



Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi

<https://dergipark.org.tr/tr/pub/yyufbed>



Araştırma Makalesi

Sosyal Ağlarda Merkezilik Ölçütleri Kullanılarak Makine Öğrenmesi ile Etkili Bireylerin Tespiti

Aybike ŞİMŞEK

Milli Savunma Üniversitesi, Kara Harp Okulu, Bilgisayar Mühendisliği Bölümü, 06420, Ankara, Türkiye

Aybike ŞİMŞEK, ORCID No: 0000-0002-1033-1597

Sorumlu yazar eposta: aysimsek@kho.msu.edu.tr

Makale Bilgileri

Geliş: 23.08.2023
Kabul: 30.11.2023
Online Nisan 2024

DOI: [10.53433/yyufbed.1348472](https://doi.org/10.53433/yyufbed.1348472)

Anahtar Kelimeler

Etkili birey,
Makine öğrenmesi,
Merkezilik ölçütleri,
Sosyal ağlar

Öz: Sosyal ağlardaki etkili bireylerin tespiti, kamuoyu şekillendirme, viral pazarlama, dedikodu yayılımını önleme gibi pratikte birçok alan için önemli bir problemdir. Bunun için her bir bireyin ne kadar etkiye sahip olduğunun, bireyin ağ üzerindeki konumuna göre tahmin edilmesi gerekmektedir. Bu amaçla, bireylerin ağ üzerindeki konumları ile ilgili bilgi veren ağ merkezilik ölçütleri literatürde sıklıkla kullanılmaktadır. Mevcut çalışmaların büyük bir kısmı, sosyal ağlardaki bireyleri etkilerine göre sıralamaya ve bu şekilde en etkili kişileri tespit etmeye çalışırlar. Öte yandan, bir sosyal ağ üzerindeki bireylerin çok küçük bir kısmı gerçekten etkili bireydir. Bu bakımdan, bütün bireyleri etkilerine göre bir sraya koymaya çalışmak yerine, etkili olabilecek bireyleri ve diğer bireyleri iki sraya ayırmak birçok uygulama için yeterlidir. Biz bu çalışmada, etkili birey tespiti problemini ikili sınıflandırma problemi olarak ele aldık. Bireylerin merkezilik ölçütlerini birer öznelik olarak belirleyip, Karar Ağacı sınıflandırıcı ile bireyleri etkili ve değil şeklinde sınıflandırdık. Deneysel çalışmalar; Karar Ağacı sınıflandırıcının, temel merkezilik ölçütlerine göre daha başarılı sonuçlar verdiğini göstermiştir.

Identification of Influencers with Machine Learning Using Centrality Measures in Social Networks

Article Info

Received: 23.08.2023
Accepted: 30.11.2023
Online April 2024

DOI: [10.53433/yyufbed.1348472](https://doi.org/10.53433/yyufbed.1348472)

Keywords

Centrality measures,
Influencer,
Machine learning,
Social networks

Abstract: Detection of influencers in social networks is an essential problem for many areas in practice, such as public opinion shaping, viral marketing, and preventing the spread of rumors. For this, it is necessary to estimate how much influence each individual has according to their position on the network. For this purpose, network centrality measures, which provide information about the position of individuals on the network, are frequently used in the literature. Most existing studies try to rank individuals on social networks according to their influence and thus identify the most influential people. On the other hand, a tiny percentage of individuals on a social network are influencers. In this regard, instead of trying to rank all individuals according to their influence, it is sufficient for many applications to divide potential influencers and other individuals into two classes. In this study, we considered the detection problem of influencers as a binary classification problem. We determined the centrality measures of individuals as features and classified the individuals as the influencers and the others with the Decision Tree classifier. Experimental studies have shown that the Decision Tree classifier gives more successful results than the basic centrality measures.

