Tasarım ölçümlerinin (metriklerin) olası hatalar ile ilişkisi
Lojistik Regresyon (*Logistic Regression*)

Makine öğrenmesine dayalı yöntemlerden biri de lojistik regresyon yöntemidir.

Ölçümler (metrikler) ile yazılım kalitesi (karşılaşılan hatalar) arasındaki ilişki ne kadar kuvvetlidir?

Ölçümler yardımıyla ileride hataya neden olacak sınıflar tahmin edilebilir mi?

Konuyu incelerken aşağıdaki yazıdaki deneylerden yararlanacağız:

- L. C. Briand, J. Wüst, J. W. Daly, and D. Victor Porter, "Exploring the relationships between design measures and software quality in object-oriented systems," *Journal of Systems and Software*, vol. 51, no. 3, pp. 245-273, May 2000.

Yazı uzun süre önce yayımlanmasına karşın burada sözü edilen matematiksel yöntemler geçerlidir ve başka projeler üzerinde de uygulanabilirler.

Matematiksel yöntemlerle ilgili kaynak:

- David W. Hosmer, Jr., Stanley Lemeshow, Rodney X. Sturdivant, "Applied Logistic Regression", 3/e, Wiley, 2013.

  (Daha eski baskıları da kullanılabilir)

2012 - 2022 Feza BUZLUCA

Sınanacak olan Hipotezler

- **H-IC** (import coupling measures): Başka sınıflara bağımlılığı yüksek olan sınıfta hata çıkma olasılığı daha yüksektir.
- **H-EC** (export coupling measures): Kendisine başka sınıfların çok bağımlığı olduğu sınıfta hata çıkma olasılığı daha yüksektir.
- **H-COH** (cohesion measures): Uyumu düşük olan sınıflarda hata çıkma olasılığı daha yüksektir.
- **H-Depth** (measures DIT, AID): Türetim ağacında daha derinde olan sınıflarda hata çıkma olasılığı daha yüksektir.

  DIT (depth of inheritance tree)

  AID (average inheritance depth of a class)
- **H-Ancestors** (measures NOA, NOP, NMI): Çok sayıda atası (öncülü) olan sınıflarda hata çıkma olasılığı daha yüksektir.

  NOA (number of ancestors)

  NOP (number of parents)

  NMI (number of methods inherited)

2012 - 2022 Feza BUZLUCA

Sınanacak olan Hipotezler (devamı)

- **H-Descendents** (measures NOC, NOD, CLD): Çok sayıda çocuğu/torunu (ardılı) olan sınıflarda hata çıkma olasılığı daha yüksektir.

  NOC (number of children)

  NOD (number of descendents)

  CLD (class-to-leaf depth)

- **H-OVR** (measures NMO, SIX): İçinde çok sayıda metot örtülen (*override*) sınıflarda hata çıkma olasılığı daha yüksektir.

  NMO (number of methods overridden)

  SIX (specialization index) = NMO * DIT / (NMO+NMA+NMI)

  NMA (number of methods added), NMI (number of methods inherited)

- **H-SIZE** (measure NMA and the size measures): Büyük sınıflarda hata çıkma olasılığı daha yüksektir.

Çalışmada kullanılan metriklerin tanımları için ilgili yazıdaki Tablo 20-22'ye bakınız.

2012 - 2022 Feza BUZLUCA

Veriler üzerinde önişlemler

Değerlendirmeye başlamadan önce verilerin (ölçüm değerleri) ne kadar sağlıklı olduğu ile ilgili araştırmalar ve çeşitli önişlemler yapılır.

1. Ölçümlerin (metrik değerlerinin) dağılımlarının ve varyanslarının incelenmesi:

Çok değişim göstermeyen ölçümler sınıflar hakkında ayırt edici bilgi taşımazlar.

Örnek çalışmada sıfırdan farklı en az beş sonuç üretmiş olan ölçmeler değerlendirmeye alınmıştır.

2. Uçdeğerlerin (*outlier*) belirlenmesi:

Aynı değişkenle ilgili bazı ölçüm sonuçları (değişik sınıflar için) diğerlerinden çok farklı olabilir.

Bu tür uç değerler oluşturulan hata öngörü modelini olumsuz etkileyebilir.

