



**BULANIK DİLSEL ÖZETLEMENİN OLASILIKLI BELİRSİZLİĞİ
DİKKATE ALAN BİR GENİŞLEMESİ**

Sena AYDOĞAN

**YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

ARALIK 2015

Sena AYDOĞAN tarafından hazırlanan “BULANIK DİLSEL ÖZETLEMENİN OLASILIKLI BELİRSİZLİĞİ DİKKATE ALAN BİR GENİŞLEMESİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Endüstri Mühendisliği Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Diyar AKAY

Endüstri Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Başkan : Doç. Dr. Suat ÖZDEMİR

Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Üye : Doç. Dr. Saadettin Erhan KESEN

Endüstri Mühendisliği, Selçuk Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

Tez Savunma Tarihi: 18/12/2015

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Metin GÜRÜ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
 - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
 - Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
 - Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
 - Bu tezde sunduğum çalışmanın özgün olduğunu,
- bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

.....
Sena Aydoğan

18.12.2015

BULANIK DİLSEL ÖZETLEMENİN OLASILIKLI BELİRSİZLİĞİ DİKKATE ALAN BİR GENİŞLEMESİ

(Yüksek Lisans Tezi)

Sena AYDOĞAN

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Aralık 2015

ÖZET

Dilsel özetleme tanımlayıcı amaçlı veri madenciliği tekniklerinden birisidir. Büyük hacimdeki sayısal verinin doğal dil temelli insanların kolayca anlayabileceği bir yapıda ifade edilmesini sağlar. Gerçek dünya bilgisi çeşitli belirsizlikler içerir ve olasılık teorisi, olabilirlik teorisi veya kanıt teorisi gibi sıklıkla tercih edilen yöntemlerle modellenir. Mühendislik, ulaşım, karar verme gibi alanlarda birden fazla belirsizlik çeşidi bir arada bulunur ve bu farklı belirsizliklerin birlikte modellenmesi bilgiyi daha doğru yansıtacağından önemlidir. Z-sayısı hem olasılığı hem olabilirliği dikkate alan yeni bir kavramdır. Z-sayısı iki bileşenden oluşur, (A, B) . A değişkenin değerini bir bulanık küme ile ifade ederken, B A'nın kesinliğini yine bir bulanık küme ile gösterir. Dilsel özetleme yönteminde en önemli adım elde edilen özetin doğruluk derecesinin hesaplanmasıdır. Z-sayısında B'nin A'yı sağlama derecesi dilsel özetlemedeki doğruluk derecesi hesaplamasıyla çok benzerdir. Literatürde dilsel özetlerin değerlendirilmesi için skalar ve bulanık kardinalite temelli yöntemler geliştirilmiştir. Ancak farklı belirsizliklerin bir arada olduğu verilerin özetlenmesinde kullanılacak bir yöntem mevcut değildir. Bu nedenle dilsel özetleme ve Z-sayısı benzerliğinden yola çıkılarak dilsel özetlerin değerlendirilmesinde olasılık ve olabilirlik teorisini bir arada kullanan Z-sayısı temelli yeni bir yaklaşım önerilmiştir. Önerilen yaklaşımda tip-II niceleyici cümleler için farklı değişkenlere ait ortak olasılık dağılımını elde etmede kopulalardan faydalanılmıştır. Veritabanındaki değişkenler arasındaki korelasyon da, örneğin, bağımsız, pozitif ilişkili veya negatif ilişkili, hesaplama kopula kullanımı ile dâhil edilmiştir. Önerilen yöntem örnek bir verisetinde test edilmiş ve mevcut dilsel özetleri değerlendirme yöntemleri ile karşılaştırılmıştır.

Bilim Kodu : 906.1.071

Anahtar Kelimeler : Veri madenciliği, Dilsel özetleme, Bulanık küme, Z-sayısı, Belirsizlik, Kopula

Sayfa Adedi : 81

Danışman : Doç. Dr. Diyar AKAY

AN EXTENSION OF FUZZY LINGUISTIC SUMMARIZATION CONSIDERING
PROBABILISTIC UNCERTAINTY

(M. Sc. Thesis)

Sena AYDOĞAN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

December 2015

ABSTRACT

Linguistic summarization is one of the data mining techniques for descriptive purposes. It provides to express large volumes of data in natural language based forms that could be easily understood by humans. The real world information covers various forms of uncertainties and is typically modeled with the preferred methods such as probability theory, possibility theory or evidence theory. In areas such as engineering, transportation and decision-making, there is apparently more than one variety of uncertainty at hand, and joint modeling of distinct uncertainties is important to reflect the information more accurately. Z-number is a new concept taking into account both probability and possibility. A Z-number consists of two components, (A,B). A represents the value of the variable with a fuzzy set and B represents the certainty of A with also a fuzzy set. Most important step in linguistic summarization is to determine the truth degree of the generated summary. The degree to which A satisfies B in Z-number concept is very similar with the calculating the truth degree in linguistic summarization. In literature, various methods have been proposed for linguistics summarization based on scalar cardinality and fuzzy cardinality. However, there isn't any method for linguistic summarization in which different uncertainties exist together in the data. Hence, starting from the similarity between linguistic summarization and Z-number, a new method for linguistic summarization based on Z-number, considering probabilistic and possibilistic uncertainties together, is proposed. In the proposed method, the copulas are employed to obtain the joint probability distribution of the different variables for type-II quantified sentences. The correlation, i.e., independent, positive correlated or negative correlated, between the variables in the database is also incorporated in the computations with copulas. The proposed method is tested on a sample dataset and compared with the existing methods in the literature.

Science Code : 906.1.071

Key Words : Data mining, Linguistic summarization, Fuzzy set, Z-number, Uncertainty, Copula

Page Number : 81

Supervisor : Assoc. Prof. Dr. Diyar AKAY

TEŞEKKÜR

Çalışmalarım boyunca değerli yardımları ve katkılarıyla beni yönlendiren danışman hocam Doç. Dr. Diyar AKAY'a, çalışmalarından ve bilgilerinden faydalandığım Arş. Gör. Dr. Fatih Emre BORAN'a ve manevi destekleriyle beni hiçbir zaman yalnız bırakmayan çok değerli aileme teşekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. BELİRSİZLİK VE KUSURLU BİLGİ	3
3. DİLSEL ÖZETLEME.....	7
3.1. Veri Madenciliği	7
3.2. Bulanık Kümeler	8
3.3. Dilsel Özetlemenin Temel Tanımlamaları	15
3.4. Dilsel Özetlerin Değerlendirilmesi	18
3.4.1. Tip-I niceleyici cümlelerin değerlendirilmesinde doğruluk derecesinin sağlaması gereken özellikler	19
3.4.2. Tip-II niceleyici cümlelerin değerlendirilmesinde doğruluk derecesinin sağlaması gereken özellikler	20
3.5. Metotlar.....	21
3.5.1. Skalar kardinalite temelli yöntemler	21
3.5.2. Bulanık kardinalite temelli yöntemler.....	31
3.5.3. Diğer yöntemler	39
3.6. Uygulamalar.....	39
4. Z-SAYILARI İLE DİLSEL ÖZETLEME	43

	Sayfa
4.1. Z-sayıları	43
4.2. Dilsel Özetleme	46
4.2.1. Tip-I niceleyici cümle için Z-sayısı temelli yöntem	46
4.2.2. Tip-II niceleyici cümle için Z-sayısı temelli yöntem	48
5. SONUÇ VE ÖNERİLER	63
KAYNAKLAR	65
EKLER.....	79
EK-1. Çalışan Veriseti	80
ÖZGEÇMİŞ	81

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. t-norm ve t-konorm operatörlerinin sağlaması gereken özellikler.....	12
Çizelge 3.2. Temel t-norm operatörleri	13
Çizelge 3.3. Temel t-konorm operatörleri.....	13
Çizelge 3.4. Dilsel özetlemede kullanılan notasyonlar	15
Çizelge 3.5. Tip-I niceleyici cümlelerin değerlendirilmesi için önerilen yöntemlerin sağladığı özellikler	33
Çizelge 3.6. Tip-II niceleyici cümlelerin değerlendirilmesi için önerilen yöntemlerin sağladığı özellikler	33
Çizelge 3.7. Yöntemler arasındaki ilişkiler.....	33
Çizelge 3.8. $w_g(S_g)$ ve S 'nin kısıtlama seviyesi ve $gc(S / w_g(S_g))$, E ve B 'nin dereceli sayıları	36
Çizelge 3.9. Çeşitli yarı bulanık niceleyici örnekleri.....	38
Çizelge 3.10. Dilsel özetleme ile ilgili çalışmalar (zaman serileri)	40
Çizelge 3.11. Dilsel özetleme ile ilgili çalışmalar (uygulamalar).....	41
Çizelge 4.1. Çalışanların yaşları ve <i>genç</i> kümesine aitlik dereceleri.....	47
Çizelge 4.2. Eğitim seviyesine ait olasılık ve kümülatif olasılık değerleri.....	58
Çizelge 4.3. Aylık gelire ait olasılık ve kümülatif olasılık değerleri	58
Çizelge 4.4. Eğitim seviyesinin <i>az</i> , <i>orta</i> ve <i>çok</i> bulanık kümelerine aitlik dereceleri	59
Çizelge 4.5. Aylık gelirin <i>düşük</i> , <i>orta</i> ve <i>yüksek</i> bulanık kümelerine aitlik dereceleri ..	60
Çizelge 4.6. Skalar kardinalite, bulanık kardinalite temelli ve Z-sayısı temelli yöntemlere göre elde edilen sonuçlar	60
Çizelge 4.7. Belirsizlik, kapsama, uygunluk, uzunluk ve toplam geçerlilik derecesine göre elde edilen sonuçlar	61
Çizelge 4.8. Kapsama, fayda, aykırılık ve basitlik derecelerine göre elde edilen sonuçlar	62

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Kusurlu bilginin sınıflandırılması.....	3
Şekil 2.2. Farklı teoriler arasındaki bağlantılar.....	4
Şekil 3.1. (a) Kesikli bulanık küme (b) Sürekli bulanık küme	9
Şekil 3.2. Bulanık kümelerde öz, destek, sınır, geçiş noktaları ve yükseklik	10
Şekil 3.3. Üçgensel bulanık sayı	11
Şekil 3.4. Dilsel değişken örneği	12
Şekil 3.5. (a) “yaklaşık 3” mutlak niceleyicisi, (b) “çoğu” göreceli niceleyicisi	16
Şekil 3.6. (a) Almanya’nın güneyi (b) “Almanya’nın güneyinin en az %60’ı bulutludur” $Z_Q = 0$ (c) “Almanya’nın güneyinin en az %60’ı bulutludur” $Z_Q = 1$	30
Şekil 3.7. (a) Almanya’nın güneyi (b) “Almanya’nın güneyinin yaklaşık %10’u bulutludur” $Z_Q = 1$ (c) “Almanya’nın güneyinin yaklaşık %10’u bulutludur” $Z_Q = 1$	30
Şekil 3.8. (a) Almanya’nın güneyi (b) “Almanya’nın güneyinin en az %60’ı bulutludur” $Y_Q = 0,1$ (c) “Almanya’nın güneyinin en az %60’ı bulutludur” $Y_Q = 0,6$	31
Şekil 4.1. Tip-I niceleyici cümle değerlendirme işlem adımları	47
Şekil 4.2. <i>Genç</i> bulanık kümesinin aitlik fonksiyonu	47
Şekil 4.3. Tip-II niceleyici cümle değerlendirme işlem adımları	50
Şekil 4.4. Eğitim seviyesinin <i>az</i> , <i>orta</i> , <i>çok</i> bulanık kümelerine aitlik fonksiyonları	59
Şekil 4.5. Aylık gelirin <i>düşük</i> , <i>orta</i> ve <i>yüksek</i> bulanık kümelerine aitlik fonksiyonları	59

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklamalar
A	Herhangi bir küme
X	Evrensel küme
$\mu_A(x_i)$	x_i elemanın A kümesine aitlik derecesi
A_α	A kümesinin α kesmesi
A_{α^+}	A kümesinin güçlü α kesmesi
$\sup$	Supremum
a	Dilsel değişkenin adı
$ET(a)$	a değişkeninin etiket kümesi
G	Sözdizimsel kural
$\mathcal{M}$	Anlamsal kural
$\otimes$	t-norm operatörü
$\oplus$	t-konorm operatörü
$\neg A_1$	A_1 kümesinin tümleyeni
Ng	Tümleyen
$\mathbb{D}$	Veritabanı
Y	Nesneler kümesi
y_m	Veritabanındaki m . nesne
v_k	Veritabanındaki k . özellik
$v_k^m \equiv v_k(y_m)$	m . kayıtın k . özelliğe ait değeri
d_m	m . nesnenin tüm özellikleri
S_k	Özetleyici
Q	Niceleyici
$w_g(S_g)$	Öncül özetleyici

Simgeler	Açıklamalar
T_1	Doğruluk derecesi
Q_{mut}	Mutlak niceleyici
$Q_{gör}$	Göreceli niceleyici
$ X $	X 'deki eleman sayısı
$ S $	S 'deki eleman sayısı
$\exists$	Var niceleyicisi
$\forall$	Tümü niceleyicisi
$\emptyset$	Boş Küme
T_2	Belirsizlik derecesi
T_3	Kapsama derecesi
T_4	Uygunluk derecesi
T_5	Uzunluk derecesi
T	Toplam geçerlilik derecesi
T_c	Yeterli kapsama derecesi
T_u	Fayda derecesi
T_o	Aykırlık derecesi
T_s	Özetin uzunluğu
C_1	Cümle
$\mathbf{sim}_1$	İki bulanık küme arasındaki benzerlik
$\mathbf{sim}_2$	İki gerçek değer arasındaki benzerlik
Λ	α kesmelerinin oluşturduğu sonlu küme
$\rho(\alpha)$	Λ kümesinde tanımlı elemanı evrensel kümede tanımlı bir altkümeye götüren fonksiyon
$\Re$	Dereceli sayılarda denklik bağıntısı
FN	Bulanık sayı
(A, B)	Z-sayısı
(x, A, B)	Z-değerleme

Simgeler**Açıklamalar** **$G(p)$** p 'nin Z-değerlemeyi sağlama derecesi **$p(x_i)$** x_i 'nin olasılık değeri **$(S | w_g(S_g))$** $w_g(S_g)$ sağlandığında S 'nin koşullu ifadesi **$f_{v_1, v_2}(x_i, y_j)$** $v_1 = x_i$ ve $v_2 = y_j$ 'nin birlikte ortaya çıkma olasılığı **$F_{v_1}(x_i)$** v_1 ile ilişkili kümülatif dağılım fonksiyonu**M**

Minimum kopula

 Π

Çarpım kopula

w

Lukasiewicz kopula

C

Kopula

 ρ

Korelasyon katsayısı

Kısaltmalar**Açıklamalar****AHP**

Analitik Hiyerarşi Prosesi

BK

Bulanık Küme

CWW

Kelimelerle Hesaplama

NLP

Doğal Dil İşleme

OWA

Sıralı Ağırlıklı Ortalama Birleştirme Operatörü

1. GİRİŞ

Bilişim teknolojilerinin gelişmesiyle birlikte veriye ulaşmak, saklamak ve paylaşmak kolaylaşmıştır. Eldeki büyük hacimdeki veriden ziyade bu verilerden elde edilecek bilginin önemi daha fazladır. Bu nedenle verinin işlenmesi yoluyla bilgi elde edilmesini sağlayan veri madenciliği önemi giderek artan bir araştırma alanı olarak ortaya çıkmıştır [1].

Veri madenciliği, verinin elde edilmesi ve ön işleme işlemlerinin yapılmasının ardından çeşitli algoritmalar ve teknikler kullanılarak anlamlı ve faydalı bilginin elde edilmesi ve bu bilginin kullanılarak bir eylemin gerçekleşmesine kadar geçen süreci ifade etmek için kullanılır [2]. Bir insan kaynakları veritabanında yer alan veriden tanımlayıcı bilgilere ulaşmak için kullanılan yöntemlerden biri özetlemedir. En bilinen özetleme yöntemi istatistiksel özetlemedir. Ancak standart sapma, varyans gibi istatistiksel sonuçlar insanlar tarafından kolay anlaşılmadığı için bu yöntem dezavantajlıdır. İnsanlar günlük yaşamda daha az kesin olmasına rağmen rakamlar yerine doğal dili tercih etmektedirler. Çünkü dilsel ifade biçimindeki bilgiler daha kolay anlaşılakta ve hatırlanmaktadır. Bu yüzden dilsel önermeler türetme amacı ile verinin özetlenmesini sağlayan dilsel özetleme yöntemi ilk olarak Yager tarafından önerilmiştir [3]. Veritabanından elde edilen “*Çoğu az eğitilmiş çalışanın aylık geliri düşüktür.*” önermesi dilsel özetleme süreci ile elde edilen bir örnek cümledir. Bu önerme insanlar tarafından kolayca anlaşılabilir. Ancak burada önemli olan “*az eğitilmiş*” ve “*düşük gelirli*” gibi ifadelerin nasıl ölçeklendirileceğidir. Bu tür dilsel terimleri sayısal olarak modellemek için bilinen en iyi yol bulanık kümelerdir [4]. Bulanık kümelerle dilsel özetleme işleminde veritabanında yer alan özellikler bulanık kümelerle etiketlenmektedir.

Çoğu gerçek dünya bilgisi kesin değildir ve bu yüzden kesin fonksiyonlarla tanımlanamaz. Belirsiz bilgiyi göstermek için çeşitli teoriler önerilmiştir. Bunlardan önemli olan iki teori olasılık teorisi ve olabilirlik teorisidir. Olabilirlikli belirsizlik genellikle bir olabilirlik dağılımı ile tanımlanır ve $\mu(x)$ ilgilenilen değişkenin x değerini almasının olabilirliğini ifade eder. Olasılıklı belirsizlik ise bir olasılık dağılımı ile tanımlanır ve $p(x)$ ilgili değişkenin x değerini almasının olasılığını ifade eder. Dilsel özetler olasılık ve olabilirlik temelli yöntemlerle ayrı ayrı değerlendirilmektedir. Ancak farklı belirsizlikleri içeren

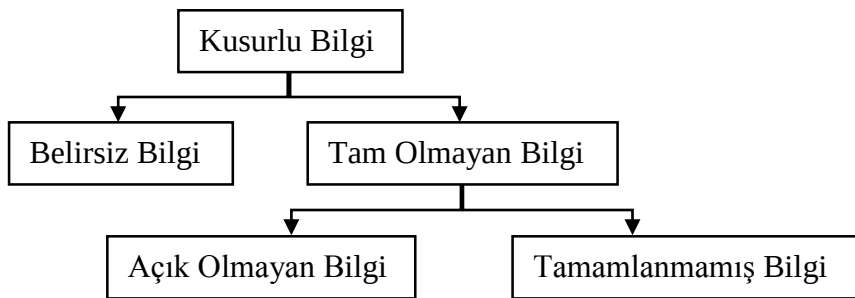
alanlarda belirsizliđi modellemede tek bir teori kullanmak yetersiz kalacaktır. Bu nedenle bilgiyi daha dođru modellemek iin bu teorilerin ortak kullanımına ihtiya vardır.

Kullanılan bilginin faydalı olabilmesi iin gvenilir olması gerekliliđi ile ilgili olarak Zadeh tarafından Z-sayısı kavramı ortaya ıkarılmıřtır [5]. Z-sayısı (A, B) sıralı ikilisinden oluřmaktadır ve kısaca deđiřkenin A deđerini almasının kesinliđini B ile tanımlayan yeni bir kavramdır. Deđiřkenin A deđerini almasının B kesinliđi ile sađlanması olabirliđi dilsel zetlerin deđerlendirilmesiyle benzerlik tařımaktadır. Bu tez alıřmasında hem olasılık teorisini hem olabirlik teorisini dikkate alan yeni bir metot arayıřı amacıyla dilsel zetler Z-sayılarına geniřletilmiřtir. Bylece deđiřkenlerdeki belirsizliđi ele almada Z-sayılarında olduđu gibi hem olasılık hem olabirlik deđerlerinden faydalanılan yeni bir metot nerilmiřtir. nerilen yntemin olasılık tabanında olasılık dađılımlarını birleřtirmek iin kopula ynteminden, olabirlik tabanında bulanık kmelerin kesiřimi iin t-norm operatrlerinden faydalanılmıřtır. Kopula, ok deđiřkenli dađılım fonksiyonlarını bunların tek boyutlu marjinal dađılım fonksiyonuna birleřtirmesi olarak tanımlanır [6].

alıřmanın kalan kısmı řu řekildedir: 2. blmde belirsizlik ve kusurlu bilgi kavramları ile belirsizliđin eřitleri aıklanmıřtır. 3. blmde dilsel zetleme iřlemi, elde edilen dilsel zetleri deđerlendirme yntemleri, deđerlendirme yntemlerinin sađlaması gereken zellikler ve dilsel zetleme uygulamalarına yer verilmiřtir. 4. blmde Z-sayıları ile dilsel zetlemenin benzerliđi aıklanarak Z-sayısı temelli dilsel zetleme metodu nerilmiřtir ve metotta kullanılan kopula kavramı ile ortak dađılım elde etme konuları aıklanmıřtır. nerilen yntem rnek bir veri setinde test edilmiřtir ve literatrde mevcut yntemlerle karřılařtırılmıřtır. Son blmde alıřmanın sonularına ve gelecek alıřma nerilerine yer verilmiřtir.

2. BELİRSİZLİK VE KUSURLU BİLGİ

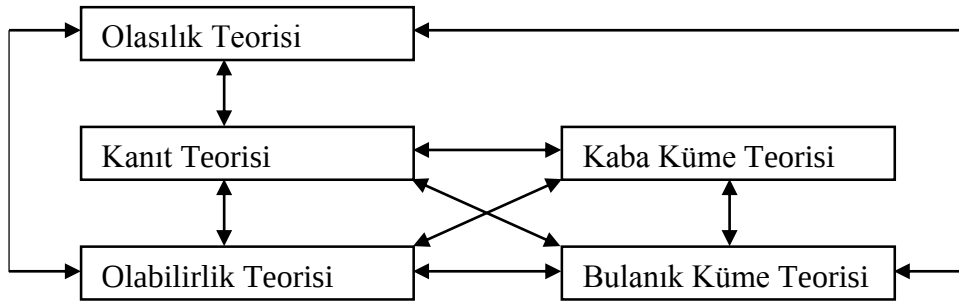
Belirsizlik (uncertainty) kusurlu bilgiyi (imperfect information) tanımlamak için kullanılır [7, 8]. Smets mükemmel bilginin tam (precise) ve kesin (certain) olduğunu söylemiştir [9]. Nesneler sınıfı tek bir nesneden oluşuyorsa yani tekilse tamdır ve önermenin güven derecesi 1'e eşitse kesindir. Ancak gerçek dünyada mükemmel bilgi asla elde edilemez. Bilgi nesnel veya öznel olabilir. Örneğin “zarın 4 gelme olasılığı $1/6$ 'dır.” bilgisi nesnel iken, “Bence zarın 4 gelmesi olasılığı 0,2'dir.” bilgisi öznelidir. Kusurlu bilginin sınıflandırılması Şekil 2.1'de gösterilmiştir. Kusurlu bilgi belirsiz (uncertain) ve tam olmama (imprecise) olmak üzere iki çeşittir. Tamamlanmamış (incomplete) bilgi ve açık olmayan (vague) bilgi tam olmamanın çeşididir. Belirsiz bilgi güven derecesi 1'den küçük olan bilgidir. Tam olmayan bilgi birden fazla nesne ile ilgili olan bilgidir. Güven derecesi bilinmeyen bilgi tamamlanmamış bilgidir. Yanlış değişken ile karakterize edilmiş bilgi açık olmayan bilgidir ve üyeliği bulanık tanımlanır. Gerçek durumlarda belirsizlik ve tam olmama bir arada bulunur. Örneğin, “Hedef 0,7 güven derecesi ile 3. nesnedir” bilgisi tamdır ancak belirsizdir, “Hedefle ilgili toplam güven hedef uzunluğunun 28-32 m arasında olduğudur” bilgisi tam değildir ancak kesindir. “Hedefin 19. nesne olması olasıdır.” bilgisi tamamlanmamış bilgiye, “Gözlemlenen hedef büyüktür.” bilgisi açık olmayan bilgiye örnektir.



Şekil 2.1. Kusurlu bilginin sınıflandırılması

Kusurlu bilginin çeşitlerini ele almada bazı teoriler mevcuttur. Her bir teori bilgiyi daha doğru modellemeyi amaçlamaktadır. En bilinen ve en çok kullanılan teoriler olasılık teorisi, kanıt teorisi [10, 11], bulanık küme teorisi [4], olabilirlik teorisi [12] ve kaba küme teorisidir [13]. Farklı teoriler arasındaki ilişki diyagramı Şekil 2.2'de verilmiştir. Olasılık teorisi belirsizliğin en bilinen ve en eski teorisidir. Nesnel olasılık 17. yy'da, öznel olasılık 20. yy'da geliştirilmiştir. Öznel olasılık inanç derecesi olarak yorumlanabilir. Olasılık teorisi

belirsizlik problemlerini etkili bir şekilde ele alır; ancak, tam olmama durumu probleme dâhil olduğunda olasılık teorisi ile modelleme yapmak yetersiz kalır. Kanıt teorisi, olasılık teorisinin genelleştirilmiş halidir ve öznel inancın daha iyi bir gösterimini sağlar. Kanıt teorisi olasılık teorisinin aksine hem tam olmama hem belirsizliği göstermede kullanılabilir ve iki durumu içeren problemlerde olasılık teorisine göre daha gerçekçi sonuçlar sağlar. Bulanık küme teorisi açık olmayan bilgiyi modellemek için önerilmiştir. Çünkü açık olmayan bilgi olasılık teorisi veya kanıt teorisi ile modellenemez. Olabilirlik teorisi, belirsizliği göstermede alternatif bir yöntemdir, eksik bilgiyi ele almak için geliştirilmiştir. Kaba küme teorisi belirsizliği yok sayarak tam olmayan bilgiyi ele alır.



Şekil 2.2. Farklı teoriler arasındaki bağlantılar

Belirsizliği gösteren klasik teoriler bilginin farklı kusurlarını ele alır. Bu teorileri rakip olarak düşünmektense tamamlayıcı olarak düşünmek faydalı sonuç elde etmeye yarar. Bazı durumlarda teoriler arasında dönüşüm uygulamak gerekli olabilir ancak bu dönüşümler her zaman bilgi kaybına sebep olacaktır. Teoriler arasındaki ilişkiler kısaca açıklanmıştır. Olasılık teorisi tam olmamayı içermediğinden kanıt teorisi olasılık teorisinin genel hali olarak görülebilir. Olasılık teorisi sadece nesne kümesini karakterize ederken, kanıt teorisi nesnelerin tüm alt kümelerini karakterize eder. Olasılık teorisini kanıt teorisine dönüştüren ve ters dönüşüm yapan çeşitli metotlar mevcuttur. Olabilirlik teorisi kanıt teorisinin içinde yer alabilir. Çünkü olabilirlik ve gereklilik dereceleri iç içe geçmiş altkümeler için mümkünlük ve inanç derecelerine karşılık gelir. Ancak bu teoriler farklı belirsizlikleri tanımladıkları için aynı değildirler, bu yüzden bazı dönüşüm yöntemleri bu iki teori arasında da mevcuttur. Kanıt teorisi hem olasılık teorisi hem olabilirlik teorisinin genelleştirilmesi olarak görülebilir. Ancak bu teoriler farklı tip bilgileri gösterdiğinden bilginin parçalarını birleştirmek için olasılık ve olabilirlik teorilerini eş zamanlı kullanmak tamamlayıcı olur ve faydalı olabilir. Olabilirlik teorisi bulanık küme teorisinin temelinde geliştirildiği için bulanık küme teorisi ve olabilirlik teorisi arasında güçlü bir ilişki vardır; bir teori ile

modellenen bilgi diğeriyle de modellenabilir. Bulanık küme teorisi ile kanıt teorisi arasındaki ilişki bulanık küme teorisi-olabilirlik teorisi ve olabilirlik teorisi-kanıt teorisi dönüşümleriyle elde edilebilir. Bulanık küme teorisi ve olasılık teorisi aynı kusurlu bilgiyi dikkate almazlar. Olasılık teorisi belirsiz elemanın kesin kümeye aitliği ile ilgilenirken bulanık küme teorisi kesin elemanın belirsiz kümeye aitliği ile ilgilenir. Olabilirlik teorisi ise bu iki teori arasında yer alır. Kaba kümeler bulanık kümelerden çok daha geneldir. Kaba küme teorisi ve olabilirlik teorisi farklı bilgiyi ele aldığından rakip teori olarak görülmemelidirler, iki teorinin hibrit kullanımı ile daha kapsamlı belirsiz bilgi durumları modellenenebilir. Kaba küme ile modellenenebilen bilgi kanıt teorisi ile de modellenenebilir. Kaba kümenin alt ve üst sınırları inanç ve mümkünlik değerlerine dönüştürülebilir, ancak ters dönüşüm mümkün değildir.