1. Giriş

Sosyal ağlardaki etkili bireylerin tespiti birçok uygulama için önemlidir. Bu uygulamalardan biri olan Etki Maksimizasyonu (EM) problemi, bir bilginin ya da etkinin bir ağ üzerindeki en fazla sayıdaki bireye ulaşabilmesi için yayılımın hangi bireylerden başlaması gerektiğinin tespiti olarak tanımlanabilir. Ağ üzerinde maksimum sayıda kullanıcıya bir bilgiyi ulaştırmak ya da bu kullanıcıları etkilemek, viral pazarlama, reklam, kamuoyu oluşturulması gibi uygulamalar için önemlidir. EM için temel 2 yaklaşım vardır. İlk yaklaşım bireylerin tek tek seçilmesidir (Kempe ve ark., 2003). Bu yaklaşım açgözlü bir yaklaşımdır. Ne var ki, düğümlerin etkilerini hesaplama işi ağır Monte-Carlo simülasyonları gerektirir (Zhang ve ark., 2021). Bu nedenle düğümlerin etkilerini işaret eden (tahmin etmemizi sağlayan) merkezilik ölçütleri kullanılır. Bu ölçütlerden bazıları Derece, Yakınlık, Aradalık ve Özvektör ölçütleridir. Bu ölçütleri kullanan yöntemlerin genel yaklaşımı şudur: düğümleri bir ölçüte göre sırala, ilk k adet düğümü etkili birey olarak al. Ancak problem bu şekilde basit olarak çözülemez. Bu amaçla, literatürde IM probleminin Kombinatoriyal Optimizasyon problemi olarak ele alınması ve k adet düğümün eş zamanlı olarak seçilmesi önerilmiştir (Borgatti, 2006). Gerçekten de IM problemi submodular olduğu için, rastgele 2 düğümün tek başlarına oluşturdukları etkiyle, beraber alındıklarında oluşturdukları etki aynı olmayabilir (Kempe ve ark., 2003). Bundan dolayı k adet düğümü eş zamanlı olarak seçmek ve etkisini değerlendirmek gerekir. Buradaki zorluk ise, problemin büyüklüğüdür. n adet düğümün bulunduğu bir ağdan k-adet bireyi seçmek için $\binom{n}{k}$ kadar farklı olasılık vardır. Problem kaba kuvvet yaklaşım ile makul sürelerde çözülemeyeceği için bazı optimizasyon teknikleri bu probleme uygulanmıştır (Wang ve ark., 2019; Azaouzi ve ark., 2021). Optimizasyon teknikleri genel olarak daha iyi sonuç verirler ancak açgözlü yaklaşımlara göre daha yavaş çalışırlar (Li ve ark., 2017). Optimizasyon algoritmalarının yakınsama hızını artırmak için problemin boyutu küçültülmelidir. Bununla birlikte, bir sosyal ağdaki gerçek etkili bireylerin sayısı toplam düğüm sayısına göre çok azdır (Guo & Wu, 2021).

Bu çalışmada, bu noktadan yola çıkarak bir sosyal ağdaki bireyleri etkili olanlar ve etkili olmayanlar şeklinde sınıflandırarak problem büyüklüğünün azaltılması amaçlanmıştır. İlk olarak, bireylerin merkezilik ölçütleri ve IC (Independent Cascade) (Kempe ve ark., 2003) modeline göre gerçek etkinlikleri hesaplanmıştır. Daha sonra, en yüksek etkiye sahip %1 ve %5 bireyler Etkili Birey olarak etiketlenip iki farklı veri seti elde edilmiştir. Böylece, ilk veri setinde %1 Etkili Birey ve %99 diğer bireyler; ikinci veri setinde ise %5 Etkili Birey ve %95 diğer bireyler olmak üzere iki sınıf elde edilmiştir. Böylece bireylerin merkezilik ölçütleri birer öznelik; Etkili Birey olup olmama durumu ise sınıf olmuştur. Son olarak, Karar Ağacı yöntemi ile bireyler sınıflandırılmıştır. Elde edilen sonuçlar, literatürde sıklıkla kullanılan merkezilik ölçütleri ile kıyaslanmıştır. Sınıflandırma performansı açısından her iki veri setinde de Karar Ağacının daha iyi sonuçlar verdiği görülmüştür.

Literatürde Etkili Bireylerin tespiti ile ilgili merkezilik ölçütlerinin kullanıldığı çalışmalara bakıldığında, iki ana yaklaşım olduğu görülmektedir. Bunlardan birincisi, var olan merkezilik ölçütlerini bir algoritma içinde kullanmak veya yeni merkezilik ölçütleri geliştirmek; ikincisi ise var olan merkezilik ölçütlerinden bileşik merkezilik ölçütleri elde etmektir.

