



پردازش و شناسایی گفتار

پردازش و شناسایی گفتار

نوشین ریاحی

عضو هیئت علمی دانشگاه الزهراء (س)



انتشارات دانشگاه الزهراء (س)



پردازش و شناسایی گفتار

نوشین ریاحی

عضوهیئت علمی دانشگاه الزهراء (س)

ناشر: انتشارات دانشگاه الزهراء (س)

طراحی صفحات:

کارگاه گرافیک فرگاهی

نوبت چاپ: یکم، ۱۳۹۶

شمارگان: ۱۰۰۰ نسخه

شابک: ۹۷۸-۶۲۲-۶۱۱۴-۰۱-۱

قیمت: ۴۵۰۰۰ تومان

کلیه حقوق برای دانشگاه الزهراء (س) محفوظ است.

سرشناسه: ریاحی، نوشین، ۱۳۴۴ -
عنوان و نام پدیدآور: پردازش و شناسایی گفتار/ نوشین ریاحی.
مشخصات نشر: تهران: دانشگاه الزهراء (س)، انتشارات، ۱۳۹۷.
مشخصات ظاهری: ۴۷۲ ص:، مصور، جدول، نمودار.
شابک: ۹۷۸-۶۲۲-۶۱۱۴-۰۱-۱
وضعیت فهرست نویسی: یا یادداشت: واژه نامه.
یادداشت: کتابنامه، ص: ۴۵۸. موضوع: گفتارپردازی
موضوع: Speech processing systems موضوع: گفتار
موضوع: Speech شناسه افزوده: دانشگاه الزهراء
شده سه افزوده: Alzahra University
رده بندی کنگره: ۹۱۳۹۷/TK۷۸۸۲
رده بندی دیویی: ۵۴/۵۶
شماره کتابشناسی ملی: ۵۲۲۱۷۵۳

مراکز پخش:

ونک، دانشگاه الزهراء (س) تلفن: ۸۵۶۹۲۷۶۹-۸۸۰۴۸۹۳۳

بلوار کشاورز، نیش ۱۶ آذر، انتشارات دانشگاه الزهراء (س)

وبگاه فروش اینترنتی: <http://research.alzahra.ac.ir>

فهرست مطالب

۱۷	پیش‌گفتار مؤلف
۲۱	فصل اول. مقدمه‌ای بر بازشناسی گفتار
۲۳	مقدمه
۲۳	۱-۱. بازشناسی گفتار چیست؟
۲۴	۲-۱. مشکلات بازشناسی گفتار
۲۵	۳-۱. پیچیدگی‌های مطرح در سیستم‌های بازشناسی گفتار
۲۸	۴-۱. روش‌های بازشناسی گفتار
۳۰	۵-۱. کاربردهای سیستم‌های بازشناسی خودکار گفتار
۳۲	۶-۱. جمع‌بندی
۳۲	تمرین‌ها
۳۴	منابع
۳۵	فصل دوم. سیگنال گفتار: تولید، دریافت
۳۷	مقدمه
۳۷	۱-۲. فرایند تولید و دریافت گفتار در انسان
۳۹	۲-۱-۱. فرایند تولید گفتار
۴۲	۲-۱-۲. محاسبه فرکانس تشدید برای مجرای صوتی

۴۴	۳-۱-۲. اندام‌های مؤثر در تولید گفتار
۴۸	۲-۲. دریافت سیگنال صدا
۴۸	۱-۲-۲. مدل فیزیولوژیکی گوش انسان
۵۰	۲-۲-۲. نحوه عملکرد گوش داخلی و میانی انسان
۵۲	۳-۲. جمع‌بندی
۵۳	تمرین‌ها
۵۴	مراجع