Gerçek dünya problemlerinde birden fazla belirsizlik çeşidi bir arada bulunmaktadır. Örneğin, ulaşım sistemlerinin analizinde veri eksik, yaklaşık veya tutarsız olabilir ve model belirsizdir. Bunların ötesinde, ulaşım analizlerinin ana odağı olan insan davranışı dikkate değer ölçüde açıklanamayan çeşitlilikler içerir. Ulaşım çalışmaları bir sistemin tahmini, tespiti, kontrolü ve optimizasyonu ile ilgilidir. Analizin önemli elemanları veri, model, amaç ve sonuçlardır. Veri olasılıksal analizlere elverişli istatistiksel bir elemandır, modeldeki nedensellik genellikle dil tabanlı kurallarla ve çıkarımlar benzerlik mantığıyla gösterilir, amaçlar genellikle çok spesifik değildir, süreç deterministik olmasına rağmen sonuçlar yaklaşık olarak yorumlanır [14]. Bu elemanların birlikteliği ulaşım sistemlerinin modellenmesinde birden fazla teori kullanımını gerektirmektedir.

Ulaşım sistemlerinde olduğu gibi karar verme alanında da bilgi (maliyet, zaman vb.) algıya, yaklaşık değerlere veya tahmini değerlere bağlı olduğunda hem bireysel hem de ortak seviyelerde belirsizlik bir arada bulunur [14].

Askeri alanda hedef belirleme problemlerinde (arkadaş, tarafsız veya düşman) farklı kaynaklardan gelen verilerin birleştirilmesi söz konusudur. Bazı kaynaklardaki belirsizlik olasılık teorisiyle modellenirken, bazı kaynaklardaki bilginin olabilirlik teorisiyle modellenmesi daha gerçekçi sonuç verebilir [15]. Bu gibi belirsizliğin farklı tiplerinin bir arada olabileceği alanlara mühendislik, ekonomi, telekomünikasyon gibi alanlar da eklenebilir. Bu yüzden birlikte modellenmesi gerçek dünyaya daha doğru bir yaklaşım sunacağından oldukça önemlidir.

3. DİLSEL ÖZETLEME

Bu bölümde veri madenciliği, bulanık kümeler ve dilsel özetlemenin temel tanımlamaları yapılmıştır, dilsel özetleri değerlendirme metotları açıklanmıştır ve dilsel özetleme konusunda literatürde yapılan uygulamalara yer verilmiştir.

3.1. Veri Madenciliği

Veri madenciliği, verinin elde edilmesi ve ön işleme işlemlerinin yapılmasının ardından çeşitli algoritmalar ve teknikler kullanılarak anlamlı ve faydalı bilginin elde edilmesi ve bu bilginin kullanılarak bir eylemin gerçekleşmesine kadar geçen süreci ifade etmek için kullanılır [2]. Veri madenciliğinde iki temel amaç söz konusudur. Bunlardan birincisi tahmin yapmak, ikincisi ise tanımlayıcı bilgileri ortaya çıkarmaktır [1]. Tahmin edici modellerde veri tabanındaki bir özelliğin diğer özelliklere bağlı öngörülmesi vardır. Tahmin edici modeller sınıflandırma ve regresyon modelleridir. Sınıflandırma modelinde tahmin edilecek bağımlı değişken kategorik değer alırken, regresyon modelinde sürekli ya da kategorik değer alabilmektedir. Tanımlayıcı modellerde amaç, veri tabanında yer alan özellikler arasında var olan ilişkileri özetleyen örüntülerin elde edilmesidir [16]. Tanımlayıcı modeller birliktelik kuralları, kümeleme, aykırı değer analizi ve özetleme modelleridir. Birliktelik kuralları eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır. Kümeleme tanımlanan veriler ışığında, nesnelerin birbirlerine olan benzerlik ve farklılıklarına göre verilerin kümelenmesi işlemidir. Aykırı değer analizinde veri kümesinden tamamen farklı veya onunla tutarsızlık gösteren örneklerin analizi vardır. Özetleme işleminde ise amaç veriden bilgi çıkarımıdır. Özetleme işleminde ilk yapılan çalışmalar istatistik temellidir. İstatistik temelli özetlemede özellikler ortalama, standart sapma ve medyan gibi sınırlı değerler ile tanımlanmaktadır ve insanlar tarafından kolay bir şekilde anlaşılammamaktadır. Günlük hayatta insanlar doğal dil ifadelerini daha kolay anlayabilmektedir. Örneğin, insan kaynakları veri tabanından “Çalışanların ortalama geliri 1078,45 liradır.” gibi bir ifade çıkarılsın. Özeti anlamak kolay olsa bile insanlar “Çalışanların ortalama geliri bin lira civarındadır.” veya “Çalışanların ortalama geliri düşüktür.” gibi dilsel ifadelerin kullanıldığı özetleri tercih ederler. Bu nedenle insanlar tarafından kolayca anlaşılabilir olacak özetler sunan bulanık kümeleme temelli *Dilsel Özetleme* yöntemi ortaya çıkmıştır [3].

3.2. Bulanık Kümeler

Klasik küme teorisinde elemanların evrensel kümenin altkümeleri olan A gibi bir kümeye ait olmaları veya ait olmamaları kesin sınırlarla belirlidir ve 0-1 üyelik fonksiyonuyla gösterilir. Eğer bir eleman A kümesinin elemanı ise üyelik fonksiyonu değeri 1'e eşittir. Fakat belirsizliğin ve kesinsizliğin olduğu dünyamızda olayların iki değerli 0-1 mantıkla açıklanması tam olarak doğru bir yaklaşım değildir. Bu nedenle Zadeh klasik küme kavramının gerçek dünyada, özellikle insanları içeren kısmen karmaşık sistemlerle uğraşırken yetersiz kalmasından yola çıkarak, niteliklerin (küme elemanlarının) ikili üyelik fonksiyonuyla ifade edildiği klasik kümeler yerine, dereceli üyelik fonksiyonlarıyla ifade edildiği bulanık kümelerle tanımlanmasını önermiştir [4].

Tanım 2.1. Bulanık Küme

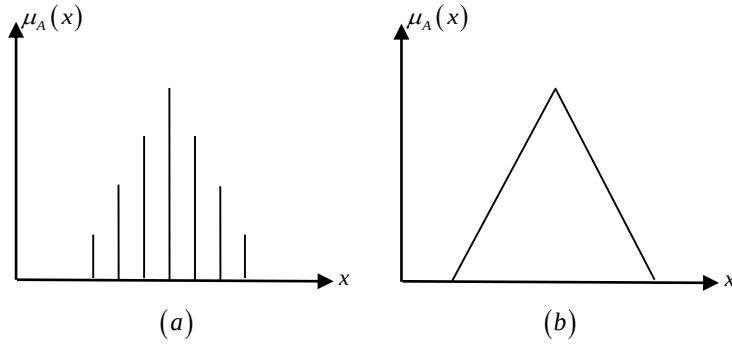
A kümesi, X evrensel kümesinde bir bulanık kümedir. Bu durumda A bulanık kümesinin üyelik fonksiyonu $\forall x \in X$ için $\mu_A : X \rightarrow [0,1]$ şeklinde tanımlanır. $\mu_A(x_i)$, x_i elemanın A kümesine aitlik derecesini ifade eder. Evrensel küme kesikli ise A bulanık kümesi Eş. 3.1 ile, evrensel küme sürekli ve sonsuz ise A bulanık kümesi Eş. 3.2 ile gösterilir.

$$A = \left\{ \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n} \right\} = \left\{ \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i} \right\} \quad (3.1)$$

$$A = \left\{ \int_X \frac{\mu_A(x)}{x} \right\} \quad (3.2)$$

Eş. 3.1'de gösterilen “+” işareti cebirsel toplam işareti değildir, elemanların birleşimini gösterir. Aynı şekilde Eş. 3.2'de gösterilen integral işareti de sürekli değer alan elemanların birleşimini gösterir.

Şekil 3.1'de bir bulanık kümeye ait elemanların üyelik fonksiyonlarının kesikli ve sürekli formda gösterimleri verilmiştir [17].



Şekil 3.1. (a) Kesikli bulanık küme (b) Sürekli bulanık küme

Bulanık kümelerde önemli kavramlardan iki tanesi alfa (α) kesmesi ve güçlü alfa (α^+) kesmesidir. X evrensel kümesinde bir bulanık A kümesi ve $\alpha \in [0,1]$ olmak üzere α kesmesi ve α^+ kesmesi sırasıyla Eş. 3.3 ve Eş. 3.4'te gösterilmiştir.

$$A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\} \quad (3.3)$$

$$A_{\alpha^+} = \{x \in X \mid \mu_A(x) > \alpha\} \quad (3.4)$$

α (α^+) kesmesiyle A bulanık kümesinin elemanlarının üyelik dereceleri α 'ya eşit ya da büyük (α 'dan büyük) olan A_α (A_{α^+}) kesin kümesine dönüşmektedir [17, 18].

X evrensel kümesindeki bir bulanık A kümesinin özü, desteği, sınırı, geçiş noktaları ve yüksekliği Eş. 3.5 ile Eş. 3.9 arasındaki eşitliklerde ve Şekil 3.2'de gösterilmiştir [17, 18].

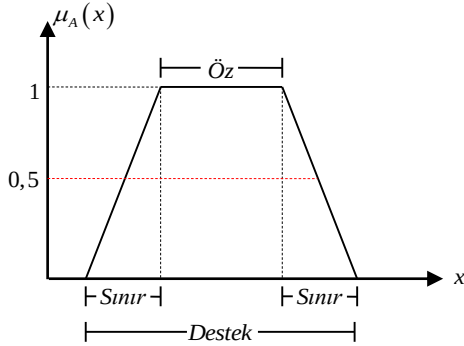
$$\text{Öz}(A) = \{x \in X \mid \mu_A(x) = 1\} \quad (3.5)$$

$$\text{Destek}(A) = \{x \in X \mid \mu_A(x) > 0\} \quad (3.6)$$

$$\text{Sınır}(A) = \{x \in X \mid 0 < \mu_A(x) < 1\} \quad (3.7)$$

$$\text{Geçiş Noktaları}(A) = \{x \in X \mid \mu_A(x) = 0,5\} \quad (3.8)$$

$$\text{Yükseklik}(A) = \sup_{x \in X} \{\mu_A(x)\} \quad (3.9)$$

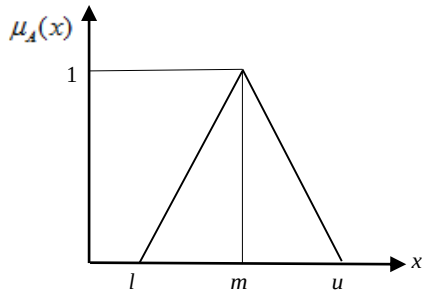


Şekil 3.2. Bulanık kümelerde öz, destek, sınır, geçiş noktaları ve yükseklik

X evrensel kümesinde bir bulanık A kümesinin normal olması için Eş. 3.10’da verilen şartı sağlaması gerekir [17, 18].

$$\text{Yükseklik}(A) = \sup_{x \in X} \{\mu_A(x)\} = 1 \quad (3.10)$$

X evrensel kümesinde bir bulanık alt kümenin bir bulanık sayı olabilmesi için konveks ve normal olması gerekir. Gerçek sayılar kümesinde tanımlanan bulanık sayılar çeşitli bulanık küme tipleri içerisinde özel öneme sahip alt kümelerdir. Bulanık sayılar, verilen gerçek sayıya yaklaşık sayıların kümesi olduğu için belirsizliğin olduğu mühendislik uygulamalarında ve karar vermede sıklıkla kullanılmaktadır. Bulanık sayılar üyelik fonksiyonlarının şekillerine göre isimlendirilir. Literatürde çok sayıda üyelik fonksiyonu tanımlanmakla birlikte işlem yükünün az olması nedeniyle en fazla kesikli, üçgen, gauss ve yamuk fonksiyonlar kullanılmaktadır [18-20]. Örnek olması amacıyla Şekil 3.3’te üçgensel bulanık sayının gösterimi verilmiştir. Üçgensel bulanık sayı için üyelik fonksiyonu Eş. 3.11 ile tanımlanmıştır [18].



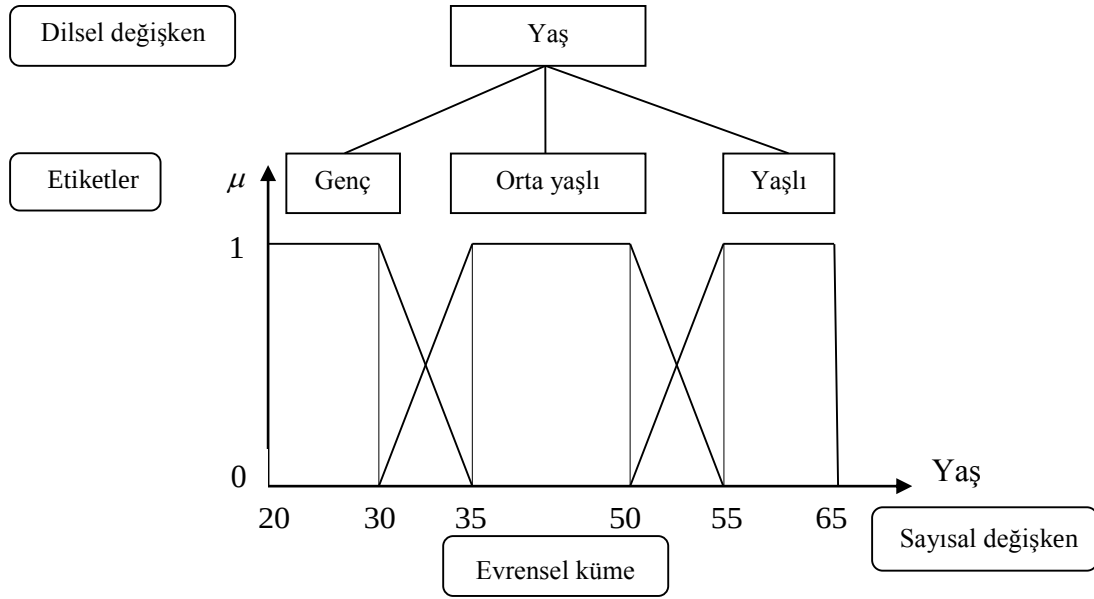
Şekil 3.3. Üçgensel bulanık sayı

$$\mu_A(x) = \begin{cases} 0 & x < l \\ (x-l)/(m-l) & l \leq x \leq m \\ (u-x)/(u-m) & m \leq x \leq u \\ 0 & x > u \end{cases} \quad (3.11)$$

Tanım 2.2. Dilsel Değişken

Bir dilsel değişken $(a, ET(a), X, G, \mathcal{M})$ fonksiyonuyla tanımlanabilir. Burada a dilsel değişkenin adını; $ET(a)$, a değişkeninin etiket kümesini ve X evrensel kümeyi göstermektedir. G sözdizimsel kuraldır ve $ET(a)$ kümesinde yer alan etiketleri üretir. $\mathcal{M}$ anlamsal kuraldır ve her A etiketinin (bulanık küme) anlamını gösterir. Örneğin Şekil 3.4'te "yaş" değişkeni a ile temsil edilmektedir. Bu değişkenin $ET(a)$ etiket kümesi ise "genç", "orta yaşlı" ve "yaşlı" dilsel etiketlerinden oluşabilir. Dilsel terimler temel değişken olan yaşa göre atanan bulanık sayılarla ifade edilirler. Yaş için tanımlanan "genç", "orta yaşlı" ve "yaşlı" etiketlerinin gösterimi Şekil 3.4'te verilmiştir [18, 21]. Bulanık kümelerin üyelik fonksiyonlarının oluşturulmasında oylama, direk değerlendirme, ters değerlendirme, aralık tahmini, üyelik derecesi örnekleme ve ikili karşılaştırma gibi başlıca yöntemler kullanılmaktadır [22-24].

Bulanık kümelerle işlem yaparken t-norm ve t-konorm operatörlerinden faydalanılır. $\otimes$ operatörü t-norm olarak, $\oplus$ operatörü t-konorm olarak adlandırılır. t-norm ve t-konorm operatörlerinin sağlaması gereken özellikler Çizelge 3.1'de verilmiştir.



Şekil 3.4. Dilsel değişken örneği

Çizelge 3.1. t-norm ve t-konorm operatörlerinin sağlaması gereken özellikler

	$\otimes$ t-norm	$\oplus$ t-konorm
Sınır Şartı	$\otimes(x_1, 1) = x_1$	$\oplus(x_1, 0) = x_1$
Monotonluk	$y_1 \leq y_2$ olduğu takdirde $\otimes(x_1, y_1) \leq \otimes(x_1, y_2)$	$y_1 \leq y_2$ olduğu takdirde $\oplus(x_1, y_1) \leq \oplus(x_1, y_2)$
Simetri	$\otimes(x_1, y_1) = \otimes(y_1, x_1)$	$\oplus(x_1, y_1) = \oplus(y_1, x_1)$
Birleşme	$\otimes(x_1, \otimes(x_2, y_2)) = \otimes(\otimes(x_1, x_2), y_2)$	$\oplus(x_1, \oplus(x_2, y_2)) = \oplus(\oplus(x_1, x_2), y_2)$

Literatürde tanımlanan bazı temel t-norm operatörleri Çizelge 3.2’de ve t-konorm operatörleri Çizelge 3.3’te gösterilmiştir.

Tanım 2.3. Bulanık Kümelerde Kesişim

A_1 ve A_2 , X evrensel kümesinde tanımlanmış bulanık kümeler ve bu kümelerin üyelik fonksiyonları $\mu_{A_1}(x)$ ve $\mu_{A_2}(x)$ ile gösterilsin.

A_1 ve A_2 bulanık kümelerinin kesişimi $A_1 \cap A_2$ ve üyelik fonksiyonu $\mu_{A_1 \cap A_2}(x)$ Eş. 3.12 ile tanımlanır.

$$\otimes : [0,1] \times [0,1] \rightarrow [0,1]$$

$$\mu_{A_1 \cap A_2}(x) = \otimes(\mu_{A_1}(x), \mu_{A_2}(x)) \quad \forall x \in X \quad (3.12)$$

Çizelge 3.2. Temel t-norm operatörleri

Operatör adı	Fonksiyon
Minimum	$\min(x_1, y_1)$
Çarpım	$x_1 y_1$
Lukasiewicz	$\max\{0, x_1 + y_1 - 1\}$
Drastik	$\begin{cases} x_1, & \text{Eğer } y_1 = 1 \\ y_1, & \text{Eğer } x_1 = 1 \\ 0, & \text{d.d.} \end{cases}$
Hamacher	$\begin{cases} 0, & \text{eğer } x_1 = y_1 = 0 \\ \frac{x_1 y_1}{x_1 + y_1 - x_1 y_1}, & \text{d.d.} \end{cases}$

Çizelge 3.3. Temel t-konorm operatörleri

Operatör adı	Fonksiyon
Maksimum	$\max(x_1, y_1)$
Çarpım	$x_1 + y_1 - x_1 y_1$
Lukasiewicz	$\min\{1, x_1 + y_1\}$
Drastik	$\begin{cases} x_1, & \text{Eğer } y_1 = 0 \\ y_1, & \text{Eğer } x_1 = 0 \\ 1, & \text{d.d.} \end{cases}$

Tanım 2.4. Bulanık Kümelerde Birleşim

A_1 ve A_2 , X evrensel kümesinde tanımlanmış bulanık kümeler ve bu kümelerin üyelik fonksiyonları $\mu_{A_1}(x)$ ve $\mu_{A_2}(x)$ ile gösterilsin.

A_1 ve A_2 bulanık kümelerinin birleşimi $A_1 \cup A_2$ ve üyelik fonksiyonu $\mu_{A_1 \cup A_2}(x)$ Eş. 3.13 ile tanımlanır.

$$\oplus: [0,1] \times [0,1] \rightarrow [0,1]$$

$$\mu_{A_1 \cup A_2}(x) = \oplus(\mu_{A_1}(x), \mu_{A_2}(x)) \quad \forall x \in X \quad (3.13)$$

Tanım 2.5. Bulanık Kümenin Tümleyeni

A_1 , X evrensel kümesinde tanımlanmış bulanık küme ve bu kümenin üyelik fonksiyonu $\mu_{A_1}(x)$ ile gösterilsin. A_1 kümesinin tümleyeni $\neg A_1$ ve üyelik fonksiyonu Eş. 3.14'teki gibi tanımlanır.

$$Ng: [0,1] \rightarrow [0,1]$$

$$\mu_{\neg A_1}(x) = Ng(\mu_{A_1}(x)) \quad \forall x \in X \quad (3.14)$$

Bir fonksiyonun tümleyen olabilmesi için aşağıda verilen şartları sağlaması gerekir. Bu özellikler $\forall x, y \in [0,1]$ için;

- $Ng(0) = 1$ ve $Ng(1) = 0$
- Eğer $x \leq y$ ise $Ng(y) \leq Ng(x) \quad \forall x, y \in [0,1]$

Eğer bulanık tümleyen aşağıda verilen iki özelliği de sağlıyorsa katıdır.

- Ng 'nin sürekli bir fonksiyon olması
- Eğer $x < y$ ise $Ng(y) < Ng(x) \quad \forall x, y \in [0,1]$

Eğer bulanık tümleyen aşağıdaki özelliği sağlıyorsa kuvvetli bulanık tümleyendir.

- $Ng(Ng(x)) = x \quad \forall x \in [0,1]$

$Ng(x) = 1 - x$ ve $Ng(x) = \sqrt{1 - x^2}$ kuvvetli bulanık tümleyenlerdir [17, 18].

3.3. Dilsel Özetlemenin Temel Tanımlamaları

Tanımlayıcı tekniklerden en basiti istatistiksel özetlemedir. Günlük hayatta insanlar doğal dil kullanılarak elde edilen özetleri istatistiksel metotlarla elde edilen özetlerden daha az kesin olmasına rağmen daha iyi anlayabilmektedir. Doğal dilde azı, birkaçı, çoğu gibi belirsiz anlamlara sahip bazı niceleyiciler bulunmaktadır. Bu dilsel terimler literatürde genellikle dilsel özetleme proseslerinde kullanıldığından Zadeh [21, 25, 26] tarafından önerilen bulanık niceleyici kavramı dilsel özetleme çalışmalarının temelini oluşturmaktadır. Dilsel özetleme çalışmaları verinin özetlenmesine yeni bir yaklaşım olarak Yager tarafından geliştirilmiştir [3]. Veri tabanının dilsel özetlenmesindeki amaç, özlü ve kolay anlaşılabilen bilgi çıkarmaktır.

Bulanık kümelerle dilsel özetlemede kullanılan temel semboller Çizelge 3.4'te tanımlamaları ve örnekleriyle birlikte verilmiştir.

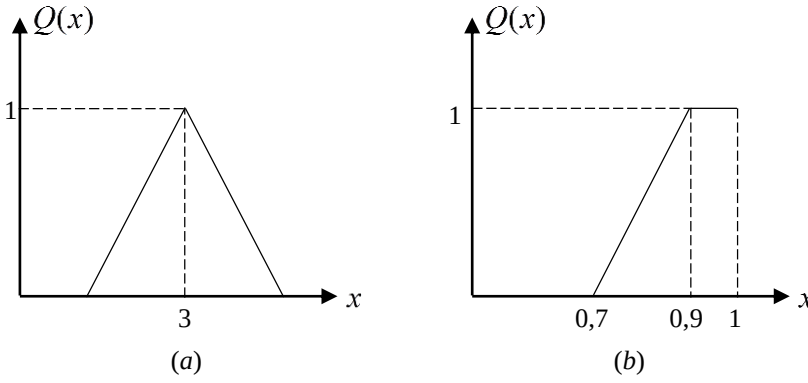
Çizelge 3.4. Dilsel özetlemede kullanılan notasyonlar

Sembol	Açıklama	Örnek
$\mathbb{D}$	Veri tabanı	İnsan kaynakları veri tabanı
Y	Nesneler kümesi	Çalışanlar
y_m	Veri tabanındaki m . nesne	Veri tabanındaki m . çalışan
v_k	k . özellik	Yaş, aylık kazanç
X_k	v_k 'nın tanım kümesi	[35-60] yaş, [5-10] bin TL kazanç
$v_k^m \equiv v_k(y_m)$	m . kayıtın k . özelliğe ait değeri	m . çalışanın yaşı
d_m	m . nesnenin tüm özellikleri	<35 yaş, 5 bin TL, ...>
S_k	Özetleyici	Genç, orta vb.
Q	Niceleyici	Çoğu, yarısı, hepsi vb.
$w_g(S_g)$	Öncül özetleyici	Genç, orta vb.
T_1	Doğruluk derecesi	[0-1] aralığında bir değer

Veritabanı $\mathbb{D}$ (Eş. 3.15), Y kümesine ait nesnelerle ilgili verilerin toplamından oluşmaktadır.

$$\mathbb{D} = \left\{ \langle v_1^1, v_2^1, \dots, v_K^1 \rangle, \langle v_1^2, v_2^2, \dots, v_K^2 \rangle, \dots, \langle v_1^M, v_2^M, \dots, v_K^M \rangle \right\} \equiv \{d_1, d_2, \dots, d_M\} \quad (3.15)$$

Dilsel özetleme çalışmalarının çoğunda Zadeh tarafından önerilen mutlak ve göreceli niceleyicilere dayalı 2 özet cümle yapısı kullanılmaktadır [26]. Mutlak niceleyiciler negatif olmayan tamsayılar üzerinde bir olabilirlik dağılımı olarak tanımlanır ve “yaklaşık 3”, “2 ile 4 arasında” niceleyicileri örnek verilebilir. Göreceli niceleyiciler $[0,1]$ aralığında bir olabilirlik dağılımı olarak tanımlanır ve “en az yarısı”, “çoğu” niceleyicileri örnek verilebilir [27]. Şekil 3.5’te mutlak niceleyici ve göreceli niceleyici gösterilmiştir. Bunlara ek olarak, Q niceleyicisi azalmayansa, $Q(0) = 0$ ve $Q(1) = 1$ ise tutarlı niceleyicidir [28, 29].



