



DOĞAL DİL İŞLEME

Doğal Dil İşlemenin Günlük Hayattaki Uygulamaları

ZEYNEP KIR

Zeynepkr47@gmail.com

İçindekiler Tablosu

NPL'nin sosyal medya analizi ve duygu analizi üzerindeki etkisi	2
Sesli asistanlar (Alexa, Siri) ve konuşma tanıma teknolojileri.	2
Makine çevirisi ve çok dilli iletişimdeki rolü.	4
Otomatik metin özetleme ve öne çıkarılmış içerik oluşturma.....	6
KAYNAKÇA.....	8



NPL'nin sosyal medya analizi ve duygu analizi üzerindeki etkisi

Doğal Dil İşleme (NLP), bilgisayarların insan dilini anlamasına ve işlemesine olanak tanıyan bir disiplindir. NLP'nin sosyal medya analizi ve duygu analizi üzerindeki etkisi büyük bir öneme sahiptir. NLP'nin sosyal medya üzerindeki etkileri:

Metin Madenciliği (Text Mining): NLP, büyük miktarlardaki metin verilerini analiz etmek için kullanılır. Sosyal medyadaki kullanıcı yorumları, paylaşımlar ve tweet'ler gibi metin verileri, NLP algoritmalarıyla işlenerek önemli bilgiler çıkarılabilir.

Duygu Analizi (Sentiment Analysis): NLP, duygu analizi konusunda oldukça etkili bir araçtır. Sosyal medyada paylaşılan metinler üzerinden, kullanıcıların oluşturulan veri setleri ile duygusal durumları belirlenebilir. Bu, markaların ürünleri veya hizmetleri hakkındaki genel kamusal görüşleri anlamalarına yardımcı olabilir.

Trend Analizi: NLP, sosyal medya platformlarında popüler olan konuları ve trendleri analiz edebilir. Bu, şirketlerin pazarlama stratejilerini oluştururken veya içerik üretirken kullanabilecekleri değerli bilgiler sunar.

Müşteri Geri Bildirimleri: Sosyal medya üzerindeki müşteri yorumları ve geri bildirimleri, NLP veri setleri kullanılarak analiz edilebilir. Bu, şirketlerin müşteri memnuniyetini anlamalarına ve geliştirmeleri gereken alanları belirlemelerine yardımcı olabilir.

Kriz Yönetimi: Sosyal medyada ortaya çıkan kriz durumlarına hızlı bir şekilde tepki vermek için NLP kullanılabilir. Kullanıcıların paylaşımları ve duygusal tonları analiz edilerek, şirketler olası sorunlara karşı daha hızlı ve etkili bir şekilde yanıt verebilirler.

Kişisel Asistanlar ve Chatbotlar: Sosyal medya üzerindeki etkileşimlerde kullanıcı deneyimini artırmak için NLP, chatbotlar ve kişisel asistanlar aracılığıyla entegre edilebilir. Bu, kullanıcılara daha etkili ve kişiselleştirilmiş hizmetler sunmada yardımcı olabilir.

Sosyal Ağ Analizi: NLP, sosyal medyadaki ilişkileri ve etkileşimleri analiz ederek, belirli kullanıcı grupları veya topluluklar hakkında bilgi sağlayabilir. Bu bilgiler, hedef kitleye daha iyi odaklanmak için kullanılabilir.

NLP'nin sosyal medya analizi ve duygu analizi üzerindeki bu etkileri, şirketlerin müşteri odaklı stratejiler geliştirmelerine, kriz durumlarına daha etkili bir şekilde müdahale etmelerine ve genel olarak sosyal medyadaki dinamikleri daha iyi anlamalarına yardımcı olabilir.

Sesli asistanlar (Alexa, Siri) ve konuşma tanıma teknolojileri.