Bu nedenle uç değerlerin belirlenerek incelenmesi ve eğer gerekli ise veri kümesinden ayıklanmaları gerekir.

Tek değişkenli (*univariate*) ve çok değişkenli (*multivariate*) uç değerler (*outlier*) üzerinde analizler yapılır.

Hata öngörüsü üzerinde etkili olan uç değerler (*influential outlier*) (varlığı yokluğu fark yaratan) veri kümesinden çıkartılır.

2012 - 2022 Feza BUZLUCA

Hata öngörü modelinin (*Prediction model*) oluşturulması

Bu modelde sınıflardan toplanan metrikler **bağımsız değişkenler** (*predictor, covariate*), o sınıfta hata olma olasılığı ise **bağımlı değişken** (*dependent*) olacaktır.

Model üç adımda oluşturulacaktır.

1. Tek değişkenli lojistik regresyon (*univariate logistic regression*)
2. Çok değişkenli lojistik regresyon (*multivariate logistic regression*)
3. Modelin değerlendirilmesi (*model evaluation (goodness of fit)*)

Tek değişkenli lojistik regresyon (*univariate logistic regression*)

Her bağımsız değişkenin (metrik) tek başına bağımlı değişken (sınıfta hata olması) ile beklenen doğrultuda ilişkisi araştırılır.

Amaç bir metriğin beklendiği şekilde hata çıkma olasılığını etkileyip etkilemediğini görmektir.

Burada daha önce verilen hipotezlerin geçerliliği sınanır.

Çok değişkenli lojistik regresyon (*multivariate logistic regression*)

Çok değişkenden oluşan hata öngörü modelinin kurulması amaçlanır.

Birden fazla metriğin uygun bir birleşiminin (kombinasyonunun) hata öngörüsünde nasıl kullanılacağı belirlenir.

2012 - 2022 Feza BUZLUCA

Lojistik regresyon

Veri kümesine uygun olarak aşağıdaki gibi bir denklemi oluşturmak amaçlanır.

Logit transformassion:

$$ln\left[\frac{p(X)}{1-p(X)}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_mx_m$$

$$p(X) = \frac{e^{(b_0+b_1x_1+b_2x_2+b_3x_3+\cdots+b_mx_m)}}{1 + e^{(b_0+b_1x_1+b_2x_2+b_3x_3+\cdots+b_mx_m)}}$$

Burada $x_1, x_2, x_3, \ldots, x_m$ bağımsız değişkenlerdir. $X = \{x_1, x_2, x_3, \ldots, x_m\}$

Bizim konumuzda bunlar sınıflardan elde edilen ölçümlerdir (metrikler).

$p(X)$, incelenen bir olayın, $X = \{x_1, x_2, x_3, \ldots, x_m\}$ değerlerine bağlı olarak meydana gelme olasılığıdır.

Bizim örneğimizde $p(X)$ sınıfta hata olma olasılığıdır.

Örneğin tek değişkenli bir regresyon modelinde DIT=3 değeri için p(DIT=3) = 0.4 sonucu elde ediliyorsa DIT=3 olan bir sınıfta hata olma olasılığı 0.4 demektir.

Diğer bir deyişle DIT=3 olan sınıfların %40'ında hata vardır.

2012 - 2022 Feza BUZLUCA

Lojistik regresyon (devamı)

Modelin oluşturulması:

Regresyon modeli oluşturulurken amaç doğru katsayıları ($b_0, b_1, b_2, b_3, \dots, b_m$) belirlemektir.

Lojistik regresyon modelinin doğru olması için bağımsız değişkenler (x_i) ile $p(X)$ arasında *monotonik* ilişki olmalıdır.

Katsayılar belirlenirken maksimum olabilirlik (*maximum likelihood*) yöntemi kullanılır.

Olabilirlik fonksiyonu olarak veri kümesi üzerinde yapılan her bir gözlemde (denemede) elde edilen olasılıkların çarpımı kullanılır.

Her bir gözlemde x_i'ler bilinenler, b_i'ler ise bilinmeyenlerdir.

Örnek çalışmadaki bir gözlemde bir sınıfın metrik değerleri (x_i'ler) ve o sınıfta hata olup olmadığı bilgileri elde edilir.

Bu olasılıkların çarpımlarının kullanılabilmesi için her bir gözlemin diğerlerinden istatistiksel olarak bağımsız olması gerekir.