Şekil 3.5. (a) “yaklaşık 3” mutlak niceleyicisi, (b) “çoğu” göreceli niceleyicisi

Mutlak niceleyici ve göreceli niceleyici literatürde yaygın olarak kullanılmasına rağmen farklı tip bilgiyi sunmada bazen başarısız olmaktadır. Eğer özellikler bulanık kümelerle modelleniyor ise daha fazla bilgi içeren özet yapılarının mutlak ve göreceli bulanık niceleyicilerle modellenmesi mümkün değildir. “Yaklaşık beşi hariç tüm öğrenciler çalışandır” gibi bir ifade mutlak ve göreceli bulanık niceleyicilerle modellenemez. Bu nedenle farklı tip bilgiyi gösteren farklı tip nicelendirilmiş cümlelerin oluşturulmasında klasik niceleyiciler ve bulanık niceleyiciler arasında kalan yarı bulanık niceleyici kavramı geliştirilmiştir [30].

Dikkate alınan 2 çeşit özet cümle yapısı aşağıda kısaca tanımlanmıştır;

1. Tip-I niceleyici cümle yapısı: " Q $Y(nesneler)$ $S = \{S_1, S_2, \dots, S_K\}$ *dır*." $[T_1]$
2. Tip-II niceleyici cümle yapısı: " Q $w_g(S_g)$ $Y(nesneler)$ $S = \{S_1, S_2, \dots, S_K\}$ *dır*." $[T_1]$

Burada Q dilsel niceleyiciyi (örneğin, yarısı, çoğu vb.), Y nesneler kümesini (örneğin, çalışanlar), S_k özetleyiciyi, $w_g(S_g)$ öncül özetleyiciyi (örneğin, genç çalışan, düşük gelir vb.) ve T_1 özetin veriyi ne kadar sağladığının derecesini gösterir. Özet veriyi ne kadar iyi yansıtıyorsa T_1 değeri o kadar yüksektir. Tip-I niceleyici cümleler $w_g(S_g) = X$ olan tip-II niceleyici cümlelerin özel bir halidir.

Mutlak niceleyici (Q_{mut}) ve göreceli niceleyicilerin ($Q_{gör}$) her ikisi de her iki tip cümle yapısında kullanılabilmesine rağmen, değerlendirme metotları açısından tip-I niceleyici cümleler mutlak niceleyicilerle, tip-II niceleyici cümleler göreceli niceleyicilerle ilişkilendirilmiştir, çünkü;

- " $Q_{gör} Y(nesneler) S' dir.$ " tip-I niceleyici cümlesi, $Q_{mut}(i) = Q_{gör}(i/|X|)$ olduğundan " $Q_{mut} Y(nesneler) S' dir.$ " tip-I niceleyici cümlesi ile aynı sonucu sağlar.
- " $Q_{mut} w_g(S_g) Y(nesneler) S' dir.$ " tip-II niceleyici cümlesi " $Q_{mut} Y(nesneler) S \cap w_g(S_g)' dir.$ " tip-I niceleyici cümlesi ile aynı sonucu sağlar. Örneğin, "Yaklaşık 30 iyi öğrenci gençtir." ile "Yaklaşık 30 öğrenci genç ve iyidir." cümleleri değerlendirildiğinde aynı sonucu sağlar [27].

Temel doğruluk derecesi T_1 hesaplaması Z_Q yöntemine göre niceleyici cümlelerin Q niceleyicisinin mutlak veya göreceli olmasına göre sırasıyla Eş. 3.16 ve Eş. 3.17 ile hesaplanır [26].

$$Z_Q(S) = \mu_{Q_{mut}} \left(\sum_{m=1}^M \mu_S(d_m) \right) \quad (3.16)$$

$$Z_Q(S|w_g(S_g)) = \mu_{Q_{gör}} \left(\frac{\sum_{m=1}^M (\mu_S(d_m) \otimes \mu_{w_g(S_g)}(v_g^m))}{\sum_{m=1}^M \mu_{w_g(S_g)}(v_g^m)} \right) \quad (3.17)$$

Oluşturulan niceleyici cümlede birden fazla özetleyici bulunuyorsa bunların kesişimi Eş. 3.18'deki gibi t-norm operatörüyle elde edilir ve doğruluk derecesi hesaplamasına dahil edilir.

$$\mu_s(d_m) = (\mu_{s_1}(v_1^m) \otimes \mu_{s_2}(v_2^m) \otimes \dots \otimes \mu_{s_k}(v_k^m)) \quad (3.18)$$

Tip-I niceleyici cümleye “Çoğu çalışan gençtir.”[0,5] ifadesi örnek verilebilir. Burada “çoğu” niceleyiciyi, “çalışan” nesneleri, “genç” özetleyiciyi ve “0,5” ise doğruluk derecesini göstermektedir. Tip-II niceleyici cümleye “Çoğu genç çalışanın aylık geliri düşüktür.” [0,7] ifadesi örnek verilebilir. Burada “çoğu” niceleyiciyi, “genç” öncül özetleyiciyi, “çalışan” nesneleri, “düşük gelir” özetleyiciyi ve “0,7” ise doğruluk derecesini göstermektedir.

Elde edilecek özetlerin sayısı niceleyici, öncül özetleyici, özetleyici ve veri tabanındaki özellik sayısının mümkün kombinasyonlarının sayısı kadardır ve çok sayıda özet oluşabilir. Bu nedenle yüksek kalitede cümlelerin belirlenmesi önemlidir. Kaliteli cümlelerin belirlenmesi işlem hem özet üretmeden önce hem de sonra yapılabilir. Özetler üretilmeden yapılması hesaplama zamanını azaltırken, sonra yapılması bu özetleri okuyacak kullanıcının zamanını azaltır [31]. Değerlendirilecek özetlerin sayısını azaltmada çeşitli teknikler vardır. Örneğin, bir niceleyicinin bir başkası tarafından içerildiği durumda dar anlamdaki niceleyici ile oluşturulan cümle düşük doğruluk derecesine sahipse daha geniş anlamdaki niceleyici dikkate alınır, eğer değilse dar anlamdaki niceleyicili cümle korunur ve diğeri ile hesaplama yapılmaz [32]. Başka bir çözüm doğruluk derecesi belirli bir değerin altında olan özetleri elemektir [33]. Diğer bir yöntem ise kullanıcı rehberli özetlemedir. Bu yöntemde kullanıcılar özette yer alacak önermeleri seçerler ve sadece o cümlelerin doğruluk derecesi hesaplanır [34-36]. Yüksek kalitede cümle belirlenmesinin özet oluşturmadan önce yapılmasının bir yolu genetik algoritma kullanımıdır. Genetik algoritma ile dilsel özetlemede amaç birerleme yapmadan doğruluk derecesi belirli bir eşik değerden yüksek olan dilsel özetleri elde etmektir. Bu sayede kullanışsız ve daha az önemli olan dilsel özetler elimine edilir [37-39].

3.4. Dilsel Özetlerin Değerlendirilmesi

Dilsel özetleme işleminin en önemli adımı özetlerin değerlendirilmesidir. Niceleyici cümleler doğruluk derecesinin hesaplanmasıyla değerlendirilir. Eğer doğruluk derecesi

güvenilir olarak hesaplanmaz ise elde edilen özetler gerçeği yansıtmaz ve yanlış sonuçlar çıkarılmasına neden olur. Bu nedenle doğruluk derecesinin doğru ve güvenilir bir şekilde hesaplanması oldukça önemlidir. Delgado ve diğerleri [29] ve Barro ve diğerleri [40] tip-I ve tip-II niceleyici cümlelerin değerlendirilmesinde kullanılan doğruluk derecesinin sağlaması gereken özellikleri önermişlerdir. Delgado ve diğerlerinin önerdiği özellikler alt başlıklarda verilmiştir [29].

3.4.1. Tip-I niceleyici cümlelerin değerlendirilmesinde doğruluk derecesinin sağlaması gereken özellikler

Özellik 1.1: (Kesin durum) S kesin küme ve Q göreceli niceleyici ise, doğruluk derecesi $Q(|S|/|Y|)$ 'e, Q mutlak niceleyici ise $Q(|S|)$ 'e eşit olmalıdır.

Özellik 1.2: Niceleyicinin “tümü” ($\forall$) ve “var” ($\exists$) olduğu durumlarda, değerlendirme işlemi bulanık mantıkla tutarlı olmalıdır. $Q = \exists$ olduğu durumda $\bigoplus_{m=1}^M \mu_s(d_m)$ ile, niceleyicinin $Q = \forall$ olduğu durumda $\bigotimes_{m=1}^M \mu_s(d_m)$ ile gösterilebilmeli ve değerlendirilebilmelidir.

Özellik 1.3: Değerlendirme işlemi, niceleyicilerin kapsama durumuna göre tutarlı olmalıdır. Örneğin, $Q \subseteq Q'$ olduğu durumda değerlendirme işlemi $Q(S) \leq Q'(S)$ olmalıdır.

Özellik 1.4: Değerlendirme işleminin hesaplama karmaşıklığı $O(M)$ ile $O(M \log M)$ arasında olmalıdır.

Özellik 1.5: Değerlendirme işlemi katı olmamalıdır. Örneğin, $H = \{p/q \mid p \in \{0, \dots, M\}, q \in \{1, \dots, M\}\}$ kümesi üzerinde tanımlı bir niceleyici, $Q \neq \emptyset$ ve $Q \neq H$ olduğu durumda bir S bulanık kümesi varsa doğruluk derecesi $\{0,1\}$ 'e eşit olmamalıdır.

Özellik 1.6: Değerlendirme işleminde herhangi bir niceleyici kullanılabilirdir. Negatif olmayan tamsayılar ya da $[0,1]$ aralığı üzerinde herhangi bir olabilirlik dağılımı bu duruma örnektir.

3.4.2. Tip-II niceleyici cümlelerin değerlendirilmesinde doğruluk derecesinin sağlanması gereken özellikler

Özellik 2.1: (Kesin durum) Eğer $w_g(S_g)$ ve S kesin kümeler ve Q göreceli niceleyici ise, doğruluk derecesi $Q(|S \cap w_g(S_g)|/|w_g(S_g)|)$ ’ye eşit olmalıdır.

Özellik 2.2: $w_g(S_g) = X$ olduğu durumda tip-II niceleyici cümle yapısı tip-I niceleyici cümle yapısına dönüştüğünden değerlendirme metodu tip-I cümleler için de geçerli olmalıdır.

Özellik 2.3: Değerlendirme işleminin hesaplama karmaşıklığı $O(M)$ ile $O(M \log M)$ arasında olmalıdır.

Özellik 2.4: Eğer $w_g(S_g) \subseteq S$ ve $w_g(S_g)$ normal bulanık küme ise, cümlelerin doğruluk derecesi $Q(1)$ ’e eşit olur.

Özellik 2.5: Eğer $w_g(S_g) \cap S = \emptyset$ ise, doğruluk derecesi $Q(0)$ ’a eşit olur.

Özellik 2.6: Niceleyicinin “tümü” ve “var” olduğu durumlarda, değerlendirme işlemi bulanık mantıkla tutarlı olmalıdır. Niceleyici $Q = \exists$ olduğu durumda Eş. 3.19 ile, $Q = \forall$ olduğu durumda Eş. 3.20 ile gösterilebilmelidir.

$$\bigoplus_{m=1}^M \left(\bigotimes_{k \in \{K\} \setminus g} \mu_{S_k}(v_k^m) \otimes \mu_{w_g(S_g)}(v_g^m) \right) \quad (3.19)$$

$$\bigotimes_{m=1}^M \left(\mu_{w_g(S_g)}(v_g^m) \rightarrow \bigotimes_{k \in \{K\} \setminus g} \mu_{S_k}(v_k^m) \right) \quad (3.20)$$

Özellik 2.7: Değerlendirme işleminde herhangi bir niceleyici kullanılabilirdir. [0-1] aralığı üzerinde herhangi bir olabilirlik dağılımı bu duruma örnektir.

Özellik 2.8: Değerlendirme işlemi katı olmamalıdır. Örneğin, $H = \{p/q \mid p \in \{0, \dots, M\}, q \in \{1, \dots, M\}\}$ kümesi üzerinde tanımlı bir niceleyici, $Q \neq \emptyset$ ve $Q \neq H$ olduğu durumda, S ve $w_g(S_g)$ bulanık kümesi varsa doğruluk derecesi $\{0,1\}$ 'e eşit olmamalıdır.

3.5. Metotlar

Dilsel özetleri değerlendirme yöntemleri temel olarak 3 gruba ayrılmaktadır: skalar kardinalite temelli yöntemler (sonucu skalar bir değer), bulanık kardinalite temelli yöntemler (sonucu bulanık bir küme) ve diğer yöntemler. Literatürde en yaygın yaklaşımlar skalar kardinalite ve bulanık kardinalite temelli olanlardır. Ancak daha sonraları bulanık kardinalite temelli yöntemlerden elde edilen sonucun bulanık sayı (normal ve konveks bulanık küme) olması gerektiği dikkate alınmıştır [41]. Bunun üzerine alternatif yöntemler ortaya çıkmıştır. Bu bölümde literatürde yer alan değerlendirme yöntemleri 3 başlık altında incelenmiştir.

3.5.1. Skalar kardinalite temelli yöntemler

Dilsel Özetleme işleminde tip-I ve tip-II niceleyici cümlelerin değerlendirilmesinde kullanılan bazı doğruluk dereceleri skalar kardinalite temellidir. İlk olarak Zadeh hem tip-I hem de tip-II niceleyici cümlelerin değerlendirilmesi için Σ -count temelli Z_Q yöntemini (Bkz. Eş. 3.16 ve Eş. 3.17) önermiştir [26]. Daha sonra Yager sadece tutarlı ve göreceli niceleyicilerin kullanıldığı tip-I ve tip-II niceleyici cümleler için OWA (Sıralı Ağırlıklı Ortalama) operatörü temelli Y_Q yöntemini (Eş. 3.21 ve Eş. 3.23) önermiştir [42, 43]. Bosc ve Lietard ise tip-I niceleyici cümlelerin değerlendirilmesi için Yager'ın [42] önerdiği OWA operatörü temelli yöntemin daha genel hali olan Choquet integrali temelli C_Q yöntemini önermişlerdir [44, 45].

$$Y_Q(S) = \sum_{i=1}^M w_i b_i \quad (3.21)$$

$$w_i = Q(i/M) - Q((i-1)/M), i \in \{1, 2, \dots, M\} \text{ ve } Q(0) = 0 \quad (3.22)$$

b_i , S bulanık kümesine ait elemanların üyelik dereceleri büyükten küçüğe sıralandığında i . en büyük değerdir.

$$Y_Q(S|w_g(S_g)) = \sum_{i=1}^M w_i j_i \quad (3.23)$$

$$w_i = Q(P_i) - Q(P_{i-1}), i \in \{1, 2, \dots, M\} \quad (3.24)$$

$$P_i = \frac{1}{t} \sum_{j=1}^i e_j \text{ ve } t = \sum_{j=1}^M e_j \quad (3.25)$$

j_i , $\neg w_g(S_g) \oplus S$ bulanık kümesinin üyelik dereceleri büyükten küçüğe sıralandığında i . en büyük üyelik derecesidir.

Kacprzyk ve Yager [46], Wu ve Mendel [47] tip-I ve tip-II niceleyici cümlelerin değerlendirilmesinde sadece doğruluk derecesinin kullanılmasının yeterli olmadığını savunmuşlardır. Doğruluk derecesine ek olarak Kacprzyk [48], Kacprzyk ve Strykowski [49] tarafından belirsizlik derecesi, kapsama derecesi, uygunluk derecesi ve uzunluk derecesi tanımlanmıştır. Kacprzyk ve Zadrozny bu değerlerin ağırlıklı birleştirilmesiyle elde edilen toplam geçerlilik derecesini tanımlayarak niceleyici cümleleri değerlendirmişlerdir [50].

Belirsizlik derecesi oldukça önemli bir kalite ölçümüdür. “hemen hemen tüm kış günleri soğuktur.” Gibi bir özetin doğruluk derecesi 1’e eşit olabilir fakat herkes tarafından bilinen bir ifade olduğu için içerdiği bilgi önemli değildir. Belirsizlik derecesi (T_2) Eş. 3.26 ile tanımlanmıştır.

$$T_2 = 1 - \sqrt[k]{\prod_{k=1,2,\dots,K} in(S_k)} \quad (3.26)$$

in bulanıklık derecesidir ve Eş. 3.27 ile tanımlanmıştır.

$$in(S_k) = \frac{kard \{v \in X_k : \mu_{S_k}(v_k) > 0\}}{kard X_k} \quad (3.27)$$

Kapsama derecesi (T_3) öncül özetleyicideki kaç nesnenin aynı zamanda özetleyici tarafından kapsandığını ölçen bir kalite ölçüsüdür ve Eş. 3.28 ile gösterilir.

$$T_3 = \frac{\sum_{m=1}^M t_m}{\sum_{m=1}^M h_m} \quad (3.28)$$

t_m ve h_m sırasıyla özetleyici ve öncül özetleyicinin kapsadığı nesne sayıları Eş. 3.29 ve Eş. 3.30 ile gösterilmiştir.

$$t_m = \begin{cases} 1, & \mu_S(v_g^m) > 0 \text{ ve } \mu_{w_g}(v_g^m) > 0 \\ 0, & d.d. \end{cases} \quad (3.29)$$

$$h_m = \begin{cases} 1, & \mu_{w_g}(v_g^m) > 0 \\ 0, & d.d. \end{cases} \quad (3.30)$$

Uygunluk derecesi (T_4) özetin ilginç olup olmadığını ölçer. Özet K tane özete bölünür ve hepsinde özetleyici aynıdır. Daha sonra kaç nesnenin parçalanmış özetlerde yer aldığı Eş. 3.31 ile ve T_4 Eş. 3.33 ile hesaplanır.

$$r_k = \frac{\sum_{m=1}^M t_{k,m}}{M} \quad k = 1, 2, \dots, K \quad (3.31)$$

$t_{k,m}$ Eş. 3.32 ile hesaplanır.

$$t_{k,m} = \begin{cases} 1, & \mu_{S_k}(v_k^m) > 0 \\ 0, & d.d. \end{cases} \quad (3.32)$$

$$T_4 = \left| \prod_{k=1}^K r_k - T_3 \right| \quad (3.33)$$

Uzun özetleri anlamak zor olduğundan özetin uzunluğu (T_5) önemli bir ölçüttür ve Eş. 3.34 ile gösterilmiştir.

$$T_5 = 2(0,5^{kard\ S}) \quad (3.34)$$

$kard\ S$ özetinde yer alan özetleyicilerin sayısıdır.

Toplam geçerlilik derecesi (T) 5 kalite ölçütünün ağırlıklı ortalaması alınarak Eş. 3.35'teki gibi hesaplanır.

$$T(T_1, T_2, T_3, T_4, T_5; w_1, w_2, w_3, w_4, w_5) = \sum_{i=1}^5 w_i T_i \quad (3.35)$$

Wu ve Mendel eğer/ise yapısını kullanarak dilsel özetleme işlemini gerçekleştirmişlerdir [47]. Eğer/ise yapısındaki kanonik form

$$\text{“Eğer } X_1 \text{ } S_1 \text{ ise, } X_2 \text{ } S_2 \text{ 'dir. } [T] \text{”} \quad (3.36)$$

şeklinde. Bu yapıda X_1 ve X_2 v_k 'nın tanım kümesinde bulunan elemanlar, S_1 ve S_2 özetleyiciler ve T doğruluk derecesidir. Bu kanonik forma örnek olarak “Eğer çalışanların yaşı genç ise, aylık kazançları ortadır. [0,7]” verilebilir. Bu örnekte “çalışanların yaşı” ve “aylık kazançları” veritabanında yer alan nesnelere ait özellikler, “genç” ve “orta” ise özetleyicilerdir.

Eş. 3.36 ile tanımlanan kanonik form, Eş. 3.37 ile tanımlanan forma eşdeğerdir.

$$\text{"Tüm } S_1 \text{ } Y(\text{nesneler}) \text{ } S_2 \text{ 'dir.}[T]\text{"} \quad (3.37)$$

Doğruluk derecesine ek olarak Wu ve Mendel genelliği, yararlılığı ve yeniliği dikkate alan 3 kalite ölçümünü önererek cümleleri değerlendirmişlerdir [47]. Genelliği temsil eden yeterli kapsama derecesi (T_c) özetin yeterli veri ile desteklenip desteklenmediğini ölçer. T_c 'yi hesaplamak için ilk olarak Eş. 3.38 ile verilen kapsama oranı hesaplanır.

$$r = \frac{\sum_{m=1}^M t_m}{M} \quad (3.38)$$

t_m Eş. 3.39 ile hesaplanır.

$$t_m = \begin{cases} 1, & \mu_{s_1}(v_1^m) > 0 \text{ ve } \mu_{s_2}(v_2^m) > 0 \\ 0, & d.d. \end{cases} \quad (3.39)$$

Genellikle çok küçük bir değer olduğundan r doğrudan değerlendirmede kullanılamamaktadır. Bu yüzden r_1 ve r_2 (örneğin, $r_1 = 0,02$ ve $r_2 = 0,15$ olan $0 \leq r_1 < r_2$ şartını sağlayan reel sayılar) ile belirlenen bir fonksiyon ile T_c Eş. 3.40'taki gibi hesaplanır.

$$T_c = f(r) = \begin{cases} 0, & r \leq r_1 \\ 2 \left(\frac{r - r_1}{r_2 - r_1} \right)^2, & r_1 < r < \frac{r_1 + r_2}{2} \\ 1 - 2 \left(\frac{r_2 - r}{r_2 - r_1} \right)^2, & \frac{r_1 + r_2}{2} \leq r < r_2 \\ 1, & r \geq r_2 \end{cases} \quad (3.40)$$

Fayda derecesi (T_u), çıkarılan özetin ne derece faydalı olduğunu belirtmektedir. Bir özetin faydalı olabilmesi için verilen dilsel özetin “doğruluk derecesi yüksek” ve “yeterli kapsama derecesine sahip olmalı” şartlarını sağlaması gerekir. Fayda derecesi doğruluk derecesi ve yeterli kapsama derecesi ile Eş. 3.41’deki gibi ölçülür.

$$T_u = \min(T_1, T_c) \quad (3.41)$$

Aykırılık derecesi (T_o), özetten elde edilen bilginin yeni ya da beklenmedik olduğuyla ilgilidir. Özeti yeni ya da beklenmedik olabilmesi için yeterli kapsama derecesinin düşük ve doğruluk derecesinin ya çok yüksek ya da çok düşük olması gerekir. Aykırılık derecesi Eş. 3.42’de tanımlanmıştır.

$$T_o = \begin{cases} \min(\max(T_1, 1 - T_1), 1 - T_c), & T_1 > 0 \\ 0, & T_1 = 0 \end{cases} \quad (3.42)$$

Basitlik derecesi (T_s) özetin uzunluğunu ölçmektedir. Özette yer alan öncüllerin ve ardılların sayısına (l) göre basitlik derecesi hesaplanmaktadır. Basitlik derecesinin formülü Eş. 3.43 ile verilmiştir.

$$T_s = 2^{2-l} \quad (3.43)$$

Bulanık dilsel özetlerin yorumlanabilirliği de önemli bir konu haline gelmiştir. Bu konuda Lesot ve diğerleri hem cümlelerin hem de cümleler kümesi olan özetlerin yorumlanabilirliği üzerinde çalışmışlardır [31]. İlk olarak cümle seviyesinde değerlendirme yapılmaktadır; cümlelerin kalitesi belirlenir, kaliteli cümleler seçilir ve dilsel çeviri yapılır. Cümleler kümesinin (özetlerin) değerlendirilmesi ise bireysel cümlelerin değerlendirilmesine ve bütün olarak yorumlanabilmesine bağlıdır. Cümle seviyesinde yorumlama hem ölçülen kalite derecesine hem de dilsel ifadeye bağlıyken, özet seviyesinde yorumlamada bunlara ek olarak tutarlılık, gereksiz olmama ve taşınan bilgi gibi farklı özellikler de dikkate alınmaktadır.

Çelişmemelik ve çift olumsuzlama özelliklerini sağlayan bir özet tutarlıdır. Çelişmemelik özelliğine göre çelişen terimlere sahip iki cümlelerin doğruluk dereceleri toplamı 1'e eşit olmalıdır. Örneğin, C : "Çoğu genç insan uzundur.", C_1 : "Azı genç insan uzundur." ve C_2 : "Çoğu genç insan kısadır." cümleleri dikkate alınsın. C_1 ve C_2 cümleleri C ile çelişkilidir ve doğruluk dereceleri $T_1(C_1) = T_1(C_2) = 1 - T_1(C)$ olmalıdır. Çift olumsuzlama özelliği cümleye iki çelişki uygulandığında doğruluk derecesinin değişmemesi gerektiği anlamına gelmektedir. Yani, C : "Azı genç insan kısadır." cümlesi ile bu cümlelerin çift olumsuzlaması "Çoğu genç insan uzundur." cümlesinin benzer anlamlara sahip olduğu dikkate alınır.

Aynı anlamlara sahip farklı cümleler özetinde yer aldığı zaman özet gereksizce uzamaktadır. Çift olumsuzlama gereksizliğe bir örnektir. Özetinde çift olumsuzlama durumu olduğunda iki cümleden biri özette çıkarılmalıdır. Diğer gereksizlik durumları içerilme ve benzerliktir. İçerilme bir cümlelerin öncül özetleyicisi, özetleyicisi veya niceleyicisi bir başka cümlelerin öncül özetleyicisi, özetleyicisi veya niceleyicisi tarafından içerilmesi anlamına gelir. Örneğin, "%80'den fazla" niceleyicisinin desteği "çoğu" niceleyicisinin desteği tarafından içerildiği durumda, "%80'den fazla genç insan iyi gelirlidir." cümlesi "Çoğu genç insan iyi gelirlidir." cümlesi tarafından içerilir. Cümlelerin benzerliği de bir gereksizlik ölçüsü olarak dikkate alınmaktadır. $C_1 = "Q_1 w_g(S_g)_1 Y S_1"$ ve $C_2 = "Q_2 w_g(S_g)_2 Y S_2"$ cümleleri ve doğruluk dereceleri dikkate alındığında iki cümlelerin benzerliği Eş. 3.44 ile tanımlanır.

$$\text{sim}(C_1, C_2) = \min\left(\text{sim}_1(Q_1, Q_2), \text{sim}_1(S_1, S_2), \text{sim}_1(w_g(S_g)_1, w_g(S_g)_2), \text{sim}_2(T_1(C_1), T_1(C_2))\right) \quad (3.44)$$

sim_1 iki bulanık küme arasındaki benzerliği ve sim_2 iki gerçek değer arasındaki benzerliği ölçer.

İki cümlelerin benzerliği kullanıcı tarafından belirlenen bir eşik değerden yüksek olduğunda cümlelerden birisi özette çıkarılır.

Taşınan bilgi sadece cümlelerin taşıdığı bilgiler değil aynı zamanda onlar arasındaki ilişkidir. Cümleler arasındaki ilişkiye dayalı bilginin kaynakları cümle yorumlama ve ardında yatan anlamdır. Cümleler arasındaki nedensellik onların anlamlarını geliştirmeye

izin verir. Örneğin, " $Q_1 A Y(nesneler) B$ 'dir." ve " $Q_2 B Y(nesneler) C$ 'dir." cümleleri yorumlandığında Q_1 ve Q_2 bulanık sayılarının çarpımı olarak tanımlanan Q ile " $Q Y(nesneler) A$ ve C 'dir." formunda yeni bir cümle çıkarılabilir. Ardında yatan anlam özelliği şu şekilde örneklenebilir: Tüm B kipleri ile " $Q B Y(nesneler) A$ 'dır." cümleleri incelendiğinde, her B için doğruluk derecesi yüksek çıkıyorsa " $Q B Y(nesneler) A$ 'dır." cümleleri " $Q Y(nesneler) A$ 'dır." cümlesi ile değiştirilebilir. Bu durum aynı zamanda " $\text{Özellik}(B)$ 'ye bağlı kalmadan $Q Y(nesneler) A$ 'dır." formunda yorumlanabilir. Örneğin, B kipi boyut değişkeni ile ilgili ise "Boyutuna bağlı kalmadan çoğu kitap az satar." cümlesi bu forma örnektir.