۵۵ فصل سوم. توصیف صوتی واج‌ها و مشخصات آن‌ها

۵۷	مقدمه
۵۷	۱-۳. واج‌شناسی و آواشناسی
۵۸	۲-۳. نمایش گفتار در حوزه زمان و فرکانس
۶۱	۳-۳. سازوکار صوتی و تقسیم‌بندی واج‌ها
۶۵	۱-۳-۳. مصوت‌ها
۷۱	۲-۳-۳. حروف نیمه مصوت
۷۳	۳-۳-۳. صامت‌ها
۷۴	۱-۳-۳-۳. صامت‌های خیشومی
۷۵	۲-۳-۳-۳. آواهای سایشی بدون صدا و صدادار
۷۷	۳-۳-۳-۳. انسدادی‌های صدادار و بی صدا
۷۸	۴-۳-۳-۳. نجوا
۷۹	۵-۳-۳-۳. انفجاری-سایشی
۷۹	۴-۳. واج‌شناسی و آواشناسی فارسی
۷۹	۱-۴-۳. واج‌های واکدار در فارسی
۸۱	۱-۴-۳. واج‌های همخوان در فارسی
۸۲	۵-۳. جمع‌بندی

۸۳	تمرین‌ها
۸۵	منابع
۸۷	فصل چهارم. پردازش گفتار
۸۹	مقدمه
۹۰	۴-۱. مرحله پردازش گفتار
۹۰	۴-۱-۱. پیش‌تأکیدکردن
۹۱	۴-۱-۲. قاب‌بندی
۹۳	۴-۱-۳. پنجره‌گذاری
۹۷	۴-۱-۴. محاسبه بردار ویژگی‌ها
۹۸	۴-۱-۵. محاسبه مشتقات بردار ویژگی‌ها
۹۸	۴-۲. آشکارسازی گفتار (آشکارسازی نقاط پایانی)
۱۰۰	۴-۲-۱. مشکلات موجود در آشکارسازی دقیق نقاط پایانی
۱۰۱	۴-۲-۲. آشکارسازی گفتار در سیستم بازشناسی گفتار ساده
۱۰۳	۴-۳. سیستم تشخیص گفتار مبتنی بر قانون
۱۰۳	۴-۳-۱. رویکرد صوتی-آواشناختی
۱۰۷	۴-۳-۲. طبقه‌بندی اصوات صدا دار
۱۰۹	۴-۳-۳. طبقه‌بندی‌کننده تمامی اصوات
۱۱۰	۴-۳-۴. مثال‌هایی از برچسب‌گذاری صوتی-آواشناختی
۱۱۲	۴-۳-۵. مسائل موجود در رویکرد صوتی-آواشناختی
۱۱۳	۴-۴. جمع‌بندی
۱۱۴	تمرین‌ها
۱۱۶	منابع

۱۱۷ فصل پنجم. تعیین ویژگی‌های سیگنال گفتار

۱۱۹	مقدمه
۱۲۰	۱-۵. استخراج ویژگی‌های گفتار در حوزه زمان
۱۲۰	۱-۱-۵. دامنه کوتاه مدت
۱۲۰	۱-۱-۵. انرژی
۱۲۲	۱-۱-۵. نرخ عبور از صفر (ZCR)
۱۲۵	۱-۱-۵. خودهمبستگی
۱۲۸	۱-۱-۵. تابع میانگین تفاضل دامنه‌ها
۱۲۹	۲-۵. استخراج ویژگی‌های گفتار در حوزه فرکانس
۱۲۹	۱-۲-۵. تبدیل فوریه
۱۳۲	۲-۲-۵. تحلیل پیشگویی خطی
۱۴۲	۳-۲-۵. تحلیل کپسترال و محاسبه مشتق زمانی ضرایب کپسترال
۱۵۰	۴-۲-۵. ضرایب کپسترال مبتنی بر معیار مل (MFCC)
۱۵۲	۵-۲-۵. ضرایب پیشگویی خطی ادراکی (PLP)
۱۵۵	۳-۵. تخمین گام
۱۵۷	۴-۵. کوانتیزاسیون برداری
۱۶۲	۵-۵. جمع بندی
۱۶۲	تمرین‌ها
۱۶۶	منابع

۱۶۷ فصل ششم. تعیین تابع فاصله

۱۶۹	مقدمه
۱۶۹	۱-۶. تعیین فاصله میان الگوها
۱۷۰	۲-۶. توابع محاسبه فاصله (تفاوت)
۱۷۱	۱-۲-۶. چگونگی ادراک انسان از اختلاف میان صداها
۱۷۴	۲-۲-۶. طیف سیگنال گفتار

