

# Vector Semantics

Dr.Furkan Göz

Kasım 27, 2023

- N-gram veya metin sınıflandırma yöntemleri: kelimeler string
- Kelimelerin anlamlarını nasıl temsil edebiliriz?
  - Bir model soru-cevaplama, metin özetleme gibi anlama dayalı problemleri çözmekte bize faydalı çıkarımlar yapmamızı sağlamalıdır.
- Kelimelerin anlamla ilgili yapısını inceleyen dilbilimsel alana lexical semantic (sözcüksel anlambilim) denir.

- **Madde:** mouse (İsim)
  - hayvan... (word sense)
  - bir imleci kontrol eden el ile kullanılan cihaz... (word sense)
- **Anlam:**
  - mouse - lemma (orjinal form)
    - mouse, mice kelimesi için lemma;
    - sözlükler çekimli formlar için ayrı tanımlar içermez, örneğin mice gibi.
    - Benzer şekilde, sing kelimesi için sing, sang, sung formlarının lemma'sı sing'dir.
    - Birçok dilde fiiller için lemma olarak mastar hali kullanılır.

- Her lemma birden çok anlam taşıyabilir;
  - lemma "mouse", kemirgen veya imleç kontrol cihazına atıfta bulunabilir
- Bu anlamlardan her birini "mouse" kelimesinin bir **word sense**'i (kelime anlamı) olarak adlandırılır
- Lemmaların birden fazla anlamının olması, yorumlamayı zorlaştırır
  - mouse → kemirgen veya imleç kontrol cihazı.
- Word sense disambiguation (Kelime anlamı çözümlemesi), bir kelimenin belirli bir bağlamda hangi anlamda kullanıldığının belirlenmesi görevidir.
- Kelime anlamları arasında birçok farklı ilişki olabilir.
  - eşanlamlılık, zıtlık, benzerlik, üst/alt kavramlar, ...

- Bir kelimenin anlamı, başka bir kelimenin bir anlamıyla özdeş veya neredeyse özdeş olduğunda, bu iki kelimenin anlamlarının (synonym) eş anlamlı olduğunu söyleyebiliriz.
  - Örnekler (eş anlamlılar): couch/sofa, big/large, car/automobile, ...
- İki kelime anlamı, anlamın bir özelliği açısından zıt olduklarında zıt anlamlı olduğunu (antonym) söyleyebiliriz.
  - Örnekler (antonimler): dark/light, short/long, fast/slow, rise/fall, hot/cold, up/down, in/out.
- Zıt anlamlılar, anlam olarak farklılık gösterir, ancak yapı olarak benzerlik gösterir.
- Bu nedenle, eş anlamlıları zıt anlamlılardan otomatik olarak ayırt etmek zor olabilir çünkü benzer bağlamlarda yer alırlar.

- Kelimelerin çok fazla eş anlamlısı olmasa da, çoğu kelimenin birçok benzer kelimesi vardır.
- Kedi, köpek kelimesinin eş anlamlısı değildir, ancak kediler ve köpekler kesinlikle benzer kelimelerdir.
- Benzer kelimeler anlamın bazı öğelerini paylaşır.
- kedi, köpek : evcil hayvan
- Hangi kelimeler benzerdir?  
otomobil - SUV - bisiklet - domates
- **Words Similarity** (Kelime benzerliği): semantic tasks
- Kelime benzerliği, ifade veya cümle benzerliğini belirlemede kullanılır (soru cevaplama sistemleri, özetleme gibi)

- Kelime benzerliği değerlerini elde etme yöntemlerinden biri, insanlardan bir kelimenin diğerine ne kadar benzediğini değerlendirmelerini istemektir.
- SimLex-999 veri seti (Hill ve diğerleri, 2015), farklı kelime çiftlerinin 0 ile 10 arasında değişen benzerliklerini gösterir.

word1	word2	similarity
vanish	disappear	9.8
behave	obey	7.3
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.3

- Word relatedness: benzerlik dışında birbiriyle ilintili kelimeler
- Word Relatedness or Word association
- *Coffee* ve *cup* kelimesi hemen hemen hiç ortak özellik paylaşmadığı için benzer değildir.
- Ancak *coffee* ve *cup* kelimeleri, bir fincandan kahve içme olayında birliktelik gösterdikleri için ilintilidir
- *Scalpel* ve *surgeon* benzer değildir ancak olay bağlamında ilintilidirler



Kelimeler arasındaki yaygın ilinti türlerinden biri, aynı anlamsal alana ait olmalarıdır. Anlamsal alan, belirli bir anlamsal alanını kapsayan ve birbirleriyle yapılandırılmış ilişkiler taşıyan sözcükler kümesidir.