Skalar kardinalite temelli yöntemlerin değerlendirmede kullanılması bazı durumlarda tutarsız sonuç vermektedir. Bunun sebebi küçük üyelik derecelerinin toplamının büyük üyelik dereceleri ile farkının ayırt edilememesinden kaynaklanmaktadır. Örnek 2.1. ve Örnek 2.2. ile skalar kardinalitenin tutarsız sonuç verdiği gösterilmiştir.

Örnek 2.1.

$w_g(S_g) = \left\{ \frac{1}{y_1}, \frac{0}{y_2} \right\}$ ve $S = \left\{ \frac{0}{y_1}, \frac{1}{y_2} \right\}$, $Y = \{y_1, y_2\}$ evrensel uzayında iki bulanık küme olsun.

Minimum operatörü t-norm olarak kullanıldığında bu kümelerin kesişimi $S \cap w_g(S_g)$ aşağıda verilmiştir.

$$S \cap w_g(S_g) = \left\{ \frac{0}{y_1}, \frac{0}{y_2} \right\}$$

$w_g(S_g)$ ve S farklı kümeler olup $Q(a) = a$ için $T = 0$ olmaktadır [51, 52].

Örnek 2.2.

$$w_g(S_g) = \left\{ \frac{1}{y_1}, \frac{0,01}{y_2}, \frac{0,01}{y_3}, \dots, \frac{0,001}{y_{1000}} \right\}$$

$$S = \left\{ \frac{0,01}{y_1}, \frac{1}{y_2}, \frac{0,01}{y_3}, \dots, \frac{0,001}{y_{1000}} \right\}$$

$Y = \{y_1, y_2, \dots, y_{1000}\}$ evrensel uzayında iki bulanık küme olsun. Minimum operatörü t-norm olarak kullanıldığında bu kümelerin kesişimi $S \cap_{w_g}(S_g)$ aşağıda verilmiştir:

$$S \cap_{w_g}(S_g) = \left\{ \frac{0,01}{y_1}, \frac{0,01}{y_2}, \frac{0,01}{y_3}, \dots, \frac{0,01}{y_{1000}} \right\}$$

$w_g(S_g)$ ve S farklı kümeler olmalarına rağmen $Q(a) = a$ alındığında $T = \frac{10}{10,99} = 0,91$

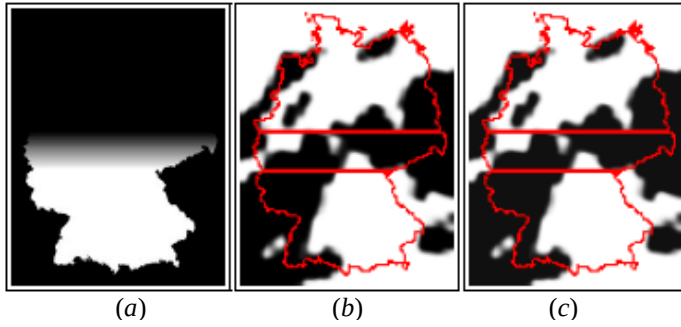
gibi büyük bir değer çıkmaktadır. Bunun nedeni küçük üyelik dereceleri ile büyük üyelik dereceleri arasındaki farkın ayırt edilememesidir. Ayrıca küçük üyelik dereceleri hesaplamada etkin rol oynamaktadır. 0,01 gibi 0'a yakın değerler nedeniyle T değeri çok büyük çıkmaktadır. Bu örnek skalar kardinalitenin doğruluk derecesini hesaplamada kullanılmasının tutarsız sonuçlar ürettiğini göstermektedir [51, 52].

Knoll ve Glöckner tip-I ve tip-II niceleyici cümlelerin değerlendirilmesi için önerilen bazı yöntemlerin tutarsız sonuç ürettiklerini meteoroloji veri seti üzerinde göstermişlerdir (Örnek 2.3.) [53].

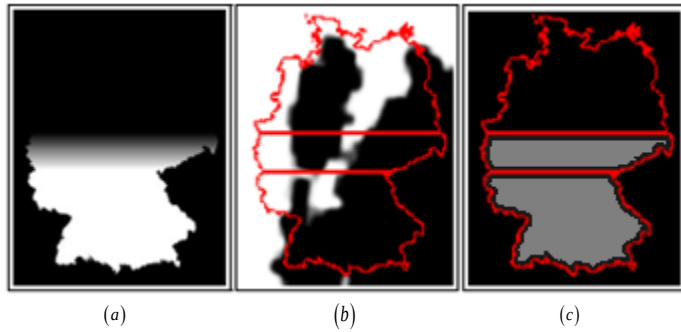
Örnek 2.3.

Şekil 3.6a'da Almanya'nın güney kesimi gösterilmiştir. Beyaz ile gösterilen alanın güney Almanya'ya ait olma derecesi bire eşit iken gri ve grinin tonu olan renklerin aitlik derecesi sıfır bir aralığında bir değere eşittir. Siyah renkle gösterilen alanın aitlik derecesi sıfıra eşittir. Şekil 3.6b ve Şekil 3.6c'de beyaz alan havanın bulutlu olduğunu göstermektedir. "Almanya'nın güneyinin en az %60'ı bulutludur" ifadesi ele alınsın. Şekil 3.6b için bu ifadenin doğruluk derecesi $Z_Q = 0$ 'a eşit iken, Şekil 3.6c için doğruluk derecesi $Z_Q = 1$ 'e eşittir. İki resim arasında çok az bir farklılık olmasına rağmen doğruluk derecesindeki değişim oldukça fazladır. Bu durumun nedeni Zadeh tarafından önerilen yöntemin Özellik 2.8'i sağlamamasıdır. Çok az değişimin olduğu bir durumda doğruluk derecesi büyük oranda değişmektedir.

Şekil 3.7’de “Almanya’nın güneyinin yaklaşık %10’u bulutludur” ifadesi ele alınmıştır. Şekil 3.7b’de $Z_Q = 1$ ’e eşittir ve bu beklenen bir değerdir; çünkü Almanya’nın güneyinin %10’luk bir kesimi bulutludur. Şekil 3.7c’deki durumda da $Z_Q = 1$ ’e eşittir. Bu durum beklenmeyen bir durumdur; çünkü Almanya’nın güneyi çok az bulutludur. Diğer bir deyişle bu durumun bulutlu olma kümesine üyeliği 0,1’e eşittir. Şekil 3.7c’deki resimden “Almanya’nın güneyinin yaklaşık %10’u bulutludur.” gibi bir ifadenin çıkarılması yanlıştır; çünkü Almanya’nın güneyinin tamamı az bulutludur. Şekil 3.7c’deki durumda $Z_Q = 1$ ’e eşit olmasının nedeni doğruluk derecesini hesaplamada fazla sayıdaki küçük üyelik derecelerinin toplamının az sayıdaki büyük üyelik derecelerinin toplamına eşit olmasıdır.



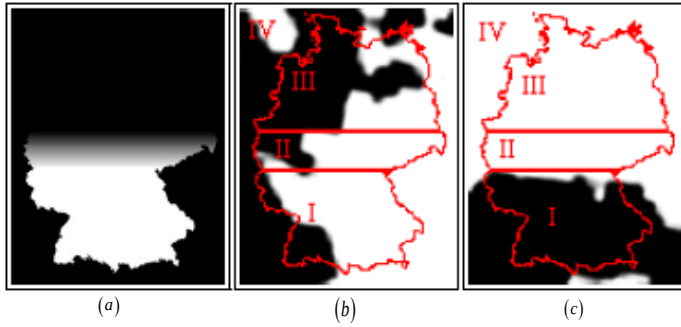
Şekil 3.6. (a) Almanya’nın güneyi (b) “Almanya’nın güneyinin en az %60’ı bulutludur” $Z_Q = 0$ (c) “Almanya’nın güneyinin en az %60’ı bulutludur” $Z_Q = 1$



Şekil 3.7. (a) Almanya’nın güneyi (b) “Almanya’nın güneyinin yaklaşık %10’u bulutludur” $Z_Q = 1$ (c) “Almanya’nın güneyinin yaklaşık %10’u bulutludur” $Z_Q = 1$

Şekil 3.8’de “Almanya’nın güneyinin en az %60’ı bulutludur” ifadesinin doğruluk derecesi Yager’ın önerdiği OWA operatörü temelli yöntem (Bkz. Eş. 3.23) kullanılarak hesaplanmıştır. Şekil 3.8b’deki durum için doğruluk derecesi 0,1’e eşit çıkmıştır. Şekil

3.8c'deki durum için ise doğruluk derecesi 0,6'ya eşit çıkmıştır. Şekil 3.8b'den de anlaşılacağı gibi Almanya'nın güneyinin en az %60'lık kesimi bulutludur. Bu nedenle doğruluk derecesinin 1'e eşit çıkması beklenmektedir. Şekil 3.8c'deki durum için doğruluk derecesinin 0'a yakın bir değer çıkması beklenmektedir. Ayrıca Şekil 3.8c'deki durum için hesaplanan doğruluk derecesi Şekil 3.8b'deki durum için hesaplanan doğruluk derecesinden daha büyüktür.



Şekil 3.8. (a) Almanya'nın güneyi (b) “Almanya'nın güneyinin en az %60'ı bulutludur” $Y_Q = 0,1$ (c) “Almanya'nın güneyinin en az %60'ı bulutludur” $Y_Q = 0,6$

3.5.2. Bulanık kardinalite temelli yöntemler

Doğruluk derecesinin hesaplanmasında skalar kardinalite kullanımı bazı durumlarda tutarsız sonuçlar üretmektedir. Bu nedenle doğruluk derecesinin hesaplanmasında skalar kardinalite yerine bulanık kardinaliteyi kullanan yöntemler önerilmiştir. Delgado ve diğerleri tip-I niceleyici cümleler için E tipi bulanık kardinalitenin kullanıldığı G_Q metodunu, tip-I ve tip-II niceleyici cümleler için ED tipi bulanık kardinalitenin kullanıldığı olasılık temelli GD_Q metodunu ve Z tipi bulanık kardinalitenin kullanıldığı olabilirlik temelli ZS_Q metodunu önermişlerdir [29]. GD_Q metodu Eş. 3.45 ile, ZS_Q metodu Eş. 3.46 ile verilmiştir.

$$GD_Q(S|w_g(S_g)) = \sum_{\alpha_i \in \Gamma(S|w_g(S_g))} (\alpha_i - \alpha_{i+1}) \times Q \left(\frac{|S \cap w_g(S_g)|_{\alpha_i}}{|w_g(S_g)|_{\alpha_i}} \right) \quad (3.45)$$

$$ZS_Q(S|w_g(S_g)) = \max_{\alpha_i \in \Gamma(S|w_g(S_g))} \min \left(\alpha_i, Q \left(\frac{|S \cap w_g(S_g)|_{\alpha_i}}{|w_g(S_g)|_{\alpha_i}} \right) \right) \quad (3.46)$$

Eş. 3.45 ve Eş. 3.46’da verilen $\Gamma(S|w_g(S_g))$, $S \cap w_g(S_g)$ ve $w_g(S_g)$ kümelerinde bulunan elemanların üyelik derecelerinin birleşimiyle oluşan bir kümedir (Eş. 3.47). $w_g(S_g)$ bulanık kümesi normal bulanık küme olmak zorundadır. Eğer normal bulanık küme değilse normalize edilir ve aynı katsayı ile $S \cap w_g(S_g)$ bulanık kümesi de normalize edilmelidir.

$$\Gamma(S|w_g(S_g)) = \Gamma(S \cap w_g(S_g)) \cup \Gamma(w_g(S_g)) \quad (3.47)$$

$\Gamma(S|w_g(S_g)) = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ kümesi α kesmelerinden oluşan bir küme olup $1 = \alpha_1 > \alpha_2 > \dots > \alpha_m > \alpha_{m+1} = 0$ şartını sağlamaktadır.

Örnek 2.4.

Örnek 2.2. ile aynı veri kullanılarak problem bulanık kardinalite temelli yöntemler kullanılarak çözülmüştür. Elde edilen sonuçlar $ZS_Q(S|w_g(S_g)) = 0,01$, $GD_Q(S|w_g(S_g)) = 0,01$ olarak bulunmuştur. Bulanık kardinalite temelli yöntemler kullanıldığında sonuç 0,01 gibi küçük bir değere eşit olmuştur.

Örnek 2.4. ile bulanık kardinalitenin doğruluk derecesi hesaplamada skalar kardinaliteye göre daha tutarlı sonuçlar ürettiği gösterilmiştir.

Tip-I ve tip-II niceleyici cümlelerin değerlendirilmesi için önerilen skalar kardinalite temelli ve bulanık kardinalite temelli yöntemlerin sağladıkları özellikler Çizelge 3.5 ve Çizelge 3.6’da verilmiştir. Çizelge 3.7’de var olan yöntemler arasındaki ilişkiler gösterilmiştir.

Çizelge 3.5. Tip-I niceleyici cümlelerin değerlendirilmesi için önerilen yöntemlerin sağladığı özellikler

	1.1	1.2	1.3	1.4	1.5	1.6
$Z_Q(S)$	✓	×	✓	$O(M)$	×	✓
$Y_Q(S)$	✓	✓	✓	$O(M \log M)$	✓	×
$GD_Q(S)$	✓	✓	✓	$O(M \log M)$	✓	✓
$ZS_Q(S)$	✓	✓	✓	$O(M \log M)$	✓	✓

O: Zaman karmaşıklığı, M: Nesne sayısı

✓: Özellik sağlanıyor. ×: Özellik sağlanmıyor.

Çizelge 3.6. Tip-II niceleyici cümlelerin değerlendirilmesi için önerilen yöntemlerin sağladığı özellikler

	2.1	2.2	2.3	2.4	2.5	2.6- $\exists$	2.6- $\forall$	2.7	2.8
$Z_Q(S w_g(S_g))$	✓	✓	$O(M)$	✓	✓	×	×	✓	×
$Y_Q(S w_g(S_g))$	✓	✓	$O(M \log M)$	×	×	×	✓	×	×
$GD_Q(S w_g(S_g))$	✓	✓	$O(M \log M)$	✓	✓	✓	?	✓	✓
$ZS_Q(S w_g(S_g))$	✓	✓	$O(M \log M)$	✓	✓	✓	✓	✓	✓

✓: Özellik sağlanıyor. ×: Özellik sağlanmıyor.

?: Özelliğin sağlanıp sağlanmadığı ispat edilmemiştir.

Çizelge 3.7. Yöntemler arasındaki ilişkiler

	Olasılıksal		Olabilirlikli	
	Yöntem	Kardinalite	Yöntem	Kardinalite
Tip-II niceleyiciler	$GD_Q(S w_g(S_g))$	$ER(S w_g(S_g))$	$ZS_Q(S w_g(S_g))$	$ES(S w_g(S_g))$
Tip-I Göreceli ve mutlak niceleyiciler	$\downarrow$ $GD_Q(S)$	$\downarrow$ $ES(S)$	$\downarrow$ $ZS_Q(S)$	$\downarrow$ $Z(S)$
Tip-I Göreceli ve tutarlı niceleyiciler	$\downarrow$ Choquet (Yager-OWA)		$\downarrow$ Sugeno	

Örnek 2.5.

$w_g(S_g)$ ve S , $Y = \{y_1, y_2\}$ evrensel uzayında iki bulanık küme:

$$w_g(S_g) = \left\{ \frac{1}{y_1}, \frac{0,99}{y_2} \right\} \quad S = \left\{ \frac{1}{y_1}, \frac{0,01}{y_2} \right\}$$

ve $Q = \forall$ olsun. İki bulanık küme ve t-norm minimum alındığında bu kümelerin kesişimi $S \cap w_g(S_g)$ kümesi aşağıda verilmiştir.

$$S \cap w_g(S_g) = \left\{ \frac{1}{y_1}, \frac{0,01}{y_2} \right\}$$

Problem olasılık ve olabilirlik temelli bulanık kardinalite yöntemleriyle çözüldüğü takdirde $GD_{\forall}(S|w_g(S_g)) = 0,02$ ve $ZS_{\forall}(S|w_g(S_g)) = 1$ sonuçları elde edilir. Sonuçlardan görüldüğü gibi iki durum arasındaki fark oldukça yüksektir. Bu örnekte görüldüğü gibi $ZS_{\forall}(S|w_g(S_g)) = 1$ sonucu tutarlı bir sonuç değildir. Çünkü bu yöntemde maksimum ve minimum operatörlerinin kullanılmasıyla $w_g(S_g)$ ve S kümelerinde bulunan bazı elemanlar hakkındaki bilgiler göz ardı edilmektedir [28].

Bulanık kardinalite temelli yöntemlerden elde edilen sonucun bulanık sayı olması gerektiği savunulmuştur [54, 55]. Chamorro-Martinez ve diğerleri bazı durumlarda bu sonucun elde edilemediğini Örnek 2.6. ile göstermişlerdir [41].

Örnek 2.6.

$A = \left\{ \frac{1}{x_1}, \frac{0,5}{x_2}, \frac{0,5}{x_3} \right\}$ bulanık kümesi dikkate alınsın. Sadece x_1 A kümesine kesinlikle ait

olduğundan A 'nın kardinalitesi 1 değerini alabilir veya eğer A 'daki elemanların kabul kriterini genişletirsek x_2 ve x_3 de A 'ya ait olacağından A 'nın kardinalitesi 3 değerini alabilir. Ancak $x_2 \in A$ ise $x_3 \in A$ olacağından A 'nın kardinalitesi 2 değerini alamaz. Bu yüzden kardinalite konveks değildir ve sonuç olarak bulanık sayı elde edilememektedir.

Literatürde tip-I niceleyici ve tip-II niceleyici cümlelerin değerlendirilmesinde bulanık kardinalite temelli farklı yöntemler önerilmiştir. Sanchez ve diğerleri [54, 55] ve Delgado ve diğerleri [56] α kesmeleri yerine bulanık kümeleri kullanan kısıtlama seviyesini önermişlerdir. Kısıtlama seviyesi kullanılarak tip-II niceleyici cümlelerin değerlendirilmesi için Sanchez ve diğerleri niceleyici ve kardinalite arasındaki uyumluluğu hesaplayan bir yöntem önermişlerdir [54, 55].

Bulanık kardinalite temelli yöntemlerden elde edilen sonucun bulanık sayı olması gerektiği savunulduğu için alternatif yöntemler ortaya çıkmıştır [54, 55]. Rocacher bulanık kardinaliteye karşılık gelen dereceli doğal tamsayıları tanımlamıştır [57]. Lietard ve Rocacher [58], Rocacher ve Bosc [59] tip-II niceleyici cümleleri değerlendirmek için dereceli sayı temelli bir yöntem önermişlerdir. Böylece bulanık kardinalite temelli bu yöntem ile sonuçta bulanık sayı elde edilmiştir. Dereceli sayılar kısıtlama seviyesinin özel bir halidir [55]. Ancak iki yöntem tamamen aynı değildir. Dereceli sayı temelli değerlendirme yönteminde kardinalitesi kesin (0 veya 1) olan kısıtlamanın her seviyesinde cümle değerlendirilir ve daha sonra bu farklı değerlendirmeler tek bir değerde birleştirilirken, kısıtlama seviyesi temelli yöntemde kardinalitesi $[0,1]$ aralığında olan kısıtlamaların her seviyesinde cümle değerlendirilerek son bir değerde birleştirilir. Örnek 2.7. ile bu farklılık gösterilmiştir [41].

Örnek 2.7.

"Çoğu $w_g(S_g)$ $Y(nesneler)$ S 'dir." cümlesi dikkate alınsın. Nesnelerin $w_g(S_g)$ ve S bulanık kümelerine aitlik dereceleri aşağıdaki gibidir;

$$w_g(S_g) = \left\{ \frac{1}{y_1} + \frac{0,8}{y_2} + \frac{0,6}{y_3} + \frac{0,4}{y_5} \right\}$$

$$S = \left\{ \frac{0,9}{y_1} + \frac{0,6}{y_3} + \frac{0,5}{y_4} \right\}$$

$\Lambda_{w_g(S_g)} = \{1; 0,8; 0,6; 0,4\}$ ve $\Lambda_S = \{1; 0,9; 0,6; 0,5\}$ elde edilir. “Çoğu” niceleyicisi Eş. 3.48 ile tanımlandığında Çizelge 3.8 iki kümenin kısıtlama seviyelerini, dereceli sayı gösterimini ve değerlendirmeleri göstermektedir.

$$Q_{\text{çoğu}}(x) = \begin{cases} 0 & x \leq 0,5 \\ 1 & x \geq 0,75 \\ 4(x - 0,5) & \text{dd} \end{cases} \quad (3.48)$$

Çizelge 3.8. $w_g(S_g)$ ve S 'nin kısıtlama seviyesi ve $gc(S / w_g(S_g))$, E ve B 'nin dereceli sayıları

α	$\rho_{w_g(S_g)}(\alpha)$	$\rho_S(\alpha)$	$gc(S / w_g(S_g))$	E	B
1	$\{x_1\}$	$\emptyset$	0	0	0
0,9	$\{x_1\}$	$\{x_1\}$	1	1	1
0,8	$\{x_1, x_2\}$	$\{x_1\}$	1/2	0	0
0,6	$\{x_1, x_2, x_3\}$	$\{x_1, x_3\}$	2/3	2/3	1
0,5	$\{x_1, x_2, x_3\}$	$\{x_1, x_3, x_4\}$	2/3	2/3	1
0,4	$\{x_1, x_2, x_3, x_5\}$	$\{x_1, x_3, x_4\}$	1/2	0	0

$gc(S / w_g(S_g))$, E ve B değerleri sırasıyla Eş. 3.49, Eş. 3.50 ve Eş. 3.51 ile hesaplanmıştır.

$$\mathfrak{R}_{|S/w_g(S_g)|}(\alpha) = \frac{|S_\alpha \cap w_g(S_g)_\alpha|}{|w_g(S_g)_\alpha|} \quad (3.49)$$

$$\mathfrak{R}_E(\alpha) = Q\left(\frac{|\rho_{A \wedge D}(\alpha)|}{|\rho_D(\alpha)|}\right) = Q\left(\frac{|\rho_A(\alpha) \cap \rho_D(\alpha)|}{|\rho_D(\alpha)|}\right) = Q\left(\mathfrak{R}_{|A/D|}(\alpha)\right) \quad (3.50)$$

$$\mathfrak{R}_B(\alpha) = \begin{cases} 1 & Q\left(\mathfrak{R}_{|A/D|}(\alpha)\right) \geq \alpha \\ 0 & \text{dd} \end{cases} \quad (3.51)$$

Kısıtlama seviyesi temelli yöntem (E) ve dereceli sayı temelli yöntem (B) için final değerler Eş. 3.52 ve Eş. 3.53 ile hesaplanmıştır.

$$T(E) = \sum_{\alpha_i \in \Lambda_E} (\alpha_i - \alpha_{i+1}) \cdot \mathfrak{R}_E(\alpha_i) \quad (3.52)$$

$$T(B) = \sum_{\alpha_i \in \Lambda_B} (\alpha_i - \alpha_{i+1}) \cdot \mathfrak{R}_B(\alpha_i) \quad (3.53)$$

Dereceli sayıların her seviyede kesin bir değerlendirme sağladığını görebiliriz. Kısıtlama seviye temelli yöntem ve dereceli sayı temelli yöntem için birleştirilmiş final değerler $T(E) = 0,46$ ve $T(B) = 0,3$ olarak hesaplanmıştır.

Dilsel özetler genellikle niceleyici cümleler veya eğer/ise kural yapılarıyla elde edilir ve niceleyici cümlelerde çoğu metotta mutlak niceleyici veya göreceli niceleyici kullanılır. Ancak bazı bilgiler bu niceleyicilerle ifade edilemez. Örneğin, “3 kişi hariç tüm öğrenciler uzundur.”, “Sarışınlardan yaklaşık 3 kat daha fazla uzun insan vardır.” gibi cümleler mutlak veya göreceli niceleyicilerle modellenemez. Bu nedenle Glöckner yarı bulanık niceleyici kavramını tanıtmıştır [60-65]. Klasik niceleyiciler, kesin olguları ele alır; ancak, bulanık dilsel niceleyiciler bulanık olguları ele alır ve doğruluk derecesi $[0,1]$ aralığında bir değere eşit olur. Yarı bulanık niceleyiciler kesin olguları $[0,1]$ aralığında bir değere atar. Bir yarı bulanık niceleyici klasik niceleyici ile bulanık niceleyici arasında bir kavramdır. Bu yüzden yarı bulanık niceleyici farklı tip özet oluşturmada esnek bir yoldur. Yarı bulanık niceleyiciler konusunda yapılan çeşitli çalışmalar ile yarı bulanık niceleyicilerin farklı tip cümle yapılarını modellemede kullanılabileceği gösterilmiştir ve farklı yapıda yarı bulanık niceleyiciler tanımlanmıştır [30, 66-68]. Yarı bulanık niceleyiciler 4 ana kategoriye ayrılır: birli yarı bulanık niceleyici, ikili yarı bulanık niceleyici, üçlü yarı bulanık niceleyici, dörtlü yarı bulanık niceleyici. $Q_{X,a,tip}^{FN}$ veya $Q_{X,a,tip}^v$ notasyonları çeşitli yarı bulanık niceleyici tiplerini tanımlamada kullanılır. FN bir bulanık sayıdır, $v \in [0,1]$ ’dir, X evrensel kümedir, a özellik sayısıdır ve tip farklı tip (örneğin, kardinalite temelli, oransal temelli, karşılaştırma temelli) yarı bulanık niceleyici oluşturmayı sağlayan bir parametredir. Oransallık söz konusu olduğunda v parametresi kullanılır. Çeşitli yarı bulanık niceleyici örnekleri Çizelge 3.9’da verilmiştir.