İlk gruba giren çalışmalara (Salavati ve ark., 2019; Sheng ve ark., 2020; Wen ve ark., 2020; Yang ve ark., 2020; Zhao ve ark., 2020a; Zhuang ve ark., 2021) örnek olarak verilebilir. Bu çalışmaların amacı özetle, her bir bireyi diğerinden ayırt edebilecek bir algoritma ya da yeni merkezilik ölçütleri geliştirmektir. Böylece bireyler, kendileri için hesaplanan merkezilik ölçütlerine göre büyükten küçüğe doğru sıralanabilir ve en yüksek değere sahip olan bazı bireyler Etkin Birey olarak kabul edilebilir. İkinci gruba giren çalışmalara (Keng ve ark., 2020; Şimşek, 2021; Şimşek & Meyerhenke, 2020; Sabah & Şimşek, 2023) örnek olarak verilebilir. Bu çalışmaların amacı da ilk gruptaki çalışmalarla aynıdır. Farklı olarak, birden çok merkezilik ölçütünü birleştirmeyi (birlikte kullanmayı) önermektedirler.

Yukarıda bahsedilen çalışmalar Etkin Birey tespiti problemini, her bir bireyi diğerinden ayırt etme, yani bireyleri etkilerine göre sıralama problemi olarak ele almış; probleme, sınıflandırma problemi olarak yaklaşmamışlardır. Bu çalışmalardan farklı olarak makine öğrenmesi tekniklerinin kullanıldığı çalışmalar da yeni yeni yapılmaya başlanmıştır. Rezaei ve ark., Etkin Birey tespiti problemini bir regresyon problemi olarak ele almış; bireylerin derece ve bağlılık gibi merkezilik ölçütleri ile gerçek etkinlik düzeyleri arasında bir ilişki aramışlardır (Rezaei ve ark., 2023).

Çizelge 1. Literatür taramasının özeti

Yazar, Yıl	Yaklaşım	Amaç
Salavati ve ark., 2019	Var olan merkezilik ölçütlerini bir algoritma içinde kullanmak	Bireylerin ayırt edilebilirliğini arttırmak için düğümleri sıralamak
Sheng ve ark., 2020	Var olan merkezilik ölçütlerini bir algoritma içinde kullanmak	Etkili birey tespiti
Wen ve ark., 2020	Yeni merkezilik ölçütü geliştirmek	Etkili birey tespiti
Yang ve ark., 2020	Yeni merkezilik ölçütü geliştirmek	Etkili birey tespiti
Zhao ve ark., 2020a	Var olan merkezilik ölçütlerini bir algoritma içinde kullanmak	Etkili birey tespiti
Zhuang ve ark., 2021	Yeni merkezilik ölçütü geliştirmek	Etkili birey tespiti
Keng ve ark., 2020	Var olan merkezilik ölçütlerinden bileşik merkezilik ölçütleri elde etmek	Bireylerin ayırt edilebilirliğini arttırmak
Şimşek, 2021	Var olan merkezilik ölçütlerinden bileşik merkezilik ölçütleri elde etmek	Bireylerin ayırt edilebilirliğini arttırmak için düğümleri sıralamak
Şimşek ve Meyerhenke, 2020	Var olan merkezilik ölçütlerinden bileşik merkezilik ölçütleri elde etmek	Bireylerin ayırt edilebilirliğini arttırmak için düğümleri sıralamak
Sabah & Şimşek, 2023	Var olan merkezilik ölçütlerinden bileşik merkezilik ölçütleri elde etmek	Bireylerin ayırt edilebilirliğini arttırmak için düğümleri sıralamak
Rezaei ve ark., 2023	Makine Öğrenmesi, regresyon	Bireylerin ayırt edilebilirliğini arttırmak için düğümleri sıralamak
Zhao ve ark., 2020b	Makine Öğrenmesi, Naive Bayes, Karar Ağacı, Rasgele Orman, Destek Vektör Makinesi sınıflandırıcıları	Bireylerin etkinliklerine göre birçok sınıfa ayrılması