Sesli asistan, insan konuşmasını tanımak, kelimeleri çevirmek, doğru yanıt vermek ve istenen eylemleri gerçekleştirmek için ses tanıma teknolojisini, yapay zekayı ve doğal dil işlemeyi kullanan bir uygulama veya programdır. Sesli asistanlar, müşterilerin arama yapma ve çevrimiçi komut verme şeklini önemli ölçüde değiştirdi. Ayrıca sesli asistan teknolojisi, akıllı telefonlar, hoparlörler ve giyilebilir cihazlar gibi günlük cihazlarımızı akıllı uygulamalara dönüştürdü.

A. Doğal Dil Üretimi ve İşleme

İnsanların doğal karşıladığı çok karmaşık yazma ve konuşma süreci, bilgisayar bilimcilerinin çözmesi ve kopyalaması gereken bir zorluktu. Doğal dil işleme ve doğal dil oluşturmanın artan karmaşıklığı, yapay zekanın alt kümeleri, verileri alıyor, işliyor ve sanki bir insan gerçekten

konusuyor veya yazıyormuş gibi görünen doğal dil yaratıyor. Doğal dil işleme (NLP), bir makinenin insanlar tarafından üretilen içeriği "okuma" veya "anlama" yeteneğidir; Yazarak veya konuşarak. Doğal dil üretimi (NLG), bir makinenin, insanlar tarafından anlaşılabilir şekilde yazılı veya sözlü dilde içerik oluşturma yeteneğini ifade eder.

B. Makine Öğrenimi Alexa ve Siri'nin Öğrenmesine Yardım Edin

Makine öğrenimi, makinelere verilere erişim izni verilen ve daha sonra bu verilerle ilgili ne düşünecekleri ve ne yapacakları konusunda insanlar tarafından programlanmaları gerekmeyen bu verilerden öğrenebilecekleri yapay zeka uygulamasıdır. veri.

Alexa veya Siri, isteğinize yanıt verirken her hata yaptığında, bir dahaki sefere iyileştirme yapmak için orijinal sorguya nasıl yanıt verdiğini temel alarak aldığı verileri kullanır. Bir hata yapılmışsa bu veriyi alır ve ondan öğrenir. Yanıt olumluysa sistem bunu da not eder.

Dijital ses asistanlarının patlayıcı büyümesinden veri ve makine öğrenimi sorumludur. Sahip oldukları deneyimler ve biriktirdikleri verilerle daha iyi olmaya devam ediyorlar.

C. Sesle Etkinleşen Dijital Asistanlar Nasıl Çalışır?

Artık yapay zekanın perde arkasında doğal dil üretimi, işleme ve makine öğrenimi yoluyla Alexa ve Siri'nin sonuçlarını güçlendirmeye yardımcı olduğunu anladığımıza göre, gerçekte nasıl çalışıyor?

Alexa veya Siri'den istekte bulunduğunuzda kullandığınız cihazdaki mikrofona komutunuzu kaydeder. Bu kayıt internet üzerinden buluta gönderilir. Yani internet hizmetiniz yoksa dijital sesli asistanlarınız çalışmayacaktır. Alexa ile konuşuyorsanız kayıt Alexa Voice Services'e (AVS) gönderilir. Bu bulut tabanlı servis kaydı inceleyecek ve isteğinizi yorumlayacaktır. Daha sonra sistem cihaza ilgili bir yanıt gönderecektir.

"Bugün hava nasıl olacak?" diye sorduysanız, ilgili yanıt, günün tahminini detaylandıran ses olacaktır. Bunların hepsi bir anda oluyor.

Size günün tahminini söylemek, Alexa'nın yapabileceği 70.000'den fazla beceriden yalnızca biridir. Şirketin bu kadar kısa sürede bu olağanüstü miktara ulaşması ve Alexa'nın başarabileceği becerilerin sayısını artırmaya devam edebilmesi, Amazon'un geliştiricilere sistemi güçlendirmek için beceriler geliştirmeye devam edebilmeleri için AVS'ye ücretsiz erişim sağlamasıyla sağlandı. . Şu anda Alexa ile çalışabilen 28.000'den fazla akıllı ev cihazı bulunmaktadır.

