Konuşma tanıma

Seni çok iyi anlıyorum, dostum

Bilgisayarlı konuşma tanıma yeni kıyılara yelken açıyor: Teknoloji her kullanıcıya ayak uyduruyor, sistemler içerikleri anlıyor ve diyalog yeteneği kazanıyor.

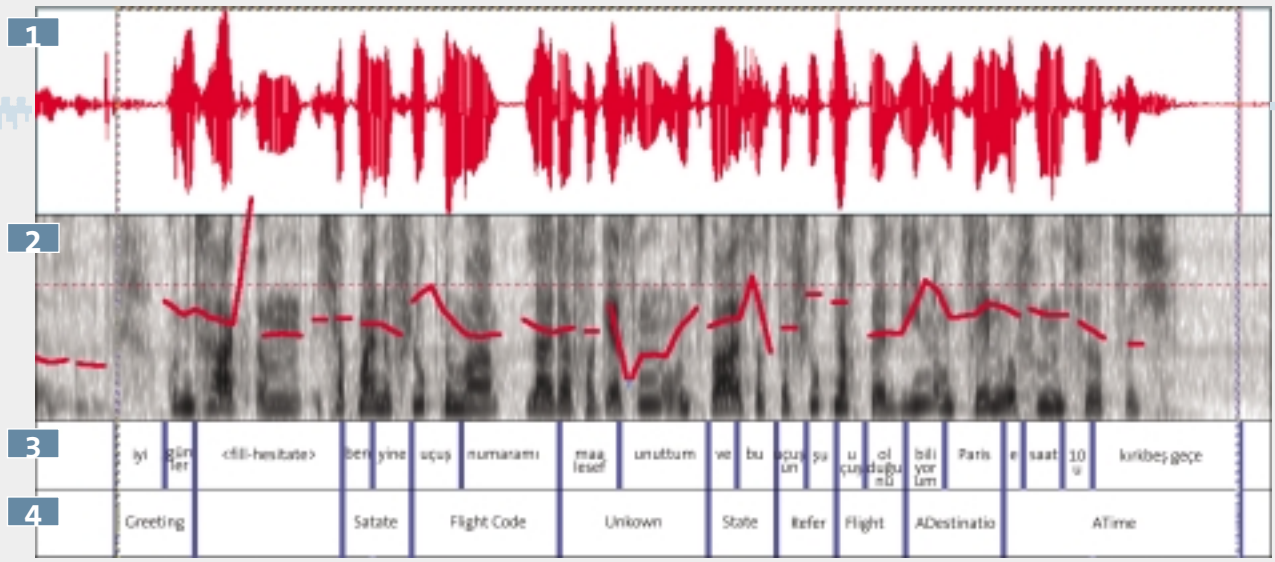
Neredeyse sihirden farksız: Sürüş simulasyon cihazındaki deneğe, yalnızca aracın hangi işlevlerini sesle ya da el işaretleriyle yönetmesi gerektiği söyleniyor. Elektronik ölçüm cihazlarıyla donatılmış vasıtaya binişi sırasında, bu kişiye başka hiçbir talimat verilmiyor. 7 serisi BMW'nin direksiyonunu, denek kendisi kullanmak zorunda. Dev perdedeki sanal manzara üzerinde seyreden yolculuk, deneğin tüm dikkatini toplamasını gerekiyor. Sürücü, istediği her şeyi konuşarak ve el işaretleriyle ifade edebiliyor. "Sesi aç. Başka kanala geç. Bayern üç. Bir sonraki." Radyoyu yönetmek için hangi sözleri kullanırsa kullansın, işler yolunda gidiyor. Doğru kanallar ayarlanıyor, ses seviyesi sürücünün keyfine göre belirleniyor. Klimayı ayarlamak içinse gevşek bir el hareketi yetiyor da artıyor bile. Peki, otomobilin üzerindeki bilgisayar bir gecede kuantum sıçramasıyla mı bu hale geldi, yoksa işin içinde sihirli bir değnek mi var? Gerçekten de Münih Teknik Üniversitesi'nde bu alanda kullanılan yöntemin adı "Oz Büyücüsü". Ancak sihri gerçekleştiren, bir camın ardındaki kontrol odasında oturan ve çok sayıda monitörden araçtaki ve çevresindeki olayları gözlemleyen deney yöneticisi. Arabayı kullanan denek bilgisayarın tüm emirleri yerine getirdiğini zannetse de, aslında onun dileklerini yerine getiren (üstelik de bunu elle yapan) camın gerisindeki bu bilim adamı. Araştırmacılar bu tarz deneyler yürüterek, insanların baştan itibaren yeteneklerinden emin olduğunda bir bilgisayarı nasıl kullanacağını su yüzüne çıkarmak istiyorlar. Bu testler akıllı makineleri geliştirmeyi ve onları insana yaklaştırmayı hedefliyor. İnsan ile makine arasındaki iletişimin çeşitli biçimlerinin araştırıldığı Münih'teki kurum gibi laboratuarlar, konuşma tanıma alanındaki yeni gelişmelerin başını çekiyor. Profesör Gerhard Rigoll, bilgisayar üzerinden konuşma tanımayı daha verimli ve gündelik hayata uygun hale getirmek için güncel yaklaşımlardan birini, "Kelime hazinesini büyük ölçüde kısıtlıyor, buna karşın daha fazla anlam gündeme getiriyoruz", şeklinde açıklıyor. Rigoll, Münih Teknik Üniversitesi'nde insan-makine iletişimi kürsüsünde ordinaryüs.