Bireylerin gerçek etkinliklerini belirlemek için bir salgın modeli olan SIR (Susceptible-Infected-Recovered) kullanmışlardır. SIR modeli, virüs gibi ajanların bir topluluk üzerindeki yayılımını modellemede sıklıkla kullanılan bir modeldir. Rezaei ve ark.'nın çalışması da diğer çalışmalar gibi Etkin Birey tespiti problemini, bireyleri etkilerine göre sıralama problemi olarak ele almıştır (Rezaei ve ark., 2023). Diğer bir çalışma Zhao ve ark. tarafından yapılmış olan ve Etkin Birey tespiti problemini bir sınıflandırma problemi olarak ele alan çalışmadır. Bu çalışmada bireylerin gerçek etkinliklerini belirlemek için SIR modelini kullanmışlar; düğümleri gerçek etkinliklerine göre birçok sınıfa ayırmışlar (etiketlemişler) ve Naive Bayes, Karar Ağacı, Rasgele Orman, Destek Vektör Makinesi gibi sınıflandırıcılarla bireyleri sınıflandırmışlardır (Zhao ve ark., 2020b). Literatür özeti Çizelge 1'de verilmiştir.

Literatür taraması sonucunda şunlar söylenebilir:

- Etkili Birey tespiti problemini sınıflandırma problemi olarak ele alan çalışma sayısı çok azdır. Sınıflandırma problemi olarak ele alan çalışmalar da ikili sınıflandırma değil çoklu sınıflandırma problemi olarak ele almışlardır.
- Çalışmaların neredeyse tamamı yayılım modeli olarak SIR yayılım modelini kullanmışlardır. SIR yayılım modeli, virüs gibi bulaşarak yayılan ajanların bir ağ üzerindeki yayılımını modellemek için sıklıkla kullanılır. Ne var ki; viral pazarlama, kamuoyu şekillendirme gibi uygulamalarda reklam/bilgi, bir bireyden diğerine bulaşmamakta, bireyler tarafından benimsenmektedir. Dolayısı ile SIR modeli bu tip uygulamalar için uygun değildir.

Bu çalışmanın literatüre katkılarını şu şekilde sıralayabiliriz:

- Etkili Birey tespiti problemi ikili sınıflandırma problemi şeklinde ele alındı. Giriş kısmında da bahsedildiği gibi, ağdaki kişileri etkili ve etkili olmayan şekilde iki sınıfa ayırmak, etki maksimizasyonu amacıyla kişiler arasından seçim yapan optimizasyon algoritmaları için problem kümesini küçültmek demektir.
- Etkili Birey tespiti problemini sınıflandırma problemi olarak ele alan ve etki modelleme için SIR modelini kullanan mevcut çalışmaların aksine etki modelleme için IC (Independent Cascade) modeli kullanıldı. IC modeli, bir etkinin (bilgi, reklam vb.) ağ üzerinde bir kişiden diğer kişiye geçişini bulaşma değil benimseme şeklinde modellemektedir.

2. Materyal ve Yöntem

2.1. IC modeli

IC (Independent Cascade) modeli, bir bilginin/etkinin bir ağ üzerinde kişilerden kişilere benimseme yoluyla geçişini ve bu şekilde ağ üzerinde yayılımını modellemek amacı ile üretilmiş bir modeldir. Her bir kişinin komşularından (Twitter veri seti için, takip ettiği kişilerden) etkilenmesi için belirli bir olasılık vardır. Bununla birlikte her bir kişinin, etkileyebilme potansiyeline sahip olduğu kişileri (Twitter veri seti için, takipçilerini) etkilemesi için birer şansı vardır. Etki yayılımı bir kişiden başladığında, olasılığa bağlı olarak, komşulardan bir kısmı etkilenir. Bir sonraki zaman diliminde etki, yeni etkilenen kişilerden yayılmaya devam eder. Böylece, yayılımın herhangi bir aşamasında etkilenen yeni bir kişi olmayana kadar yayılım devam eder. Yayılım bittiğinde, etkilenen kişilerin sayısının, ağdaki toplam kişi sayısına oranı etkinin ne kadar yayıldığını gösterir. IC ile ilgili detaylı bilgi için (Kempe ve ark., 2003) incelenebilir.