۱۷۶	۳-۲-۶. توابع فاصله
۱۷۶	۱-۳-۲-۶. فاصله طیفی لگاریتمی
۱۸۰	۲-۳-۲-۶. فاصله های کپسترال
۱۸۰	۳-۳-۲-۶. فاصله های کپسترال دارای وزن و بالابرد
۱۸۳	۴-۲-۶. اعوجاج شباهت ایتاکورا-سایتو
۱۸۴	۵-۲-۶. اعوجاج نسبت شباهت
۱۸۵	۶-۲-۶. گونه های مختلف اعوجاج های شباهت
۱۸۹	۷-۲-۶. اعوجاج طیفی با استفاده از اندازه پیچش یافته فرکانسی
۱۹۷	۳-۶. افزودن مشخصه های دینامیکی طیفی به اندازه اعوجاج
۱۹۸	۴-۶. جمع بندی
۱۹۸	تمرین ها
۲۰۰	منابع

فصل هفتم. انتخاب طبقه بند

۲۰۱	مقدمه
۲۰۳	۰۱-۷. روش پیچش زمانی پویا (DTW)
۲۰۳	۰۱-۷. پیدا کردن بهترین مسیر تنظیم بین یک جفت الگوبا روش DTW
۲۰۸	۰۲-۷. پیاده سازی روش پیچش زمانی پویا
۲۲۱	۰۲-۷. مدل های مخفی مارکوف
۲۲۳	۰۱-۲-۷. مدل مارکوف
۲۲۴	۰۲-۲-۷. مدل مخفی مارکوف
۲۲۶	۰۳-۷. شبکه های عصبی مصنوعی
۲۳۲	۰۱-۳-۷. مدل پرسپترون چند لایه
۲۳۳	۰۲-۳-۷. یادگیری در شبکه عصبی مصنوعی
۲۳۵	۰۱-۲-۳-۷. الگوریتم پس انتشار خطا
۲۳۶	

۲۴۰	۴-۷. جمع بندی
۲۴۱	تمرین ها
۲۴۵	منابع

۲۴۷ فصل هشتم. بازشناسی گفتار با کلمات گسسته

۲۴۹	مقدمه
۲۵۰	۱-۸. مدل مخفی مارکوف
۲۵۴	۱-۱-۸. حل مسئله ۱ و متدهای پیشرو و پسرو
۲۵۸	۲-۱-۸. حل مسئله ۲ و متد Viterbi
۲۶۳	۳-۱-۸. حل مسئله ۳ و متدهای BaumWelch و segmentalk-means
۲۷۰	۲-۸. ملاحظات عملی در استفاده از HMM ها
۲۷۰	۱-۲-۸. برآوردهای اولیه
۲۷۲	۲-۲-۸. توپولوژی مدل
۲۷۶	۳-۲-۸. ضوابط آموزش: یکنواخت کردن پارامترها
۲۷۷	۴-۲-۸. چگونه مشکل سکوت بین کلمات را حل کنیم
۲۷۸	۳-۸. سیستم بازشناسی گفتار با کلمات گسسته
۲۷۹	۴-۸. جمع بندی
۲۸۰	تمرین ها
۲۸۴	منابع

۲۸۵ فصل نهم. بازشناسی گفتار با کلمات پیوسته

۲۸۷	مقدمه
۲۸۸	۱-۹. ساختار یک سیستم بازشناسی گفتار با کلمات پیوسته
۲۹۱	۲-۹. ایجاد مدل آوایی
۲۹۱	۱-۲-۹. آموزش مدل CDHMM برای واج ها

۲۹۳	۲-۲-۹. الگوریتم Embedded Baum-Welch
۲۹۶	۳-۹. تعیین واحد آوایی برای بازشناسی گفتار با کلمات پیوسته-اثر بافت صحبت
۲۹۷	۱-۳-۹. روش های مختلف به کارگیری اطلاعات وابسته به بافت در بازشناسی گفتار با کلمات پیوسته
۲۹۸	۱-۱-۳-۹. روش های فونتیک به کارگیری اطلاعات وابسته به بافت
۳۰۸	۲-۱-۳-۹. به کاربردن مدل قطع شده برای سه واجی ها
۳۰۹	۴-۹. رمزگشای آکوستیکی و رمزگشای لغوی
۳۱۵	۵-۹. جمع بندی
۳۱۶	تمرین ها
۳۱۷	منابع

