Rapor Taslak

Rapor: Makine Öğrenimi ve Derin Öğrenme Yöntemleri ile Duygu Analizi

1. Makine Öğrenmesi Algoritmaları Kullanılarak Twitter Metinlerinin Duygu Analizi (Belge 1)

Bu çalışma, Twitter platformundaki metinleri analiz ederek kullanıcıların duygusal durumlarını tespit etmeyi amaçlamaktadır. Araştırmada makine öğrenimi algoritmalarının bu analizlerde nasıl kullanıldığı detaylı şekilde açıklanmıştır.

- Veri Toplama ve Ön İşleme: Twitter API kullanılarak belirli anahtar kelimeler üzerinden tweetler toplanmış ve bu veriler siyasi olaylar, ürün lansmanları gibi farklı konulara göre kategorize edilmiştir. Veri ön işleme aşamasında tweetlerdeki özel karakterler ve gereksiz kelimeler temizlenmiş, kelimeler köklerine indirgenmiştir.

- Algoritmalar:

- Destek Vektör Makineleri (SVM), yüksek doğruluk oranı ile en başarılı sonuçları vermiştir.

- Naive Bayes, hızlı ve basit bir model olmasına rağmen belirli bir başarı elde etmiştir.

- Karar Ağaçları, anlaşılır bir yapı sunsa da aşırı uyum riski taşımaktadır.

- Sonuçlar: SVM algoritması %90’ın üzerinde bir doğruluk oranı elde ederek en başarılı model olmuştur. Duygu sınıflandırmasında anahtar kelimelerin belirleyici olduğu ve bu tür analizlerin sosyal medya etkileşimlerini artırmak için kullanılabileceği vurgulanmıştır.

2. DLSTA: Derin Öğrenme Destekli Anlamsal Metin Analizi (Belge 2)

Bu makale, duygu analizi için Derin Öğrenme Destekli Anlamsal Metin Analizi (DLSTA) yöntemini önermektedir.

- Yöntem: DLSTA, doğal dil işleme (NLP) ve kelime gömme teknikleri kullanarak metinlerin anlamsal ve sözdizimsel özelliklerini analiz eder. Bu yöntem, özellikle metin kök duygu analizi (textual root emotion analysis) üzerine yoğunlaşmaktadır.

- Algoritma Performansı:

- DLSTA, %97,22 duygu tanıma oranı ve %98,02 sınıflandırma doğruluğu ile diğer yöntemlere kıyasla üstün performans sergilemiştir.

- Özellikle SVM tabanlı bir platform üzerinde çalışan DLSTA, yedi farklı duygu sınıfını (üzüntü, sürpriz, sevinç, öfke, korku, tiksinti, nötr) başarıyla tespit etmektedir.

- Sonuçlar: DLSTA, mevcut yöntemler arasında sınıflandırma doğruluğu ve performans açısından en yüksek sonuçları elde etmiştir. Yöntem, duygusal analizde yüksek performansı ve tutarlılığıyla öne çıkmaktadır.

3. TTL: Transformer Transfer Learning Duygu Algılama Modeli (Belge 3)

Bu çalışma, büyük veri setleri üzerinde eğitilmiş Transformer Transfer Learning (TTL) modelini önermektedir.

- Yöntem: TTL modeli, önce büyük bir kendiliğinden rapor edilen duygu veri seti üzerinde, ardından anotatör tarafından derecelendirilen daha küçük veri setlerinde eğitilmiştir. Bu iki aşamalı süreç, duygu tespiti doğruluğunu artırmayı hedeflemektedir.

- Veri Setleri:

- TTL modeli, 3,654,544 veri noktasından oluşan büyük bir kendiliğinden rapor edilen veri seti ve 10 farklı duygu veri seti kullanılarak eğitilmiştir. Veri setleri, Ekman’ın yaygın altı temel duygusunu (öfke, tiksinti, korku, sevinç, üzüntü ve sürpriz) içerir.

- Performans:

- TTL modeli, RoBERTa-large modelini kullanarak duygu tespiti alanında en yüksek F1 puanını (%0.84) elde etmiştir. Anotatör değerlendirmeli veri setlerinde ortalama %0.87, kendi bildirimli veri setlerinde ise %0.79 başarı göstermiştir.