D. Makine Öğrenimi Muazzam Büyüme Sağladı

Alexa ve Siri daha akıllı olmaya devam ediyor. Alexa, insanların takip eden soruları yanıtladığı gibi konuşmayı bir sorudan diğerine nasıl aktaracağını öğrendi. Ve eğer Alexa'nın yapmasını istediğiniz becerinin tam adını bilmiyorsanız, yaklaşırsanız muhtemelen istediğiniz şeyi çağırabilecektir. Ayrıca Alexa Hunches ve akıllı eve bağlı cihazlar aracılığıyla asistan, ışıklar gibi düzenli bir düzenin takip edilmemesi durumunda sizi uyarabilecek. açık bırakılmak veya bir kapının kilidi açılmak ve bunu sizin için tamir etmeyi teklif etmek. Apple'ın Siri'si artık arka plan gürültüsünü ve yüksek sesli müziği "uyandırmak" için ayırabiliyor. Bunlar, yalnızca birkaç yıl öncesine göre daha konuşkan ve yetenekli bir sesli asistana doğru büyük atılımları temsil ediyor. Bu sıçramalar makine öğrenimi sayesinde mümkün.

Makine çevirisi ve çok dilli iletişimdeki rolü.

Yeni çağın önemli bir sembolü, makine çevirisi teknolojisine bir derlemin, doğal dil işleme yoluyla makine çevirisine dönüştürülmüş veritabanının girmesidir. Son yıllarda, derlem tabanlı makine çevirisi sistemler hızla gelişti ve olağanüstü sonuçlar elde etti. Warren weaver, 1947 gibi erken bir tarihte, ‘çeviri’ başlıklı momerandumunda şifreleri çözme yöntemini kullanarak makine olarak adlandırılan bu yöntem esasen bir tür doğal dil işlemedir. Makine çevirisini çözmek için istatistik tabanlı bir yöntem kullanmak istemiştir. Makine çevirisi için matematiksel modeller ortaya koydular. İstatistiksel tabanlı makine çevirisi şekilde gösterildiği gibi gürültülü bir kanal problemi olarak ele alınır.