2.2. Veri seti

Bu çalışmada SNAP (Stanford Network Analysis Platform) veri setlerinden ego-Twitter veri seti kullanılmıştır. ego-Twitter veri seti 81306 düğüm (kişi) ve 1768149 kenara (takip) sahiptir (McAuley & Leskovec, 2012). Yönlü ve ağırlıksız bir graf olan ego-Twitter veri setinde kenar yönleri takip ilişkisini gösterir. Yani kenarlar takip edenden takip edilene doğrudur. Ancak, etki yayılımını modelleyebilmek için takip ilişkisini etkileme ilişkisine dönüştürmek gereklidir. Bunun için kenarların yönlerini ters çevirmek yeterlidir. Bu çalışmada da orijinal veri setindeki kenarların yönleri ters çevrilmiştir.

2.3. Merkezilik ölçütleri

Bu çalışmada kullanılan merkezilik ölçütleri aşağıda anlatılmaktadır. Bu ölçütler temel merkezilik ölçütleridir ve sıklıkla kullanılmaktadır. Bu ölçütler iki amaçla kullanılmıştır. Birincisi, her bir düğüm için hesaplanarak düğümlerin öznitelikleri olarak ele alınmışlardır. İkincisi, sınıflandırıcının rakipleri olarak kullanılmıştır.

Derece (çıkış): Yönlü bir ağ üzerindeki bir düğümden çıkan yönlü kenarların sayısı olarak hesaplanır. Kullanılan, kenar yönleri ters çevrilmiş ego-Twitter gibi bir ağ için bir kişinin doğrudan etkileyebildiği takipçi sayısı olarak düşünülebilir.

Aradalık: Ağ üzerindeki bütün düğüm çiftleri arasındaki en kısa yolların kaçta kaçının belirli bir düğüm üzerinden geçtiğidir. Bir etki ağ üzerinde yayılırken kısa yollar üzerinden yayılma olasılığı daha yüksek olduğundan, aradalık değeri yüksek olan düğümlerin etki yayılımında daha aktif oldukları söylenebilir.

Yakınlık: Bir düğümden ağ üzerindeki diğer düğümlere olan en kısa yollarının bir ölçütü olarak tanımlanabilir. Yakınlık değeri yüksek olan bir düğüm, diğer düğümlere daha kısa yollar üzerinden ulaşabilir. Bu da, bir etkiyi daha hızlı yayabilme durumu ile ilişkilendirilmektedir.

Özvektör: Derece merkezilik ölçütünün genişletilmiş hali olarak düşünülebilir. Örnek vermek gerekirse; bir kişinin takipçi sayısı onun derece merkezilik ölçütü ile ifade edilirken, takipçilerini takip edenlerin ve onları da takip edenlerin ve böylece bütün çevresinin sayılarının toplamı üzerinden hesaplanan ölçüt Özvektör merkeziliği ile ifade edilir. Kısaca, bir kişi ne kadar etkili takipçilere sahipse o kadar etkilidir.

Katz: Özvektör merkezilik ölçütü ile benzerdir. Farklı olarak, düğümden uzaklaştıkça düğümün etkisinin de azalacağını ifade eder. Örnek vermek gerekirse; bir kişi takipçilerini büyük oranda etkileyebilir. Ancak, takipçilerinin takipçilerini aynı oranda etkileyemeyecektir. Ve böylece kendi takipçilerinden uzaklaştıkça etkisi de azalacaktır. Özvektör merkeziliği bu durumu hesaba katmazken, Katz merkeziliği katar.

Merkezilik ölçütleri ile ilgili detaylı bilgi için (Saxena & Iyengar, 2020) incelenebilir.

2.4. Problemin sınıflandırma problemine dönüştürülmesi, karar ağacı ve merkezilik ölçütlerinin sınıflandırıcı olarak kullanılması