Maksimizasyon işlemi yapılırken kolaylık amacıyla olabilirlik fonksiyonunun doğal logaritması (ln) kullanılır (*Log likelihood*) (Türev almada kolaylık).

Olabilirlik fonksiyonunu maksimize eden katsayılar (b_i) seçilir.

2012 - 2022 Feza BUZLUCA

Çok değişkenli lojistik regresyon (*multivariate logistic regression*)

Çok değişkenden oluşan hata öngörü modelinin kurulması amaçlanır.

Birden fazla metriğin uygun bir birleşiminin (kombinasyonunun) hata öngörüsünde nasıl kullanılacağı belirlenir.

Bağımsız değişkenler (metrikler) seçilirken iki konuya dikkat edilir:

- Değişken sayısının mümkün olduğu kadar az tutulması,
- Birbirine bağımlı (*highly correlated*) değişkenlerin azaltılması.

Böylece model daha kolay anlaşılır ve daha genel (o deneydeki verilerden bağımsız) olabilir.

Bağımsız değişkenler seçilirken bu çalışmada çok sayıda değişken olduğu için ileriye doğru ekleme (*forward selection*) yöntemi kullanılmıştır.

Her adımda, o anda modelde olmayan değişkenler sınanır ve en anlamlı sonuç veren modele eklenir.

Yeni bir değişken eklendiğinde modeldeki değişkenlerden birinin modele katkısı azalırsa (α_{exit}=0.10) bu değişken modelden çıkarılır.

Ekleme işleminin durması: Eğer dışarıda kalan değişkenlerden hiçbirinin modele olan katkısı belli bir değeri geçmiyorsa (α_{enter}=0.05) değişken ekleme işlemi sonlandırılır.

2012 - 2022 Feza BUZLUCA

Çok değişkenli lojistik regresyon (devamı)

Seçilen değişkenler arasındaki bağımlılığın (*multi-collinearity*) incelenmesi:

Değişkenler arası bağımlılık modelin anlaşılırlığını ve kullanışlılığını azaltır.

Bu analizi yapmak için Belsley tarafından ortaya koyulan yöntem önerilmiştir.

Buna göre modeldeki değişkenlere PCA yöntemi uygulanır.

λ_{max}: En büyük özdeğer; λ_{min}: En küçük özdeğer olmak üzere aşağıdaki gibi bir koşul değeri (*conditional number*) hesaplanır.

$$koşul\ değeri = \sqrt{\lambda_{max}/\lambda_{min}}$$

Koşul değerinin (*conditional number*) büyük olması bağımlılığa (*multi-collinearity*) işaret eder.

Belsley'e göre bu değer 30'u aştığında değişkenler incelenip modelde düzeltme yapılmalı.

2012 - 2022 Feza BUZLUCA

Bir bağımsız değişkenin bağımlı değişken üzerindeki etkisi:

Bu etkiyi görmek için incelenen değişken nedeniyle olasılıklar oranında (*odds ratio*) meydana gelen değişim hesaplanır.

Olasılıklar oranı:

$$odds\ ratio\ OR(x_i) = \frac{p(x_i)}{1 - p(x_i)}$$

Bu oran örnek çalışmada, belli bir metrik değeri elde edildiğinde bir sınıfta hata olma olasılığının hata olmama olasılığına oranını gösterir.

Örneğin belli bir X değeri için OR(X)=2 ise bu durumda sınıfta hata olma olasılığı olmama olasılığının 2 katıdır.

Olasılıklar oranında (*odds ratio*) meydana gelen değişim:

$$\Delta OR(x_i) = \frac{OR(x_i + \sigma)}{OR(x_i)}$$

Burada σ, x_i ölçümünün standart sapmasıdır.

$\Delta OR(x_i)$ oranı, x_i değerinin standart sapma kadar artması veya azalması sonucu olasılıklar oranında meydana gelen değişikliği gösterir.

Bu değer x_i bağımsız değişkeninin (metrik) bağımlı değişken (sınıfta hata olması) üzerinde ne kadar etkili olduğunu gösterir.

2012 - 2022 Feza BUZLUCA

Modelin değerlendirilmesi

Model, bilinen bir veri kümesi kullanılarak oluşturulur (katsayılar bulunur).

