightarrow \Psi$ karar kuralı kesin olarak adlandırılır. Aksi takdirde karar kuralı, kesin olmayan şekilde adlandırılır. Karar kurallarının bir kümesi,

$$\text{Dec}(\Phi, \Psi) = \{\Phi_i \rightarrow \Psi_i\}_{i=1}^n \quad (35)$$

şeklindedir.

Eğer (35) de verilen ifadenin tüm karar kuralları aşağıdaki özellikleri sağlarsa $n \geq 2$ için bir karar algoritması olarak adlandırılır.

- $\text{supp}(\Phi_i, \Psi)_i \neq \emptyset$ (her $1 \leq i \leq n$ için)
- Karşılıklı olarak bağımsız, örneğin, her $\Phi_i \rightarrow \Psi_i$ ve $\Phi_j \rightarrow \Psi_j$ için, $\Phi_i = \Psi_j$ veya $\Phi_i \wedge \Psi_j = \emptyset$ ve $\Phi_j = \Psi_i$ veya $\Psi_i \wedge \Phi_j = \emptyset$.
- $|\bigvee_{i=1}^n \Phi_i| = |\bigvee_{i=1}^n \Psi_i| = U$.

Eğer, $\text{Dec}(\Phi, \Psi) = \{\Phi_i \rightarrow \Psi_i\}_{i=1}^n$ bir karar algoritması ise,

$$\text{Dec}(\Psi, \Phi) = \{\Psi_i \rightarrow \Phi_i\}_{i=1}^n \quad (36)$$

ifadesi $\text{Dec}(\Phi, \Psi)$ 'nin ters karar algoritması olarak adlandırılır.

4.1. Karar Algoritmalarının Özellikleri

$\text{Dec}(\Phi, \Psi)$ bir karar algoritması ve $\Phi \rightarrow \Psi$, karar algoritmasındaki bir karar kuralı olsun. $D(\Phi)$ ve $C(\Psi)$, Φ 'nin tüm kararların kümesini ve $\text{Dec}(\Phi, \Psi)$ 'deki Ψ 'nin tüm koşullarının kümesini gösterir (Pawlak, 2002c). Her karar algoritması aşağıdaki olasılıksal özelliklere sahiptir.

$$\sum_{\Phi' \in C(\Psi)} \text{cov}(\Phi', \Psi) = 1 \quad (37)$$

$$\sum_{\Psi' \in D(\Phi)} \text{cer}(\Phi, \Psi') = 1 \quad (38)$$

$$\pi(\Psi) = \sum_{\Phi' \in C(\Psi)} \text{cer}(\Phi', \Psi) \cdot \pi_s(\Phi') = \sum_{\Phi' \in C(\Psi)} \sigma(\Phi', \Psi) \quad (39)$$

$$\pi(\Phi) = \sum_{\Psi' \in D(\Phi)} \text{cov}(\Phi, \Psi') \cdot \pi_s(\Psi') = \sum_{\Psi' \in D(\Phi)} \sigma(\Phi, \Psi') \quad (40)$$

$$\text{cer}(\Phi, \Psi) = \frac{\text{cov}(\Phi, \Psi) \cdot \pi(\Psi)}{\sum_{\Psi' \in D(\Phi)} \sigma(\Phi, \Psi')} = \frac{\sigma(\Psi, \Phi)}{\pi(\Phi)} \quad (41)$$

$$\text{cov}(\Phi, \Psi) = \frac{\text{cer}(\Phi, \Psi) \cdot \pi(\Phi)}{\sum_{\Phi' \in C(\Psi)} \sigma(\Phi', \Psi)} = \frac{\sigma(\Phi, \Psi)}{\pi(\Psi)} \quad (42)$$

Burada, $\pi(\Psi)$ ve $\pi(\Phi)$ aşağıdaki gibi tanımlanır.

$$\pi(\Psi) = \frac{\text{card}(|\Psi|)}{\text{card}(U)} \quad (43)$$

$$\pi(\Phi) = \frac{\text{card}(|\Phi|)}{\text{card}(U)} \quad (44)$$

(37) ve (42) formülleri olasılıksal bir çerçeveye sahiptir. (39) ve (40) formülleri toplam olasılık formülleri olarak anlaşılabilir. (41) ve (42) formülleri ise Bayes kuramı biçimindedir. Daha öncede belirtildiği gibi eğer, güç, kesinlik ve kapsama faktörleri, olasılıklar olarak yorumlarsak o halde, bu formüller karar algoritmasındaki veri arasında bazı ilişkileri açıklar. Bununla beraber, bu faktörleri bir gerçeklik derecesi olarak deterministik bir yolla yorumlamak da mümkündür.

