

Tez Konuları

Yüksek Lisans:

- Tweet normalization: bozuk yazılmış tweet'lerin imla kurallarına uygun hale getirilmesi
 - o Dilara Torunoğlu and Gülşen Eryiğit. A Cascaded Approach for Social Media Text Normalization of Turkish., 5th Workshop on Language Analysis for Social Media (LASM) at EACL, Gothenburg, Sweden, April 2014
- Daha az etiketli veri kullanımı
 - o Semi supervised learning : etiketsiz verileri de kullanmak
 - o Active learning: etiketlenen verileri seçmek
 - o Active sampling: ilk veri kümesi ile bir sonraki etiketli veri kümesi belirlenir. Kararına güvenmediğin örnekleri (mesela içinde şu şu kelimeler olan tweet leri) elde et ve etiketleyerek eğitim kümesine ekle.
 - o Uygulama alanı: Sentiment analysis vb.

Yüksek Lisans / Doktora:

- Curriculum learning
 - o Zorluğu aşamalı olarak artan eğitim kümeleriyle öğrenmek
 - o Zorluk nasıl tanımlanır?
 - o http://ronan.collobert.com/pub/matos/2009_curriculum_icml.pdf
- Kelimelerin anlamsal bir uzaydaki koordinatlarını bulmak
 - o Word2vec iyileştirme ve yeni boyutlar ekleme, ilişki türünü belirleme
 - o <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
- Hayat bilgisi veri tabanları oluşturma / kullanma
 - o Sürekli çalışan / büyüyen / iyileşen bir CSDB
 - o Uygulama alanları: story understanding, story generation, text paraphrasing, textual entailment
 - o "TURKISH COMMONSENSE DATABASE (CSDB) AND CSOYUN (A GAME WITH A PURPOSE)", Serkan ÖZCAN, M. Fatih Amasyali, Sigma Journal of Engineering and Natural Sciences, Year: 2014, Volume: 32, Issue: 2, (JUN 2014), pp:116-127, [PDF](#)
 - o <http://rtw.ml.cmu.edu/rtw/>
- Data Augmentation / Generation
 - o Yeni veriler üretmek
 - o Yeni verileri eğitimi iyileştirmek için kullanmak
 - o Generative Models
 - o Generative Adversarial Networks
 - <https://www.quora.com/What-are-Generative-Adversarial-Networks>

- <https://arxiv.org/abs/1406.2661>
- ML algoritmalarını kandırmak
 - <http://www.popsci.com/byzantine-science-deceiving-artificial-intelligence>
 - <https://arxiv.org/abs/1412.1897>
 - <http://www.kdnuggets.com/2016/04/tricking-deep-learning.html>
- Robotik
 - Visual odometry
 - Multi robot exploration
 - Manipulation, task planning

Doktora:

- Bongard problemlerini çözen / üreten sistemler
 - "Bongard Problems for Representation Learning", "Temsil Öğrenme için Bongard Problemleri", Mehmet Fatih Amasyalı, UBMK 2016, October 20-23, Tekirdağ, Turkey, 2016 [PDF](#)
- Neuro dynamics, ensemble dynamics, ML algorithm dynamics, learning curve prediction
- Multi-tasking: Birden fazla problemi aynı anda öğrenmek

Tez Başvurularınız için

Yukarıdaki konularda ya da ilgili konularda (Yapay Öğrenme, Doğal Dil İşleme, Robotik) aşağıdaki formatta hazırlayacağınız bir ön çalışmayı mfatihamasyali@gmail.com adresine gönderip randevu alınız.

Format

Nedir: çözeceğiniz problemi tanımlayınız. Sistemin girişi ve çıkışı

Önemi: bu problem neden önemlidir? Çözünce ne olacak? Ne gibi uygulama alanları olacak?