Problemi sınıflandırma problemine dönüştürmek için ilk olarak her bir düğüm (birey) için yukarıda anlatılan merkezilik ölçütleri hesaplandı. Bu ölçütler her bir düğümün özneliği olarak belirlendi. Daha sonra, her bir düğümün gerçek etkisini hesaplamak için düğümler tek tek alınarak IC modelinde etkinin başlatıcısı olan düğüm olarak kullanıldı. Böylece, bir düğüm tek başına etki yayıcısı olarak belirlendiğinde, etkinin ağ üzerinde ne oranda yayıldığı (ağdaki toplam düğümlerin kaçta kaçını etkilediği) hesaplandı. Son olarak, sırası ile en yüksek etkiye sahip %1'lik ve %5'lik düğümler Etkili Düğüm; diğer düğümler ise Diğer olarak etiketlendi ve 2 adet veri seti elde edildi. %1 ve %5 değerleri bu amaçla literatürde sıklıkla kullanılan değerlerdir (Zengin Alp & Gündüz Öğüdücü, 2018). Bu veri setlerine Birinci ve İkinci veri setleri diyeceğiz. Birinci veri setinde 814 Etkili Düğüm, 80492 Diğer Düğüm; ikinci veri setinde ise 4066 Etkili Düğüm, 77240 Diğer Düğüm bulunmaktadır.

Sınıflandırıcı olarak Kullanılan Karar Ağacı, Python ve Scikit-learn (Pedregosa ve ark., 2011) modülü kullanılarak kodlanmıştır. Veri setleri dengesiz oldukları için 10-fold Stratified çapraz doğrulama uygulanmıştır.

Derece, Aradalık, Özvektör ve Katz merkezilik ölçütleri de ayrı ayrı sınıflandırıcı olarak kullanılmış ve sonuçlar elde edilmiştir. Merkezilik ölçütlerinin sınıflandırıcı olarak kullanılması şu şekildedir: bir düğümün merkezilik ölçütü ne kadar büyükse etkinliğin de o kadar büyük olması beklenir. İlk olarak düğümler gerçek etki değerlerine göre sıralanır. Elde edilen sıralamaya karşılık gelen sınıf listesi gerçek (true) liste olarak alınır. Daha sonra düğümler bir merkezilik ölçütüne göre (örneğin Derece) büyükten küçüğe doğru sıralanır ve elde edilen sıralamaya karşılık gelen sınıf listesi tahmin (predicted) liste olarak alınır. Bu iki liste kullanılarak oluşturulan karışma matrisi (confusion matrix) ile performans ölçütleri hesaplanır.

3. Bulgular

Kullanılan veri setleri dengesiz oldukları için performans karşılaştırması adına Precision, Recall ve F1 Score değerleri her bir sınıf için (Diğer ve Etkili Birey sınıfları için) ayrı ayrı verilmiştir. Ayrıca, Balanced Accuracy değeri hesaplanmıştır.

Birinci veri seti için elde edilen sonuçlar Çizelge 2'de, İkinci veri seti için elde edilen sonuçlar Çizelge 3'de verilmiştir. Çizelge 2 ve Çizelge 3'de, Etkili Birey sınıfının tespitindeki en başarılı sonuçlar koyu yazılarak belirtilmiştir. Çalışmanın amacı, Etkili Bireyleri diğerlerinden ayırt etmek olduğu için (diğer bir deyişle, hedef sınıf Etkili Birey sınıfı olduğu için) sınıflandırıcıların özellikle Etkili Birey sınıfı üzerindeki başarılarına odaklanılmıştır.

Çizelge 2. Birinci veri seti için elde edilen sınıflandırma sonuçları

	Precision		Recall		F1 Score		Balanced Accuracy
	Diğer	Etkili Birey	Diğer	Etkili Birey	Diğer	Etkili Birey	
Karar Ağacı	0.996	0.713	0.997	0.663	0.996	0.686	0.830
Derece (çıkış)	0.995	0.598	0.995	0.598	0.995	0.598	0.797
Aradalık	0.992	0.292	0.992	0.292	0.992	0.292	0.642
Yakınlık	0.989	0	0.989	0	0.989	0	0.494
Özvektör	0.991	0.143	0.991	0.143	0.991	0.143	0.567
Katz	0.992	0.251	0.992	0.251	0.992	0.251	0.622