Kelime hazinesi ne kadar küçükse anlama işlemi de o kadar başarılı

Sloganımız uzmanlaştırma: Öteki dil tanıma programları 100.000 sözcüğü aşan kelime kapasiteleriyle boğuşur ve %100'lük isabet kotasına hiçbir zaman ulaşamazken, özel programlar 1.000 kelimeden daha küçük bir sözcük dağarcığıyla yetiniyor. Bu programlar söylenenin içeriğini anlamayı ve konuşmanın amacını kestirmeyi deniyor. Hedef,

UÇUŞ YARDIM YAZILIMI: BİLGİSAYAR KONUŞMA İÇERİĞİNİ NASIL ANALİZ EDİYOR?

Seslerle sözcüklerin eşleştirilmesi zorlu bir iş. Konuşulan tınılardan, sesbirimlerden, sözcüklerden, tümce parçalarından ya da tam cümlelerden yazılı bir metin oluşturma işlemini, bilim adamları "yazıya dökme" olarak niteliyor. Kullandıkları bir araçla kontrolü ellerinde bulunduruyorlar. Yazılım, konuşulan şeyin doğru olarak tanınıp tanınmadığını gösteriyor. Son adımda ise kararlaştırılan sözcüklere bir anlam atanıyor; bu, insan ile makine arasındaki diyalogun temelini oluşturuyor.



1 Frekans diyagramı: Bu, tanınması gereken sayısallaştırılmış orijinal veriyi gösteriyor. Burada, konuşmayı mikrofon girişi sayesinde sayısal ortama aktaran bir ses kartı kullanılıyor.

2 Tayf: Burada, bir araya gelerek dili oluşturan tüm elemanlar görünüyor. Yazılım, her on milisaniyede bir sinyalin kısa bir kesitini alıp birçok ayırt edici özelliği hesaplıyor ve tınıların tanınmasını sağlayan bir vektör haline sokuyor. Örneğin kırmızı renk, konuşma melodisini karakterize eden temel frekansa denk.

3 Konuşma tanıma: Önce hazırlanmış verilere uygun sesbirimler, yani kısa ses unsurları atanıyor, sonra bunlardan oluşan uygun sözcükler aranıyor.

4 Anlamsal tanıma: Kullanılan sözcükler, cümle yapısı ve bağlam, sisteme konuşulan cümlenin içeriğini çözümleme imkanı tanıyor.