Bu veri kümesindeki sınıfların metrik değerleri ve hangi sınıflarda hata olduğu bilinmektedir.

Hem modeli oluştururken hem de modeli sınarken aynı veri kümesi kullanılabilir ancak bu durumda sonuçlar gerçektekinden daha iyi çıkabilir.

Verilerde değişiklik yaratmak için k-katlı çapraz sağlama yöntemi (*k-fold cross validation*) kullanılabilir.

Bu durumda veri kümesi k gruba bölünür.

Model k-1 gruptaki veri ile oluşturulur; sınama kalan diğer grup ile yapılır.

Bu işlem k defa her grup üzerinde sınama yapılacak şekilde tekrar edilir.

Hatırlatma; model metrik değerlerine bağlı olarak bir sınıfın hatalı olma olasılığını vermektedir p(x).

Modelin verdiği p(x) değeri önceden belirlenen bir p_0 eşik değerinden büyükse o sınıf "hataya sahip" olarak tahmin edilmiş olur. ($p(x) > p_0 \rightarrow$ sınıf hatalı)

Bu çalışmada p_0 eşik değeri seçilirken veri kümesindeki gerçekten hatalı sınıf sayısı ile hatalı olarak tahmin edilebilecek sınıf sayısının eşit (yakın) çıkması sağlanmaya çalışılmıştır.

2012 - 2022 Feza BUZLUCA

Modelin değerlendirilmesi (devamı)

Modelin gerçekten hatalı olan sınıfları tahmin etmekteki başarısı aşağıdaki ölçümlerle değerlendirilir:

- **Tamlık** (Eksiksizlik) (*Completeness*) (*Recall*):

Tamlık= Hatalı olduğu öngörülebilen sınıf sayısı / toplam hatalı sınıf sayısı

Eğer bir sınıfta birden fazla hata varsa tamlık hata sayılarına göre de tanımlanabilir:

Tamlık= Hatalı olduğu öngörülebilen sınıflardaki toplam hata sayısı /
Sistemdeki toplam hata sayısı

Tamlık değeri küçükse bazı hataların gözden kaçırıldığı anlaşılır.

Bu değerin büyük olması ise hataların çoğunun tahmin edildiği anlamına gelir.

Ancak bu değerin tek başına kullanılması anlamlı değildir.

2012 - 2022 Feza BUZLUCA

Modelin değerlendirilmesi (devamı)

- **Doğruluk** (*Correctness*) (*Precision*):

Hatalı olarak tahmin edilen sınıf sayısını (tamlık) arttırmak için sınıfları hatalı olarak değerlendirmede kullanılan eşik değeri p_0 küçültülebilir.

Ancak bu durumda hatalı olmadığı halde yanlışlıkla hatalı olarak tahmin edilen sınıf sayısı da artacaktır (*false positive*).

Doğruluk = Hatalı olduğu doğru olarak öngörülen sınıf sayısı /
Hatalı olduğu tahmin edilen toplam sınıf sayısı

Bu değerin yüksek olması istenir. Aksi durumda test yapanlar ve sınıfların bakımını yapanlar aslında hatalı olmayan sınıflarla zaman harcamış olurlar.

2012 - 2022 Feza BUZLUCA

Modelin değerlendirilmesi (devamı)

- **Kappa (Cohen)** :

Bu ölçü iki değerlendiricinin (*rater*) aynı konuda verdikleri kararların birbirleriyle ne kadar uyumlu olduğunu gösterir.

K (Kappa) -1,+1 arası değerler alır.

K, 1'e yakın ise iki değişken (verilen kararlar) uyumludur.

K, 0'a yakın ise iki değişken (verilen kararlar) arasındaki uyum rastlantısaldır.

K, 0'dan küçükse iki değişken (verilen kararlar) arasındaki uyum rastlantısaldan da azdır.

Bu çalışmada Kappa ölçüsü modelin değerlendirme sonucu ile gerçekte sınıfların hatalı olması arasındaki uyumu belirlemek için kullanılmıştır.

Genellikle K > 0.7 uyum sınırı olarak kullanılır.

2012 - 2022 Feza BUZLUCA

Sonuçlar

Tek değişkenli (*univariate*) analizler :

- **Bağımlılık (*Coupling*) Ölçümleri:**

Yeteri kadar değişim gösteren ölçümler (metrikler) hata üzerinde etkili olmuştur.