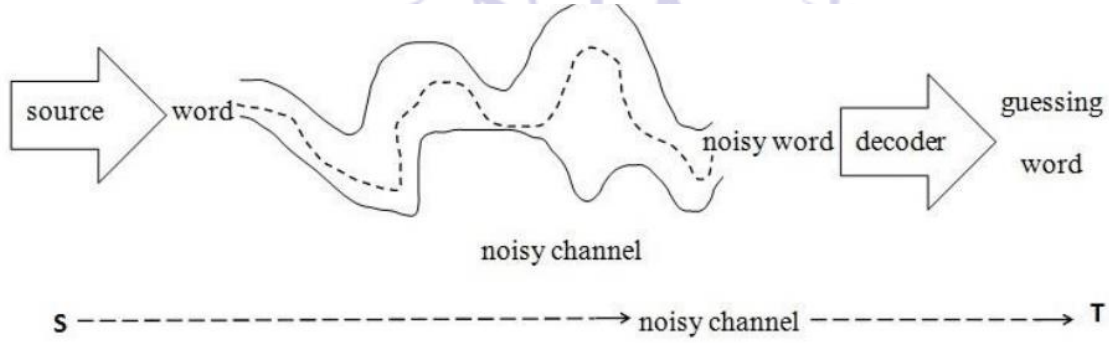
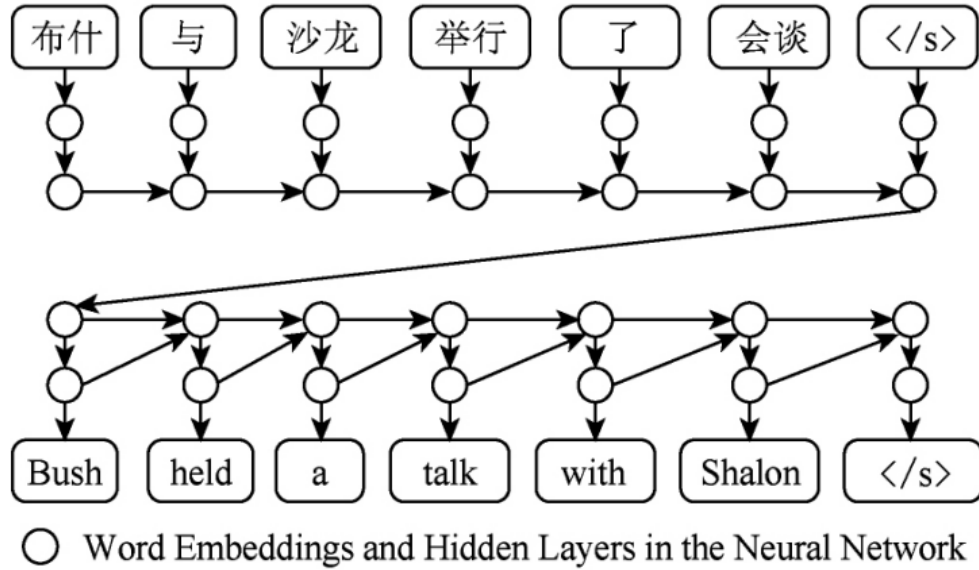


Fig 1. noisy channel model

Şekil 1

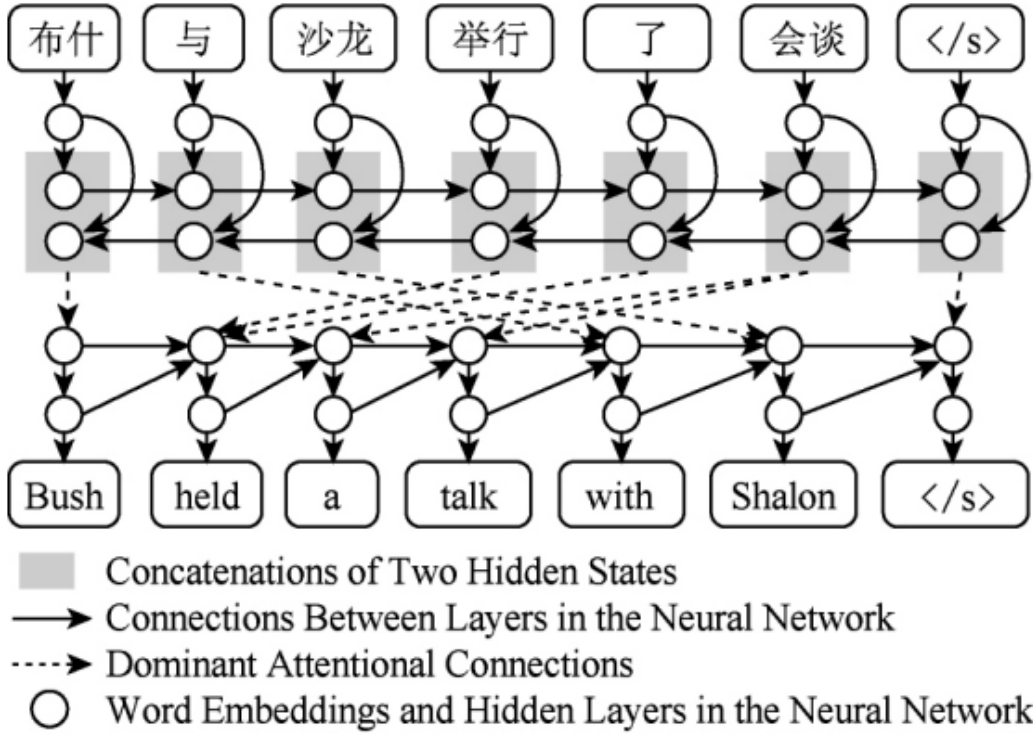


Şekil 2

Uçtan uca nöral makine çevirisinin temel fikri, nöral ağlar aracılığıyla doğal diller arasında otomatik çeviriyi doğrudan uygulamaktır. Bu nedenle, nöral makine çevirisi genellikle