insan ile makine arasında, çok kısıtlı bir konu üzerine de olsa, hakiki bir diyalog. Rigoll, örnek uygulama olarak bir bilgisayarın kullanıcıyla diyaloga geçtiği (yukarıdaki grafiğe bakınız.) otomatik bir uçuş yardım sisteminden söz ediyor. Sistem yalnızca kullanıcının ne dediğinin farkına varmakla kalmıyor, terimleri teker teker anlamlarına göre düzenliyor. İçerik, daha giriş aşamasında anlamsal bakımdan inceleniyor. Ancak bu, yalnızca adı geçen anahtar sözcükler gerçekten uçuş danışmanlığı ile ilgiliyse gerçekleşiyor. Bilgisayar, anlamadığı bir şey olduğunda yanlış yorumlamaktansa, planlı bir biçimde, etraflıca soruyor. Gerhard Rigoll bürosundaki siyah deri koltuğa bir güzel kuruluyor ve ve önemli bir bölümü kendi tarihini de içeren konuşma tanıma tarihi üzerine çene çalmaya başlıyor. 45 yaşındaki bilim adamı, yirmi yıldan fazla bir zamandır konuşma tanıma (speech recognition) konusuyla uğraşıyor. "Başlangıçta ben de günün birinde bir bilgisayarla diyaloga girebileceğimizden kuşkuluydum", diye anlatıyor Rigoll. Rigoll 1986 yılında ABD'ye IBM'in araştırma kısmına gitmiş. O yıllar konuşma tanıma konusunun patlak verdiği ve ilk ticari ürünlerin piyasaya çıktığı yıllarmış. Hidden-Markov denilen modellerin konsepti (yukarıya bkz.) geliştiricilerin eline dili istatistiksel karakteristiklere bakarak tanıyabilen bir araç veriyor. Bu sırada örneğin frekanslar gibi belirli karakteristiklerin üretilme olasılığı ortaya çıkarılıyor. Bu karakteristik özelliklere bakılarak da sözcükler belirleniyor. İstatistik ile dil tanımanın yolu açılmadan önce, sayısallaştırılmış sinyalleri seslere ayırma denenmiş. "Ama dil, duyu organlarımızın derininde bulunan bir süreç. Bu yüzden dili "Eğer – Öyleyse" kuralları haline indirgemek kolay değil," diyor Rigoll. Hidden-Markov modelleri dil tanıma alanında çığır açmış bulunuyor. IBM, bunu temel alan dikte sistemi ViaVoice ile piyasanın önderi haline gelmiş ve diğer firmalar da bu alana yönelmekte hiç gecikmemiş.

Pratikte iletişim sorunları

Ancak coşku geçmişte kalmış. Çoğu işyeri konuşma tanıma sisteminden vazgeçiyor; çünkü hata oranı, rahat bir kullanıma izin vermeyecek kadar yüksek. Bunu, bekleme odasında kırmızı bir mekanik daktilo bulunduran Rigoll da biliyor. Kendi ofisindeki deneyleri çok geçmeden askıya almış. Bilim adamı Rigoll'u şimdi ilgilendiren, algoritmaların "acımasızca hassaslaştırılması ve iyileştirilmesi." Ma- →

BİLGİSAYAR KONUŞMAYI NASIL TANIYOR?

»Veritabanları, algoritmalar ve istatistik

Sürekli konuşmada sözcükler çoğu zaman aralarında boşluklar olmadan, peş peşe sıralanır. Bir bilgisayar için bu ses örneğini sözcüklere ayrıştırmak külfetli bir iş.

1. Basamak: Sesbirimlerini (Fonemleri) tanımak. Tam bir sözcük birçok sesbirimden ibaret (Sesbirim: anlam ifade eden, ancak kendisi anlam taşımayan en küçük dilsel birim). Normal konuşma hızında, bir sesbirim 10 ila 40 milisaniye uzunluğunda. Konuşma tanıma işleminde, yaklaşık 10 milisaniyelik aralıklarla sesin kısa süreli tayfları oluşturuluyor. Sistem buradan ayırt edici değerleri tek tek hesaplıyor ve bunları bir karakteristik vektöründe bir araya getiriyor. Karakteristik vektörlerinin zamansal dizilişi, hangi sözcüğün konuşulmuş olduğunu saptamayı mümkün kılıyor. Bunun için, karakteristik vektörleri depolanmış referans örnekleriyle karşılaştırılıyor.

Usability-Lab: Programla eğitim insanla bilgisayarın daha iyi anlaşabilmesine yardımcı oluyor.

2. Basamak: Hidden-Markov modelleri. Bu karşılaştırmaları optimal tanımada olabildiğince hızlı gerçekleştirebilmek için Markov zincirleri denilen yapıları temel alan istatistiksel bir işlemden yararlanılıyor. Bunlar bir sesbirimden diğerine geçiş olasılıklarını belirten zincirler. Bir alıştırma aşamasından sonra, bilinmeyen bir örneği tanımada, modelin bu akışı üretebilmesi olasılığı hesaplanıyor. Bu hesaplama tekrar tekrar yapılıyor. Bu, yüksek bir hesaplama gücünü şart kılıyor.

3. Basamak: Bi- ve Trigram. Bir konuşma tanıma programının daha da yüksek bir tanıma doğruluğuna erişebilmesi için, Hidden-Markov modellerinin yanı sıra başka bir istatistik yöntemi daha mevcut. Dikte sırasında sürekli hesaplamalarını yürüten Bi- ya da Trigram istatistiği yoluyla bir bağlam sınaması gerçekleştiriliyor. Böylece sistem gitgide daha fazla konuşmacıya ve konuşmacıların bireysel dil tarzına uyum sağlıyor.