Sonuçlar H-IC (*Import coupling*) hipotezini desteklemektedir.

Metot çağırılarıyla ilgili bağımlılık metriklerinin ΔOR değeri diğerlerine göre yüksektir.

Demek ki metot çağırılması şeklindeki bağımlılıklar hatalar üzerinde etkili olmaktadır.

Ele alınan bazı bağımlılık metriklerinin hatalar üzerinde etkili olmadığı fikrine varılmıştır.

Etkili olmayan bu metrik değerlerinin ortalamaları ve standart sapmaları da diğerlerine göre düşük değerlerdedir.

H-EC (*Export coupling*) ile ilgili bir kanıt bulunamıyor, çünkü sınıfa dışarıdan gelen bağımlılıklarla ilgili metriklerin anlamlılık değerleri düşük çıkıyor.

Sadece OCAEC α=0.05 düzeyinde anlamlıdır; onun regresyon katsayısı negatif çıkıyor.

Bunun anlamı dışarıdan bağımlı olan sınıfların hata olasılıklarının düşmesidir.

2012 - 2022 Feza BUZLUCA

Tek değişkenli (*univariate*) analizler (devamı) :

- **Uyum (*Cohesion*) Ölçümleri:**

Uyum ile ters orantılı değerler üreten LCOM1-5 metriklerinin katsayıları pozitif,

Uyum ile düz orantılı değerler üreten LCC, TCC gibi metriklerinin katsayıları ise negatif çıkıyor.

Demek ki sınıftaki uyum azaldıkça hata çıkma olasılığı artmaktadır.

Sadece ICH metriğinin katsayısı beklenmedik şekilde pozitif çıkmıştır. "Uyumlu sınıfta daha çok hata olur."

Ancak yazarlar ICH'nin uyumdan çok karmaşıklık ölçtüğü görüşündedir.

Sadece LCOM3, Coh ve ICH metrikleri α=0.05 sınır değerine göre anlamlıdır.

ICH uyumu ölçmemektedir.

Bu nedenlerle değerler H-COH hipotezini ("uyumlu sınıfta daha az hata olur") sadece zayıf bir şekilde desteklemektedir.

LCOM1,3,4,5 Coh ve ICH metrikleri α=0.25 düzeyinde anlamlı oldukları için çok değişkenli regresyonda kullanılacaklardır.

LCOM2 en düşük anlamlı metrik çıkmıştır. Zaten bu metrik araştırmacılar tarafından eleştirilmektedir.

2012 - 2022 Feza BUZLUCA

Tek değişkenli (*univariate*) analizler (devamı) :

- **Kalıtım (*Inheritance*) Ölçümleri:**

Tüm kalıtım ölçümleri α=0.05 düzeyinde anlamlı çıkmıştır.

H-Depth hipotezi DIT ve AID ölçümleri ile desteklenmiştir.

Bir türetim ağacında derinde olan sınıflarda hata olma olasılığı daha yüksektir.

H-Ancestors hipotezi NOA, NOP ve NMI ölçümleri ile desteklenmiştir.

Bir sınıfın çok atası varsa ve çok sayıda metodu kalıtımla almışsa hata olma olasılığı daha yüksektir.

NOA, NOP, NMI ölçümleri DIT ve AID ile aynı PC'de (temel bileşen) yer aldıklarından H-Depth ve H-Ancestors arasında bir ayrım yapmak zordur.

İkisi de büyük olasılıkla aynı bilgiyi veriyordur.

H-Descendents hipotezi NOC, NOD ve CLD ölçümleri ile incelendiğinde çok sayıda çocuğu olan sınıfların daha az hata içerdiği görülmüştür.

Nedeni, yazılım geliştirenlerin bu sınıflara daha dikkatli yaklaşması olabilir.

Diğer taraftan bu ölçümlerin ΔOR değerleri çok düşüktür. Bu nedenle bu ölçümlerin hata olasılıkları üzerindeki etkisi de azdır.

2012 - 2022 Feza BUZLUCA

Tek değişkenli (*univariate*) analizler (devamı) :

- **Kalıtım (*Inheritance*) Ölçümleri** (devamı) :

H-OVR hipotezi, NMO ve SIX ölçümleri ile desteklenmektedir.