۳۱۹ فصل دهم. روش های جست وجود در سیستم تشخیص گفتار با کلمات پیوسته

۳۲۱	مقدمه
۳۲۲	۱-۱۰. روش های جست وجود
۳۲۵	۱-۱-۱۰. روش جست وجوی اول عمق
۳۲۶	۲-۱-۱۰. روش جست وجوی اول سطح
۳۲۷	۳-۱-۱۰. روش جست وجوی اول بهترین یا A^*
۳۲۸	۴-۱-۱۰. روش جست وجوی شعاعی
۳۳۰	۲-۱۰. ارزیابی روش های جست وجود
۳۳۱	۳-۱۰. الگوریتم های جست وجوی قابل استفاده در بازشناسی گفتار
۳۳۳	۱-۳-۱۰. ترکیب احتمالات مدل های زبانی و آکوستیکی
۳۳۸	۲-۳-۱۰. جست وجوی ویتربی همگام با زمان شعاعی
۳۴۲	۳-۳-۱۰. دیکدینگ پشته ای (جست وجوی A^*)
۳۴۶	۴-۱۰. جمع بندی

۳۴۶	تمرین‌ها
۳۴۹	منابع

۳۵۱ فصل یازدهم. یادگیری عمیق

۳۵۳	مقدمه
۳۵۳	۱-۱۱. شبکه‌های عصبی کم‌عمق و عمیق
۳۵۵	۲-۱۱. شبکه‌های عصبی عمیق، توانایی استخراج ویژگی‌ها
۳۵۶	۳-۱۱. انواع شبکه‌های عصبی عمیق
۳۵۹	۴-۱۱. آموزش شبکه‌های باور عمیق، دید کلی
۳۶۰	۵-۱۱. معرفی مدل RBM
۳۶۲	۶-۱۱. آموزش مدل‌های RBM با استفاده از روش بیشینه کردن لگاریتم درست‌نمایی
۳۶۳	۷-۱۱. واگرایی متقابل (CD)
۳۶۶	۸-۱۱. مراحل آموزش شبکه‌های باور عمیق
۳۶۸	۹-۱۱. تعداد متغیرهای هر لایه در شبکه‌های عمیق
۳۶۹	۱۰-۱۱. جمع‌بندی
۳۶۹	تمرین‌ها
۳۷۰	منابع

۳۷۱ فصل دوازدهم. دادگان گفتاری

۳۷۳	مقدمه
۳۷۳	۱-۱۲. دادگان گفتاری فارسی دات
۳۷۳	۲-۱۲. دادگان گفتاری فارسی دات بزرگ
۳۷۴	۳-۱۲. دادگان گفتاری فارسی دات تلفنی (مونولوگ)
۳۷۴	۴-۱۲. دادگان گفتاری فارسی دات تلفنی بزرگ (محاوره‌ای)

۳۷۵	۵-۱۲. دادگان اعداد و ارقام منفصل و پیوسته فارسی
۳۷۵	۶-۱۲. دادگان گفتاری TIMIT
۳۷۶	۷-۱۲. دادگان گفتاری TIDIGITS
۳۷۶	۸-۱۲. دادگان گفتاری AURORA ۲
۳۷۸	۹-۱۲. دادگان گفتاری SwitchBoard
۳۷۹	۱۰-۱۲. دادگان گفتاری CHiME
۳۷۹	۱۱-۱۲. جمع بندی
۳۸۰	تمرین ها
۳۸۱	منابع

۳۸۳ فصل سیزدهم. ارزیابی سیستم های بازشناسی گفتار و معیار اطمینان

۳۸۵	مقدمه
۳۸۵	۱-۱۳. ارزیابی سیستم های بازشناسی گفتار
۳۸۶	۲-۱۳. معیار اطمینان یا CM
۳۸۸	۱-۲-۱۳. معیار اطمینان با استفاده از ترکیبی از مجموعه ویژگی ها
۳۹۱	۲-۲-۱۳. معیار اطمینان به عنوان احتمال پسین
۳۹۷	۳-۲-۱۳. معیار اطمینان براساس ارزیابی گفتار
۳۹۹	۴-۲-۱۳. محاسبه تخمین اطمینان به روش هم ترازوی خروجی تشخیص دهنده واج و تشخیص دهنده کلمه
۴۰۰	۵-۲-۱۳. محاسبه تخمین اطمینان به صورت تشکیل لیست N تا بهترین
۴۰۱	۳-۱۳. کارایی، قابلیت و محدودیت های CM
۴۰۳	۴-۱۳. جمع بندی
۴۰۳	تمرین ها
۴۰۵	منابع