- Sonuçlar: TTL modeli, hem anotatör hem de kendiliğinden rapor edilen veri setlerinde güçlü bir performans göstermiştir. Özellikle büyük veri setlerinin eğitilmesi sayesinde yüksek tahmin doğruluğuna ulaşmıştır.

Genel Değerlendirme ve Karşılaştırma:

- Veri Seti ve Yöntemler: Her üç çalışma da farklı veri setleri ve yöntemler kullanarak duygu analizi gerçekleştirmiştir. İlk çalışma makine öğrenimi algoritmaları ile odaklanırken, ikinci ve üçüncü çalışmalar derin öğrenme ve transfer öğrenme yöntemlerine dayanmaktadır.

- Algoritma Performansı:

- SVM, tüm çalışmalarda etkili bir sınıflandırma aracı olarak öne çıkmıştır. Ancak DLSTA ve TTL gibi derin öğrenme modelleri, makine öğrenimi algoritmalarına kıyasla daha yüksek doğruluk oranları sunmuştur. Özellikle DLSTA %98,02 gibi çok yüksek bir doğruluk oranı elde etmiştir.

- Kullanım Alanları: Duygu analizi, pazarlama, sosyal araştırmalar, marka yönetimi ve stratejik karar alma gibi birçok alanda kullanılabilecek güçlü bir araçtır. TTL ve DLSTA gibi modeller, büyük veri setlerinde ve daha karmaşık duygu analizlerinde kullanılmaya uygundur.

Sonuç

Bu raporda, üç farklı çalışmada kullanılan duygu analizi yöntemleri karşılaştırılmıştır. DLSTA ve TTL gibi derin öğrenme tabanlı modeller, makine öğrenimi algoritmalarına göre daha yüksek doğruluk ve performans sunmaktadır. Bu tür yöntemler, sosyal medya ve büyük veri analitiği gibi alanlarda etkili sonuçlar vermektedir.

Makale Başlığı:  
The Social Media Sentiment Analysis Framework: Deep Learning for Sentiment Analysis on Social Media  
  
Özet  
  
1. Konusu  
Makale sosyal medya üzerinde duygu analizini (sentiment analysis) gerçekleştiren bir çerçeve geliştirmeyi ele almaktadır. Sosyal medyadaki kullanıcıların görüş ve duygularını anlamak için kullanılan bu yöntem özellikle Twitter ve Facebook gibi platformlardan alınan veriler üzerinde çalışmaktadır. Bu çalışma duygu analizi için özelleştirilmiş bir duygu sözlüğü oluşturarak doğruluğu artırmayı hedeflemiştir. Araştırma farklı derin öğrenme algoritmalarının etkinliğini de incelemiştir.  
  
2. Ne Yapılmış?  
Araştırma sosyal medya duygu analizi (SMSA) çerçevesi geliştirmiştir. Bu çerçeve duygu analizi için bir duygu sözlüğü oluşturmayı ve bu sözlüğü derin öğrenme modelleriyle entegre etmeyi içerir. Çalışma derin öğrenme algoritmalarının çeşitli parametre ayarları altında duygu çıkarımı üzerindeki etkisini değerlendirmiştir.  
  
3. Nasıl Yapılmış?  
- Veri Toplama: İlk adımda sosyal medya paylaşımları toplanmış ve negatif/pozitif duygu ifadelerini içeren özel bir duygu sözlüğü oluşturulmuştur.  
- Ön İşleme: Doğal dil işleme (NLP) teknikleri kullanılarak veriler önişlemden geçirilmiş kelime köklerine ayrılmış gereksiz kelimeler çıkarılmış ve kelimeler etiketlenmiştir.  
- Derin Öğrenme Uygulaması: İşlenen veriler derin öğrenme modellerine (örn. DNN) aktarılmış ve bu modeller kullanılarak duygular sınıflandırılmıştır.  
- Performans Değerlendirmesi: Farklı algoritmaların doğruluk verimlilik ve dayanıklılık açısından performansları karşılaştırılmıştır.  
  