Çizelge 3.9. Çeşitli yarı bulanık niceleyici örnekleri

Yarı bulanık niceleyici	Örnek
$Q_{X,1,oran_hariç_tümü}^{FN}(S) = FN\left(\frac{ X - S }{ X }\right)$	Yaklaşık %30 hariç tüm kişiler öğrencidir. (X :kişiler, S :öğrenci)
$Q_{X,2,oran_kıyas}^{FN,v}(S_1, S_2) = \begin{cases} FN(S_1 / S_2), & S_2 \neq \emptyset \\ FN(\infty), & S_1 \neq \emptyset \wedge S_2 = \emptyset \\ FN(1), & S_1 = \emptyset \wedge S_2 \neq \emptyset \end{cases}$	Kadınların yaklaşık iki katı kadar erkek vardır. (S_1 :erkek, S_2 :kadın)
$Q_{X,3,oran_kıyas}^{FN}(S_1, S_2, S_3) = FN\left(\frac{\frac{ S_1 \cap S_2 }{ S_1 }}{\frac{ S_1 \cap S_3 }{ S_1 }}\right) = FN\left(\frac{ S_1 \cap S_2 }{ S_1 \cap S_3 }\right)$	Koşan kadınlardan yaklaşık 2 kat ya da daha fazla bisiklete binen kadın vardır. (S_1 :kadın, S_2 :bisiklete binen, S_3 : koşan)
$Q_{X,4,oran_kard}^{FN}(S_1, S_2, S_3, S_4) = \begin{cases} FN\left(\frac{\frac{ S_1 \cap S_2 }{ S_1 }}{\frac{ S_3 \cap S_4 }{ S_3 }}\right) & \frac{ S_1 \cap S_2 }{ S_1 } \neq 0 \wedge \frac{ S_3 \cap S_4 }{ S_3 } \neq 0 \\ FN(\infty) & \frac{ S_1 \cap S_2 }{ S_1 } \neq 0 \wedge \frac{ S_3 \cap S_4 }{ S_3 } = 0 \\ v \in [0,1] & dd \end{cases}$	Koşan erkeklerden yaklaşık 2 kat ya da daha fazla bisiklete binen kadın vardır. (S_1 :kadın, S_2 :bisiklete binen, S_3 : erkek, S_4 :koşan)

Yarı bulanık niceleyiciler klasik niceleyiciler gibi sadece kesin olguları ele alırken bulanık niceleyiciler gibi $[0,1]$ aralığında doğruluk derecesi üretir [66]. Yarı bulanık niceleyici tanımlamak bulanık niceleyici tanımlamaktan daha kolay olmasına rağmen, bulanık niceleyici cümlelerin değerlendirilmesi çok zordur. “Kişilerin yaklaşık %70’i öğrencidir.” cümlesi yarı bulanık niceleyici ile değerlendirilebilirken, “Kişilerin yaklaşık %70’i sarışındır.” cümlesinin yarı bulanık niceleyicilerle değerlendirilmesi için niceleyici bulanıklaştırma mekanizmasına ihtiyaç duyulur [30]. Niceleyici bulanıklaştırma mekanizmaları yarı bulanık niceleyici ve bulanık niceleyicinin tamamlayıcı görev yapmasını sağlayan bir çerçevedir [69]. Literatürde yarı bulanık niceleyiciler tabanında bulanık niceleyicilerin tanımlanmasına izin veren iki niceleyici bulanıklaştırma mekanizması tanımlanmıştır. Bunlar Glöckner [60, 61, 65] tarafından tanımlanan M mekanizması ve Diaz-Hermida ve diğerleri [70] tarafından tanımlanan olasılıksal F^I mekanizmasıdır. Niceleyici bulanıklaştırma mekanizmaları sadece mutlak ve göreceli niceleyiciler ile sınırlı

olmadığından avantajlıdır [27]. Tip-I ve tip-II niceleyici cümleler bu yaklaşımın özel bir hali olarak görülebilir.

3.5.3. Diğer yöntemler

Tip-I ve tip-II niceleyici cümleleri değerlendirmede skalar kardinalite ve bulanık kardinalitenin yanısıra olabilirlik teorisine dayanan doğruluk dereceleri de vardır. Tip-I niceleyici cümlelerin değerlendirilmesinde Yager [71] OWA operatörü temelli olmayan (non-OWA) doğruluk derecesini ve Bosc ve Lietard [44, 45] sadece göreceli niceleyicilerin kullanıldığı cümlelerin değerlendirilmesinde Sugeno integrali temelli doğruluk derecesini önermişlerdir. OWA operatörü temelli olmayan metod Sugeno integrali temelli metodun genel bir halidir. Tip-II niceleyici cümlelerin değerlendirilmesinde Vila ve diğerleri [72] orness ölçüsünü kullanarak $Q = \exists$ ve $Q = \forall$ uç durumlarının konveks kombinasyonu ile doğruluk derecesini hesaplamışlardır, Martin ve diğerleri [51] ve Martin ve Yun [52] birliktelik kurallarının değerlendirilmesi için kütle ataması temelli doğruluk derecesini önermişlerdir ve Lietard [73] niceleyici cümleleri değerlendirmede küme merkezli fonksiyonel bir metod önermiştir. Kütle ataması temelli yöntem aynı zamanda eğer/ise yapısındaki dilsel özetlerin değerlendirilmesinde de kullanılabilir.

3.6. Uygulamalar

Veritabanlarını dilsel olarak özetleme işlemi Yager [3] tarafından önerilmesinden sonra bu konuyla ilgili çeşitli çalışmalar yapılmıştır. Dilsel özetleme işlemi başta zaman serilerinin özetlenmesi olmak üzere, karar vericilerin karar vermesine yardımcı karar destek sistemlerinde, doküman sınıflandırılmasında, karar vericilerin uzlaşması sürecinde, insan hareketlerin incelenmesinde, internet trafiğinde ve kullanıcı profilini dikkate alarak televizyon kanalı tavsiye eden sistem gibi birçok farklı alana uygulanmıştır [74-80].

Zaman serilerinin dilsel özetlenmesi alanında yapılan çalışmalar Çizelge 3.10'da verilmiştir. Zaman serilerinin özetlenmesinin dışında, dilsel özetleme işlemi çeşitli alanlara başarılı bir şekilde uygulanmıştır. Diğer alanlarda yapılan uygulamalar Çizelge 3.11'de verilmiştir.

Çizelge 3.10. Dilsel özetleme ile ilgili çalışmalar (zaman serileri)

Yazar	BK	Kalite Ölçümleri	Uygulama Alanı
Kacprzyk ve Wilbik [81]	Tip-1	Doğruluk derecesi (Z_Q), özgüllük derecesi	Fon yatırımı
Kacprzyk ve Wilbik [82]	Tip-1	Doğruluk derecesi (Z_Q), bulanıklık derecesi	
Wilbik ve Kacprzyk [83]	Tip-1	Doğruluk derecesi (Z_Q), özgüllük derecesi, odak derecesi	
Kacprzyk ve diğerleri [84-89]	Tip-1	Doğruluk derecesi (Z_Q)	
Kacprzyk ve diğerleri [90, 91]	Tip-1	Sugeno integrali	
Kacprzyk ve diğerleri [92]	Tip-1	Düzeltilmiş Sugeno integrali	
Kacprzyk ve diğerleri [93]	Tip-1	Choquet integrali	
Kacprzyk ve diğerleri [94]	Tip-1	OWA operatörü	
Kacprzyk ve diğerleri [95]	Tip-1	Sugeno ve Choquet integralleri	
Kacprzyk ve Wilbik [96], Wilbik ve Kacprzyk [97]	Tip-1	Doğruluk derecesi (Z_Q)	Zaman serilerinin karşılaştırılması
Chiang ve diğerleri [98]	Tip-1	Doğruluk derecesi (Z_Q)	Bilgisayar işlemci kullanımı
Ortega ve diğerleri [99, 100]	Tip-1	Doğruluk derecesi (GD_Q)	Ham petrol fiyatı
Ortega ve diğerleri [32, 101, 102]	Tip-1	Doğruluk derecesi (GD_Q)	Sağlık merkezine gelen hasta sayısı
Van der Heide ve Trivino [103]	Tip-1	Uygunluk derecesi Geçerlilik derecesi	Enerji tüketim verisi
Boran ve Akay [104]	Tip-2	Doğruluk Derecesi (Z_Q)	Avrupa brent petrol piyasa fiyatı
Almeida ve diğerleri [105]	Tip-1	Doğruluk Derecesi (Z_Q), Benzememezlik Ölçüsü	Septik şok hastası verisi
Banaee ve diğerleri [106]	Tip-1	Doğruluk Derecesi (Z_Q)	Psikolojik zaman serisi verisi
Boran ve diğerleri [107]	Tip-2	Doğruluk Derecesi ($\tilde{T}_{BAY}, \tilde{T}_{WM}, \tilde{T}_N$)	Avrupa brent petrol piyasa fiyatı
Hatipoğlu ve diğerleri [108]	Tip-1	Doğruluk Derecesi (GD_Q)	Avrupa brent petrol piyasa fiyatı
Kacprzyk ve Zadrozny [109]	Tip-1	Doğruluk Derecesi (Z_Q)	Web sürücü günlükleri
Dziedzic ve diğerleri [110]	Tip-1	Doğruluk Derecesi (Z_Q , GD_Q)	Fon Yatırımı
Dziedzic ve diğerleri [111]	Tip-1	Doğruluk Derecesi (Z_Q)	Fon Yatırımı
Hryniewicz ve Kaczmarek [112]	Tip-1	Doğruluk Derecesi (Z_Q)	Kısa zaman serilerinin tahmini
Kacprzyk ve Zadrozny [113]	Tip-1	Doğruluk Derecesi (Z_Q)	Web sürücü günlükleri

Çizelge 3.11. Dilsel özetleme ile ilgili çalışmalar (uygulamalar)

Yazar	BK	Kalite Ölçümleri	Uygulama Alanı
Kacprzyk ve Zadrozny [75, 77]	Tip-1	Doğruluk (Z_Q), belirsizlik, kapsama uygunluk derecesi, özet uzunluğu	Karar destek sistemleri/Doğal dil üretimi
Kacprzyk ve Zadrozny [76]	Tip-1	Doğruluk derecesi (Z_Q)	Dokümanların kategorize edilmesi
Kacprzyk ve diğerleri [78]	Tip-1	Doğruluk derecesi (Z_Q)	Karar verme
Anderson ve diğerleri [74]	Tip-1	Güven derecesi	İnsan hareketlerinin dilsel özetlenmesi
Pouzols ve diğerleri [80]	Tip-1	Doğruluk derecesi (Z_Q)	İnternet trafik akışı
Pigeau ve diğerleri [79]	Tip-1	İlgi derecesi	Tv-kanalı öneren sistem
Eciolaza ve diğerleri [114], Eciolaza ve Trivino [115], Eciolaza ve diğerleri [116]	Tip-1	Kardinalite temelli ölçüm	Sürüş davranışlarının incelenmesi
Alvarez-Alvarez ve diğerleri [117-119], Gonzalez-Villanueva ve diğerleri [120], Mendez-Nunez ve Trivino [121], Menendez ve Trivino [122], Trivino ve diğerleri [123], Trivino ve Sugeno [124]	Tip-1	Kardinalite temelli ölçüm	Finansal raporlar, Mars yüzeyinin incelenmesi, karayolu trafik akışı, enerji tüketimi, insan yürüyüş analizi
Yager [125]	Tip-1	OWA operatörü	Sosyal ağlar
Carrasco ve Villar [126]	Tip-1	OWA operatörü	Turizm verilerinin özetlenmesi
Chamorro-Martinez ve diğerleri [41]	Tip-1	Dereceli sayılar	Görüntü işleme
Anderson ve diğerleri [127], Ros ve diğerleri [128]	Tip-1	Benzerlik ölçüsü	İnsan davranışlarının incelenmesi
Wilbik ve Keller [129, 130], Wilbik ve diğerleri [131, 132]	Tip-1	Doğruluk derecesi (Z_Q), odak derecesi	Dilsel özetlerin benzerliklerinin karşılaştırılması
Kacprzyk ve Zadrozny [133]	Tip-1	Doğruluk derecesi (Z_Q)	Veri madenciliği/Apriori algoritması
Kacprzyk ve diğerleri [39]	Tip-1	Doğruluk derecesi (Z_Q)	Zaman serileri/Genetik algoritma
Pilarski [35]	Tip-1	Doğruluk derecesi (Z_Q)	Dilsel özetlerin indirgenmesi
Mendel ve Korjani [134]	Tip-1	Tutarlılık derecesi	Demokrasi analizi
Wijayasekara ve Manic [135]	Tip-1	Doğruluk derecesi (T_{SOM}), yeterli kapsama derecesi	Sınıflandırma problemleri
Wijayasekara ve Manic [135]	Tip-1	Doğruluk derecesi (T_{SOM}), yeterli kapsama derecesi	Sınıflandırma problemleri

Çizelge 3.11. (devam) Dilsel özetleme ile ilgili çalışmalar (uygulamalar)

Yazar	BK	Kalite Ölçümleri	Uygulama Alanı
Niewiadomski [136]	Tip-2	Doğruluk (Z_Q), belirsizlik, kapsama ve uygunluk derecesi, özet uzunluğu	İnsan kaynakları
Niewiadomski [137]	Tip-2	Doğruluk (Z_Q), belirsizlik, kapsama ve uygunluk derecesi, özet uzunluğu	İnsan kaynakları
Wu ve Mendel [47]	Tip-2	Doğruluk (Z_Q), belirsizlik, kapsama ve uygunluk derecesi, özet uzunluğu, aykırı değer	Diyabet veritabanı
Tre ve diğerleri [138]	Tip-1	Geçerlilik Derecesi	Kayıp bilgi ile dilsel özetleme (örnek işçi verisi)
Popescu ve diğerleri [139]	Tip-1	Doğruluk Derecesi (T_p)	Duyguların özetlenmesi
Niewiadomski [140]	Tip-2	Bilgi Vericilik	Metinsel tanımlar üreten bilgisayar sistemi
Korjani ve Mendel [141]	Tip-2	Tutarlılık Derecesi	Demokrasi analizi
Conde-Clemente ve diğerleri [142]	Tip-1	Geçerlilik Derecesi	Görme engellilerin kendi profil fotoğrafını çekebilmesi
Donis-Diaz ve diğerleri [38]	Tip-1	Doğruluk (Z_Q), Belirsizlik, Kapsama, Uygunluk Derecesi	Akma kopma basıncı verisi
Iqbal ve diğerleri [143]	Tip-1	Ölçeklendirilmiş Bulanık Ağırlık	Öğrenci performansı izleme
Kacprzyk ve Zadrozny [113]	Tip-1	Doğruluk Derecesi (Z_Q)	Firmaların yenilikçiliği
Niewiadomski ve Superson [144]	Tip-1	Kardinalite temelli ölçüm	Yaş ve cinsiyete göre özetleme sonuçlarının karşılaştırılması

4. Z-SAYILARI İLE DİLSEL ÖZETLEME

Bu bölümde önce Z-sayıları tanıtılmış daha sonra Z-sayıları ile dilsel özetleme bağlantısı kurularak dilsel özetlemede tip-I ve tip-II niceleyici cümleler için olasılık ve olabilirliği birlikte dikkate alan Z-sayısı temelli yeni bir değerlendirme yöntemi önerilmiştir.

4.1. Z-sayıları

Kullanılan bilginin faydalı olabilmesi için güvenilir olması gereklidir. Bilginin güvenilirliği ile ilgili olarak Zadeh tarafından Z-sayısı kavramı ortaya çıkarılmıştır [5]. Bir Z-sayısı sıralı iki bileşenden oluşur, (A, B) . Z-sayısının ilk bileşeni A , x değişkeninin alabileceği değerler üzerinde bir sınırlandırmadır ve x değişkeninin X bölgesinde bir bulanık altkümesidir. Z-sayısının ikinci bileşeni B , ilk bileşenin güvenilirliğidir ve genellikle birim aralıkta bir bulanık altkümedir. A , değişkenin değerini gösterirken, B doğruluk veya olasılığın kesinliği hakkında bir fikir sunmaktadır. Örneğin, (yaklaşık 45 dk, çok kesin) ifadesi bir Z-sayıdır. Bu örnekte değişkenin değeri “yaklaşık 45 dk” bulanık kümesi ile tanımlanmıştır ve değişkenin “yaklaşık 45 dk” değerini almasının güvenilirliği “çok kesin” bulanık kümesi ile tanımlanmıştır. Yani, “yaklaşık 45 dk” A değerini, “çok kesin” B değerini göstermektedir.

(x, A, B) sıralı üçlüsüne Z-değerleme denir [145]. Bir Z-değerleme x değişkeninin A değerini almasının kesinliğinin B değerine eşit olduğu bilgisini ifade eder. Örneğin, (Merve’nin yaşı, genç, kesin) ifadesi bir Z-değerlemedir. Bu örnekte değişken “Merve’nin yaşı”dır, değişkenin değeri “genç” bulanık kümesi ile ifade edilmiştir ve değişkenin “genç” değerini almasının güvenilirliği “kesin” bulanık kümesi ile ifade edilmiştir. Yani, “Merve’nin yaşı” x değerini, “genç” A değerini ve “kesin” B değerini göstermektedir. Z-değerlemeler bir araya geldiğinde Z-bilgisini oluşturur.

Bir Z-değerleme “ $x \in (A, B)$ ’dir.” ifadesine eşittir [145]. Burada Z-sayısı x değişkeni hakkında bilgi sağlar. x ’in A değerini almasının olasılık değeri B ’ye eşittir. Bir önceki örnek için, Merve’nin genç olma olasılığı “kesin” bulanık kümesiyle ifade edilmektedir. Zadeh [5] tarafından önerildiği gibi, bu Z-değerleme x üzerinde bir sınırlandırma olarak görülebilir ve $\text{Prob}(x \in A) = B$ şeklinde yorumlanabilir.

Zadeh'in bulanık altkümenin olasılığı tanımına göre, $\text{Prob}(x = A)$ Eş. 4.1 ile ifade edilebilir [146].

$$\text{Prob}_p(x = A) = \int_x A(x)p(x)dx \quad (4.1)$$

Eş. 4.1'i kullanarak $\text{Prob}_p(x = A)$ 'nın B 'yi sağlama derecesini yani, p 'nin Z -değerlemeyi sağlama derecesi olan $G(p)$ 'yi Eş. 4.2 ile elde ederiz.

$$G(p) = B(\text{Prob}_p(x = A)) = B\left(\int_x A(x)p(x)dx\right) \quad (4.2)$$

Günlük hayattaki muhakemeler ve karar verme genellikle Z -bilgisine bağlıdır. Z -bilgisi kavramı özellikle ekonomi, yönetim, karar analizi gibi alanlarda oldukça önemlidir. Kullanılan bilginin güvenilirliği alınan kararların etkili ve doğru olmasında kilit önem taşımaktadır. Bu yüzden Zadeh bilginin güvenilirliğini modellemek üzere Z -sayıları kavramını önermiştir ve iki Z -sayısı ile aritmetik hesaplamaların nasıl yapılacağını göstermiştir [5]. Yager Z -sayılarının yeni gösterim ve uygulamalarını sunarak önerilen Z -sayıları yaklaşımının karar verme gibi süreçlerdeki faydasını göstermiştir [145, 147]. Aynı zamanda Z -bilgisinin Dempster-Shafer inanç yapısı ile alternatif bir formülasyonunu önermiştir. Z -sayısı hakkındaki teoriler olgunlaşmadığı için Z -sayılarını klasik bulanık sayılara dönüştürme ihtiyacı doğmuş ve bunun için bulanık kümenin bulanık beklentisine göre Z -sayılarını klasik bulanık sayılara dönüştüren bir metod önerilmiştir [148]. Z -sayıları hakkında yapılan çalışmaların çoğu karar verme ile ilgilidir. İlk olarak Kang ve diğerleri belirsiz ortamda karar verme için Z -sayılarını kullanarak bir çok ölçütlü karar verme modeli önermişlerdir [149]. Daha sonra Azadeh ve diğerleri en iyi üniversite değerlendirmesi için kriter arayan bir Z -AHP modeli geliştirmişlerdir [150]. Soroudi ve Amraee enerji sistemlerine uygulanan karar verme sürecindeki belirsizliği modelleme tekniklerine Z -sayısını kullanan yeni bir teknik önermişlerdir ve diğer belirsizlik modelleme teknikleriyle karşılaştırmışlardır [151]. Aliev ve Zeinalova [152] karar verme probleminin bilgi yapısını formüle etmede, Cai [153] yeşil strateji karar uygulamasında, Lorkowski ve diğerleri [154] adil fiyat belirlemedeki belirsizliği modellemede Z -sayılarını kullanmışlardır. Mohamad ve

diğerleri ise önerdikleri Z-sayısı temelli karar verme prosedürünü üretimde risk analizi problemi üzerinde göstermişlerdir [155]. Xiao Z-sayısının ikinci bileşeni olan güvenilirliği bulanık sayının belirsiz bölgesi olarak dikkate almıştır ve erkek müşterilerin giyimi üzerine bir çok ölçütlü karar verme problemine önerdiği modeli uygulamıştır [156]. Zeinalova önerdiği Z-sayısı temelli metodu bir yatırım problemi üzerinde göstermiştir [157]. Sotoudeh Anvari ve Sadi-Nezhad çok ölçütlü karar vermede kriterlerin göreceli ağırlıklarını Z-sayıları ile ifade etmişlerdir [158]. Tavakkoli-Moghaddam ve diğerleri tesis yer seçimi problemi için Z-PROMETHEE yöntemini sunmuşlardır [159].

Literatürde Z-sayıları ile ilgili karar verme alanı dışında farklı uygulamalar da mevcuttur. Z-sayıları CWW (Computing with Words, Kelimelerle Hesaplama) için yeni bir bulanık kuramsal yaklaşım olmuştur. CWW iki seviyeden oluşur: “seviye-1” kelimelerin nicelendirilmesine odaklanırken, “seviye-2” doğal dil ifadelerinin kesinleştirilmesine odaklanır. Banarjee ve Pal seviye-2 CWW’ye Z-sayılarını dahil ederek sübjektif doğal dil anlayışı sağlamışlardır [160]. Banarjee ve Pal [161] ve Pal ve diğerleri [162] Z-sayısının NLP (Doğal Dil İşleme) ve CWW’yi birleştiren bir boyut olduğunu keşfetmişlerdir, Z-sayılarının CWW’ye katkılarını listelemişlerdir ve Z-sayısını kullanan bir CWW algoritması geliştirmişlerdir. Patel ve diğerleri kurdukları Z-sayısı modelini CWW için bir uzman sisteme entegre etmişlerdir [163].

Silah sistemlerinin etkinliğinin objektif ve doğru bir şekilde değerlendirilmesinde Zhang ve diğerleri Z-sayılarından faydalanmışlardır [164]. Azadeh ve diğerleri güvenilen bir birimin gelecek güvenini tahmin etmede yaptıkları bir çalışmada Z-sayılarını kullanmışlardır [165]. Salari ve diğerleri Z-sayısı temelli bir kazanılmış değer yönetimi tekniği önermişlerdir, böylece projenin performans gelişimi ve bitiş maliyeti için güvenilir bir değerlendirme sağlamışlardır [166]. Tadayon ve Tadayon Z-sayılarının yaklaşık değerlendirilmesi için bir metot önermişlerdir [167]. Aliev ve diğerleri önce Z-sayıları için aritmetik teoriler önermişlerdir, daha sonra bu teorileri kullanarak Z-sayıları ile doğrusal programlama yaklaşımı önermişlerdir [168, 169]. Aliev ve Memmedova [170] pilates egzersizinin psikolojik etkilerini modellemede Z-sayılarını kullanmışlardır. Görüldüğü gibi literatürde enerjiden, psikolojiye çeşitli çalışmalar yapılmıştır. Ancak dilsel özetleme ile ilgili herhangi bir çalışmaya rastlanmamıştır. Alt bölümlerde dilsel özetleme ile Z-sayılarının ilişkisi, tip-I ve tip-II niceleyici cümleler için Z-sayısı temelli doğruluk derecesi hesaplaması gösterilmiştir.

4.2. Dilsel Özetleme

Z-sayıları ile dilsel özetleme arasında güçlü bir ilişki olduğu Yager tarafından gösterilmiştir [145]. Dilsel özetler veri setindeki bilginin insan dostu bir yolla ifade edilmesini sağlar ve dilsel özetin doğruluğu bir önceki bölümdeki hesaplamalarla elde edilir. S özetleyicisi bir bulanık altkümedir ve bazı dilsel kavramlara karşılık gelir. Q niceleyicisi birim aralıkta bir fonksiyondur. Z-sayısında S özetleyicisi A değerine, Q niceleyicisi B değerine karşılık gelir. Z-değerlemede doğruluk derecesi hesaplaması dilsel özetleme için olasılık ve olabilirliği bir araya getiren bir doğruluk derecesi hesaplaması anlamına gelmektedir.

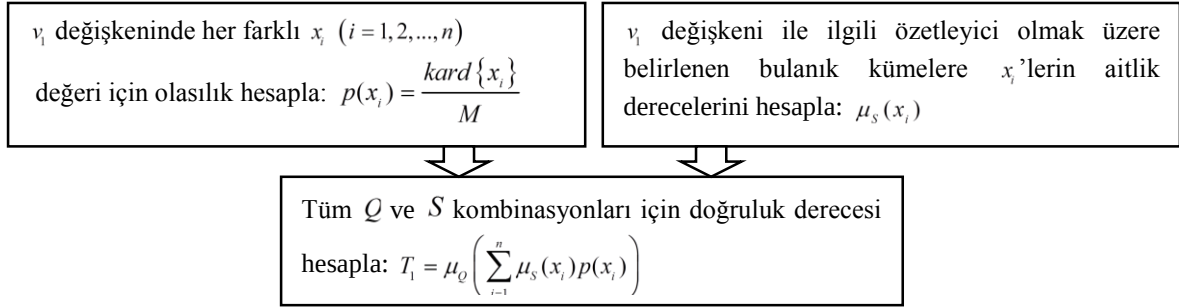
4.2.1. Tip-I niceleyici cümle için Z-sayısı temelli yöntem

Tip-I niceleyici cümleler " $Q Y(nesneler) S = \{S_1, S_2, \dots, S_K\}$ dir." $[T_1]$ yapısındaki cümlelerdir. Z-sayısı bakış açısıyla S A 'ya, Q B 'ye karşılık gelmektedir. Dilsel özetlemede bir miktar verimiz mevcuttur ve bu verinin olasılık dağılımı bilindiğinden dilsel özetin doğruluk derecesi Z-sayısının bu olasılık dağılımı ne derece sağladığı ile ölçülür. Tip-I niceleyici cümlenin doğruluk derecesi Eş. 4.3 ile hesaplanabilir.

$$T_1 = \mu_Q \left(\sum_{i=1}^n \mu_S(x_i) p(x_i) \right) \quad (4.3)$$

Bu eşitlikte Q niceleyiciyi, x_i ($i = 1, 2, \dots, n$) v_1 değişkeninin aldığı farklı değerleri, $\mu_S(x_i)$ x_i değerinin S bulanık kümesine aitlik derecesini, $p(x_i)$ x_i 'nin olasılık değerini göstermektedir. İşlem adımları Şekil 4.1 ile verilmiştir.

Tip-I niceleyici cümlenin değerlendirilmesi için önerilen yöntem Örnek 3.1. ile açıklanmıştır.



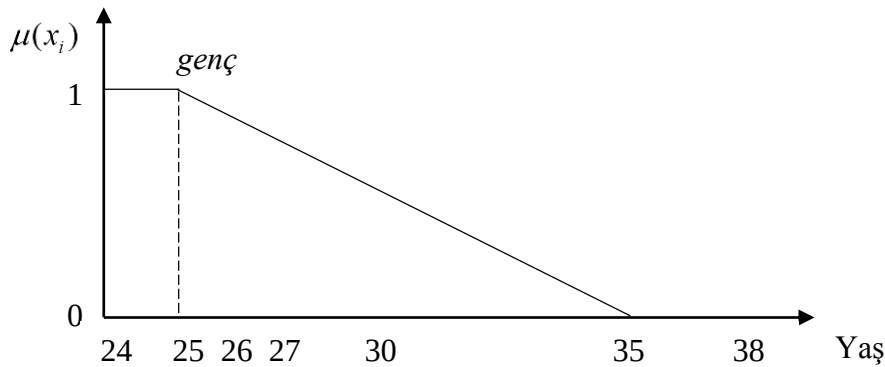
Şekil 4.1. Tip-I niceleyici cümle değerlendirme işlem adımları

Örnek 3.1.

Çalışan veri tabanında yaş değişkeni dikkate alınsın. Çalışanların yaşları ve yaşların genç bulanık kümesine olan aitlik dereceleri Çizelge 4.1’de ve genç bulanık kümesinin üyelik fonksiyonu Şekil 4.2’de verilmiştir.