kine ile iletişimde insanı her şeyin ölçüsü yapmak ne derece değerliyse, sistemlerden daha fazla tanıma performansı elde etmek için insan o derece gereksizdir. "Bizde bugün hiçbir insan artık bilgisayar ile konuşmuyor", diye açıklıyor Rigoll. Araştırmacıların bilgisayara bir metin okuması ve akabinde yazılımın ne kadarını hatasızca tanıyabildiğini sınadığı zamanlar geride kalmış bulunuyor. Artık en küçük ilerlemeleri de görünür kılan yeniden üretilebilir test koşullarını standartlaştırılmış veritabanları sağlıyor. İngilizce uluslararası standart olsa da, algoritmaların hassaslaştırılmasında dilin hiçbir rolü yok. İşin tamamı, büyük ağırlığını programlamanın oluşturduğu, son derecede zahmetli bir prosedür. Araştırmacılar, bir şey elde edip edemediklerini bazen saatler, aşırı durumlarda ise haftalar süren hesaplamalar sonucunda görmek zorunda kalıyorlar. "Tüm istatistik yöntemlerinin altında yatan soru şu: Belirli karakteristik noktalar gözlemlendiğinde, hangi sözcükler hangi olasılıkla konuşulmuş oluyor?" diye özetliyor Profesör Günther Ruske. Münih Teknik Üniversitesi'nde araştırmalar yürüten Ruske, zorunlu soyutlamaların hakkından geliyor.

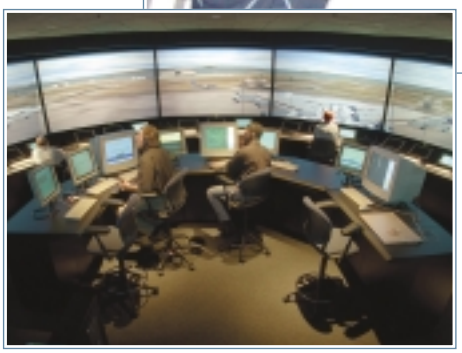
> »Günümüzde algoritmaların acımasızca özelleştirilmesi ve optimize edilmesi ön planda
>
> Profesör Gerhard Rigoll, Münih Teknik Üniversitesi

Eğrilerin incelenmesi hedefe götürüyor

Dilde örneği tanıma, tıpkı el yazısı, harf ya da yüz tanıma sistemlerinde olduğu gibi işliyor. Birçok görüş arasından ayırt edici olanlar süzülüyor. Ruske, verimli algoritmaların önemini, "Ne de olsa bir insanın bir evin ev olduğunu anlaması için dünyadaki bütün evleri görmüş olması gerekmez," diyerek açıklıyor. Otomatik konuşma tanıma alanında yalnızca insanlara bakmak çok az şey kazandırmış. Bu iş için çizilmiş eğriler daha fazla şey ifade ediyor. Ruske gerçekler üzerinde duruyor: "Bence bu ilkel bir yaklaşım, ama sonuçlar çok iyi olduğu için kullanıyoruz." Bilim adamları tayfsal maksimumlara bakarak sesli harfleri kolayca tanıyabiliyorlar. Örneğin "A" sesi spektrumda hep çok önde yer alıyor ve böylece bir nirengi noktası görevini üstleniyor. Bu rezonansın yerinin belirli bir insanda tam olarak nerede bulunduğunu ise, Ruske'nin ses üretim yoluna taktığı isimle, pasif "boru" tayin ediyor. Bu yer ses üretim yolunun uzunluğuna bağımlı ve ağız ve dil hareketleriyle kaydırılıyor. Logaritmik kaydırma vasıtasıyla bu sesli harf uzunlukları sistemde standartlaştırılıyor ve böylece yazılımların konuşmacıyla uyumu sağlanıyor. "Saniyeler içinde bir konuşmacıya adapte olan bir sistem bir düş olurdu," diyerek güncel sınırları gösteriyor Ruske. Kelime hazinesi kısıtlaması olmadığında ve sistemin bir konuşmacıya uyumu sağlanmadığında tanımanın ne kadar zor olduğunu şimdiden kullanılan dil tanımanın üst uç uygulamaları gösteriyor. Makine için örneğin bir Amerikan