Tarihçe / mevcut durum / referanslar: bu problemi çözmek için neler yapılmış (bu konuda en fazla referans almış, görece yeni, 5 makale neler?), mevcut çözümlerin tarihçesi nasıl? Birbirleri üzerine nasıl gelişim göstermişler? Mevcut yöntemlerin birbirlerine göre avantaj ve dezavantajları neler? Problemin hangi yönlerinde ilerlemeler kaydedilmiş? Mevcut çözümlerin kısıtları, yapamadıkları şeyler neler?

Önerdiğiniz araştırma yönleri: planınız, varsayımlarınız, araştırma sorularınız, probleme bakış açınız, ön çalışmalarınız, mevcut yöntemlerin hangi eksikliklerini nasıl gidermeyi planlıyorsunuz?

Başarı ölçümleri: mevcut veri kümeleri, başarı kriterleri

Genel araştırma yönleri

daha hızlı, daha ölçeklenebilir, daha doğru, daha basit/anlaşılır, daha genel, daha az hiperparametre içeren,

Genel Tavsiyeler

Tezler araştırma süreçleridir. Süreç boyunca birçok hipotez üretilir / yoklanır / elenir. Çeşitli varsayımlar eklenir / silinir. Bu sebeple başlangıçtaki araştırma yönleri / varsayımlar süreç içinde defalarca değişebilir, genelde de değişir. Ama gelişim için bir başlangıç noktası ve yön tanımlamak gerekir. Yukarıdaki formatta yapacağınız çalışma bu başlangıç noktasını ve olası yönleri belirlemenize, konuyu sevip sevmeyeceğinize karar vermenize yardımcı olacaktır.

Büyük sorulara tek denemede cevap bulmak pek olası değildir. Amaç büyük sorulara cevap olabilecek küçük ve iyi sorular tanımlamaktır.

Giderek iyileşen sorular:

Şu algoritmayı nasıl iyileştirebilirim?

Şu algoritmayı nasıl hızlandırırım?

Şu algoritmanın hangi bileşeni hızını azaltıyor?

Şu algoritmanın hızını azaltan durumlar nelerdir? Bunların ortaya çıkma oranları nasıldır?

Bu durumları nasıl tanımlarım?

Bu durumlarda algoritmayı nasıl hızlandırırım?

Bu durumları paralelleştirmek mümkün mü?

Çözümüm arama uzayının hangi boyutları daha önemli?

Hızlandırırken doğruluktan ne kadar ödün verebilirim?

Çözümünüzü değerlendirirken:

Bulduğunuz çözümün matematiksel altyapısı nasıl? Çözümünüzü diğerlerinden daha iyi olduğunu deneysel olarak göstermenin yanında, belirli varsayımlar altında matematiksel olarak kanıtlayabiliyor musunuz?

Çözüm ararken:

Mevcut çözümlerin tarihçesi nasıl? Birbirleri üzerine nasıl / nerelerde gelişim göstermişler?

Mevcut çözümlerin varsayımları neler?

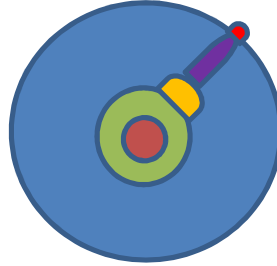
Varsayımların etkileri neler?

Varsayımlayı genelleştirmek sıkılaştırmak nasıl sonuç verir?

Başka tür varsayımlar tanımlamak mümkün mü?

Teziniz şeklen*

İnsanlığın bilgisi, ilkokul, ortaokul, lise, lisans, lisansüstü dersleriniz, araştırma makaleleri okumak, teziniz



* <http://matt.might.net/articles/phd-school-in-pictures/> den uyarlanmıştır.

Tavsiye Yazılar:

<http://www.cs.virginia.edu/~robins/YouAndYourResearch.html>

<https://terrytao.wordpress.com/career-advice/>

Bazı Veri Kümeleri ve Uygulama Alanları:

<http://www.kdnuggets.com/2015/12/tour-real-world-machine-learning-problems.html>

27 Ekim 2016 v.1

Mehmet Fatih Amasyalı