Çizelge 4.1. Çalışanların yaşları ve genç kümesine aitlik dereceleri

#	1	2	3	4	5	6	7	8	9	10
Yaş	24	24	24	26	27	30	35	38	25	27
$\mu_{genç}$	1	1	1	0,9	0,8	0,5	0	0	1	0,8



Şekil 4.2. Genç bulanık kümesinin aitlik fonksiyonu

Çalışanların yaş değerlerinden sıklıklarına bakılarak bir olasılık fonksiyonu belirlenebilir. Buna göre, $p(24) = 0,3$; $p(25) = 0,1$; $p(26) = 0,1$; $p(27) = 0,2$; $p(30) = 0,1$; $p(35) = 0,1$; $p(38) = 0,1$ elde edilir.

Veriden elde edilebilecek olası bir özet “Çoğu çalışan gençtir.” için doğruluk derecesi şöyle hesaplanır;

$$\begin{aligned} T_1 &= \mu_Q \left(p(24) \cdot \mu_{genç}(24) + p(25) \cdot \mu_{genç}(25) + \dots + p(38) \cdot \mu_{genç}(38) \right) \\ &= \mu_Q (0,3*1 + 0,1*1 + \dots + 0,1*0) = \mu_Q (0,7) \end{aligned}$$

Dilsel özetlemede temel doğruluk derecesi hesaplama yöntemi olan Z_Q ’ya göre ise (Bkz. Eş. 3.16);

$$T_1 = \mu_Q \left(\frac{1+1+1+0,9+0,8+0,5+0+0+1+0,8}{10} \right) = \mu_Q (0,7)$$

Aslında Z_Q yöntemi şöyle de ifade edilebilir;

$$\begin{aligned} T_1 &= \mu_Q \left(\frac{1+1+1+0,9+0,8+0,5+0+0+1+0,8}{10} \right) \\ &= \mu_Q \left(3 \times \frac{1}{10} + 1 \times \frac{1}{10} + 1 \times \frac{0,9}{10} + 1 \times \frac{0,8}{10} + 1 \times \frac{0,8}{10} + 1 \times \frac{0,5}{10} + 1 \times \frac{0}{10} + 1 \times \frac{0}{10} \right) \\ &= \mu_Q (0,3*1 + 0,1*1 + 0,1*0,9 + 0,1*0,8 + 0,1*0,8 + 0,1*0,5 + 0,1*0 + 0,1*0) \\ &= \mu_Q (0,7) \end{aligned}$$

Buradan da görüldüğü gibi her iki yöntem eşdeğerdir. Yani sonuç her zaman aynı çıkacaktır. Ancak Z-sayısı temelli önerilen gösterimde olasılık ve olabilirlik birlikte formülde görüldüğünden yorumlama açısından daha faydalıdır.

4.2.2. Tip-II niceleyici cümle için Z-sayısı temelli yöntem

" $Q w_g(S_g)Y(nesneler) S = \{S_1, S_2, \dots, S_K\}$ dir." $[T_1]$ yapısındaki cümlelerdir. Z-sayısı bakış açısıyla cümle değerlendirilmek istendiğinde $(S | w_g(S_g))$ koşullu ifadesi A ’ya, Q B ’ye karşılık gelmektedir. Tip-II niceleyici cümlenin doğruluk derecesi Eş. 4.4 ile hesaplanabilir.

$$T_1 = \mu_Q \left(\frac{\sum_{i=1}^n \sum_{j=1}^t (\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j)) f_{v_1, v_2}(x_i, y_j)}{\sum_{i=1}^n (\mu_{w_g(S_g)}(x_i) p_{v_1}(x_i))} \right) \quad (4.4)$$

Bu eşitlikte Q niceleyiciyi, x_i ($i=1,2,...,n$) v_1 değişkeninin aldığı farklı değerleri, y_j ($j=1,2,...,t$) v_2 değişkeninin aldığı farklı değerleri, $\mu_{w_g(S_g)}(x_i)$ x_i 'nin $w_g(S_g)$ bulanık kümesine aitlik derecesini, $\mu_S(y_j)$ y_j 'nin S bulanık kümesine aitlik derecesini, $f_{v_1, v_2}(x_i, y_j)$ $v_1 = x_i$ ve $v_2 = y_j$ 'nin birlikte ortaya çıkma olasılığını, $p_{v_1}(x_i)$ $v_1 = x_i$ 'nin olasılığını ifade etmektedir. İşlem adımları Şekil 4.3 ile verilmiştir.

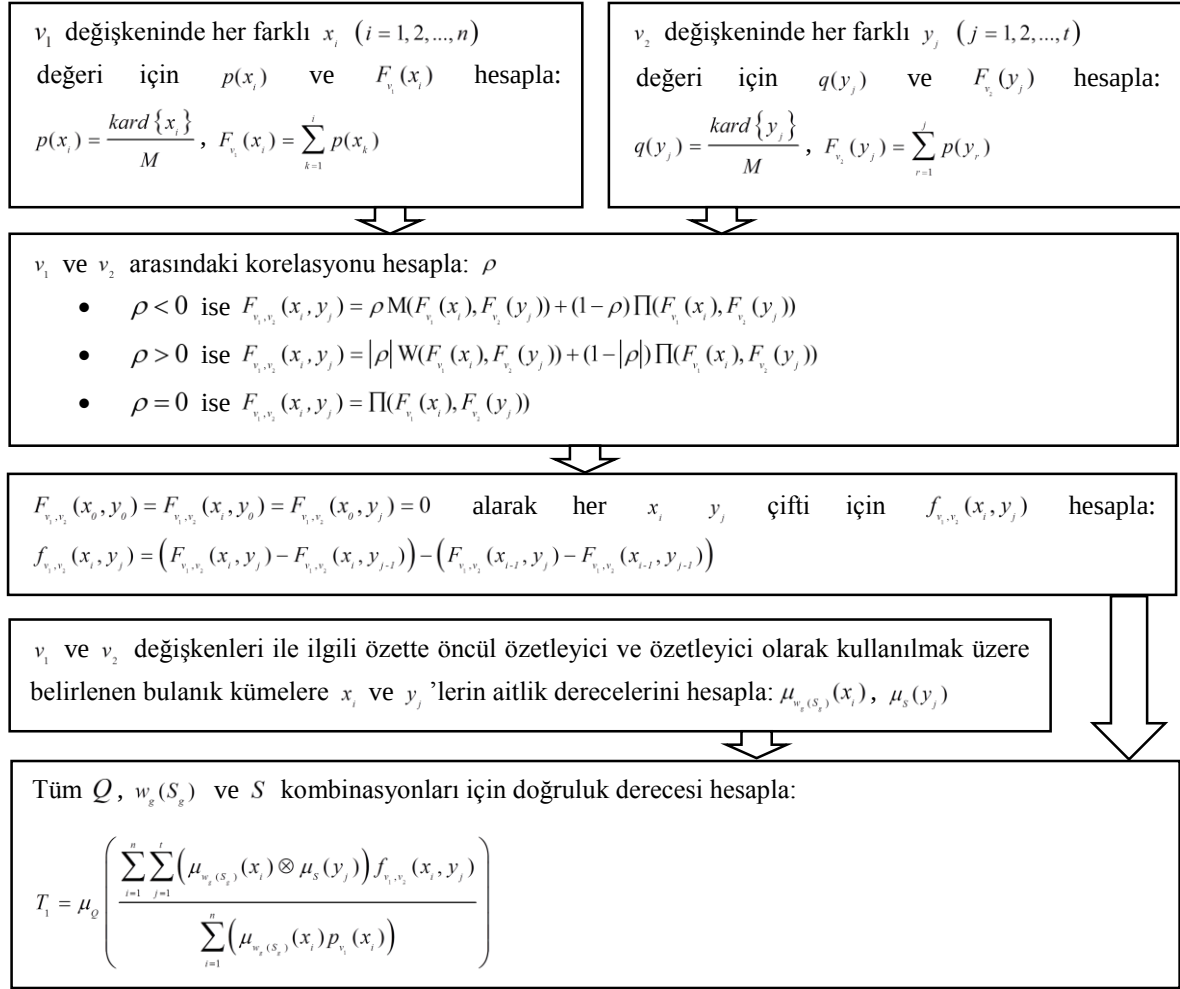
Mevcut verimizde farklı değişkenlere ait marjinal dağılımlar bilinmektedir, ancak ortak dağılım ($f_{v_1, v_2}(x_i, y_j)$) bilinmemektedir. Ortak dağılım elde etmek için kopula kullanımına ihtiyaç duyulur. Kopula ve ortak dağılım elde etme konusu alt bölümlerde incelenmiştir.

Kopula

Kopula, çok değişkenli dağılım fonksiyonlarını bunların tek boyutlu marjinal dağılım fonksiyonuna birleştirmesi olarak tanımlanır [6]. v_1 ve v_2 değişkenlerinin kopula kullanımı ile birleştirilmesinden önce ilgili değişkenlere ait kümülatif dağılım fonksiyonları aşağıda tanımlanmıştır.

$x_i < x_{i+1}$ olan $X = \{x_1, ..., x_n\}$ sıralı uzayında v_1 bir rassal değişken olsun. $v_1 = x_i$ olma olasılığı p_i ile gösterilsin. v_1 ile ilişkili kümülatif dağılım fonksiyonu Eş. 4.5 ile gösterilmiştir.

$$F_{v_1}(x_i) = \text{Prob}(v_1 \leq x_i) = \sum_{j=1}^i p_j \quad (4.5)$$



Şekil 4.3. Tip-II niceleyici cümle değerlendirme işlem adımları

Bu eşitlikten $F_{v_1}(x_n) = 1$ olduğunu ve ayrıca, $F_{v_1}(x_i) - F_{v_1}(x_{i-1}) = \text{Prob}(v_1 = x_i) = p_i$ olduğunu görebiliriz.

$y_j < y_{j+1}$ olan $Y = \{y_1, \dots, y_t\}$ sıralı uzayında farklı bir v_2 değişkeni dikkate alındığında $v_2 = y_j$ olma olasılığı q_j ile gösterilsin. v_2 ile ilişkili kümülatif dağılım fonksiyonu Eş. 4.6 ile gösterilmiştir.

$$F_{v_2}(y_j) = \text{Prob}(v_2 \leq y_j) = \sum_{i=1}^j q_i \quad (4.6)$$

Bu eşitlikten $F_{v_2}(y_t) = 1$ olduğunu ve ayrıca, $F_{v_2}(y_j) - F_{v_2}(y_{j-1}) = \text{Prob}(v_2 = y_j) = q_j$ olduğunu görebiliriz.

Bu iki değişken verildiğinde, ortak kümülatif dağılım $F_{v_1, v_2}(v_1 \leq x_i, v_2 \leq y_j) = \text{Prob}(v_1 \leq x_i, v_2 \leq y_j)$ Eş. 4.7 ile gösterilir.

$$\text{Prob}(v_1 \leq x_i, v_2 \leq y_j) = \sum_{j=1}^i \sum_{i=1}^j \text{Prob}(v_1 = x_j, v_2 = y_i) \quad (4.7)$$

Eş. 4.7’de $\text{Prob}(v_1 = x_j, v_2 = y_i)$, $v_1 = x_j$ ve $v_2 = y_i$ ’nin birlikte ortaya çıkma olasılığını ifade eder.

Sklar teoremi ortak kümülatif dağılım fonksiyonu ve bununla ilişkili marjinal kümülatif dağılım fonksiyonları arasındaki ilişkinin temel bir sonucunu sağlar [171, 172]. Sklar teoremine göre her bir ortak kümülatif dağılım fonksiyonu $F_{v_1, v_2}(x_i, y_j)$ bir kopula C kullanılarak ifade edilebilir ve Eş. 4.8 ile gösterilebilir [171].

$$F_{v_1, v_2}(x_i, y_j) = C[F_{v_1}(x_i), F_{v_2}(y_j)] \quad (4.8)$$

Eş. 4.8 $\text{Prob}(v_1 \leq x_i, v_2 \leq y_j) = C[\text{Prob}(v_1 \leq x_i), \text{Prob}(v_2 \leq y_j)]$ anlamına gelmektedir.

İki boyutlu bir kopula fonksiyonunun $C : [0,1]^2 \rightarrow [0,1]$ sağlaması gereken şartlar aşağıdadır [6];

- Tüm $x, y \in [0,1]$ için $C(x, 0) = C(0, y) = 0$
- Tüm $x, y \in [0,1]$ için $C(x, 1) = x$ ve $C(1, y) = y$
- Tüm $x_1 \leq x_2$ ve $y_1 \leq y_2$ için, $C(x_2, y_2) - C(x_2, y_1) - C(x_1, y_2) + C(x_1, y_1) \geq 0$

Kopulaların bazı önemli özellikleri aşağıdadır [6, 173];

- Monotonluk: eğer $x_2 \geq x_1$ ve $y_2 \geq y_1$ ise, $C(x_2, y_2) \geq C(x_1, y_1)$ olur.
- Kopulaların konveks kombinasyonları bir kopuladır: eğer $j = 1, \dots, n$ için C_j kopula

ise ve $w_j \geq 0$ olmak üzere $\sum_{j=1}^n w_j = 1$ ise, $C^*(x, y) = \sum_{j=1}^n w_j C_j(x, y)$ de bir kopuladır.

Kopulalar bir birleşim operatörü olarak görülebilir [174, 175]. Üç önemli kopula örneği minimum, çarpım ve lukasiewiczdir;

- M Min: $C(x, y) = \text{Min}(x, y)$
- Π Çarpım: $C(x, y) = xy$
- W Lukasiewicz: $C(x, y) = \text{Max}(x + y - 1, 0)$

Kopula teoresinde önemli bir sonuç Fréchet-Hoeffding sınırlarıdır [6]. Herhangi bir iki boyutlu kopula C için alt sınır ve üst sınır Eş. 4.9 ile gösterilebilir. Yani minimum en büyük kopuladır ve lukasiewicz en küçük kopuladır.

$$\text{Max}(x + y - 1, 0) \leq C(x, y) \leq \text{Min}(x, y) \quad (4.9)$$

Kopulalar ve Bölüm 3.2’de tanımlanan t-normlar arasında çok yakın bir ilişki vardır [174, 175]. t-norm çok değerli mantıktaki “ve” operatörünün genel bir halidir ve aşağıdaki gibi tanımlanır;

$\text{Max}(x + y - 1, 0) \leq T(x, y) \leq \text{Min}(x, y)$ olan herhangi bir t-normun bir kopula olduğu gösterilebilir [175]. $\lambda \geq 1$ olan Yager sınıfı t-normların hepsi

$T_\lambda(x, y) = 1 - \min[1, ((1-x)^\lambda + (1-y)^\lambda)^{\frac{1}{\lambda}}]$ kopuladır [176]. Burada $\lambda = 1$ iken $T_\lambda(x, y) = \text{Max}(x + y - 1, 0)$ ve $\lambda \rightarrow \infty$ iken $T_\lambda(x, y) = \text{Min}(x, y)$ elde edilir.

Çoğu durumda, marjinal kümülatif dağılım fonksiyonları hakkında bilgimiz olurken ortak kümülatif dağılım hakkında bilgimiz olmaz. Sklar teoremi bize marjinal kümülatif dağılım fonksiyonlarından ortak bir kümülatif dağılım fonksiyonu modellememizi sağlar.

İki marjinal kümülatif dağılım fonksiyonu verildiğinde uygun bir kopula seçilirken, ortak kümülatif dağılım fonksiyonunun yaklaşık bir modelini elde edebiliriz. Trivedi ve Zimmer bu problemi özellikle finans modelleme alanında detaylı bir şekilde incelemişlerdir [177].

İki marjinalden bir ortak dağılım elde etmede önemli bir kavram iki değişken arasındaki korelasyondur. Korelasyon katsayısı $\rho \in [-1,1]$ iki rassal değişken arasındaki ilişkiyi göstermede kullanılır ve korelasyonun gücü ve yönünün bir ölçüsüdür. Pozitif korelasyon iki değişkenin aynı doğrultuda hareket ettiğini gösterir. Yani bir değişken büyük değerler alırken diğer değişken de genellikle büyük değerler alır ve benzer şekilde bir değişken küçük değerler alırken diğer değişken de genellikle küçük değerler alır. Negatif korelasyon iki değişkenin zıt doğrultuda hareket ettiğini gösterir. Yani bir değişken büyük değerler alırken diğer değişken genellikle küçük değerler alır. Korelasyon değerleri sıfıra yakınsa genellikle değişkenler ilişkisizdir. Değişkenleri birleştirmede Sklar teoremi kullanılırken değişkenler arasındaki korelasyon kopula seçimiyle ele alınır [177-179]. Güçlü pozitif korelasyon, $\rho \rightarrow 1$, Min kopula ile modellenenir, güçlü negatif korelasyon, $\rho \rightarrow -1$, Lukasiewicz kopula ile modellenenir. Korelasyonsuz ilişki, $\rho \rightarrow 0$ Çarpım kopula ile modellenenir. Daha genel olarak, korelasyon dereceleri bu uç noktaların arasında farklı kombinasyonlarla veya diğer parametrelili kopulaların kullanımı ile modellenenir.

Literatürde çok sayıda kopula aileleri üretilmiştir [6]. Bunlardan çok kullanılan bir tanesi Fréchet ailesidir, $\alpha, \beta \in [0,1]$ ve $\alpha + \beta \leq 1$ olmak üzere Eş. 4.10 ile tanımlanır.

$$C_{\alpha,\beta}(x, y) = \alpha M(x, y) + (1 - \alpha - \beta) \Pi(x, y) + \beta W(x, y) \quad (4.10)$$

Çoğu uygulamada Fréchet kopula korelasyon hakkındaki bilgiyi içerdiği için çok kullanışlı ve erişilebilir bir form sağlar [180]. Fréchet kopulanın α ve β olmak üzere iki parametresi vardır. Burada eğer $\alpha = 1$ ve $\beta = 0$ ise güçlü pozitif korelasyon durumu $C_{\alpha,\beta}(x, y) = M(x, y)$ oluşur. Eğer $\beta = 1$ ve $\alpha = 0$ ise güçlü negatif korelasyon durumu $C_{\alpha,\beta}(x, y) = W(x, y)$ oluşur. $\alpha = \beta = 0$ durumunda bağımsızlık durumu olan $C_{\alpha,\beta}(x, y) = \Pi(x, y)$ oluşur.

ρ 'nun alabileceği farklı değerler için bir yaklaşım şöyledir [180]. Eğer $\rho > 0$ ise $\beta = 0$ ve $\alpha = \rho$ olarak alınır. Bu durumda $C_{\alpha,\beta}(x, y) = \rho M(x, y) + (1 - \rho) \Pi(x, y)$ elde edilir. Eğer $\rho < 0$ ise $\alpha = 0$ ve $\beta = |\rho|$ olarak ele alınır ve $C_{\alpha,\beta}(x, y) = |\rho| W(x, y) + (1 - |\rho|) \Pi(x, y)$ elde edilir. Eğer $\rho = 0$ ise o halde $\alpha = \beta = 0$ alınır ve $C_{\alpha,\beta}(x, y) = \Pi(x, y)$ elde edilir.

Ortak dağılım elde etme

Sklar teoremi sadece özel tip olaylar için marjinal olasılıktan doğrudan ortak olasılık elde etmeye izin verir. Özellikle, $v_1 \leq x_j$ ve $v_2 \leq y_i$ tipinde olayları içeren marjinal kümülatif dağılım fonksiyonlarından ortak bir kümülatif dağılım fonksiyonu tanımlanmasına izin verir. Sklar teoreminin diğer tip olayların ortak olasılığını elde etmede de kullanılması daha faydalı olacağından Yager ve Alajlan $X \times Y$ 'de herhangi bir noktada ortak olasılığın Teorem 3.1. ile hesaplanabileceğini önermişlerdir [180]. Teorem 3.1. ve ispatı aşağıda verilmiştir.

Teorem 3.1.

$$\text{Prob}(v_1 = x_i, v_2 = y_j) = (F_{v_1, v_2}(x_i, y_j) - F_{v_1, v_2}(x_i, y_{j-1})) - (F_{v_1, v_2}(x_{i-1}, y_j) - F_{v_1, v_2}(x_{i-1}, y_{j-1})) \quad (4.11)$$

İspat

$$F_{v_1, v_2}(x_i, y_j) - F_{v_1, v_2}(x_i, y_{j-1}) = \text{Prob}(v_1 \leq x_i, v_2 \leq y_j) - \text{Prob}(v_1 \leq x_i, v_2 \leq y_{j-1})$$

$$= \sum_{k=1}^i \sum_{r=1}^j \text{Prob}(v_1 = x_k, v_2 = y_r) - \sum_{k=1}^i \sum_{r=1}^{j-1} \text{Prob}(v_1 = x_k, v_2 = y_r)$$

$$= \sum_{k=1}^i \text{Prob}(v_1 = x_k, v_2 = y_j)$$

$$F_{v_1, v_2}(x_{i-1}, y_j) - F_{v_1, v_2}(x_{i-1}, y_{j-1}) = \text{Prob}(v_1 \leq x_{i-1}, v_2 \leq y_j) - \text{Prob}(v_1 \leq x_{i-1}, v_2 \leq y_{j-1})$$

$$= \sum_{k=1}^{i-1} \sum_{r=1}^j \text{Prob}(v_1 = x_k, v_2 = y_r) - \sum_{k=1}^{i-1} \sum_{r=1}^{j-1} \text{Prob}(v_1 = x_k, v_2 = y_r)$$

$$= \sum_{k=1}^{i-1} \text{Prob}(v_1 = x_k, v_2 = y_j)$$

$$\sum_{k=1}^i \text{Prob}(v_1 = x_k, v_2 = y_j) - \sum_{k=1}^{i-1} \text{Prob}(v_1 = x_k, v_2 = y_j) = \text{Prob}(x_i, y_j) \quad \square$$

Teorem 3.1.'de elde edilen $\text{Prob}(x_i, y_j)$ ifadesi Eş. 4.4'teki $f_{v_1, v_2}(x_i, y_j)$ ifadesine karşılık gelmektedir.

Lemma 3.1.

$$\sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) = p_{v_1}(x_i)$$

İspat

$$\begin{aligned} \sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) &= (F_{v_1, v_2}(x_i, y_1) - F_{v_1, v_2}(x_i, y_0)) - (F_{v_1, v_2}(x_{i-1}, y_1) - F_{v_1, v_2}(x_{i-1}, y_0)) \\ &+ (F_{v_1, v_2}(x_i, y_2) - F_{v_1, v_2}(x_i, y_1)) - (F_{v_1, v_2}(x_{i-1}, y_2) - F_{v_1, v_2}(x_{i-1}, y_1)) \\ &+ \dots \\ &+ (F_{v_1, v_2}(x_i, y_t) - F_{v_1, v_2}(x_i, y_{t-1})) - (F_{v_1, v_2}(x_{i-1}, y_t) - F_{v_1, v_2}(x_{i-1}, y_{t-1})) \end{aligned}$$

$$F_{v_1, v_2}(x_i, y_0) = F_{v_1, v_2}(x_{i-1}, y_0) = 0 \quad \text{olduğundan} \quad \sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) = F_{v_1, v_2}(x_i, y_t) - F_{v_1, v_2}(x_{i-1}, y_t)$$

elde edilir. Korelasyon katsayısının sıfırdan büyük olması, sıfırdan küçük olması ve sıfıra eşit olması gibi üç farklı durum için inceleme yapılır.

$\rho > 0$ ise

$$F_{v_1, v_2}(x_i, y_t) = \rho \cdot \min\{F_{v_1}(x_i), F_{v_2}(y_t)\} + (1 - \rho) \cdot F_{v_1}(x_i) \cdot F_{v_2}(y_t)$$

$$F_{v_1, v_2}(x_{i-1}, y_t) = \rho \cdot \min\{F_{v_1}(x_{i-1}), F_{v_2}(y_t)\} + (1 - \rho) \cdot F_{v_1}(x_{i-1}) \cdot F_{v_2}(y_t)$$

$$F_{v_2}(y_t) = 1 \quad \text{olduğundan}$$

$$F_{v_1, v_2}(x_i, y_t) = \rho \cdot F_{v_1}(x_i) + (1 - \rho) \cdot F_{v_1}(x_i) = F_{v_1}(x_i)$$

$$F_{v_1, v_2}(x_{i-1}, y_t) = \rho \cdot F_{v_1}(x_{i-1}) + (1 - \rho) \cdot F_{v_1}(x_{i-1}) = F_{v_1}(x_{i-1})$$

$$\sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) = F_{v_1, v_2}(x_i, y_t) - F_{v_1, v_2}(x_{i-1}, y_t) = F_{v_1}(x_i) - F_{v_1}(x_{i-1}) = p_{v_1}(x_i) \quad \text{sağlanır.}$$

$\rho = 0$ ise

$$F_{v_1, v_2}(x_i, y_t) = F_{v_1}(x_i) \cdot F_{v_2}(y_t)$$

$$F_{v_1, v_2}(x_{i-1}, y_t) = F_{v_1}(x_{i-1}) \cdot F_{v_2}(y_t)$$

$$F_{v_2}(y_t) = 1 \text{ olduğundan}$$

$$F_{v_1, v_2}(x_i, y_t) = F_{v_1}(x_i)$$

$$F_{v_1, v_2}(x_{i-1}, y_t) = F_{v_1}(x_{i-1})$$

$$\sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) = F_{v_1, v_2}(x_i, y_t) - F_{v_1, v_2}(x_{i-1}, y_t) = F_{v_1}(x_i) - F_{v_1}(x_{i-1}) = p_{v_1}(x_i) \text{ sağlanır.}$$

$\rho < 0$ ise

$$F_{v_1, v_2}(x_i, y_t) = |\rho| \cdot \min\{F_{v_1}(x_i), F_{v_2}(y_t)\} + (1 - |\rho|) \cdot F_{v_1}(x_i) \cdot F_{v_2}(y_t)$$

$$F_{v_1, v_2}(x_{i-1}, y_t) = |\rho| \cdot \min\{F_{v_1}(x_{i-1}), F_{v_2}(y_t)\} + (1 - |\rho|) \cdot F_{v_1}(x_{i-1}) \cdot F_{v_2}(y_t)$$

$$F_{v_2}(y_t) = 1 \text{ olduğundan}$$

$$F_{v_1, v_2}(x_i, y_t) = |\rho| \cdot F_{v_1}(x_i) + (1 - |\rho|) \cdot F_{v_1}(x_i) = F_{v_1}(x_i)$$

$$F_{v_1, v_2}(x_{i-1}, y_t) = |\rho| \cdot F_{v_1}(x_{i-1}) + (1 - |\rho|) \cdot F_{v_1}(x_{i-1}) = F_{v_1}(x_{i-1})$$

$$\sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) = F_{v_1, v_2}(x_i, y_t) - F_{v_1, v_2}(x_{i-1}, y_t) = F_{v_1}(x_i) - F_{v_1}(x_{i-1}) = p_{v_1}(x_i) \text{ sağlanır.}$$

Korelasyon katsayısının tüm farklı durumları için $\sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) = p_{v_1}(x_i)$ eşitliği

doğrulanmıştır. □

Teorem 3.2.