Çizelge 3. İkinci veri seti için elde edilen sınıflandırma sonuçları

	Precision		Recall		F1 Score		Balanced Accuracy
	Diğer	Etkili Birey	Diğer	Etkili Birey	Diğer	Etkili Birey	
Karar Ağacı	0.986	0.752	0.987	0.742	0.986	0.746	0.864
Derece (çıkış)	0.982	0.663	0.982	0.663	0.982	0.663	0.823
Aradalık	0.967	0.388	0.967	0.388	0.967	0.388	0.678
Yakınlık	0.965	0.338	0.965	0.338	0.965	0.338	0.651
Özvektör	0.957	0.199	0.957	0.199	0.957	0.199	0.578
Katz	0.967	0.388	0.967	0.388	0.967	0.388	0.678

4. Tartışma ve Sonuç

DeneySEL sonuçlara göre en başarılı sınıflandırmayı Karar Ağacı vermiştir. Karar Ağacına en yakın sonuçları veren merkezilik ölçütü Derece (çıkış) merkezilik ölçütüdür. Diğer merkezilik ölçütlerinin sınıflandırma performansları ise çok daha düşüktür. Giriş bölümünde de bahsedildiği gibi, Etki Maksimizasyonu amacıyla kullanılan optimizasyon algoritmaları ağ üzerindeki bütün bireyler içerisinde toplam etkiyi maksimize edecek bir alt kümeyi seçmeyi amaçlamaktadır. Önerdiğimiz yöntem, etkili bireyleri diğerlerinden ayırarak optimizasyon algoritmalarının seçim yapması gereken kümeyi küçültmektedir. Bu da, optimizasyon algoritmalarının daha kısa sürede optimal çözüme yakınsamasını sağlayacaktır.

Çalışmanın amacı özellikle Etkili Bireyleri tespit etmek olduğu için performans ölçütlerinin ortalama değerleri yerine her iki sınıf için ayrı ayrı değerleri sunulmuştur. Ölçütlerin yalnızca ortalama değerleri verilseydi, örneğin bütün düğümleri “Diğer olarak sınıflandıran (yani hiçbir şey yapmayan) bir sınıflandırıcının sınıflandırma performansı da yüksek görülebilirdi.

Ek olarak; bu çalışmada Karar Ağacının yanı sıra Destek Vektör Makinesi, Çok Katmanlı Yapay Sinir Ağı ve Naive Bayes sınıflandırıcılar da denenmiş; ancak, sınıflandırma başarıları temel merkezilik ölçütlerinden bile kötü olduğu için sonuçlar sunulmamıştır.

Bu çalışmada, Etkili Birey tespiti problemi ikili sınıflandırma problemi olarak ele alınmış; Karar Ağacı sınıflandırıcısı probleme başarı ile uygulanmış ve literatürdeki temel merkezilik ölçütlerine göre daha başarılı sonuçlar elde edilmiştir. Halen problem üzerinde bazı açık zorluklar vardır. İlk olarak, farklı tipteki ağlar üzerinde çalışılabilir. Bu çalışmada, takip eden-takip edilen ilişkisini sunan yönlü bir veri seti olan ego-Twitter veri seti kullanılmıştır. Bundan başka, arkadaşlık ilişkisini, beraber çalışma ilişkisini sunan yönsüz ağlar üzerinde de çalışmalar yapılmalıdır.

Bu çalışmada kullanılan veri seti, bir graf veri seti olduğu için sadece merkezilik ölçütleri birer öznitelik olarak kullanılmıştır. Twitter gibi bir ortamdan toplanacak veriler ile bireylerin kişisel özellikleri de (konumları, meslekleri, yaşları vb.) birer öznitelik olarak değerlendirilebilir.

Son olarak, yüksek boyutlu veriler üzerinde başarılı sınıflandırma performansı sağlayan derin sinir ağlarının bu problem üzerinde kullanımı düşünülebilir.

Kaynakça

- Azaouzi, M., Mnasri, W., & Romdhane, B. L. (2021). New trends in influence maximization models. *Computer Science Review*, 40, 100393. doi:10.1016/j.cosrev.2021.100393
- Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational and Mathematical Organization Theory*, 12(1), 21-34. doi:10.1007/s10588-006-7084-x
- Guo, J., & Wu, W. (2021). Adaptive influence maximization. *ACM Transactions on Knowledge Discovery from Data*, 15(5), 1-23. doi:10.1145/3447396
- Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). *Maximizing the spread of influence through a social network*. International Conference on Knowledge Discovery and Data Mining, New York, USA.