5. Akış Çizgelerindeki Bağımlılıklar

$(x, y) \in \mathfrak{B}$ olsun. $\sigma(x, y) = \sigma(x) \cdot \sigma(y)$ olduğunda x ve y düğümleri birbirinden bağımsızdır. Sonuç olarak,

$$\frac{\sigma(x, y)}{\sigma(x)} = \text{cer}(x, y) = \sigma(y) \quad (45)$$

$$\frac{\sigma(x, y)}{\sigma(y)} = \text{cov}(x, y) = \sigma(x) \quad (46)$$

$$\text{cer}(x, y) > \sigma(y) \quad (47)$$

$$\text{cov}(x, y) > \sigma(x) \quad (48)$$

yukarıdaki (45) - (48) ifadelerinin gerçekleşmesi durumunda x veya y 'nin birbirine pozitif olarak bağımlı olduğu söylenebilir.

Benzer şekilde, eğer

$$\text{cer}(x, y) < \sigma(y) \quad (49)$$

$$\text{cov}(x, y) < \sigma(x) \quad (50)$$

(49) ve (50) eşitsizlikleri gerçekleşmesi durumunda, x veya y birbirine negatif olarak bağımlıdır denir.

Bağımlılık ve bağımsızlık ilişkilerinin simetrik ve istatistikte kullanılan durumlara benzer olduğu söylenebilir.

Her $(x, y) \in \mathfrak{B}$ için, $\eta(x, y)$ bağımlılık faktörü aşağıdaki gibi tanımlanır (Pawlak, 2004c).

$$\eta(x, y) = \frac{\text{cer}(x, y) - \sigma(y)}{\text{cer}(x, y) + \sigma(y)} = \frac{\text{cov}(x, y) - \sigma(x)}{\text{cov}(x, y) + \sigma(x)} \quad (51)$$

Burada,

- (a) eğer, $\eta(x, y)=0$ ise x veya y birbirinden bağımsızdır,
 (b) eğer, $-1 < \eta(x, y) < 0$ ise, x veya y negatif olarak bağımlı ve
 (c) eğer, $0 < \eta(x, y) < 1$ ise x veya y birbirine pozitif olarak bağımlıdır.

$\dots < a_n$, yani, $A, <$ ile doğrusal olarak sıralansın. O halde, $AS = (U, \mathcal{R}, V)$ yaklaşım uzayını temsil eden bir $G(A, S) = (N, \mathfrak{B}, \varphi)$ akış grafiği oluşturulabilir, burada;

1. $N = \{n_x: X \in \mathcal{R}\};$
2. $n_x \mathfrak{B} n_y$ sadece ve sadece bazı $a_i \in A, a_j \in A, v \in V_{a_i}, v' \in V_{a_j}$ için

$X = \{u \in U: a_i(u) = v\}, Y = \{u \in U: a_j(u) = v'\}$ elde ederiz ve a_j, a_1, \dots, a_n 'nin doğrusal sırasına göre a_i 'nin doğrudan mirasçısıdır.

3. Herhangi bir $n_x, n_y \in N$ için:

- (a) $\varphi(n_x, n_y) = |X \cap Y|/|U|;$
- (b) $cer(n_x, n_y) = |X \cap Y|/|X|;$
- (c) $cov(n_x, n_y) = |X \cap Y|/|Y|.$

Bu nedenle, $G(A, S) = (N, \mathfrak{B}, \varphi)$ akış grafiği, A' dan alınan özneliklerin verilen $<$ sıralamasına göre $AS = (U, \mathcal{R}, \vartheta)$ yaklaşım uzayının bir görüntüsü olarak ele alınabilir. Bu tür görüntüler ve bunların birleşimleri (füzyonları), kavram (konsept) yaklaşımları için modellerin sonuç çıkarımında kullanılabilir (Pawlak, 2005).