diziden diziye dönüşüm elde etmek için bir kodlayıcı-kod çözücü çerçevesi kullanır. Şekil 3 örnek olarak alındığında, Çince bir cümle “ 噯掬 4 解”verildiğinde, kodlayıcı-kod çözücü çerçevesi ilk olarak her bir cümle için bir vektör temsili oluşturur.

Çince kelime ve ardından tüm Çince cümle için bir vektör temsili oluşturan tekrarlayan bir sinir ağı aracılığıyla soldan sağa gider. Bunlar arasında "</s>" sonlandırıcıyı gösterir. Kaynak dil tarafından kullanılan özyinelemeli sinir ağını kodlayıcı olarak adlandırıyoruz, yani kaynak dil cümlesini yoğun, sürekli bir gerçek sayı vektörüne kodluyoruz.



Şekil 3

Bu yaklaşımın avantajı, her bir kaynak dil kelimesinin vektör temsili için sol ve sağ bağlam bilgilerini içermesidir. Hedef dil tarafında, kod çözücü her bir hedef dil kelimesiyle ilişkili kaynak dil bağlamını dinamik olarak bulur. Örneğin, ne zaman İngilizce "Bush" kelimesi generated, Çince "en alakalı olanıdır ve "Hold" ve "Last" kelimeleri alakalı olmayabilir. Yalnızca "Bush" vektör temsili için kaynak bağlam olarak hedefe aktarmak gerekir. İngilizce "held" kelimesi oluşturulduğunda, en alakalı Çince kelime "噯掬" ve "4" kelimeleridir. Bu nedenle, dikkat mekanizması bilgi aktarım şeklini değiştirir ve en alakalı bağlamı dinamik olarak hesaplayabilir, böylece uzun mesafeli bilgi aktarımı sorunlarını daha iyi çözebilir ve nöral makine çevirisinin performansını önemli ölçüde artırabilir. Bu nedenle, dikkat mekanizmasına dayalı kodlayıcı-kod çözücü modeli, nöral makine çevirisinin ana yöntemi haline gelmiş ve yaygın olarak kullanılmaktadır.

Doğal dil işleme, makine çevirisi için bir ön koşul ve şarttır. Doğal dil işleme açısından, nöral makine çevirisi sadece daha fazla genelliğe sahip olmakla kalmaz, aynı zamanda büyük veri ve büyük veri düşüncesinin gücünü de yansıtır. Yapay zeka teknolojisinin sürekli gelişmesiyle birlikte, doğal dil işleme için yeni teknolojiler çevirinin endüstriyel yapısını değiştirecek ve çeviri endüstrisi zinciri büyük bir endüstriyel değişim yaşayabilir. Şu anda, nöral

makine çevirisi, derin öğrenmenin avantajı ve yüksek boyutlu, etiketsiz ve büyük doğal dil verilerini işleme özellikleri nedeniyle yaygın olarak kullanılmaktadır.