Tip-II niceleyici cümle için Z-sayısı temelli değerlendirme yönteminde (Bkz. Eş. 4.4)

$$\frac{\sum_{i=1}^n \sum_{j=1}^t \left(\mu_{w_g(s_g)}(x_i) \otimes \mu_s(y_j) \right) f_{v_1, v_2}(x_i, y_j)}{\sum_{i=1}^n \left(\mu_{w_g(s_g)}(x_i) p_{v_1}(x_i) \right)} \text{ ifadesi her zaman } 0-1 \text{ aralığında sonuç verir.}$$

İspat

$$T_1 = \mu_Q \left(\frac{\sum_{i=1}^n \sum_{j=1}^t \left(\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j) \right) f_{v_1, v_2}(x_i, y_j)}{\sum_{i=1}^n \left(\mu_{w_g(S_g)}(x_i) p_{v_1}(x_i) \right)} \right) \text{ eşitliğinde } p_{v_1}(x_i) \text{ yerine } \sum_{j=1}^t f_{v_1, v_2}(x_i, y_j)$$

yazılır. (Lemma 3.1.)

$$\begin{aligned} & \frac{\sum_{i=1}^n \sum_{j=1}^t \left(\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j) \right) f_{v_1, v_2}(x_i, y_j)}{\sum_{i=1}^n \left(\mu_{w_g(S_g)}(x_i) p_{v_1}(x_i) \right)} = \frac{\sum_{i=1}^n \sum_{j=1}^t \left(\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j) \right) f_{v_1, v_2}(x_i, y_j)}{\sum_{i=1}^n \left(\mu_{w_g(S_g)}(x_i) \sum_{j=1}^t f_{v_1, v_2}(x_i, y_j) \right)} \\ & = \frac{\sum_{i=1}^n \sum_{j=1}^t \left(\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j) \right) f_{v_1, v_2}(x_i, y_j)}{\sum_{i=1}^n \sum_{j=1}^t \mu_{w_g(S_g)}(x_i) f_{v_1, v_2}(x_i, y_j)} \text{ elde edilir.} \end{aligned}$$

$\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j)$ ifadesinde herhangi bir t-norm operatörü kullanılabilir. En büyük t-norm operatörü Eş. 4.12 ile gösterilen t-norm sınırlara göre minimum operatörüdür [18].

$$\max(0, x + y - 1) \leq xy \leq \min(x, y) \quad (4.12)$$

$\mu_{w_g(S_g)}(x_i) \otimes \mu_S(y_j) = \min \left\{ \mu_{w_g(S_g)}(x_i), \mu_S(y_j) \right\} \leq \mu_{w_g(S_g)}(x_i)$ olduğundan pay daima paydadan küçüktür. En büyük t-norm operatörü için sağlanan bu eşitsizlik diğer t-norm operatörleri için de sağlanır. Bu yüzden elde edilecek sonuç daima 0-1 arasında değerler üretir. $\square$

Yukarıda verilen kopula ve ortak dağılım elde etme formüllerinden faydalanılarak tip-II niceleyici cümle için Z-sayısı temelli yöntem Örnek 3.2. ile açıklanmıştır.

Örnek 3.2.

100 çalışana ait eğitim seviyesi ve aylık gelir değerlerinin tutulduğu bir verisetinde (Ek-1) değişkenlerin değer aralığı şöyledir:

$$v_1 = \{x_i\} = \{8, 12, 15, 16, 17, 18, 19\}$$

$$v_2 = \{y_j\} = \{16, 20, 30, 40, 50, 60, 70, 85, 110\}$$

x_i ve y_j değerlerine ait 100 satırlık verideki sıklıklara bağlı olarak elde edilen olasılık ve kümülatif olasılık değerleri Çizelge 4.2 ve Çizelge 4.3'te verilmiştir.

Çizelge 4.2. Eğitim seviyesine ait olasılık ve kümülatif olasılık değerleri

x_i	8	12	15	16	17	18	19
p_i	0,11	0,32	0,30	0,14	0,04	0,02	0,07
$F_{v_1}(x_i)$	0,11	0,43	0,73	0,87	0,91	0,93	1,00

Çizelge 4.3. Aylık gelire ait olasılık ve kümülatif olasılık değerleri

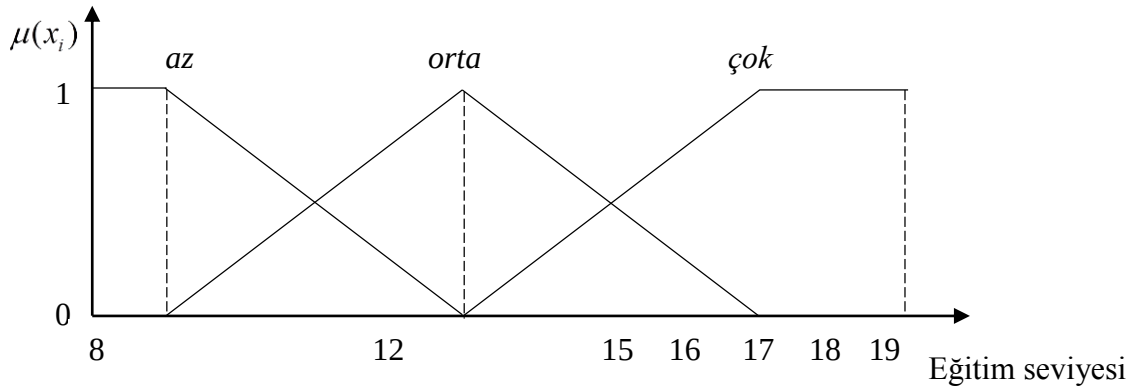
y_j	16	20	30	40	50	60	70	85	110
q_j	0,04	0,40	0,25	0,12	0,06	0,03	0,04	0,03	0,03
$F_{v_2}(y_j)$	0,04	0,44	0,69	0,81	0,87	0,90	0,94	0,97	1,00

$X \times Y$ 'de mümkün tüm durumlar için eğitim seviyesi ve aylık gelir arasındaki korelasyonu ($\rho = 0,648$) dikkate alarak hesaplama yapan Frechet kopula kullanımı ile $F_{v_1, v_2}(x_i, y_j)$ elde edilir (Eş. 4.13).

$$F_{v_1, v_2}(x_i, y_j) = C(F_{v_1}(x_i), F_{v_2}(y_j)) = 0,648 M(F_{v_1}(x_i), F_{v_2}(y_j)) + (1 - 0,648) \Pi(F_{v_1}(x_i), F_{v_2}(y_j)) \quad (4.13)$$

Daha sonra Eş. 4.11 ile $f_{v_1, v_2}(x_i, y_j)$ tüm i ve j çiftleri için elde edilir. Hesaplamalar yapılırken $F_{v_1, v_2}(x_0, y_0) = F_{v_1, v_2}(x_i, y_0) = F_{v_1, v_2}(x_0, y_j) = 0$ olarak dikkate alınmıştır.

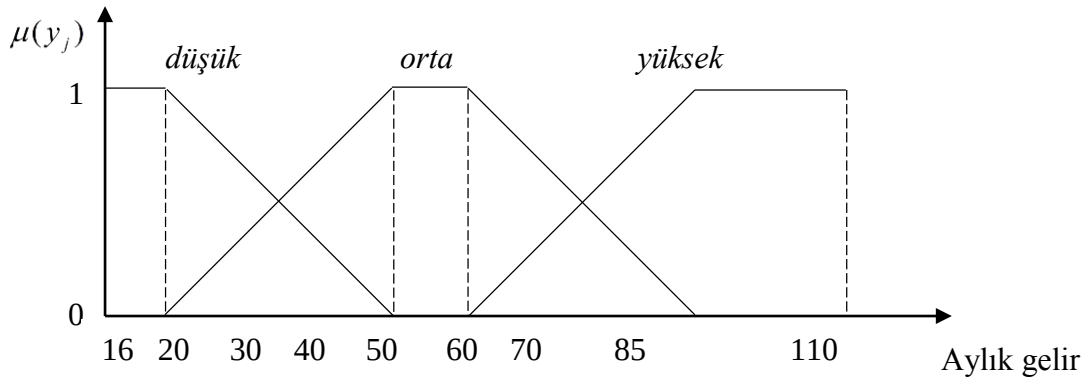
Eğitim seviyesinin az, orta ve çok bulanık kümelerine aitlik fonksiyonları Şekil 4.4 ile, x_i değerlerinin bu kümelerine aitlik dereceleri Çizelge 4.4 ve aylık gelirin düşük, orta ve yüksek bulanık kümelerine aitlik fonksiyonları Şekil 4.5 ile, y_j değerlerinin bu kümelerine aitlik dereceleri Çizelge 4.5 ile verilmiştir.



Şekil 4.4. Eğitim seviyesinin *az*, *orta*, *çok* bulanık kümelerine aitlik fonksiyonları

Çizelge 4.4. Eğitim seviyesinin *az*, *orta* ve *çok* bulanık kümelerine aitlik dereceleri

x_i	8	12	15	16	17	18	19
$\mu_{az}(x_i)$	1	0,33	0	0	0	0	0
$\mu_{orta}(x_i)$	0	0,67	0,5	0,25	0	0	0
$\mu_{çok}(x_i)$	0	0	0,5	0,75	1	1	1



Şekil 4.5. Aylık gelirin *düşük*, *orta* ve *yüksek* bulanık kümelerine aitlik fonksiyonları

Çizelge 4.5. Aylık gelirin *düşük*, *orta* ve *yüksek* bulanık kümelerine aitlik dereceleri

y_j	16	20	30	40	50	60	70	85	110
$\mu_{düşük}(y_j)$	1	1	0,67	0,33	0	0	0	0	0
$\mu_{orta}(y_j)$	0	0	0,33	0,67	1	1	0,67	0,17	0
$\mu_{yüksek}(y_j)$	0	0	0	0	0	0	0,33	0,83	1

Veriden elde edilebilecek olası bir özet “Q az eğitilmiş çalışan düşük gelirlidir.” için $w_g(S_g)$ az eğitilmiş bulanık kümesini, S düşük gelirli bulanık kümesini ifade etmektedir. Bu ifade için doğruluk derecesi şöyle hesaplanır: $X \times Y$ ’deki her eleman için $\mu_{az}(x_i) \cap \mu_{düşük}(y_j)$ ve

$f_{v_1, v_2}(x_i, y_j)$ değerleri çarpılarak toplanır ve $\mu_{az}(x_i)$ ve $p_{v_1}(x_i)$ değerlerinin çarpımlarının toplamına bölünerek elde edilen değer $Q(a) = a$ niceleyicisinde aldığı değer cümleinin doğruluk derecesidir. “Q az eğitilmiş çalışan düşük gelirlidir.” ifadesinin doğruluk derecesi 0,9039’dur.

İki değişkenle oluşturulabilecek mümkün tüm cümlelerin Z_Q (Bkz. Eş. 3.17), GD_Q (Bkz. Eş. 3.45) ve Z-sayısı temelli yöntem (Bkz. Eş. 4.4) ile hesaplanan doğruluk dereceleri Çizelge 4.6’da verilmiştir.

Çizelge 4.6. Skalar kardinalite, bulanık kardinalite temelli ve Z-sayısı temelli yöntemlere göre elde edilen sonuçlar

Özet	Z_Q	GD_Q	Z-sayısı temelli yöntem
Q az eğitilmiş çalışan düşük gelirlidir.	0,9537	0,9100	0,9039
Q az eğitilmiş çalışan orta gelirlidir.	0,1681	0,0844	0,1400
Q az eğitilmiş çalışan yüksek gelirlidir.	0	0	0,0295
Q orta eğitilmiş çalışan düşük gelirlidir.	0,9123	0,9148	0,8809
Q orta eğitilmiş çalışan orta gelirlidir.	0,3383	0,2894	0,3673
Q orta eğitilmiş çalışan yüksek gelirlidir.	0,0063	0,0049	0,0296
Q çok eğitilmiş çalışan düşük gelirlidir.	0,4675	0,3287	0,5366
Q çok eğitilmiş çalışan orta gelirlidir.	0,5610	0,4918	0,5684
Q çok eğitilmiş çalışan yüksek gelirlidir.	0,1704	0,1883	0,1420

Sonuçlardan da anlaşılacağı gibi Z-sayısı temelli yöntem skalar kardinalite temelli ve bulanık kardinalite temelli yöntemler ile çok benzer sonuçlar vermiştir. Elde edilen özetlerin doğruluk dereceleri tutarlıdır. GD_Q yöntemi dışındaki yöntemlerde verinin en çok desteklediği özet “Q az eğitilmiş çalışan düşük gelirlidir.” olmuştur. Eğitim seviyesi ve aylık gelir arasında pozitif korelasyon olduğundan bu özet beklenen bir özettir. Z-sayısı temelli önerilen yöntemde olasılıklı ve olabirlikli iki belirsizlik aynı anda hesaplamaya dâhil edildiğinden yorumlamalar açısından daha kullanışlıdır. Ayrıca heterojen belirsizliği dikkate alan bu yöntem sadece değişkenlerin bağımsız olduğu durumlar için değil değişkenler arasında bağımlılık olan durumlara da başarıyla uygulanabildiğinden diğer yöntemlerden farklıdır.

Önerilen yöntem skalar kardinalite ve bulanık kardinalite temelli yöntemlere ek olarak belirsizlik (Bkz. Eş. 3.26), kapsama (Bkz. Eş. 3.28), uygunluk (Bkz. Eş. 3.33) ve uzunluk (Bkz. 3.34) derecelerine göre ve yöntem ağırlıkları eşit alınarak toplam geçerlilik derecesine (Bkz. Eş. 3.35) göre de değerlendirilmiştir. Toplam geçerlilik derecesi hesaplanırken T_1 değeri için Z-sayısı temelli yöntem ile elde edilen sonuçlar kullanılmıştır. Olası tüm cümlelere göre elde edilen sonuçlar Çizelge 4.7’de verilmiştir.

Çizelge 4.7. Belirsizlik, kapsama, uygunluk, uzunluk ve toplam geçerlilik derecesine göre elde edilen sonuçlar

Özet	T_2	T_3	T_4	T_5	T
Q az eğitilmiş çalışan düşük gelirlidir.	0,4098	1	0,1900	1	0,7007
Q az eğitilmiş çalışan orta gelirlidir.	0,5226	0,2558	0,2742	1	0,4385
Q az eğitilmiş çalışan yüksek gelirlidir.	0,7926	0	0,1000	1	0,3844
Q orta eğitilmiş çalışan düşük gelirlidir.	0,2154	0,9079	0,0979	1	0,6204
Q orta eğitilmiş çalışan orta gelirlidir.	0,3653	0,5132	0,0168	1	0,4525
Q orta eğitilmiş çalışan yüksek gelirlidir.	0,7243	0,0316	0,0684	1	0,3708
Q çok eğitilmiş çalışan düşük gelirlidir.	0,3205	0,6667	0,1433	1	0,5334
Q çok eğitilmiş çalışan orta gelirlidir.	0,4504	0,7369	0,2069	1	0,5925
Q çok eğitilmiş çalışan yüksek gelirlidir.	0,7613	0,1754	0,0754	1	0,4308

Belirsizlik, kapsama, uygunluk, uzunluk dereceleri ve bunlarla birlikte Z-sayısı temelli yöntem ile elde edilen sonuçların ortalaması ile hesaplanan toplam geçerlilik derecesi skorlarına göre verinin yine en çok desteklediği özet “ Q az eğitilmiş çalışan düşük gelirlidir.” olmuştur.

Aynı örnek veriseti kapsama derecesi (Bkz. Eş. 3.40), fayda derecesi (Bkz. Eş. 3.41), aykırılık derecesi (Bkz. Eş. 3.42) ve basitlik derecesine (Bkz. Eş. 3.43) göre de değerlendirilmiştir. Fayda derecesi ve aykırılık derecesi hesaplanırken formüllerde T_1 değeri için Z-sayısı temelli yöntem ile elde edilen değer kullanılmıştır. Elde edilen sonuçlar Çizelge 4.8’de verilmiştir.

Kapsama derecesi skorlarına bakıldığında “ Q az eğitilmiş çalışan yüksek gelirlidir.” ve “ Q orta eğitilmiş çalışan yüksek gelirlidir.” cümlelerinin yeterli veri ile desteklenmediği sonucu çıkmaktadır. Fayda derecesinde de aynı cümleler için değer sıfır çıkmıştır. Fayda derecesi kapsama derecesine bağlı olarak ölçüldüğünden bu sonuç beklenen bir sonuçtur. Bu iki özet

cümle yeterli veri ile desteklenmediğinden skalar kardinalite, bulanık kardinalite temelli yöntemlerde olduğu gibi Z-sayısı temelli yöntemde de sıfıra yakın doğruluk derecesi vermiştir. Aykırılık derecesi elde edilen özetin ilginçliğini ölçen bir kalite ölçütüdür. Bu kalite ölçütüne göre veritabanından elde edilen “Q az eğitilmiş çalışan yüksek gelirlidir.” ve “Q orta eğitilmiş çalışan yüksek gelirlidir.” cümleleri ilginçtir, yani veriden beklenmeyen özetlerdir. Basitlik derecesine göre tüm cümleler bir öncül özetleyici ve bir özetleyiciden oluştuğu için basittir ve kolay anlaşılabilir.

Çizelge 4.8. Kapsama, fayda, aykırılık ve basitlik derecelerine göre elde edilen sonuçlar

Özet	T_c	T_u	T_o	T_s
Q az eğitilmiş çalışan düşük gelirlidir.	1	0,9039	0	1
Q az eğitilmiş çalışan orta gelirlidir.	0,8107	0,1400	0,1893	1
Q az eğitilmiş çalışan yüksek gelirlidir.	0	0	0,9705	1
Q orta eğitilmiş çalışan düşük gelirlidir.	1	0,8809	0	1
Q orta eğitilmiş çalışan orta gelirlidir.	1	0,3673	0	1
Q orta eğitilmiş çalışan yüksek gelirlidir.	0	0	0,9704	1
Q çok eğitilmiş çalışan düşük gelirlidir.	1	0,5366	0	1
Q çok eğitilmiş çalışan orta gelirlidir.	1	0,5684	0	1
Q çok eğitilmiş çalışan yüksek gelirlidir.	0,7041	0,1420	0,2959	1

5. SONUÇ VE ÖNERİLER

Özetleme tanımlayıcı amaçlı veri madenciliği tekniklerinden bir tanesidir. İstatistiksel özetleme insanların kolayca anlayabileceği ve hatırlayabileceği sonuçlar üretmediğinden, bulanık kümelerden faydalanarak insan dostu özetler üreten dilsel özetleme yöntemi ortaya çıkmıştır. Dilsel özetleme işleminde amaç, veritabanında yer alan özellikleri kelimelerle ifade ederek veriden açık, özlü ve anlaşılır bilgi üretmektir. Dilsel özetleme işleminin en önemli adımı elde edilen dilsel özetlerin değerlendirilmesidir. Değerlendirme işlemi ile tutarsız sonuçlar elde edilirse verinin desteklemediği dilsel özetler elde edilir. Literatürde bu konuda farklı yöntemler mevcuttur.

Gerçek dünya bilgisinin belirsizliğini ele almada iki önemli yaklaşım olasılık teorisi ve olabilirlik teorisi ile modelleme yapmaktır. Her iki teori de farklı tip belirsizliği ele almaktadır. Dilsel özetlerin değerlendirilmesinde olasılıksal ve olabirlikssel farklı değerlendirme yöntemleri mevcuttur, ancak her iki belirsizliği aynı anda modelleyen bir yaklaşım mevcut değildir.

Z-sayıları bilginin güvenilirliği ile ilgili yeni bir kavramdır ve değişkenin değerinin güvenilirliğini modellemek ile dilsel özetlerin doğruluğunu değerlendirmek arasında güçlü bir benzerlik mevcuttur. Bu benzerlikten yola çıkılarak dilsel özetleri değerlendirme işlemi Z-sayısı temelli bir yöntemle geliştirilmiştir. Böylece hesaplamada hem olasılık fonksiyonu hem olabilirlik dağılımı işleme katılmıştır. Ayrıca özetlenecek değişkenler arasındaki korelasyon kopula kullanımı ile değerlendirmeye dahil edilmiştir. Önerilen yöntemin sonuçları literatürde mevcut dilsel özetleri değerlendirme yöntemleri ile elde edilen sonuçlarla küçük bir örnek üzerinde karşılaştırılmıştır ve benzer sonuçlar elde edilmiştir.

KAYNAKLAR

1. Han, J., Kamber, M. and Pei, J. (2012). *Data mining: Concepts and techniques*. 4-6.
2. Vercellis, C. (2009). *Business intelligence: Data mining and optimization for decision making*. Wiley, 3-4.
3. Yager, R. R. (1982). A new approach to the summarization of data. *Information Sciences*, 28 (1), 69-86.
4. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8 (3), 338-353.
5. Zadeh, L. A. (2011). A note on z-numbers. *Information Sciences*, 181 (14), 2923-2932.
6. Nelsen, R. B. (2013). *An introduction to copulas*. Springer Science & Business Media, 1-24.
7. Krause, P. and Clark, D. (2012). *Representing uncertain knowledge: An artificial intelligence approach*. Springer Science & Business Media.
8. Klir, G. J. and Wierman, M. J. (1999). *Uncertainty-based information: Elements of generalized information theory*. Springer Science & Business Media.
9. Smets, P. (1997). Imperfect information: Imprecision and uncertainty. In (Eds.), *Uncertainty management in information systems*. Springer, 225-254.
10. Shafer, G. (1976). *A mathematical theory of evidence*. Princeton university press Princeton, 1-25.
11. Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, 325-339.
12. Negoita, C., Zadeh, L. and Zimmermann, H. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 13-28.
13. Pawlak, Z. (2012). *Rough sets: Theoretical aspects of reasoning about data*. Springer Science & Business Media, 99-102.
14. Kikuchi, S. and Chakroborty, P. (2006). Place of possibility theory in transportation analysis. *Transportation Research Part B-Methodological*, 40 (8), 595-615.
15. Florea, M. C., Jusselme, A.-L. and Bossé, É. (2007). *Fusion of imperfect information in the unified framework of random sets theory: Application to target identification*. Defence Research and Development Canada – Valcartier.
16. Tan, P. N., Steinbach, M. and Kumar, V. (2006). *Introduction to data mining*. Pearson Addison Wesley Boston, 2-3.

17. Ross, T. J. (2010). *Fuzzy logic with engineering applications* (3). John Wiley & Sons, 1-25.
18. Klir, G. J. and Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. River, NJ, USA: Prentice Hall 1-55.
19. Cheng, C. H. and Lin, Y. (2002). Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation. *European Journal of Operational Research*, 142 (1), 174-186.
20. Mendel, J. M. (1995). Fuzzy logic systems for engineering: A tutorial. *Proceedings of the IEEE*, 83 (3), 345-377.
21. Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-i. *Information Sciences*, 8 (3), 199-249.
22. Norwich, A. M. and Turksen, I. B. (1982). The construction of membership functions. *Fuzzy Sets and Possibility Theory: Recent Developments*, 61-67.
23. Türksen, I. B. (1991). Measurement of membership functions and their assessment. *Fuzzy Sets and Systems*, 40 (1), 5-38.
24. Chameau, J. and Santamarina, J. C. (1987). Membership functions i: Comparing methods of measurement. *International Journal of Approximate Reasoning*, 1 (3), 287-301.
25. Zadeh, L. A. (1978). Pruf—a meaning representation language for natural languages. *International Journal of Man-Machine Studies*, 10 (4), 395-460.
26. Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics with Applications*, 9 (1), 149-184.
27. Delgado, M., Ruiz, M. D., Sánchez, D. and Vila, M. A. (2014). Fuzzy quantification: A state of the art. *Fuzzy Sets and Systems*, 242 1-30.
28. Delgado, M., Sánchez, D. and Vila, M. A. (1999). A survey of methods for evaluating quantified sentences. *First European Society for Fuzzy Logic and Technologies Conf. (EUSFLAT'99)*, 279-282.
29. Delgado, M., Sánchez, D. and Vila, M. A. (2000). Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning*, 23 (1), 23-66.
30. Diaz-Hermida, F., Bugarin, A. and Barro, S. (2003). Definition and classification of semi-fuzzy quantifiers for the evaluation of fuzzy quantified sentences. *International Journal of Approximate Reasoning*, 34 (1), 49-88.
31. Lesot, M. J., Moyse, G. and Bouchon-Meunier, B. (2014). Interpretability of fuzzy linguistic summaries. *Fuzzy Sets and Systems*.

32. Castillo-Ortega, R., Marin, N. and Sánchez, D. (2011). Linguistic local change comparison of time series. *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, 2909-2915.
33. Kacprzyk, J. and Wilbik, A. (2009). Towards an efficient generation of linguistic summaries of time series using a degree of focus. *Fuzzy Information Processing Society, 2009. NAFIPS 2009. Annual Meeting of the North American*, 1-6.
34. Kacprzyk, J. and Zadrozny, S. (1994). Fuzzy querying for microsoft access. *Fuzzy Systems, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the Third IEEE Conference on*, 167-171.
35. Pilarski, D. (2010). Linguistic summarization of databases with quantirius: A reduction algorithm for generated summaries. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 18 (3), 305-331.
36. Rasmussen, D. and Yager, R. R. (1997). Summary sql—a fuzzy tool for data mining. *Intelligent Data Analysis*, 1 (1), 49-58.
37. Castillo-Ortega, R., Marín, N., Sánchez, D. and Tettamanzi, A. G. B. (2011). Linguistic summarization of time series data using genetic algorithms. *Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology*, 416-423.
38. Donis-Díaz, C. A., Bello, R. and Kacprzyk, J. (2015). Using ant colony optimization and genetic algorithms for the linguistic summarization of creep data. *Advances in Intelligent Systems and Computing*, 81-92.
39. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2006). Using a genetic algorithm to derive a linguistic summary of trends in numerical time series. *Evolving Fuzzy Systems, 2006 International Symposium on*, 137-142.
40. Barro, S., Bugarin, A. J., Carinena, P. and Diaz-Hermida, F. (2003). A framework for fuzzy quantification models analysis. *Ieee Transactions on Fuzzy Systems*, 11 (1), 89-99.
41. Chamorro-Martínez, J., Sánchez, D., Soto-Hidalgo, J. M. and Martínez-Jiménez, P. M. (2014). A discussion on fuzzy cardinality and quantification. Some applications in image processing. *Fuzzy Sets and Systems*, 257 85-101.
42. Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *Systems, Man and Cybernetics, IEEE Transactions on*, 18 (1), 183-190.
43. Yager, R. R. (1991). Fuzzy quotient operators for fuzzy relational data bases. *Fuzzy Engineering toward Human Friendly Systems*, 1 13-15.
44. Bosc, P. and Lietard, L. (1994). Monotonic quantified statements and fuzzy integrals. *Fuzzy Information Processing Society Biannual Conference, 1994. Industrial Fuzzy Control and Intelligent Systems Conference, and the NASA Joint Technology Workshop on Neural Networks and Fuzzy Logic*, 8-12.