8. Uygulama

Bu bölümde, akış çizgesi ve bu çizge üzerinde oluşturulan karar kurallarının Bayes teoremini sağlayan güç, kesinlik ve kapsama katsayıları ile değerlendirilmesine yönelik bir örnek uygulama verilecektir. Sigara içme ve Kanser hastalığı arasındaki ilişkinin gösterildiği örneği göz önünde bulunduralım. Bu örnekte veri analizinin olasılıklı yapısı gösterilmiş ve ayrıca, istatistiksel ve akış çizgelerine dayalı metodoloji arasındaki ilişki ortaya konmuştur. Bu bölümde Grinstead ve Snell (1997)'den alınan örnek üzerinde bir uygulama gösterilecektir. Tablo 1'de sigara içen ve içmeyen, kanser hastası olan ve olmayan 60 kişiye ait veriler gösterilmektedir.

Şekil 1'de, Tablo 1'de verilen verilere ilişkin akış çizgesi gösterilmektedir. Şekil 2'de, Şekil 1'de verilen akış çizgesi için normalleştirilmiş akış çizgesi sunulmuştur.

Tablo 1 Sigara içme ve Kanser Hastalığı

	Sigara İçmeyen	Sigara İçen	Toplam
Kanser Hastası Değil	40	10	50
Kanser Hastası	7	3	10
Toplam	47	13	60

6. Karar Uzayı

n değerli her karar algoritması ile n boyutlu Öklidyen uzayını birleştirmek mümkündür. Karar uzayı, bir karar algoritmasının kararlar arasındaki farklılıkları geometrik bir şekilde göstermek amacıyla kullanılmak üzere tasarlanmıştır. Burada, kararların değerleri çizgenin n apsisini belirler ve koşul özellik değerleri (denklik sınıfları) uzayın noktalarını belirler. Karar kurallarının güçleri, karşılık gelen noktaların koordinatları olarak görülmelidir. $\delta(x, y)$ mesafesi n boyutlu karar uzayında x ve y noktaları arasındaki mesafe

$$\delta(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (52)$$

şeklinde tanımlanır. Burada, $x = (x_1, x_2, \dots, x_n)$ ve $y = (y_1, y_2, \dots, y_n)$ karşılık gelen karar kurallarının vektörleridir (Pawlak, 2003c).

7. Akış Çizgeleri ve Kaba Kümeler

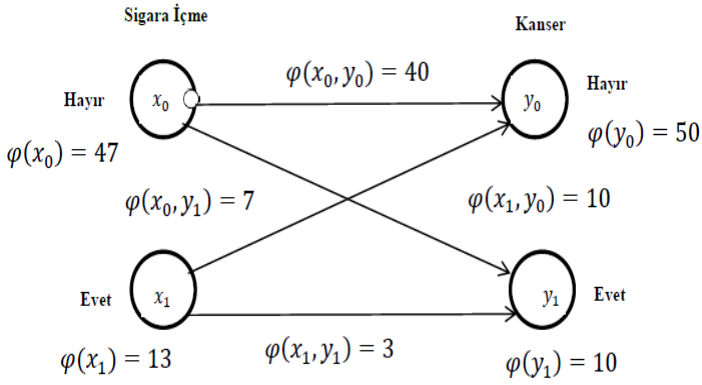
Bu bölümde akış çizgeleri ve kaba kümeler arasındaki ilişki, bazı akış çizgelerinin yaklaşım uzaylarının temsili olarak ele alınması şeklinde gösterilecektir. Bu durumu açıklamak amacıyla, bilgi sistemleri için yaklaşım uzaylarına dayalı bir örneği göz önüne alalım. $IS = (U, A)$ bir bilgi sistemi olsun. Burada, U nesnelere bir evreni ve $A, a: U \rightarrow V_a$ formunun öznelikleri kümesidir (Pawlak, 1991). Bu tür bir bilgi sistemi, $AS = (U, \mathcal{R}, \vartheta)$ yaklaşım uzayını tanımlar (Skowron & Stepaniuk, 1996). Burada, \mathcal{R}, A üzerindeki tanımlayıcılar tarafından oluşturulan bir kümeler ailesidir. Yani,

$$\mathcal{R} = \{X \subseteq U: X = \{u \in U: a(u) = v\} \text{ bazı } a \in A \text{ için, } v \in V_a\}$$

ve $\vartheta: P(U) \times P(U) \rightarrow [0, 1]$, aşağıdaki (53) ifadesi tarafından tanımlanan standart kaba dahil etme işlemidir.