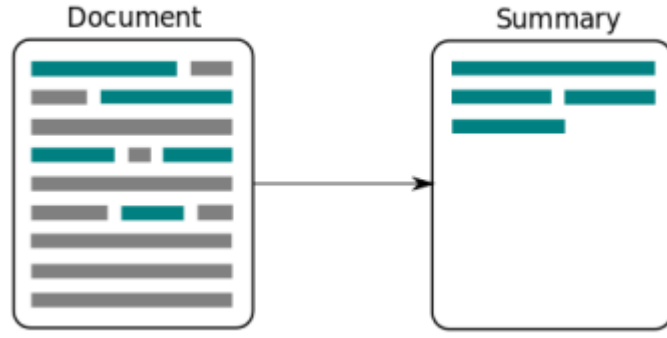
Otomatik metin özetleme ve öne çıkarılmış içerik oluşturma.

İnternetin hızlı büyümesi nedeniyle ortaya çıkan en acil sorunlardan biri aşırı bilgi yüklemesi olarak biliniyor. İlgili bilgilerin basitleştirilmesi Özet şeklinde olması birçok kişiye yardımcı olacaktır çünkü herhangi bir konuyla ilgili materyal internette bol miktarda bulunur. Büyük miktardaki metni manuel olarak özetlemek, insanlar için oldukça zorlu. Bu nedenle daha karmaşık ve güçlü özetleyicilere olan ihtiyaç artmıştır. Araştırmacılar yaklaşımları geliştirmeye çalışıyorlar. 1950'lerden bu yana, makine tarafından oluşturulan özetin insan tarafından oluşturulan özetle eşleşecek şekilde özetler oluşturmak. Bu çalışma, özetleme yaklaşımları, kullanılan teknikler, standart veri kümeleri, değerlendirme ölçütleri ve gelecekteki araştırma kapsamı gibi metin özetleme kavramlarının ayrıntılı bir durum analizini sunmaktadır. En yaygın kabul gören yaklaşımlar çıkarımsal ve soyutlayıcıdır ve bu çalışmada ayrıntılı olarak incelenmiştir. Özetin değerlendirilmesi ve gelişiminin artırılması Yeniden kullanılabilir kaynaklar ve altyapı, bulguların karşılaştırılmasına ve kopyalanmasına yardımcı olarak sonuçları iyileştirmek için rekabeti artırır. Farklı değerlendirme yöntemleri oluşturulan özetler ele alınmıştır. Son olarak bu çalışmanın sonunda metin özetlemeyle ilgili çeşitli zorluklar ve araştırma fırsatları ortaya çıkarılmıştır. Bu alanda çalışan potansiyel araştırmacılara faydalı olabilecek araştırmalardan bahsedilmektedir.

Bir metin parçasını daha kısa bir versiyona sıkıştırma görevi, önemli bilgi yönlerini ve içeriğini korurken orijinal metnin boyutunu en aza indirir anlam, özetleme olarak bilinir. Şekil 1 özetleme görevini basit bir şekilde göstermektedir. Özet, bir kaynak metnin bir özete indirgeyici bir şekilde dönüştürülmesidir çıkarma veya üretme yoluyla metin (Radev vd., 2004). Bir başka tanıma göre, "Otomatik özet, bir yazılım tarafından oluşturulan, tutarlı ve anlaşılır bir metindir.kaynak metinden önemli miktarda ilgili bilgi içerir. Sıkıştırma oranı τ orijinal belgenin uzunluğunun üçte birinden daha azdır (Hovy& Lin, 1996). Özetin uzunluğu ile kaynak belgenin uzunluğu arasındaki oran, aşağıda gösterildiği gibi sıkıştırma oranı τ ile hesaplanır:

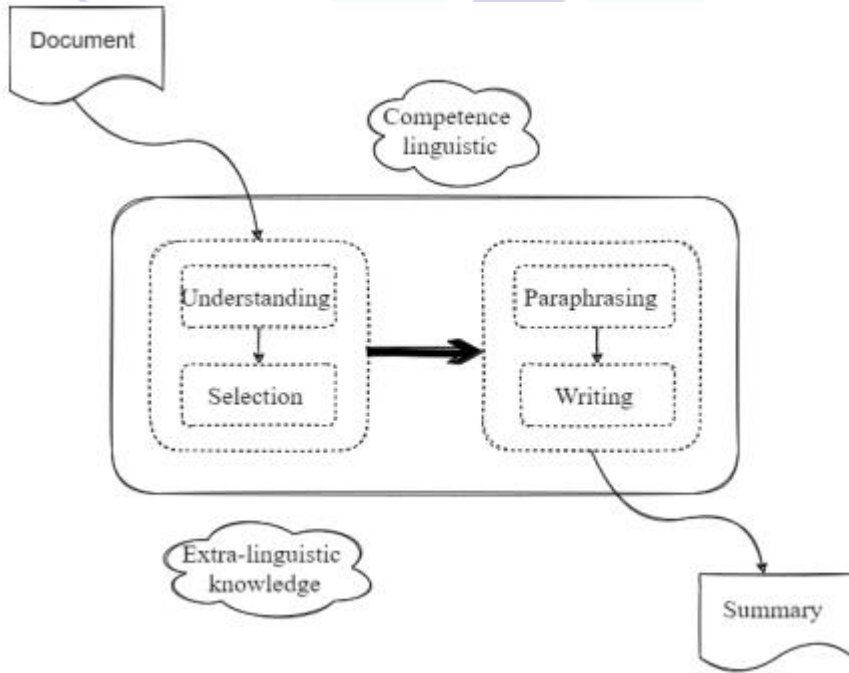
$$\tau = \frac{|summary|}{|source|}$$

Burada $| - |$ belgenin karakter, kelime veya cümle cinsinden uzunluğunu gösterir. τ yüzde olarak ifade edilebilir. Aslında, (C. Y. Lin, 1999) çalışması şunu göstermektedir. Otomatik özetleme sistemlerinin en iyi performansları, kaynak belgenin uzunluğunun $\tau = \%15$ ila $\%30$ 'u kadar bir sıkıştırma oranıyla elde edilmektedir.



Şekil 4: Girdi belgesinden özet oluşturma

Kaynak metnin anlaşılması ve kısa ve kısaltılmış bir versiyonunun oluşturulması, özetlerin insan eliyle oluşturulmasında iki süreçtir. Şekil 2 insan özetinin nasıl orijinal bir metin belgesinin özetlerini üretir. Özetleyicinin dilsel ve dil dışı yetenekleri ve bilgisi, hem metni anlamak hem de özetlemek için gereklidir. ve özetler üretmek. Her ne kadar insanlar daha iyi özetler yazabilse de (okunabilirlik, içerik, biçim ve kısalık açısından). Otomatik metin özetleme, manuel özetlemenin yerine geçmek yerine yararlı bir tamamlayıcıdır.



Şekil 5: İnsanın özet üretme süreci

Özetlerin nasıl üretilebileceğine dair yöntemlere dayalı olarak, yani sadece kaynak metinden cümleler alarak veya kaynak metni okuduktan sonra yeni cümleler üreterek veya her ikisinin bir kombinasyonu olarak özetleme üç türe ayrılabilir:

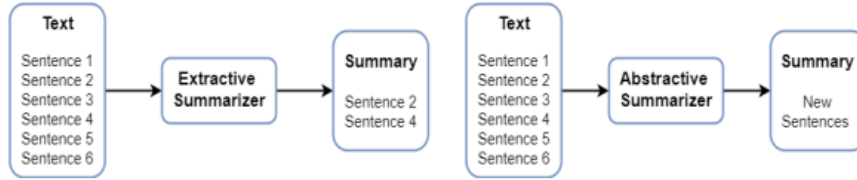
- Ekstraktif Otomatik Metin Özetleme: Ekstraktif metin özetleme, belirli bir derlemde özet çıkarma üzerine birleştirme stratejisidir

(Rau vd., 1989).

- Soyutlayıcı Otomatik Metin Özetleme: Soyutlayıcı metin özetleme, verilen derlemin yeniden yorumlanmasını ve yeni cümleler üretilmesini içerir (Zhang

ve diğerleri, 2019).