۴۰۷	فصل چهاردهم. کاهش نویز
۴۰۹	مقدمه
۴۰۹	۱-۱۴. اثر نویز و اغتشاشات محیط بر سیگنال گفتار و کارایی سیستم‌های تشخیص گفتار
۴۱۰	۲-۱۴. انواع نویز
۴۱۳	۳-۱۴. کاربردهای کاهش نویز
۴۱۴	۴-۱۴. روش‌های کاهش نویز و مقاوم‌سازی سیستم‌ها
۴۱۶	۱-۴-۱۴. ویژگی‌های مقاوم به نویز
۴۱۷	۲-۴-۱۴. بهسازی گفتار
۴۴۸	۳-۴-۱۴. روش انطباق نویز (جبران کردن مدل)
۴۵۰	۵-۱۴. ارزیابی کیفیت گفتار
۴۵۱	۱-۵-۱۴. روش‌های ارزیابی کیفی
۴۵۳	۲-۵-۱۴. روش‌های ارزیابی کمی
۴۵۵	۶-۱۴. جمع‌بندی
۴۵۵	تمرین‌ها
۴۵۸	منابع
۴۵۹	فهرست موضوعی

پیش‌گفتار مؤلف

گفتار، یکی از مهم‌ترین و پایه‌ای‌ترین وسایل ارتباطی در جوامع انسانی است. امروزه بشر به دنبال آن است که بتواند با ماشین صحبت کند و به راحتی با آن ارتباط برقرار سازد. هدف از طراحی و تولید سیستم‌های تشخیص گفتار، تبدیل سیگنال گفتار به متن یا به مجموعه دستوره‌های قابل فهم توسط ماشین، در چارچوب دایره کلمات مشخص و نرخ بازشناسی قابل قبول، روی دادگان تعریف شده است.

این کتاب حاصل هفت سال تدریس درس پردازش و شناسایی گفتار برای دانشجویان کارشناسی ارشد هوش مصنوعی است. در این درس، دانشجویان با حوزه وسیعی از مطالب درمورد مباحث صوت‌شناسی و آواشناسی، بازشناسی گفتار، سنتزگفتار، درک گفتار، پردازش زبان طبیعی، شناسایی آماری الگو، پیکره‌های متنی و زبانی، اثرنویز در سیگنال گفتار و نحوه کاهش اثر آن و ارزیابی این سیستم‌ها آشنا می‌شوند. در آغاز نوشتن کتاب دو موضوع باید مشخص می‌شد. اول آنکه کتاب با چه وسعتی مطالب مطرح شده را پوشش دهد و دوم اینکه در هر موضوعی تا چه عمقی پیش رود و چقدر وارد جزئیات شود. به دلیل تمرکز اصلی این درس بر مبحث شناسایی گفتار و لزوم حفظ پیوستگی و جامعیت مطالب، عنوان کتاب پردازش و شناسایی گفتار انتخاب شد و مطالب لازم و مرتبط با این موضوع در کتاب پوشش داده شد؛ بنابراین، این کتاب به عنوان مرجعی برای درس پردازش و شناسایی گفتار پیشنهاد می‌شود. با وجود این، در مقوله‌هایی مانند سنتزگفتار و درک گفتار، نیاز به مراجع دیگر وجود دارد. از مزایای مهم کتاب، بهره‌مندی از نمونه‌های کاربردی و وجود تمرین‌های مربوط به هر فصل است. در ضمن، طی مثال‌ها قطعه‌کدهای متلب نیز برای پردازش‌هایی روی سیگنال ورودی آمده است.