$$\vartheta(X, Y) = \begin{cases} \frac{|X \cap Y|}{|X|} & \text{Eğer } X \neq \emptyset \\ 1 & \text{Eğer } X = \emptyset \end{cases} \quad (53)$$

Dolayısıyla, $\vartheta(X, Y)$, herhangi bir $X, Y \in \mathcal{R}$ için X 'in Y 'ye dahil edildiği bir derecedir. Varsayalım ki, $A = \{a_1, \dots, a_n\}$ ve $a_1 <$



Şekil 1. Tablo 1 için Akış Çizgesi

Akış çizgesinden aşağıda verilen sonuçları elde ederiz:

- Sigara içmeyenlerin %85’inde kanser hastalığı yok ($cer(x_0, y_0) = 40/47 \approx 0.85$),
- Sigara içmeyenlerin %15’i kanser hastası ($cer(x_0, y_1) = 7/47 \approx 0.15$),
- Sigara içenlerin %77’sinde kanser hastalığı yok ($cer(x_1, y_0) = 10/13 \approx 0.77$),
- Sigara içenlerin %23’ü kanser hastası ($cer(x_1, y_1) = 3/13 \approx 0.23$).

Akış çizgesinden kansere sahip olup olmamanın aşağıdaki nedenini elde ederiz:

- Kanser hastalığı olmayanların %80’i sigara içmiyor ($cov(x_0, y_0) = 4/5 = 0.80$),
- Kanser hastalığı olmayanların %20’si sigara içiyor ($cov(x_1, y_0) = 1/5 = 0.20$),
- Kanser hastalığı olanların kişilerin %70’i sigara içmiyor ($cov(x_0, y_1) = 7/10 = 0.70$),
- Kanser hastalığı olan kişilerin %30’u sigara içiyor ($cov(x_1, y_1) = 3/10 = 0.30$).

Bu sigara içmeyen kişilerin çoğunlukla kanser hastası olmadığı, ancak sigara içmenin çoğunlukla kanserle ilişkili olmadığı anlamına gelir. Ters akış çizgesinden kanser hastalığına yakalanmamanın nedeninin sigara içmek olmadığı, ancak kanser hastalığına yakalanmanın sigara ile ilişkili olmadığı sonucuna varılır (Pawlak, 2005).

Normalize edilmiş akış çizgesindeki karar kuralları ve bu karar kurallarına ait güç, kesinlik ve kapsama faktör katsayıları aşağıda verilmektedir:

Kural 1: $x_0 \rightarrow y_0$ ($cer = 0.85, \sigma = 0.67, cov = 0.80$)

(“Eğer kişi Sigara içmiyorsa o halde Kanser hastası değil”),

Kural 2: $x_0 \rightarrow y_1$ ($cer = 0.15, \sigma = 0.12, cov = 0.70$)

(“Eğer kişi Sigara içmiyorsa o halde Kanser hastasıdır”),

Kural 3: $x_1 \rightarrow y_0$ ($cer = 0.77, \sigma = 0.17, cov = 0.20$)

(“Eğer kişi Sigara içiyorsa o halde Kanser hastası değil”),

Kural 4: $x_1 \rightarrow y_1$ ($cer = 0.23, \sigma = 0.05, cov = 0.30$)

(“Eğer kişi Sigara içiyorsa o halde Kanser hastasıdır”).

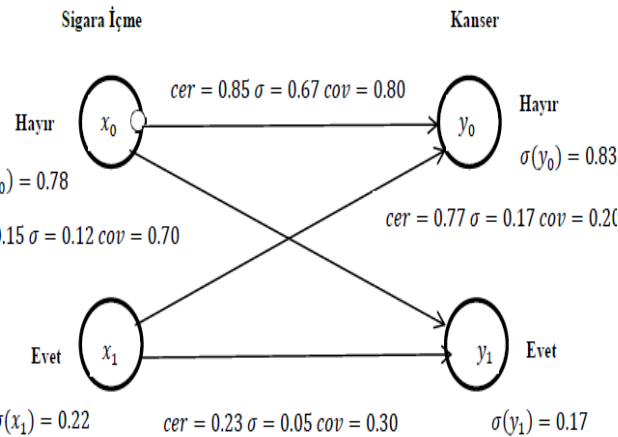
Akış çizgesi için aşağıdaki bağımlılıklar elde edilir:

$$\eta(x_0, y_0) = 0.01, \eta(x_0, y_1) = -0.06, \eta(x_1, y_0) = -0.09 \quad \text{ve} \\ \eta(x_1, y_1) = 0.15.$$