Bir sınıfta çok sayıda metot örtülüyorsa (*overriding*) hata olma olasılığı daha yüksektir.

H-SIZE hipotezi NMA ölçümleri ile desteklenmektedir.

Çok sayıda metot eklenen sınıfta hata olma olasılığı daha yüksektir.

- **Boyut (*Size*) Ölçümleri :**

Sınıfların boyutlarını ölçen tüm metrikler anlamlı çıkıyor.

H-SIZE hipotezi destekleniyor.

Bir sınıf ne kadar büyükse hata çıkma olasılığı da o kadar yüksektir.

Boyut ölçen metrikler arasında en etkilisi Stmts olarak görünüyor.

Stmts: Yürütülebilir (*executable*) ve tanımlama (*declaration*) ifadelerinin toplam sayısı

Stmts için ΔOR = 4.952

2012 - 2022 Feza BUZLUCA

Boyut ölçümleri diğer metrikler arasındaki korelasyonun araştırılması

Yazılımların boyutlarının hataların oluşmasında etkili olduğu bilinmektedir.

Bu nedenle diğer ölçümlerin (bağımlılık, uyum, kalıtım) boyuttan farklı bir bilgi verip vermediğini anlamak için boyut ölçümleri ile diğer ölçümler arasında korelasyon analizi yapmak gerekir.

Eğer aralarında yüksek korelasyon varsa zaten boyuttan başka bir şey ölçmeye gerek kalmayacaktır.

Bu çalışmada, en etkili olan boyut ölçüsü Stmts ile diğer ölçüler arasında Spearman rho (δ) katsayısı hesaplanmıştır.

Bağımlılık ölçümleri için rho genelde 0.5'ten düşüktür.

Bu ölçümler ile boyut arasında orta düzeyde bir korelasyon olduğu söylenebilir.

Uyumluluk ölçümleri için rho değerleri genelde düşüktür.

Kalıtım ölçümleri için de rho değerleri genelde düşüktür.

Sadece NMA ölçümü için δ=0.397 çıkmıştır. NMA sınıfa eklenen metot sayısı olduğu için boyutla bağlantılı olması beklenen bir durumdur.

2012 - 2022 Feza BUZLUCA

Çok değişkenli regresyon modeli (*multivariate logistic regression model*)

Bu aşamada uygun ölçüler (metrikler) kullanılarak bir hata öngörü modeli kurulmaya çalışılıyor.

Üç farklı model kurulup karşılaştırılıyor:

1. Sadece boyut ölçülerinden oluşan bir model
2. Bağımlılık, uyum, kalıtım ölçülerinden oluşan bir model
3. Tüm gruplardaki ölçüleri içerebilen bir model

Cevabı aranan soru:

Elde edilmesi daha zor olan bağımlılık, uyum, kalıtım ölçüleri ile bir model oluşturmak gerekli mi?

Sadece daha kolay toplanabilen boyut ölçüleri ile sağlıklı bir model oluşturulamaz mı?

2012 - 2022 Feza BUZLUCA

Çok değişkenli regresyon modeli (*multivariate logistic regression model*) (devamı)

Model 1: Sadece boyut ölçülerinden oluşan bir model

İleri doğru seçim yöntemiyle (*forward selection*) metrikler modele katılmıştır.

Modelde 3 metrik (bağımsız değişken) yer almıştır.

Veri kümesine model uygulandığında elde edilen sonuçlar:

Correctness: %60. 53 sınıfı hatalı buluyor, aslında 32 tanesi hatalı

Completeness: %67. Sistemdeki 234 hatanın 158 tanesini buluyor. (Bir sınıfta birden fazla hata olabiliyor)

Kappa (tahmin edilen ile gerçek hatalar arasında) = 0.25 (Düşük bir değer)

Sonuç olarak sadece boyut ölçüleri ile sağlıklı bir model oluşturulamıyor.

2012 - 2022 Feza BUZLUCA

Çok değişkenli regresyon modeli (*multivariate logistic regression model*) (devamı)

Model 2: Bağımlılık, uyum, kalıtım ölçülerinden oluşan bir model

Bağımlılık, uyum ve kalıtım ile ilgili tüm metrikleri ileri doğru seçim yöntemiyle (*forward selection*) değerlendirilerek model oluşturulmuştur.

