

Yapay Zeka'nın Etik Hukuk ilişkileri

Mehmet Fatih AMASYALI Yapay Zeka Ders Notları

YILDIZ TEKNİK ÜNİVERSİTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

içerik

- Tanımlar
- Verilerdeki önyargı
- Sorumlu kim
- İş sonrası dünya
- Manipülasyon

Mehmet Fatih AMASYALI Yapay Zeka Ders Notları

YILDIZ TEKNİK ÜNİVERSİTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Tanımlar

- Etik / Ahlak : İyi / kötü nedir? Hangi durumda ne yapılmalıdır?
- Hukuk: Toplumu düzenleyen ve devlet yaptırımıyla güçlendirilmiş bulunan kuralların, yasaların bütünü
- YZ: karar veren bir sistem
- Hukuki olan her zaman etik midir?
- Zamandan ve kültürden bağımsız değiller (kölelik bir zamanlar hukuki idi.) Gelecek?

Verilerdeki önyargı

- YZ algoritmaları veriden öğrenirler. Veride önyargı varsa onu da öğrenirler.
- Örneğin insan kaynakları bölümündeki kişiler bazı gruplara (cinsiyet / yaş vb.) ayrımcılık yaptıysa, YZ modeli de yapacaktır.

Verilerdeki önyargı

- «sarı muz» pek demeyiz.
 - «kırmızı muz» → 13.400 sonuç
 - «sarı muz» → 30.400 sonuç
 - «yeşil muz» → 33.600 sonuç
- Yazılı metinlere baksak insanlar hangisini daha çok yapıyordur? Yemek yemek / birini öldürmek
 - «yemek yedi» → 714.000 sonuç
 - «öldürdü» → 26.100.000 sonuç
- Normalde olan şeyleri söylemez yazmayız. Gerçek dünyadaki olayların frekanslarıyla yazılı metinler ya da konuşmalardaki frekansları çok farklı olur.

Verilerdeki önyargı

- Bir adam ve oğlu ciddi bir trafik kazası geçirmişler ve acildeler. Çocuğun ameliyat olması gerekiyor. Ameliyatı yapacak doktor çocuğa bakıyor ve «bu ameliyatı yapamam bu benim oğlum» diyor. Bu mümkün mü?

Verilerdeki önyargı

- Benzer zamanda olan şeyler her zaman sebep sonuç ilişkisine sahip değildir.
- İnternetin yaygınlaşması ile dünyadaki ağaç sayısının azalması
- Resmi çekilen kedi sayısındaki artış ile trafik kazalarının sayısındaki artış
- Kadınlara oy hakkı verilmesi ile dünya savaşlarının başlaması (1900 ler)

Önyargıyı ortaya çıkarmak

- X görevinde her alt grup için ayrı test sonuçları vermek:
- Yüz tanımada siyah / beyaz, kadın / erkek insanlarda ayrı testler yapmak

Bir YZ sistemi hata yaptı

- Sorumlu kim?
- YZ'ler de hata yaparlar
- Önüne atlayan bir yayaya çarpabilirler

YZ işlerimizi elimizden alacak mı?

- Global işsizlik, İnsanların çok büyük bir çoğunluğunun emeğine ihtiyaç duymayan bir sistem
- Ya da
- Önceki işgücü devirlerinde (insan → makine) olduğu gibi (sanayi devrimi de işçilerden tepki almıştı) sadece bazı mesleklerin frekansının azalıp yeni mesleklerin ortaya çıkışı

Manipölasyon

- Birisinin bir konu hakkındaki fikrini ona video izleterek / haber okutarak vb. deęiřtirmek mümkün mü?
- Tüm reklam sektörü bunun üzerine kurulu ☺
- Sosyal medyada karşınıza çıkan içerik neye göre seçiliyor?
- X'i satın aldırmaya çalışmak vs. X partisine oy verdirmeye çalışmak

Sosyal Medya*

- Optimizing clickthrough
Ne idi? learning what people want
Ne oldu? modifying people to be more predictable
Daha iyi tahmin edebilmek için daha tahmin edilebilir hale getirmek ☺

[*] Stuart Russell, [How Not to Destroy the World With AI](#), AAAI 2020 keynote talk.

Kişisel verilerinizin sahibi

- Ne amaçla, hangi verilerinizi kullandığını, kimlere vereceğini açıklamalı, size ait verileri göstermeli
- Ham veriler bir şey, işlenmiş veriler başka bir şey:
 - tıkladığınız haberler → siyasi tercihiniz
 - koordinatlarınız → rakip şirketle görüşmeye gittiniz
- Reklam amaçlı kullanacağım = ???

Kişisel veriler nereye

- Patrona
- Eşinize
- Ebeveynlerinize
- Sigorta şirketlerine (sigara içenlerin / hamburger yiyenlerin tedavi masraflarını biz mi ödeyelim?)
- Devlete (teröristlerin izlenmesi, nefret söylemleri, suçların / intiharların tahmini)
- Güvenlik

Konu sađlık olunca

- Tm verimizi paylařmak isteyebiliriz eđer erken teřhis vb. iře yarayabilir.
- Tedavi eden, nleyici sađlık hizmeti sunan kurumlara OK
- Peki sigorta řirketleriyle ?

- Seni srekli takip eden, her yaptığından anlam ıkarmaya alışan, seni ynlendirmeye alışan bir sistem
- Beni iyi tanısın ki bana uygun nerilerde bulunsun
- A markası alacakken B ye ynlendirmek, daha ucuzuna, daha kaliteliğine vb. OK
- A partisine oy atacakken B ye ynlendirmek?
- A grubu insanlara karřı fkelenmek?
- Yalanla maniplasyon

Sosyal medyadaki / telefonunuzdaki verilerinizden neler öğrenilebilir?

- Ne zaman, kiminle, nerede, ne yaptığınız (Özel hayatınız)
- Ya da ne öğrenilemez ☺

Gülmece ??

WHAT YOU PAY FOR LIFE INSURANCE COULD DEPEND ON YOUR NEXT INSTAGRAM POST[*]

Some life insurers want to monitor how you behave online as a way of assessing your health and bypassing more traditional blood and urine samples. Here are five online actions that may help you avoid the medical exam.

- Don't post photos of yourself smoking on social-media sites.
- Do post photos of yourself running. Riskier sports, like skydiving, could complicate the situation.
- Use fitness-tracking devices that indicate an interest in fitness.
- Purchase food from online meal-preparation services that specialize in healthy choices.
- Visit the gym with a phone linked to a location-tracking service. If you visit the bar, leave your phone at home.

[*] https://twitter.com/karen_ec_levy/status/1091161689335218177?s=03

Kaynaklar

- https://twitter.com/karen_ec_levy/status/1091161689335218177?s=03
- <https://www.mridvano.com/sosyal-medya-verileri-ile-suc-tahmini-mumkun-mu/>
- <https://www.youtube.com/watch?v=XR8YSRcuVLE&list=PLoROMvodv4rOhcuXMZkNm7j3fVwBBY42z&index=19>