45. Bosc, P. and Lietard, L. (1995). On the comparison of the sugeno and the choquet fuzzy integrals for the evaluation of quantified statements. *Proc. of EUFIT'95*, 709-716.
46. Kacprzyk, J. and Yager, R. R. (2001). Linguistic summaries of data using fuzzy logic. *International Journal of General Systems*, 30 (2), 133-154.
47. Wu, D. R. and Mendel, J. M. (2011). Linguistic summarization using if-then rules and interval type-2 fuzzy sets. *Ieee Transactions on Fuzzy Systems*, 19 (1), 136-151.
48. Kacprzyk, J. (1999). An interactive fuzzy logic approach to linguistic data summaries. *Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American*, 595-599.
49. Kacprzyk, J. and Strykowski, P. (1999). Linguistic summaries of sales data at a computer retailer via fuzzy logic and a genetic algorithm. *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*, 2, 1-943.
50. Kacprzyk, J. and Zadrozny, S. (2005). Linguistic database summaries and their protoforms: Towards natural language based knowledge discovery tools. *Information Sciences*, 173 (4), 281-304.
51. Martin, T., Shen, Y. and Majidian, A. (2010). Discovery of time-varying relations using fuzzy formal concept analysis and associations. *International Journal of Intelligent Systems*, 25 (12), 1217-1248.
52. Martin, T. and Yun, S. (2009). Fuzzy association rules in soft conceptual hierarchies. *Fuzzy Information Processing Society, 2009. NAFIPS 2009. Annual Meeting of the North American*, 1-6.
53. Knoll, A. and Glöckner, I. (2001). *Fuzzy quantifiers: A linguistic technique for data fusion*. Springer, 209-236.
54. Sánchez, D., Delgado, M. and Vila, M. A. (2009). Fuzzy quantification using restriction levels. In Di Gesù, V., et al. (Eds.), *Fuzzy logic and applications*. Springer Berlin / Heidelberg, 28-35.
55. Sánchez, D., Delgado, M., Vila, M. A. and Chamorro-Martínez, J. (2012). On a non-nested level-based representation of fuzziness. *Fuzzy Sets and Systems*, 192 (1), 159-175.
56. Delgado, M., Ruiz, M. D., Sánchez, D. and Serrano, J. M. (2011). A formal model for mining fuzzy rules using the rl representation theory. *Information Sciences*, 181 (23), 5194-5213.
57. Rocacher, D. (2003). On fuzzy bags and their application to flexible querying. *Fuzzy Sets and Systems*, 140 (1), 93-110.
58. Lietard, L. and Rocacher, D. (2008). Evaluation of quantified statements using gradual numbers. *Handbook of Research on Fuzzy Information Processing in Databases, Hershey, PA, USA*, 246-269.

59. Rocacher, D. and Bosc, P. (2005). The set of fuzzy rational numbers and flexible querying. *Fuzzy Sets and Systems*, 155 (3), 317-339.
60. Glockner, I. (2000). Advances in dfs theory. *TR2000-01, Technische Fakultät, Universität Bielefeld, PO-Box, 100131 33501*.
61. Glöckner, I. (1997). Dfs—an axiomatic approach to fuzzy quantification. *TR97-06, Techn. Fakultät, Univ. Bielefeld*, 2 (3), 10.
62. Glöckner, I. (1999). A framework for evaluating approaches to fuzzy quantification. *TR99-03, Techn. Fakultät, Univ. Bielefeld*.
63. Glöckner, I. (2004). Evaluation of quantified propositions in generalized models of fuzzy quantification. *International Journal of Approximate Reasoning*, 37 (2), 93-126.
64. Glöckner, I. (2009). Generalized fuzzy quantifiers and the modeling of fuzzy branching quantification. *International Journal of Intelligent Systems*, 24 (6), 624-648.
65. Glockner, I. and Knoll, A. (2001). A formal theory of fuzzy natural language quantification and its role in granular computing. In Pedrycz, W. (Eds.), *Granular computing: An emerging paradigm, studies in fuzziness and soft computing*. Physica-Verlag, 215-256.
66. Díaz-Hermida, F. and Bugarín, A. (2011). Semi-fuzzy quantifiers as a tool for building linguistic summaries of data patterns. *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - FOCI 2011: 2011 IEEE Symposium on Foundations of Computational Intelligence*, 45-52.
67. Díaz-Hermida, F., Bugarín, A., Cariñena, P., Mucientes, M. and Barro, S. (2003). Modelling fuzzy quantified statements under a voting model interpretation of fuzzy sets. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 151-158.
68. Losada, D. E., Díaz-Hermida, F. and Bugarín, A. (2006). Semi-fuzzy quantifiers for information retrieval. *Studies in Fuzziness and Soft Computing*, 195-220.
69. Glöckner, I. (2008). *Fuzzy quantifiers: A computational theory*. Springer.
70. Diaz-Hermida, F., Bugarin, A., Cariñena, P. and Barro, S. (2000). Evaluación probabilística de proposiciones cuantificadas borrosas. *Actas del X Congreso Espanol Sobre Tecnologias y Lógica Fuzzy (ESTYLF 2000)*, 477-482.
71. Yager, R. R. (1984). General multiple-objective decision functions and linguistically quantified statements. *International Journal of Man-Machine Studies*, 21 (5), 389-400.
72. Vila, M. A., Cubero, J. C., Medina, J. M. and Pons, O. (1997). Using owa operator in flexible query processing. In (Eds.), *The ordered weighted averaging operators*. Springer, 258-274.

73. Lietard, L. (2012). A functional interpretation of linguistic summaries of data. *Information Sciences*, 188 1-16.
74. Anderson, D., Luke, R. H., Keller, J. M., Skubic, M., Rantz, M. and Aud, M. (2009). Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Comput Vis Image Underst*, 113 (1), 80-89.
75. Kacprzyk, J. and Zadrozny, S. (2010). Modern data-driven decision support systems: The role of computing with words and computational linguistics. *International Journal of General Systems*, 39 (4), 379-393.
76. Kacprzyk, J. and Zadrozny, S. (2005). Fuzzy linguistic summaries in text categorization for human-consistent document-driven decision support systems In Reusch, B. (Eds.), *Computational intelligence, theory and applications*. Springer Berlin / Heidelberg, 271-279.
77. Kacprzyk, J. and Zadrozny, S. (2009). Supporting decision making via verbalization of data analysis results using linguistic data summaries In Rakus-Andersson, E., et al. (Eds.), *Recent advances in decision making*. Springer Berlin / Heidelberg, 121-143.
78. Kacprzyk, J., Zadrozny, S. and Wilbik, A. (2006). Linguistic summarization of some static and dynamic features of consensus reaching In Reusch, B. (Eds.), *Computational intelligence, theory and applications*. Springer Berlin Heidelberg, 19-28.
79. Pigeau, A., Raschia, G., Gelgon, M., Mouaddib, N. and Saint-Paul, R. (2003). A fuzzy linguistic summarization technique for tv recommender systems. *Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE International Conference on*, 1, 743-748.
80. Pouzols, F. M., Barriga, A., Lopez, D. R. and Sánchez-Solano, S. (2008). Linguistic summarization of network traffic flows. *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, 619-624.
81. Kacprzyk, J. and Wilbik, A. (2008). Linguistic summarization of time series using fuzzy logic with linguistic quantifiers: A truth and specificity based approach. In Rutkowski, L., et al. (Eds.), *Artificial intelligence and soft computing – icaisc 2008*. Springer Berlin / Heidelberg, 241-252.
82. Kacprzyk, J. and Wilbik, A. (2008). Linguistic summarization of time series using linguistic quantifiers: Augmenting the analysis by a degree of fuzziness. *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, 1146-1153.
83. Wilbik, A. and Kacprzyk, J. (2010). A multi-criteria evaluation of linguistic summaries of time series via a measure of informativeness. In Rutkowski, L., et al. (Eds.), *Artificial intelligence and soft computing*. Springer Berlin / Heidelberg, 105-113.

84. Kacprzyk, J., Wilbik, A. and Zadrony, S. (2008). Linguistic summarization of time series using a fuzzy quantifier driven aggregation. *Fuzzy Sets and Systems*, 159 (12), 1485-1499.
85. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2006). Linguistic summarization of trends: A fuzzy logic based approach. *Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, 2166-2172.
86. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2006). Capturing the essence of a dynamic behavior of sequences of numerical data using elements of a quasi-natural language. *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, 3365-3370.
87. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2006). On some types of linguistic summaries of time series. *Intelligent Systems, 2006 3rd International IEEE Conference on*, 373-378.
88. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2007). Linguistic summarization of time series under different granulation of describing features In Kryszkiewicz, M., et al. (Eds.), *Rough sets and intelligent systems paradigms*. Springer Berlin / Heidelberg, 230-240.
89. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2006). A linguistic quantifier based aggregation for a human consistent summarization of time series. In (Eds.), *Soft methods for integrated uncertainty modelling*. Springer, 183-190.
90. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2007). Analysis of time series via their linguistic summarization: The use of the sugeno integral. *Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference on*, 262-270.
91. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2006). Linguistic summaries of time series via a quantifier based aggregation using the sugeno integral. *FUZZ-IEEE*, 713-719.
92. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2007). Mining time series data via linguistic summaries of trends by using a modified sugeno integral based aggregation. *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, 742-749.
93. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2007). Linguistic summarization of time series by using the choquet integral In Melin, P., et al. (Eds.), *Foundations of fuzzy logic and soft computing*. Springer Berlin / Heidelberg, 284-294.
94. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2007). Linguistic summaries of time series via an owa operator based aggregation of partial trends. *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, 1-6.
95. Kacprzyk, J., Wilbik, A. and Zadrozny, S. (2010). An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation. *International Journal of Intelligent Systems*, 25 (5), 411-439.

96. Kacprzyk, J. and Wilbik, A. (2009). Using fuzzy linguistic summaries for the comparison of time series: An application to the analysis of investment fund quotations. *IFSA-EUSFLAT*, 1321-1326.
97. Wilbik, A. and Kacprzyk, J. (2011). On a benchmark related assessment of the performance of mutual (investment) funds. *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, 2934-2939.
98. Chiang, D. A., Chow, L. R. and Wang, Y. F. (2000). Mining time series data by a fuzzy linguistic summary system. *Fuzzy Sets and Systems*, 112 (3), 419-432.
99. Castillo-Ortega, R., Marín, N. and Sánchez, D. (2009). Fuzzy quantification-based linguistic summaries in data cubes with hierarchical fuzzy partition of time dimension. In (Eds.), *Intelligent data engineering and automated learning-ideal 2009*. Springer, 578-585.
100. Castillo-Ortega, R., Marín, N. and Sánchez, D. (2009). Linguistic summary-based query answering on data cubes with time dimension. In (Eds.), *Flexible query answering systems*. Springer, 560-571.
101. Castillo-Ortega, R., Marín, N. and Sánchez, D. (2011). Linguistic query answering on data cubes with time dimension. *International Journal of Intelligent Systems*, 26 (10), 1002-1021.
102. Castillo-Ortega, R., Marín, N. and Sánchez, D. (2010). Time series comparison using linguistic fuzzy techniques. In (Eds.), *Computational intelligence for knowledge-based systems design*. Springer, 330-339.
103. Van der Heide, A. and Triviño, G. (2009). Automatically generated linguistic summaries of energy consumption data. *2009 Ninth international conference on intelligent systems design and applications*, 553-559.
104. Boran, F. E. and Akay, D. (2014). A generic method for the evaluation of interval type-2 fuzzy linguistic summaries. *IEEE Transactions on Cybernetics*, 44 (9), 1632-1645.
105. Almeida, R. J., Lesot, M. J., Bouchon-Meunier, B., Kaymak, U. and Moyses, G. (2013). Linguistic summaries of categorical time series for septic shock patient data. *2013 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2013*,
106. Banaee, H., Ahmed, M. U. and Loutfi, A. (2013). A framework for automatic text generation of trends in physiological time series data. *2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, 3876-3881.
107. Boran, F. E., Akay, D. and Yager, R. R. (2014). A probabilistic framework for interval type-2 fuzzy linguistic summarization. *IEEE Transactions on Fuzzy Systems*, 22 (6), 1640-1653.
108. Hatipoğlu, H., Boran, F. E., Avci, M. and Akay, D. (2014). Linguistic summarization of europe brent spot price time series along with the interpretations from the perspective of turkey. *International Journal of Intelligent Systems*, 29 (10), 946-970.

109. Kacprzyk, J. and Zadrozny, S. (2013). Comprehensiveness and interpretability of linguistic data summaries: A natural language focused perspective. *2013 IEEE Symposium on Computational Intelligence for Human-Like Intelligence, CIHLI 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013*, 33-40.
110. Dziedzic, M., Kacprzyk, J. and Zadrozny, S. (2013). Bipolar linguistic summaries: A novel fuzzy querying driven approach. *9th Joint World Congress on Fuzzy Systems and NAFIPS Annual Meeting, IFSA/NAFIPS 2013*, 1279-1284.
111. Dziedzic, M., Kacprzyk, J. and Zadrozny, S. (2013). On some quality criteria of bipolar linguistic summaries. *2013 Federated Conference on Computer Science and Information Systems, FedCSIS 2013*, 643-646.
112. Hryniewicz, O. and Kaczmarek, K. (2015). Forecasting short time series with the bayesian autoregression and the soft computing prior information. *Advances in Intelligent Systems and Computing*, 79-86.
113. Kacprzyk, J. and Zadrozny, S. (2013). Comprehensiveness of linguistic data summaries: A crucial role of protoforms. *Studies in Computational Intelligence*, 207-221.
114. Eciolaza, L., Pereira, F. M. and Trivino, G. (2013). Automatic linguistic reporting in driving simulation environments. *Applied Soft Computing*, 13 (9), 3956-3967.
115. Eciolaza, L. and Trivino, G. (2011). Linguistic reporting of driver behavior: Summary and event description. *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, 148-153.
116. Eciolaza, L., Trivino, G., Delgado, B., Rojas, J. and Sevillano, M. (2011). Fuzzy linguistic reporting in driving simulators. *Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2011 IEEE Symposium on*, 30-37.
117. Alvarez-Alvarez, A., Sanchez-Valdes, D. and Trivino, G. (2011). Automatic linguistic description about relevant features of the mars' surface. *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, 154-159.
118. Alvarez-Alvarez, A., Sanchez-Valdes, D., Trivino, G., Sánchez, Á. and Suárez, P. D. (2012). Automatic linguistic report of traffic evolution in roads. *Expert Systems with Applications*, 39 (12), 11293-11302.
119. Alvarez-Alvarez, A., Alonso, J. M. and Trivino, G. (2013). Human activity recognition in indoor environments by means of fusing information extracted from intensity of wifi signal and accelerations. *Information Sciences*, 233, 162-182.
120. González-Villanueva, L., Alvarez-Alvarez, A., Ascari, L. and Trivino, G. (2014). A tool for linguistic assessment of rehabilitation exercises. *Applied Soft Computing*, 14 120-131.
121. Mendez-Nunez, S. and Trivino, G. (2010). Combining semantic web technologies and computational theory of perceptions for text generation in financial analysis. *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, 1-8.

122. Menendez, C. and Trivino, G. (2012). Selection of the best suitable sentences in linguistic descriptions of data. In Greco, S., et al. (Eds.), *Advances in computational intelligence*. Springer Berlin Heidelberg, 295-304.
123. Trivino, G., Sanchez, A., Montemayor, A. S., Pantrigo, J. J., Cabido, R. and Pardo, E. G. (2010). Linguistic description of traffic in a roundabout. *Fuzzy Systems (FUZZ)*, 2010 *IEEE International Conference on*, 1-8.
124. Trivino, G. and Sugeno, M. (2013). Towards linguistic descriptions of phenomena. *International Journal of Approximate Reasoning*, 54 (1), 22-34.
125. Yager, R. R. (2010). Concept representation and database structures in fuzzy social relational networks. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, 40 (2), 413-419.
126. Carrasco, R. A. and Villar, P. (2012). A new model for linguistic summarization of heterogeneous data: An application to tourism web data sources. *Soft Computing*, 16 (1), 135-151.
127. Anderson, D. T., Ros, M., Keller, J. M., Cuellar, M. P., Popescu, M., Delgado, M. and Vila, A. (2012). Similarity measure for anomaly detection and comparing human behaviors. *International Journal of Intelligent Systems*, 27 (8), 733-756.
128. Ros, M., Pegalajar, M., Delgado, M., Vila, A., Anderson, D. T., Keller, J. M. and Popescu, M. (2011). Linguistic summarization of long-term trends for understanding change in human behavior. *Fuzzy Systems (FUZZ)*, 2011 *IEEE International Conference on*, 2080-2087.
129. Wilbik, A. and Keller, J. M. (2012). A distance metric for a space of linguistic summaries. *Fuzzy Sets and Systems*, 208 79-94.
130. Wilbik, A. and Keller, J. M. (2013). A fuzzy measure similarity between sets of linguistic summaries. *IEEE Transactions on Fuzzy Systems*, 21 (1), 183-189.
131. Wilbik, A., Keller, J. M. and Alexander, G. L. (2012). Similarity evaluation of sets of linguistic summaries. *International Journal of Intelligent Systems*, 27 (10), 926-938.
132. Wilbik, A., Keller, J. M. and Bezdek, J. C. (2012). Generation of prototypes from sets of linguistic summaries. *Fuzzy Systems (FUZZ-IEEE)*, 2012 *IEEE International Conference on*, 1-8.
133. Kacprzyk, J. and Zadrozny, S. (2003). Linguistic summarization of data sets using association rules. *Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE International Conference on*, 1, 702-707.
134. Mendel, J. M. and Korjani, M. M. (2012). Charles ragin's fuzzy set qualitative comparative analysis (fsqca) used for linguistic summarizations. *Information Sciences*, 202 (20), 1-23.

135. Wijayasekara, D. and Manic, M. (2012). Visual, linguistic data mining using self-organizing maps. *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 1-8.
136. Niewiadomski, A. (2008). A type-2 fuzzy approach to linguistic summarization of data. *Ieee Transactions on Fuzzy Systems*, 16 (1), 198-212.
137. Niewiadomski, A. (2010). On finity, countability, cardinalities, and cylindric extensions of type-2 fuzzy sets in linguistic summarization of databases. *Ieee Transactions on Fuzzy Systems*, 18 (3), 532-545.
138. Tré, G. D., Dziedzic, M., Britsom, D. V. and Zadrozny, S. (2014). Dealing with missing information in linguistic summarization: A bipolar approach. *Advances in Intelligent Systems and Computing*, 57-68.
139. Popescu, M., Keller, J. M. and Zare, A. (2013). A framework for computing crowd emotions using agent based modeling. *2013 IEEE Symposium on Computational Intelligence for Creativity and Affective Computing, CICAC 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013*, 25-31.
140. Niewiadomski, A. (2013). Cylindric extensions of interval-valued fuzzy sets in data linguistic summaries. *Journal of Ambient Intelligence and Humanized Computing*, 4 (3), 369-376.
141. Korjani, M. M. and Mendel, J. M. (2014). Interval type-2 fuzzy set qualitative comparative analysis (it2-fsqca). *2014 IEEE Conference on Norbert Wiener in the 21st Century, 21CW 2014*,
142. Conde-Clemente, P., Alonso, J. M. and Trivino, G. (2013). Interpretable fuzzy system allowing to be framed in a profile photo through linguistic expressions. *8th Conference of the European Society for Fuzzy Logic and Technology, EUSFLAT 2013*, 463-468.
143. Iqbal, R., Doctor, F., Shah, N. and Fei, X. (2014). An intelligent framework for activity led learning in network planning and management. *International Journal of Communication Networks and Distributed Systems*, 12 (4), 401-419.
144. Niewiadomski, A. and Superson, I. (2014). On multi-subjectivity in linguistic summarization of relational databases. *13th International Conference on Artificial Intelligence and Soft Computing, ICAISC 2014*, 243-255.
145. Yager, R. R. (2012). On z-valuations using zadeh's z-numbers. *International Journal of Intelligent Systems*, 27 (3), 259-278.
146. Zadeh, L. A. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23 (2), 421-427.
147. Yager, R. R. (2012). On a view of zadeh's z-numbers. *14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012*, 90-101.

148. Kang, B., Wei, D., Li, Y. and Deng, Y. (2012). A method of converting z-number to classical fuzzy number. *Journal of Information and Computational Science*, 9 (3), 703-709.
149. Kang, B., Wei, D., Li, Y. and Deng, Y. (2012). Decision making using z-numbers under uncertain environment. *Journal of Computational Information Systems*, 8 (7), 2807-2814.
150. Azadeh, A., Saberi, M., Atashbar, N. Z., Chang, E. and Pazhoheshfar, P. (2013). Z-ahp: A z-number extension of fuzzy analytical hierarchy process. *2013 7th IEEE International Conference on Digital Ecosystems and Technologies: Smart Planet and Cyber Physical Systems as Embodiment of Digital Ecosystems, DEST 2013*, 141-147.
151. Soroudi, A. and Amraee, T. (2013). Decision making under uncertainty in energy systems: State of the art. *Renewable and Sustainable Energy Reviews*, 28, 376-384.
152. Aliev, R. A. and Zeinalova, L. M. (2014). Decision making under z-information. *Studies in Computational Intelligence*, 233-252.
153. Cai, G. P. (2014). Green strategy selection under uncertain environment. *Beijing Ligong Daxue Xuebao/Transaction of Beijing Institute of Technology*, 34, 182-185.
154. Lorkowski, J., Kreinovich, V. and Aliev, R. (2014). Towards decision making under interval, set-valued, fuzzy, and z-number uncertainty: A fair price approach. *2014 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2014*, 2244-2253.
155. Mohamad, D., Shaharani, S. A. and Kamis, N. H. (2014). A z-number-based decision making procedure with ranking fuzzy numbers method. *3rd International Conference on Quantitative Sciences and Its Applications: Fostering Innovation, Streamlining Development, ICOQSIA 2014*, 160-166.
156. Xiao, Z. Q. (2014). Application of z-numbers in multi-criteria decision making. *2014 International Conference on Informative and Cybernetics for Computational Social Systems, ICCSS 2014*, 91-95.
157. Zeinalova, L. M. (2014). Expected utility based decision making under z-information. *Intelligent Automation and Soft Computing*, 20 (3), 419-431.
158. Sotoudeh-Anvari, A. and Sadi-Nezhad, S. (2015). A new approach based on the level of reliability of information to determine the relative weights of criteria in fuzzy topsis. *International Journal of Applied Decision Sciences*, 8 (2), 164-178.
159. Tavakkoli-Moghaddam, R., Sotoudeh-Anvari, A. and Siadat, A. (2015). A multi-criteria group decision-making approach for facility location selection using promethee under a fuzzy environment. *Lecture Notes in Business Information Processing*, 145-156.
160. Banerjee, R. and Pal, S. K. (2015). On z-numbers and the machine-mind for natural language comprehension. *Studies in Fuzziness and Soft Computing*, 415-457.

161. Banerjee, R. and Pal, S. K. (2013). The z-number enigma: A study through an experiment. *Studies in Fuzziness and Soft Computing*, 71-88.
162. Pal, S. K., Banerjee, R., Dutta, S. and Sarma, S. S. (2013). An insight into the z-number approach to cww. *Fundamenta Informaticae*, 124 (1-2), 197-229.
163. Patel, P., Khorasani, E. S. and Rahimi, S. (2015). Modeling and implementation of z-number. *Soft Computing*.
164. Zhang, W. S., Li, J. L., Ran, L. and Ma, Y. (2013). Weapon system effectiveness evaluation based on z-number. *Binggong Xuebao/Acta Armamentarii*, 34 (SUPPL.2), 1-6.
165. Azadeh, A., Kokabi, R., Saberi, M., Hussain, F. K. and Hussain, O. K. (2014). Trust prediction using z-numbers and artificial neural networks. *2014 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2014*, 522-528.
166. Salari, M., Bagherpour, M. and Wang, J. (2014). A novel earned value management model using z-number. *International Journal of Applied Decision Sciences*, 7 (1), 97-119.
167. Tadayon, S. and Tadayon, B. (2014). Approximate z-number evaluation based on categorical sets of probability distributions. *Studies in Fuzziness and Soft Computing*, 117-134.
168. Aliev, R. A., Alizadeh, A. V. and Huseynov, O. H. (2015). The arithmetic of discrete z-numbers. *Information Sciences*, 290 (C), 134-155.
169. Aliev, R. A., Alizadeh, A. V., Huseynov, O. H. and Jabbarova, K. I. (2015). Z-number-based linear programming. *International Journal of Intelligent Systems*, 30 (5), 563-589.
170. Aliev, R. and Memmedova, K. (2015). Application of z-number based modeling in psychological research. *Computational Intelligence Neuroscience*, 2015 760403.
171. Schweizer, B. and Sklar, A. (2011). *Probabilistic metric spaces*. Courier Corporation.
172. Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9, 449-460.
173. Durante, F. and Sempi, C. (2010). Copula theory: An introduction. In (Eds.), *Copula theory and its applications*. Springer, 3-31.
174. Beliakov, G., Pradera, A. and Calvo, T. (2007). *Aggregation functions: A guide for practitioners*. Springer.
175. Klement, E. P., Mesiar, R. and Pap, E. (2013). *Triangular norms*. Springer Science & Business Media.
176. Yager, R. R. (1980). On a general class of fuzzy connectives. *Fuzzy Sets and Systems*, 4 (3), 235-242.

177. Trivedi, P. K. and Zimmer, D. M. (2005). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1 (1), 1-111.
178. Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance*. John Wiley & Sons.
179. Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76 (3), 639-650.
180. Yager, R. R. and Alajlan, N. (2015). On a role for copula's in jeffrey's rule with an application to decision making. *International Journal of Intelligent Systems*, 30 (11), 1117-1132.

EKLER

EK-1 Çalışan veriseti

id	educ	salary	id	educ	salary	id	educ	salary	id	educ	salary
1	15	\$50	26	15	\$30	51	12	\$30	76	15	\$20
2	16	\$40	27	19	\$60	52	15	\$40	77	12	\$20
3	12	\$20	28	15	\$30	53	18	\$70	78	12	\$20
4	8	\$20	29	19	\$110	54	12	\$20	79	16	\$30
5	15	\$40	30	15	\$30	55	12	\$20	80	16	\$50
6	15	\$30	31	12	\$30	56	15	\$20	81	12	\$20
7	15	\$30	32	19	\$110	57	15	\$30	82	12	\$20
8	12	\$20	33	15	\$40	58	15	\$20	83	12	\$20
9	15	\$20	34	19	\$85	59	15	\$20	84	8	\$20
10	12	\$20	35	17	\$85	60	12	\$30	85	15	\$20
11	16	\$30	36	8	\$30	61	8	\$20	86	15	\$30
12	8	\$20	37	12	\$20	62	16	\$40	87	12	\$20
13	15	\$20	38	15	\$30	63	17	\$50	88	19	\$70
14	15	\$30	39	16	\$30	64	16	\$50	89	19	\$60
15	12	\$20	40	15	\$16	65	8	\$20	90	8	\$16
16	12	\$40	41	12	\$20	66	19	\$70	91	12	\$16
17	15	\$40	42	15	\$30	67	16	\$40	92	8	\$20
18	16	\$110	43	12	\$20	68	16	\$40	93	12	\$20
19	12	\$40	44	8	\$20	69	16	\$50	94	12	\$20
20	12	\$20	45	12	\$30	70	15	\$40	95	12	\$20
21	16	\$30	46	15	\$20	71	17	\$85	96	8	\$30
22	12	\$20	47	12	\$30	72	16	\$50	97	17	\$40
23	15	\$20	48	12	\$30	73	12	\$20	98	8	\$30
24	12	\$16	49	15	\$30	74	15	\$30	99	12	\$20
25	15	\$20	50	16	\$60	75	15	\$20	100	18	\$70

Educ: Eğitim seviyesi, salary: aylık gelir

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : AYDOĞAN, Sena
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 01.01.1991, Ankara
 Medeni hali : Bekâr
 Telefon : 0 (312) 582 38 43
 Faks : 0 (312) 230 84 34
 e-mail : senaaydogan@gazi.edu.tr



Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi / Endüstri Mühendisliği	2015
Lisans	Gazi Üniversitesi/ Endüstri Mühendisliği	2012
Lise	Cumhuriyet Anadolu Lisesi	2008

İş Deneyimi

Yıl	Yer	Görev
2014-Halen	Gazi Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce

Yayınlar

Keleş, A., Demircan, A., Bildik, F., Aydoğan, S., Düzce, T. ve Akay, D. (2015, 14-16 Ekim). *Classification of Patients at Risk for CO Poisoning in an Emergency Department with CHAID Decision Tree*, 15. Üretim Araştırmaları Sempozyumu, İzmir, 1109-1113.

Hobiler

Gezi, Sinema



GAZİ GELECEKTİR..