- Hibrit Otomatik Metin Özetleme: Hem çıkarıcı hem de soyutlayıcı yöntemleri birleştirir. Bazı cümlelerin çıkarılması ve yeni bir cümle üretilmesi anlamına gelir. Belirli bir derlemde bir tane (Binwahan vd., 2010).



Şekil 6

Bir kitaptaki önemli cümleleri işaret etmek için kullanılan bir fosforlu kalem düşünün. Bu, çıkarımsal metin özetlemeye bir örnektir. Şimdi de hazırladığımız notları düşünün kendi kelimelerimizi kullanarak bir kitaptan. Bu bir soyutlayıcı metin özetleme örneğidir. Çıkarımsal metin özetleme, kitaptan bazı önemli kısımları kopyalayıp yapıştırmak gibidir. kaynak metinden cümleler seçerken, soyutlayıcı metin özetleme bazı anlamlı cümleleri seçer ve önceden seçilmiş cümlelerden yeni cümleler oluşturur. cümleler. Çıkarıcı ve Soyutlayıcı özetlemeyi daha iyi anlamak için Şekil 6'ya bakın. Hibrit metin özetleme, bir metin özeti üretmek için bir yaklaşımı birleştirir. Verimli bir şekilde özetleme. Hem Ekstraktif hem de Soyutlayıcı metin özetleme, Makine Öğrenimi ve NLP alanına girer. Ek olarak, soyutlayıcı metin özetleme NLG'yi kapsamaktadır. Her iki yaklaşımın incelenmesi bu makalenin ilerleyen bölümlerinde gösterilmektedir. Çıkarımsal metin özetlemenin uygulandığı kritik alanlar haber, tıbbi, kitap, yasal belge, soyut metin özetleme, müşteri yorumları, blog, tweet özetleme vb. Çıkarımsal metin özetleme modelinin artı noktası, özetlerdeki cümlelerin sözdizimsel yapının kısıtlamalarına uyması gerektiğidir. Ancak, bu Modelin eksikliği, özetlerdeki cümlelerin anlamsal olarak anlamlı olmayabilmesidir. Bu dezavantaj, özetlerdeki bitişik cümleler nedeniyle ortaya çıkar orijinal metinde her zaman bitişik değildir. Çünkü ATS modelleri kelimeler arasındaki eşdizimi öğrenir ve bu eşdizime dayalı olarak bir anahtar kelime dizisi oluşturur. Eğitimden sonra kelimeler arasında eşdizimlilik, kapsayıcı anlambilim avantajına sahiptirler. ATS modellerinin dezavantajı, aşağıdaki kriterleri karşılamamanın zor olmasıdır. Bu anahtar kelime dizisi ile sözdizimsel yapı. Nadir kelimeler, geleneksel ATS modellerindeki diğer önemli kusurdur. Nadir bir kelimenin ortaya çıkma sayısı ve Eşdizimliliği önemini belirleyecektir, ancak insanlar bir kelimenin gerekli olup olmadığını değerlendirmek için başka unsurları da kullanacaktır. Sonuç olarak, bazı durumlarda, bazı kelimeler içinde görünür.

KAYNAKÇA

1. İLHAN, N., & SAĞALTICI, D. (2020). Twitter'da Duygu Analizi. Harran Üniversitesi Mühendislik Dergisi, 5(2), 146-156. <https://doi.org/10.46578/humder.772929>
2. TOKCAER, S. (2021). Türkçe Metinlerde Duygu Analizi. Yaşar Üniversitesi E-Dergisi, 16(63), 1514-1534. <https://doi.org/10.19168/jyasar.928843>
3. <https://bernardmarr.com/are-alex-and-siri-considered-ai/>
4. <https://arxiv.org/ftp/arxiv/papers/2204/2204.01849.pdf>