- Keng, Y. Y., Kwa, K. H., & McClain, C. (2020). Convex combinations of centrality measures. *Journal of Mathematical Sociology*, 45(4), 195-222. doi:10.1080/0022250X.2020.1765776
- Li, D., Wang, C., Zhang, S., Zhou, G., Chu, D., & Wu, C. (2017). Positive influence maximization in signed social networks based on simulated annealing. *Neurocomputing*, 260, 69-78. doi:10.1016/j.neucom.2017.03.003
- McAuley, J., & Leskovec, J. (2012). *Learning to discover social circles in ego networks*. NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ..., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rezaei, A. A., Munoz, J., Jalili, M., & Khayyam, H. (2023). A machine learning-based approach for vital node identification in complex networks. *Expert Systems with Applications*, 214, 119086. doi:10.1016/j.eswa.2022.119086
- Sabah, L., & Şimşek, M. (2023). A new fast entropy-based method to generate composite centrality measures in complex networks. *Concurrency and Computation: Practice and Experience*, 35(10). doi:10.1002/cpe.7657
- Salavati, C., Abdollahpouri, A., & Manbari, Z. (2019). Ranking nodes in complex networks based on local structure and improving closeness centrality. *Neurocomputing*, 336, 36-45. doi:10.1016/j.neucom.2018.04.086
- Saxena, A., & Iyengar, S. (2020). Centrality measures in complex networks: A survey. doi:10.48550/arXiv.2011.07190
- Sheng, J., Dai, J., Wang, B., Duan, G., Long, J., Zhang, J., ..., & Guan, W. (2020). Identifying influential nodes in complex networks based on global and local structure. *Physica A: Statistical Mechanics and its Applications*, 541, 123262. doi:10.1016/j.physa.2019.123262
- Şimşek, M., & Meyerhenke, H. (2020). Combined centrality measures for an improved characterization of influence spread in social networks. *Journal of Complex Networks*, 8(1), cnz048. doi:10.1093/comnet/cnz048
- Şimşek, A. (2021). Lexical sorting centrality to distinguish spreading abilities of nodes in complex networks under the Susceptible-Infectious-Recovered (SIR) model. *Journal of King Saud University - Computer and Information Sciences*, 34(8), 4810-4820. doi:10.1016/j.jksuci.2021.06.010
- Wang, S., Liu, J., & Jin, Y. (2019). Finding influential nodes in multiplex networks using a memetic algorithm. *IEEE Transactions on Cybernetics*, 51(2), 1-13. doi:10.1109/TCYB.2019.2917059
- Wen, T., Pelusi, D., & Deng, Y. (2020). Vital spreaders identification in complex networks with multi-local dimension. *Knowledge-Based Systems*, 195, 105717. doi:10.1016/j.knosys.2020.105717
- Yang, Y., Wang, X., Chen, Y., Hu, M., & Ruan, C. (2020). A novel centrality of influential nodes identification in complex networks. *IEEE Access*, 8, 58742-58751. doi:10.1109/ACCESS.2020.2983053
- Zengin Alp, Z., & Gündüz Ögüdücü, Ş. (2018). Identifying topical influencers on twitter based on user behavior and network topology. *Knowledge-Based Systems*, 141, 211–221. doi:10.1016/j.knosys.2017.11.021
- Zhang, Z., Li, X., & Gan, C. (2021). Identifying influential nodes in social networks via community structure and influence distribution difference. *Digital Communications and Networks*, 7(1), 131-139. doi:10.1016/j.dcan.2020.04.011
- Zhao, J., Wang, Y., & Deng, Y. (2020a). Identifying influential nodes in complex networks from global perspective. *Chaos, Solitons and Fractals*, 133, 109637. doi:10.1016/j.chaos.2020.109637
- Zhao, G., Jia, P., Huang, C., Zhou, A., & Fang, Y. (2020b). A machine learning based framework for identifying influential nodes in complex networks. *IEEE Access*, 8, 65462-65471. doi:10.1109/ACCESS.2020.2984286
- Zhuang, Y.-B., Li, Z.-H., & Zhuang, Y.-J. (2021). Identification of influencers in online social networks: measuring influence considering multidimensional factors exploration. *Heliyon*, 7(4), e06472. doi:10.1016/j.heliyon.2021.e06472