Bu, x_0 ve y_0 (Kural 1) arasında hafif pozitif bağımlılık ve x_1 ile y_1 (Kural 4) arasında çok daha güçlü pozitif bağımlılık olduğu anlamına gelir. Ayrıca, x_0 ile y_1 (Kural 2) arasında negatif ilişki ve x_1 ve y_0 (Kural 3) arasında da yine negatif bir ilişki söz konusudur. Örnek uygulamadaki sonuçlar istatistiksel açıdan değerlendirildiğinde, $\sigma(x_0), \sigma(x_1)$ önsel (prior) olasılıklara, $\sigma(x_0, y_0), \dots, \sigma(x_1, y_1)$ ortak olasılık dağılımlarına, $cov(x_0, y_0), \dots, cov(x_1, y_1)$ sonsal (posterior) olasılıklara ve $\sigma(y_0), \sigma(y_1)$ ‘in de marjinal olasılıklara tekabül ettiği görülmektedir.

9. Sonuç

Bu çalışmada, akış çizgelerinin temel kavramları ve özellikleri incelenmiştir. Bunun yanında, Bayes teoremi ve Bayes ağları ile akış çizgelerinin ilişkisi gösterilmiştir. Ayrıca, akış çizgeleri ile ilgili literatür araştırması yapılmış ve uygulamada ve teorik olarak yapılan çalışmalara değinilmiştir. Literatür araştırmasında da görüldüğü üzere akış çizgeleri birçok alanda başarılı bir şekilde uygulama alanı bulmuştur. Son bölümde, bir uygulama üzerinde, karar algoritması, sonlu bir “Eğer....O halde....” şeklinde karar kuralları kümesi olarak ifade edilmiştir. Ayrıca, bu karar kurallarının taşıdığı anlamlar Bayes teoremini sağlayan güç, kesinlik ve kapsama katsayıları ile ifade edilmiş ve değerlendirilmiştir. Burada, her karar kurallarının olasılıksal özellikleri ortaya çıkardığı ve Bayes teoremi ile toplam olasılık kuralını karşıladığı görülmektedir. Bu Pawlak’ın ilgili çalışmalarında da belirttiği; Bayes teoreminin önsel (prior) ve sonsal (posterior) olasılıklara gerek olmadan doğrudan verilerden mantık yürütmeye kullanılan Bayesci



Şekil 2. Tablo 1 için Normalize Edilmiş Akış Çizgesi

çıkartım metodolojisinin de yeni bir bakış açısına yol açmaktadır. Uygulamadaki sonuçlar istatistiksel olarak incelendiğinde, $\sigma(x_0), \sigma(x_1)$ önsel olasılıklara, $\sigma(x_0, y_0), \dots, \sigma(x_1, y_1)$ ortak olasılık dağılımlarına, $cov(x_0, y_0), \dots, cov(x_1, y_1)$ sonsal olasılıklara ve $\sigma(y_0), \sigma(y_1)$ 'in de marjinal olasılıklara karşılık gelmektedir.

Kaba küme teorisine dayalı akış çizgilerinin bilgi tabanlı karar desteği sağlamada çeşitli alanlarda yararlı olduğu ve bir dizi karar kuralıyla temsil edilen bilgi akışını modellemek için güçlü bir araç olduğu çalışmada gösterilmiştir. Akış çizgilerinin farklı alanlarda uygulamalarının gelecek çalışmalarda uygulamaları yapılabilir. Özellikle, yöneticilerin karar verme süreçlerinde yol gösterici ve karar desteği sağlayan akış çizgilerinin uygulandığı çalışmaların yapılması önerilmektedir.