4. Kullanılan Metotlar  
- Derin Öğrenme Modelleri: Konvolüsyonel Sinir Ağları (CNN) Tekrarlayan Sinir Ağları (RNN) ve Derin Sinir Ağları (DNN) kullanılmıştır.  
- Özelleştirilmiş Duygu Sözlüğü: Duygu çıkarımını güçlendirmek için özel olarak oluşturulmuş bir duygu sözlüğü.  
Verilerin analiz öncesinde işlenmesi için metin tokenizasyonu kelime kökü çıkartma ve adlandırılmış varlık tanıma (NER) gibi teknikler kullanılmıştır.  
  
5. Sonuçlar  
Çalışmanın sonuçları önerilen SMSA çerçevesinin duygu analizi doğruluğunu önemli ölçüde artırdığını göstermiştir. Diğer analiz yöntemleri ile karşılaştırıldığında (TF-IDF TSA SAS) SMSA modeli daha yüksek doğruluk ve dayanıklılık sergilemiştir. Ayrıca modelin farklı veri setleri üzerinde tutarlı bir performans sergilediği gözlemlenmiştir.  
  
6. Eksikler  
- Dil Çeşitliliği: Çalışma çoğunlukla İngilizce metinler üzerinde yoğunlaşmış ve diğer dillerdeki duygu analizine yönelik sınırlamalar ortaya koymuştur.  
- Daha Fazla Parametre İncelemesi: Derin öğrenme modellerinin daha farklı parametre kombinasyonları ile test edilmesi gerekliliği belirtilmiştir.  
- Gerçek Zamanlı Uygulama: Çalışma gerçek zamanlı veri analizini tam anlamıyla ele almamış bu da gelecek araştırmalar için bir eksiklik olarak görülmüştür.  
  
7. Biz Nasıl Yaparız?  
Bu projeyi uygulamaya geçirmek için şu adımlar izlenebilir:  
1. Veri Toplama: İlk olarak Twitter ve Facebook gibi sosyal medya platformlarından verileri toplamak için veri madenciliği teknikleri kullanılabilir.Ya da kaggledan veri seti alınabilir.  
2. Özelleştirilmiş Duygu Sözlüğü Geliştirme: Kullanıcıların en çok kullandığı ifadeleri analiz ederek pozitif ve negatif duyguları ayırt edebilecek bir sözlük oluşturmak gerekir.  
3. Ön İşleme ve NLP Teknikleri: Verileri önişlemden geçirip analiz edebilmek için kelime köklerini çıkarma stop-words temizliği ve POS etiketleme gibi NLP tekniklerini kullanmak gerekir.  
4. Derin Öğrenme Modeli Kurma ve Eğitme: CNN RNN veya DNN gibi derin öğrenme modelleri kurularak bu modeller topladığımız veriler üzerinde eğitilebilir. Daha sonra farklı parametrelerle bu modellerin performansı test edilerek en iyi sonuç veren yapı belirlenebilir.  
5. Performans Analizi: Modellerin başarımını farklı ölçütlerle (doğruluk hız esneklik) değerlendirerek en uygun yaklaşımı belirlemek.  
6. Gerçek Zamanlı Uygulama: Kullanıcı geri bildirimleri toplamak ve gerçek zamanlı analiz yapabilmek için sistemin performansını optimize etmek gerekebilir.

Makale Başlığı:  
A Survey of Sentiment Analysis in Social Media  
  
Özet  
  
1. Konusu  
Bu makale, sosyal medyada duygu analizi alanındaki gelişmeleri ele alan kapsamlı bir araştırma sunmaktadır. Sosyal medya platformlarının artışıyla birlikte insanlar duygularını, görüşlerini ve düşüncelerini çevrimiçi olarak ifade etmektedir. Bu makale duygu analizinin gelişimini, kullanılan yöntemleri, uygulama alanlarını ve karşılaşılan zorlukları ele alarak mevcut durumu özetlemektedir.  
  
2. Ne Yapılmış?  
Makale, sosyal medya duygu analizinde kullanılan çeşitli yöntemleri ve teknikleri incelemiştir. Üç temel perspektiften (görev odaklı, granülerlik odaklı ve metodoloji odaklı) duygu analizi yöntemlerini kategorize etmiş ve karşılaştırmıştır. Ayrıca kullanılan araçlar, veri setleri ve sınırlamalar hakkında bilgi sunulmuştur.  
  