## Örnekler

- **Hospitals:** surgeon, scalpel, nurse, anesthetic, hospital
- **Restaurants:** waiter, menu, plate, food, chef
- **Houses:** door, roof, kitchen, family, bed

Anlamsal alanlar aynı zamanda konu modelleri ile de ilgilidir. Anlamsal alanlar ve konu modelleri, belgelerdeki topic yapılarını öğrenmek için kullanışlıdır.

Semantic frame (anlamsal çerçeve) belirli bir tür olayda rolleri belirten kelimelerdir.

- “Sam bought the book from John”
- Buyer: Sam Seller: John Goods: book

- Bir kelime başka bir kelimeye göre bir alt sınıfı belirtebilir (hyponym = subordinate)
- Bir kelime başka bir kelimeye göre daha genel bir üst sınıfı belirtebilir (hypernym = superordinate)
  - *car* (*araba*) bir *vehicle* (*taşıt*)'nin alt sınıfıdır.
  - *mango* bir *fruit* (*meyve*)'nin alt sınıfıdır.
  - *vehicle* (*taşıt*) bir *car* (*araba*)'ın üst sınıfıdır.
  - *fruit* (*meyve*) bir *mango*'nun üst sınıfıdır.

- Kelimeler: affective meanings (connotations).
  - Connotations bir kelimenin anlamının bir yazarın veya okuyucunun duyguları, hisleri, görüşleri veya değerlendirmeleriyle ilgili yönlerini ifade eder.
- Örnekler:
  - Positive connotations: *happy*
  - Negative connotations: *sad*
- Değerlendirme:
  - Positive evaluation: *great, love*
  - Negative evaluation: *terrible, hate*
- Positive ya da negative evaluation: sentiment.

Kelime anlamının farklı yönlerini (word senses, word similarity and relatedness, semantic fields and frames, connotation) başarıyla ele alan bir hesaplama modelini nasıl oluşturabiliriz?

- MÜKEMMEL bir model YOK.
- Kelime anlamını modellemek için kullanılan hesaplamalı modellerden biri WordNet gibi bir sözlük kullanımıdır:
  - WordNet, İngilizce (ve diğer dillerde) lexicall relation'ları içeren bir veritabanıdır.
- Vector semantics: Doğal Dil İşlemede (NLP) kelime anlamını temsil etmede kullanılan bir yöntem.

- Ludwig Wittgenstein, her bir kelime için sözlük anlam tanımları oluşturma olasılığına şüpheyle yaklaşmış ve bunun yerine "bir kelimenin anlamının onun dildeki kullanımı olduğunu" öne sürmüştür.
- Bu bakış açısına göre:
  - Mantıksal bir dil kullanmak yerine, kelimeleri, gerçek insanlar tarafından nasıl kullanıldığına dair temsillerle tanımlamalıyız.
  - Distributionalist approach (Dağılımsal yaklaşımda) bir kelime, çevresi veya dil kullanımındaki dağılımı ile tanımlanır.
  - Kelimeler, kullanımları ile tanımlanır.
  - Bir sözcüğün dağılımı, içinde geçtiği bağlamlar kümesidir; bu bağlamlar komşu sözcükler ya da dilbilgisel ortamlar olabilir.
  - Benzer dağılımlarda ortaya çıkan iki kelimenin muhtemelen aynı anlama sahip olduğu düşünülür.

- **Ongchoi** kelimesinin anlamını bilmiyoruz, ancak aşağıdaki cümlelerde (bağlamlarda) kullanımını görüyoruz:
  - **Ongchoi**, sarımsakla sotelenirse lezzetli olur.
  - **Ongchoi**, pirinçle yakıştır.
- **Ongchoi** kelimesinin bazı bağlam kelimeleri (lezzetli, sote, sarımsak, pirinç) aşağıdaki bağlamda da geçiyor:
  - ...sarımsakla sotelenmiş ıspanak, pirinç üzerinde...
- **Ongchoi** kelimesinin bağlam kelimelerinin ıspanak çevresinde de geçiyor olması, ıspanak ve **ongchoi** arasındaki benzerliği keşfetmemize yardımcı olabilir.
- **Ongchoi**, "su ıspanağı" anlamına gelir.

- Vector semantics:
  - Dağılımcı Sezgi: Bir kelimenin anlamını, çevresindeki diğer kelimelerin ne sıklıkta ortaya çıktığının sayarak tanımlanması
  - Vektör Sezgisi: Bir  $w$  sözcüğünün anlamını bir vektör, bir sayı listesi,  $N$  boyutlu uzayda bir nokta olarak tanımlanması
- Vector semantics'in çeşitli versiyonları vardır, her biri vektördeki sayıları biraz farklı tanımlar, ancak her durumda sayılar bir şekilde komşu kelimelerin sayımlarına dayanır.
- Vector semantics bir kelimeyi, çok boyutlu semantik bir uzayda bir nokta olarak temsil etmektedir.