Kaynakça

- Butz, C.J., Yan, W. & Yang, W., (2005). The Computational Complexity of Inference Using Rough Set Flow Graphs, Rough Sets Fuzzy Sets Data Mining and Granular Computing 2005, Lecture Notes in Artificial Intelligence, 3641, Springer Verlag, Berlin, 335 – 344.
- Butz, C.J., Yan, W. & Yang, W., (2006). An Efficient Algorithms for Inference in Rough Set Flow Graphs, Transaction on Rough Sets V, Lecture Notes in Computer Science, 4100, Springer Verlag, Berlin, 102-122.
- Chien, C.C. & Tsumoto, S., (2007). On Learning Decision Rules from Flow Graphs, In Proc. of North America Fuzzy Information Processing Society, 655 – 658.
- Chitchoreon, D. & Pattaraintakorn, P., (2008a). Knowledge Discovery by Rough Sets Mathematical Flow Graphs and its Extension, In Proc. of the 26th IAESTED Int. Conf. on AI and Applications AIA' 08, ACTA Press, 340 – 345.
- Chitchoreon, D. & Pattaraintakorn, P., (2008b). Towards Theories of Fuzzy Sets and Rough Set to Flow Graphs, In Proc. of the 2008 IEEE World Congress on Computational Intelligence, Hong Kong, 1675 – 1682.
- Chitchoreon, D. & Pattaraintakorn, P., (2010). Novel Matrix Forms of Rough Set Flow Graphs with Applicationsto Data Integration, Computers and Mathematics with Applications, 60, 2880-2897.
- Cooper, G.F., (1990). The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks, Artificial Intelligence, Vol.42 (2-3), 393 – 405.
- Czyzewski, A., & Kostek, B., (2004a). Musical Metadata Retrieval with Flow Graphs. In Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science, Springer Verlag, Berlin, 3066, 691 – 698.
- Ford, L.R. & Fulkerson, D.R., (1962). Flows in Networks, Princeton University Press. Princeton, New Jersey.
- Greco, S., Pawlak, Z. & Slowinski, R., (2004). Can Bayesian Confirmation Measures be Useful for Rough Set Decision Rules?, Engineering Applications of Artificial Intelligence, 17, 345 – 361.
- Grinstead, C.M. & Snell, J.L., (1997). Introduction to Probability, Second Revised Edition, American Mathematical Society.
- Kaur, K., Ramanna, S. & Henry, C., (2016). Measuring the Nearness of Layered Flow Graphs Application to Content Based Image Retrieval, Intelligent Decision Technology, 10, 165 – 181.
- Kostek, B. & Czyzewski, A., (2004b). Processing of Minimal Metadata Employing Pawlak's Flow Graphs, In Transactions on Rough Sets I, Vol.3100, Lecture Notes in Computer Sciences, Springer Verlag, Berlin, 279 – 298.
- Lewicki, A. & Eberbach, E., (2018). Learning Network Flow Based on Rough Set Flow Graphs and ACO Clustering in Distributed Cognitive Environments, Int.Workshop on Software Engineering for Cognitive Services (SE4COG), IEEE/ACM, 18 – 24.
- Liou, J.J.H, Chuang, Y.C. & Hsu, C.C., (2016). Improving Airline Service Quality Based on Rough Set Theory and Flow Graphs, Journal of Industrial and Production Engineering, Taylor & Francis, vol. 33, 123 – 133.
- Lisowski, K., & Czyzewski, A., (2015). Pawlak's Flow Graph Extensions for Video Surveillance Systems, Federated Conf. on Computer Science and Information Systems, 81 – 87.
- Matusiewicz, Z., & Pancerz, K., (2008). Rough Set Flow Graphs and max-* Fuzzy Relation Equations in State Prediction Problems, In Rough Sets and Current Trends in Computing in Lecture Notes in Artificial Intelligence 5306, Springer Verlag, Berlin, 359 – 368.
- Pal, S.K., & Chakraborty, D.B., (2017). Granular Flow Graph Adaptive Rule Generation and Tracking, IEEE Transactions on Cybernetics, vol.47, 4096 – 4107.
- Pattaraintakorn, P., (2009). Entropy Measures of Flow Graphs with Applications to Decision Trees, In Rough Sets and Knowledge Technology, Lecture Notes in Computer Sciences, 5589, Springer Verlag, Berlin, 618 – 625.
- Pattaraintakorn, P., Cercone, N. & Nauredomkul, K., (2006). Rule Learning: Ordinal Prediction based on Rough Set and Soft Computing, Applied Mathematics Letters, 19, 1300-1307.
- Pawlak, Z., (1982). Rough Sets, Int. Journal of Computer and Information Sciences, 11 (5) , 341 – 356.
- Pawlak, Z., (1991). Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer, Dordrecht.
- Pawlak, Z., (2002a). The Rough Set View on Bayes' Theorem, AFSS 2002, Lecture Notes in Artificial Intelligence, 2275, Springer Verlag, Berlin, 106 – 116.
- Pawlak, Z., (2002b). Rough Sets, Decision Algorithms and Bayes' Theorem, European Journal of Operational Research, 136, 181 – 189.
- Pawlak, Z., (2002c). Rough Set Theory, Journal of Telecommunications and Information Technology, Vol.3, 7 – 10.
- Pawlak, Z., (2003a). Flow Graphs and Decision Algorithms, The 9th Int. Conf. on RSDGrC 2003, Lecture Notes in Artificial Intelligence, 2639, Springer Verlag, Berlin, 1 – 10.
- Pawlak, Z., (2003b). Probability, Truth and Flow Graph, Electronic Notes in Theoretical Computer Science, Vol.82 (4), 1 – 9.
- Pawlak, Z., (2003c). Decision Algorithms and Flow Graphs: A Rough Set Approach, Journal of Telecommunications and Information Technology, 3, 98 – 101.
- Pawlak, Z., (2004a). Flow Graphs - A New Paradigm for Data Mining and Knowledge Discovery, The Proc. of. 5th International Symposium on Knowledge and Systems Science, JAIST, 147-153.
- Pawlak, Z., (2004b). Decision Rules and Flow Networks. European Journal of Operational Research, 154, 184 – 190.
- Pawlak, Z., (2004c). Data Analysis and Flow Graphs, Journal of Telecommunications and Information Technology, Vol.3, 1 – 5.