Modelde 4 adet bağımlılık metriği, 3 adet kalıtım metriği yer almıştır.

Uyumla ilgili metriklerin hiçbiri modele girememiştir.

Zaten bu metriklerin hatalar ile çok ilişkili olmadığı tek değişkenli analizlerde görülmüştü.

Veri kümesine model uygulandığında elde edilen sonuçlar:

Correctness: %81. 53 sınıfı hatalı buluyor, aslında 43 tanesi hatalı

Completeness: %94. Sistemdeki 234 hatanın 219 tanesini buluyor. (Bir sınıfta birden fazla hata olabiliyor)

Kappa (tahmin edilen ile gerçek hatalar arasında) = 0.64

Sonuç olarak hata öngörüsü yapabilmek için bağımlılık ve kalıtım ölçüleri gereklidir.

2012 - 2022 Feza BUZLUCA

Çok değişkenli regresyon modeli (*multivariate logistic regression model*) (devamı)

Model 3: Tüm ölçülerin yer alabildiği bir model

Bağımlılık, uyum, kalıtım ve boyut ile ilgili tüm metrikleri ileri doğru seçim yöntemiyle (*forward selection*) değerlendirilerek model oluşturulmuştur.

Modelde 4 adet bağımlılık metriği, 3 adet kalıtım metriği, 2 adet de boyut metriği yer almıştır.

Veri kümesine model uygulandığında elde edilen sonuçlar:

Sonuçlar Model 2'ye yakın çıkmıştır.

Correctness: %79. 53 sınıfı hatalı buluyor, aslında 42 tanesi hatalı

Completeness: %90. Sistemdeki 234 hatanın 210 tanesini buluyor. (Bir sınıfta birden fazla hata olabiliyor)

Kappa (tahmin edilen ile gerçek hatalar arasında) = 0.60

2012 - 2022 Feza BUZLUCA

Modelin Sınanması

Yapılan değerlendirmelere göre Model 2 diğerlerine göre daha iyi sonuçlar üretmiştir.

Ancak önceki denemelerde hem modeli oluştururken hem de modeli denerken aynı veri kümesi kullanılmıştır.

Modeli daha gerçekçi bir şekilde sınamak için modeli oluştururken ve denerken farklı veri kümeleri kullanılacaktır.

Bu sınama için 10 katlı çapraz sağlama (*10-fold cross validation*) yöntemi kullanılmıştır.

Veri kümesi (sınıflar) 10 eşit gruba ayrılmıştır.

Her deneyde model 9 gruptaki sınıflara göre oluşturulur; deneme diğer grup üzerinde yapılır.

Bu denemeler her grup için tekrar edilip ortalama alınır.

Sonuçlar:

Correctness: %78

Completeness: %92.

Kappa (tahmin edilen ile gerçek hatalar arasında) = 0.59

Değerler düşmüştür ama hala kullanılabilir düzeydedir.

2012 - 2022 Feza BUZLUCA

Çalışmanın Sonuçları:

- Literatürdeki ölçülerin (metriklerin) çoğu aynı bilgileri vermektedir (Temel bileşen analizi (PCA) sonucu).
- Birçok bağımlılık ve kalıtım ölçüsü sınıfta hata çıkma olasılığı ile ilişkilidir.
- Bağımlık ölçülerinden metot çağırma bağımlılığının sınıfta hata çıkma olasılığı ile ilişkisi kuvvetlidir.
- Kalıtım ölçülerinden, sınıfın türetim ağacındaki derinliği ve sınıfa eklenen metot sayısı sınıfta hata çıkma olasılığı ile sıkı bağlıdır.
- Uyumluluk ölçülerinin sınıfta hata çıkma olasılığı ile ilişkisinin zayıftır.
- Bunun iki nedeni olabilir: a) Uyumun anlamı henüz tam olarak anlaşılamamıştır, b) Bu kavramı ölçmek (sayısallaştırmak) kolay değildir.
- Bağımlılık ve kalıtım ölçüleri birlikte kullanılarak hataların öngörüsü için çok değişkenli bir model oluşturulabilir.
- Yapılan bu çalışmalar veri kümesine bağımlıdır. Bu iddiaların (sonuçların) doğrulanması için benzer çalışmaların başka projeler üzerinde de yapılması gereklidir.

BY NC ND 2012 - 2022 Feza BUZLUCA