3. Nasıl Yapılmış?  
- Görev Odaklı Analiz: Polarity sınıflandırma, valans/uyarılma düzeyleri, öznellik/nesnellik tanımlaması ve özellik/aspekt bazlı duygu analizi gibi görevler ele alınmıştır.  
- Granülerlik Odaklı Analiz: Belge, cümle ve kelime seviyesinde analiz yapılmıştır.  
- Metodoloji Odaklı Analiz: Gözetimli öğrenme, yarı-gözetimli öğrenme ve gözetimsiz öğrenme teknikleri açıklanmıştır.  
  
4. Kullanılan Metotlar  
- Gözetimli Öğrenme: SVM, karar ağaçları ve sinir ağları gibi tekniklerle duygu sınıflandırma yapılmıştır.  
- Derin Öğrenme: CNN, RNN ve Transformer modelleri gibi derin öğrenme teknikleri duygu ifadelerini daha iyi anlamak için kullanılmıştır. Özellikle BERT gibi modeller metin içindeki karmaşık ilişkileri anlamada büyük başarı sağlamıştır.  
- Gözetimsiz Öğrenme: Kümeleme algoritmaları kullanılarak metinler benzer duygu içeriklerine göre gruplanmıştır.  
- Yarı-Gözetimli Öğrenme: Gözetimsiz ve gözetimli öğrenme tekniklerini birleştirerek sınırlı etiketli veriyle eğitim yapılmasına olanak tanınmıştır.  
  
5. Sonuçlar  
Makale, derin öğrenme modellerinin ve gelişmiş gözetimli öğrenme tekniklerinin duygu analizindeki doğruluğu artırdığını göstermektedir. Özellikle sosyal medya verilerinde, bu modellerin daha doğru ve esnek olduğunu ortaya koymuştur.  
  
6. Eksikler  
- Dil Çeşitliliği ve Kültürel Farklılıklar: Mevcut yöntemler genellikle tek bir dilde çalışmakta ve farklı diller veya kültürlerde aynı performansı göstermemektedir.  
- Karmaşıklık ve Hesaplama Maliyetleri: Derin öğrenme modellerinin eğitim süreci oldukça zaman alıcı ve maliyetlidir. Bu modellerin basitleştirilmesi ve hızlandırılması gereklidir.  
- Örtük ve İronik İfadeler: İroni veya örtük ifadeleri doğru bir şekilde anlamak mevcut teknikler için hala zordur.  
  
  
**Makale Başlığı**

Sentiment Analysis Using Machine Learning: A Comprehensive Revie

Özet

1. Konusu

Bu makale, makine öğrenimi kullanarak duygu analizi üzerine kapsamlı bir inceleme sunmaktadır. Duygu analizi, metin verilerindeki duyguları yorumlamak ve sınıflandırmak amacıyla doğal dil işleme (NLP) tekniklerini kullanır. Çalışma, farklı makine öğrenimi algoritmalarının metodolojilerini, performanslarını ve gerçek dünya uygulamalarını değerlendirir. Özellikle derin öğrenme modellerinin, insan dilinin karmaşıklıklarını anlamadaki etkileri üzerinde durulmuştur.

2. Ne Yapılmış?

Makale, duygu analizi için kullanılan çeşitli makine öğrenimi tekniklerini incelemiştir. Destek vektör makineleri (SVM), rastgele ormanlar (Random Forests), sinir ağları ve BERT gibi modern modellerin duygu sınıflandırmasındaki etkinlikleri analiz edilmiştir. Çalışma, bu algoritmaların başarısını ve hangi durumlarda daha verimli olduklarını karşılaştırmalı olarak ele almıştır.

3. Nasıl Yapılmış?

- Veri İşleme: Makale, duygu analizi için verilerin nasıl ön işleme tabi tutulması gerektiğini açıklamıştır. Bu adımlar arasında tokenizasyon, durdurma kelimelerinin çıkarılması, kelime köklerine indirgeme gibi NLP teknikleri bulunmaktadır.

- Makine Öğrenimi Algoritmaları:

- Destek Vektör Makineleri (SVM):Metin verilerindeki farklı duygu sınıflarını ayırt etmek için kullanılan bir yöntemdir.