- Pawlak, Z., (2005a). Rough Sets and Flow Graphs, Int. Workshop on RSFDGrC, Lecture Notes in Artificial Intelligence, 3641, Springer Verlag, Berlin, 1 – 11.
- Pawlak, Z., (2005b). Flow Graphs and Data Mining, Transactions on Rough Sets III, Lecture Notes in Computer Sciences, 3400, Springer Verlag, Berlin, 1 – 36.
- Pawlak, Z., (2006). Decision Trees and Flow Graphs, , In Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 4259, Springer Verlag, Berlin, 268 – 277.
- Pearl, J., (2009). Causality: Models, Reasoning and Inference, Second Edition, Cambridge University Press, New York.
- Rolka, M. A., & Rolka, L., (2006). Flow Graph and Decision Tables with Fuzzy Attributes, In Proc. of the 8th Int. Conf. AI and Soft Computing, Zakopane, Lecture Notes in Artificial Intelligence, 4029, Springer Verlag, Berlin, 268 – 277.
- Skowron, A. & Stepaniuk, J., (1996). Tolerance Approximation Spaces, *Fundamenta Informaticae*, 27, 245 – 253.
- Sun, J., Liu, H. & Zhang, H., (2006). An Extension of Pawlak's Flow Graphs, In Proc. of the 1st Int. Conf. on Rough Sets and Knowledge Technology, 1191 – 199.
- Sun, J., Liu, H., Qi, C. & Zhang, H., (2008). Interpretation of Extended Pawlak's Flow Graphs Using Granular Computing, In Transaction Rough Sets VIII, Lecture Notes in Computer Sciences, 5084, Springer Verlag, 93 – 115.
- Suraj, Z., & Pancierz, K., (2007). Flow Graphs as a Tool for Mining Prediction Rule of Changes of Components, In Temporal Information Systems, Rough Sets and Knowledge Technology, Lecture Notes in Artificial Intelligence, 2475, Springer Verlag, Berlin, 468 – 475.
- Takacs, P., & Csajbok, Z.E., (2017). On the Connection of Flow Graphs and Contingency Tables, The 10th Int. Conf. on Applied Informatics, 295 – 302.
- Wong, S.K.M., Butz, C.J. & Wu, D., (2000). On the Implication Problem for Probabilistic Conditional Independency, *IEEE Transactions on Systems, Man and Cybernetics*, Vol.30 (6), 785 – 805.
- Yao, N. & Miao, D., (2019). Identification of Structures and Causation in Flow Graphs, *Information Sciences*, 486, 287 – 309.