ساختار کتاب به این ترتیب است: در فصل اول با عنوان مقدمه‌ای بر بازشناسی گفتار، عملکرد سیستم‌های بازشناسی گفتار و روش‌های موجود، مشکلات و پیچیدگی‌ها و کاربردهای این سیستم‌ها مرور می‌شود. در فصل دوم، نحوه تولید و دریافت سیگنال گفتار، اندام‌های مؤثر در تولید صدا و بخش‌های مختلف گوش و سازوکار شنیدن بررسی می‌شود. فصل سوم به توصیف صوتی واج‌ها، انواع و مشخصات آن‌ها اختصاص دارد. سیستم بازشناسی گفتار، در حالت کلی از دو بخش پردازش گفتار و رمزگشایی گفتار تشکیل شده است. در فصل چهارم، مرحله پردازش سیگنال شامل پیش‌تأکید کردن^۱، قاب‌بندی^۲، اعمال پنجره^۳، محاسبه ویژگی‌های مورد نظر، وزن‌دهی ویژگی‌های^۴ محاسبه شده همراه با مسئله آشکارسازی نقطه پایانی در سیگنال گفتار توضیح داده می‌شود. در ضمن، سیستم‌های تشخیص واج مبتنی بر قانون بررسی می‌شوند. در این سیستم‌ها، اگر گفتار ورودی در مجموعه‌ای از قواعد و قوانین جای گیرد و با آن سازگار شود، برنامه می‌تواند آن را تشخیص دهد. در فصل پنجم، ویژگی‌های مناسب برای پردازش گفتار مطرح می‌شوند و در مورد نحوه محاسبه آن‌ها توضیح داده می‌شود. در ادامه در فصل ششم، مشخصات تابع فاصله مناسب برای سیگنال‌های گفتار مطرح می‌شود و انواع تابع فاصله مورد استفاده در این زمینه معرفی می‌شوند.

یکی از قسمت‌های مهم در سیستم‌های بازشناسی گفتار، بخش مدل‌کردن و طبقه‌بندی است. در فصل هفتم با عنوان انتخاب طبقه‌بند، روش‌های پیچش زمانی پویا، شبکه‌های عصبی و مدل مخفی مارکوف - که بیشتر در سیستم‌های بازشناسی گفتار به عنوان طبقه‌بند به کار می‌روند - مطرح می‌شوند. فصل هشتم به نحوه پیاده‌سازی مدل مخفی مارکوف در یک سیستم بازشناسی گفتار با کلمات گسسته اختصاص دارد و در فصل نهم، در زمینه ساختار یک سیستم بازشناسی گفتار با کلمات پیوسته، مشکل کمبود داده‌های آموزشی، رمزگشای اکوستیکی و رمزگشای لغوی، مدل‌سازی زبانی و واژگان توضیح داده می‌شود.

-
1. Preemphasis
 2. Frame blocking
 3. Windowing
 4. Parameter weighting

در فصل دهم، ابتدا مؤلفه‌های فضای جست‌وجو در بازشناسی گفتار مطرح می‌شوند و پس از آن، روش‌های جست‌وجوی ساده و جست‌وجوی پیشرفته‌تری -که به نحوی روش‌های ساده‌تر را تکمیل کرده‌اند و در سیستم‌های بازشناسی گفتار بسیار مفیدند- توضیح داده می‌شوند. امروزه مبحث یادگیری عمیق، یکی از زمینه‌های تحقیقاتی مهم و پرطرفدار در بحث یادگیری ماشین است که در سیستم‌های بازشناسی گفتار نیز به‌طور ویژه به آن توجه شده است. در فصل یازدهم، توضیحاتی در مورد یادگیری عمیق، انواع روش‌های پیاده‌سازی آن و آموزش شبکه‌های عصبی عمیق ارائه می‌شود.

فصل دوازدهم، به معرفی دادگان‌های گفتاری فارسی و انگلیسی استاندارد اختصاص دارد که در پروژه‌های تحقیقاتی و طراحی سیستم‌های بازشناسی گفتار به‌کار می‌روند. در فصل سیزدهم، در مورد ارزیابی دقت سیستم و معیار اطمینان خروجی بحث می‌شود. فصل چهاردهم، انواع نویزها و انواع روش‌های کاهش نویز و مقاوم‌سازی سیستم‌های بازشناسی گفتار را مطرح می‌سازد.

امید است این کتاب مورد توجه اساتید بزرگوار و دانشجویان گرامی واقع شود و مؤلف را با پیشنهادها و انتقادهای خود در جهت بهبود کتاب یاری دهند.

nriahi@alzahra.ac.ir