- Rastgele Ormanlar: Karar ağaçlarını bir araya getirerek veri içindeki karmaşık desenleri yakalamaya yardımcı olur. Duygu çıkarımı için kullanılan derin öğrenme modelleri. CNN, metindeki yerel desenleri tespit ederken; LSTM, uzun metinlerdeki bağlamı anlamada başarılıdır.

- BERT ve Diğer Derin Öğrenme Modelleri:Önceden eğitilmiş dil modelleri kullanılarak duygu sınıflandırmasında önemli başarılar elde edilmiştir.

- Özellik Çıkarma: Metinlerin sayısal olarak temsil edilmesi için TF-IDF ve Bag-of-Words (BoW) gibi teknikler kullanılmıştır.

- Derin Öğrenme ve Transfer Öğrenimi:Büyük veri setleri üzerinde önceden eğitilmiş modellerin daha küçük veri setlerine adapte edilmesi amacıyla transfer öğrenim teknikleri uygulanmıştır.

4. Kullanılan Metotlar

- Destek Vektör Makineleri (SVM): İkili ve çoklu sınıflandırma için kullanılan güçlü bir yöntem.

- Rastgele Ormanlar: Birden fazla karar ağacını birleştirerek duygu sınıflandırmasında yüksek doğruluk sağlama.

- Sinir Ağları (CNN, LSTM): CNN, metindeki kelime desenlerini analiz ederken, LSTM uzun süreli bağımlılıkları tanımlayarak bağlamı anlamada başarılıdır.

- BERT ve Transformer Modelleri: Derin öğrenme modelleri arasında yer alan BERT, metin içindeki karmaşık ilişkileri anlamada öne çıkmıştır. Transfer öğrenimi ile diğer veri setlerinde daha hızlı ve doğru analiz yapma imkanı sunar.

- Ön İşleme ve Özellik Çıkarma:Tokenizasyon, kelime kökü çıkarma, TF-IDF gibi teknikler kullanılarak ham veriler analiz edilebilir forma dönüştürülmüştür.

5. Sonuçlar

Makalenin sonuçları, derin öğrenme modellerinin (özellikle BERT ve CNN) duygu analizi görevlerinde geleneksel yöntemlerden çok daha başarılı olduğunu göstermektedir. Bu modeller, metinlerdeki duygu ifadelerini daha doğru ve hızlı bir şekilde tespit edebilmiştir. Transfer öğrenimi kullanılarak daha küçük veri setlerinde bile etkili sonuçlar elde edilmiştir.

6. Eksikler

- Veri Seti Bağımlılığı: Derin öğrenme modelleri büyük veri setlerine bağımlıdır, bu nedenle bazı durumlarda yeterli veri sağlanamazsa performans düşebilir.

- Model Anlaşılabilirliği: Özellikle BERT gibi modellerin "neden" belirli bir sınıflandırma yaptığı tam olarak anlaşılmayabilir, bu da kullanıcı güvenini etkileyebilir.

- Hesaplama Gücü İhtiyacı: Derin öğrenme modelleri, yüksek hesaplama gücü ve büyük bellek kapasitelerine ihtiyaç duyar.

7. Biz Nasıl Yaparız?

Bu tür bir duygu analizi projesini gerçekleştirmek için aşağıdaki adımları izleyebiliriz:

1. Veri Toplama:Twitter, Facebook ve diğer sosyal medya platformlarından verileri toplamak ve sınıflandırmak.Kaggledan da veri alınabilir.

2. Ön İşleme: Tokenizasyon, durdurma kelimelerinin temizlenmesi, kelime kökü çıkarma gibi işlemlerle ham veriyi analiz edilebilir hale getirmek.

3. Özellik Çıkarma: TF-IDF ve kelime gömme tekniklerini kullanarak metni sayısal bir formata dönüştürmek.

4. Model Eğitimi:Derin öğrenme modelleri (CNN, LSTM, BERT) kullanarak metinleri eğitmek. İhtiyaç halinde transfer öğrenimi uygulanabilir.

5. Model Performansını Test Etme ve İyileştirme:Doğruluk, hız ve esneklik açısından modeli test ederek en iyi sonucu elde etmek için ayarlamalar yapmak.

6. Gerçek Zamanlı Analiz: Modelin canlı veriler üzerinde gerçek zamanlı analiz yapabilmesi için sistemin optimizasyonunu sağlamak.