Sürüş Simülatörü: **Münih Teknik Üniversitesi'ndeki bir usability Lab'da otomobilin kaportası altında bir motor yerine simülasyonu yöneten eski bir Pentium bilgisayarı çalışıyor (solda üstte). Kontrol odasında (solda) test sürücüsünün sesli komutları henüz manuel olarak giriliyor. Bu sürücü ve araç arasındaki otomatik bir diyalog için ön çalışma.**

uçak gemisi ile Almanya'daki bir konferans salonu arasındaki farklar ortadan kalkıyor. Her iki durumda da bir sistem saatlerce tüm iletişimi kaydediyor. Bu, ilke itibarıyla büyük bir diktafondan farksız, ancak iş içeriklerin açıklığa kavuşturulmasına geldiğinde aksaklıklar çıkıyor. Sistem olay yerinde bulunmamış olan kişilere de bilgiler veriyor. Toplantı tutanaklarının yazıya aktarılması, bir toplantıya katılmadan da bilgilenmek isteyen yöneticiler için en son moda. Mobil aygıtlar için dil tanıma çok sakıncalı bir konu. Gelecekte internette sörf yapacağımız UMTS cep tele- →



GEVEZE: İster konferans salonu, ister apron olsun, yazılım her şeyi kaydediyor. Ancak uzmanlaşma eksikliği yüzünden içerikle ilgili sorunlar var.

fonları da eninde sonunda küçük klavye ile yetinemeyecek. Gerhard Rigoll, "Bu yeni bir kullanıcı arabirimi talep ediyor ve ben mobil alanda da konuşma tanımanın sağlam bir yer edineceğinden eminim," diye iyimserliğini belli ediyor. Ancak burada talepler çok yüksek, çünkü yalnızca farklı kullanıcılar değil, kötü ton kalitesi ve rahatsız edici parazitler de hesaba katılmak zorunda. Konuşma tanıma daha fazla iyileştirilebilir ve daha sağlam yapılabilirse de, bir cep telefonu bunun için gerekli hesaplama performansını sağlamakta zorlanacak. Dağıtılmış tanıma ise bir çıkış yolu olabilir. Cep telefonu, algılanan konuşmanın yalnızca en önemli karakteristiklerini hesaplar ve ilgili verileri ser-

» Gerçekten önemli olanı tanımaktan henüz çok uzağız.

Profesör Günther Ruske, Münih Teknik Üniversitesi

vis sağlayıcıdaki bir sunucuya gönderir. Orada büyük bir bilgisayar konuşma tanımaya yönelik asıl hesaplamaları yürütür. Siemens, Münih'te daha şimdiden cep telefonu tarafında ilgili performansı sağlayacak bir yonga üzerinde çalışıyor. Bunun kullanılıp kullanılmayacağı ve şayet kullanılacaksa bunun ne zaman gerçekleşeceği henüz belli değil, çünkü henüz hiçbir cep telefonunun büyük bir bilgisayarın yardımına ihtiyacı yok. ■

MF / Garo Antikacıoğlu, agaro@chip.com.tr

LİNKLER

www.mmk.ei.tum.de/forschung/
http://verbmobil.dfki.de/Vm.Info.Phase2.html

SÖZCÜĞÜ SÖZCÜĞÜNE

»Konuşma tanımanın yarım asırlık tarihi

1952 Bell Laboratuarları telefonda konuşulan 0 – 9 arası rakamları tanıyan bir sistem sunuyor.

1959 MIT'nin ürettiği sistem sesli harflerin yüzde 93'ünü tanıyor. Yedi yıl sonra bu oran 50 sözcüğe ulaşıyor.

1962 Konuşma yeteneğine sahip ilk aygıt piyasaya çıkıyor. IBM 7772'nin sesi içi boş bir tenekeninkinden biraz daha iyice.

1968 Bilimkurgu yine yıllarca önden gidiyor. "2001 – Uzay Macerası" filminde, bilgisayar HAL astronotlarla konuşuyor.

1976 Bruce Lowerre komple cümleleri ve basit dilbilgisi yapılarını tanıyan Harpy sistemini geliştiriyor. Bunun için 50 bilgisayar yoluyla paralel işlem gerekiyor.

1977 İnşaat bankası Wüstenrot, "konuşabilen" bir sistemin Almanya'daki ilk ticari müşterisi oluyor.

1978 Texas Instruments bir dil işlemcisini bir yonga üzerine yerleştiriyor.

1986 IBM Tangora 4 gerçek zamanlı istatistiksel yapıları tanıyor.

1988 Dragon, PC için ilk konuşma tanıma yazılımını üretiyor.

1996 OS/2 Warp sesle yönetim özelliği bulunan ilk işletim sistemi.

1997 Gitgide daha fazla program piyasaya sürülüyor. Temmuz ayında 23.000 sözcük tanıyan Dragon NaturalySpeeking, Ağustos ayında ise IBM'in Via Voice'u piyasaya çıkıyor. Onu hemen ardından Philips ve Lernout & Hauspie takip ediyor.

2000 Wolfgang Wahlster dil dönüştürücü "Verbmobil"i çıkarıyor.