

ВВЕДЕНИЕ В ЯЗЫКОЗНАНИЕ

Компьютерная обработка языка

Глава 6. Познание языка





Данная лекция посвящена описанию того, как происходит обработка языка на компьютере.

Для многих целей было бы крайне полезно, если бы мы могли общаться с компьютерами точно также, как мы общаемся с людьми на человеческом языке. Сегодня компьютеры, изображенные на телевидении и в фильмах с 2013 года, являются средством, способным понимать и распознавать человеческую речь. Однако последнее поколение мобильных устройств с функцией распознавания речи, хоть и впечатляют, все же не могут поддерживать длительное лингвистическое взаимодействие.

Компьютерная лингвистика осуществляет свою деятельность во всех областях взаимодействия языка и компьютера: от фонетики до прагматики, от производства речи до ее восприятия, от устного высказывания (либо жеста) до письменного текста. Компьютерная фонетика и фонология занимаются обработкой речевых данных, главной целью которых является преобразование речи в текст с позиции понимания, и текста в речь с позиции образования. Двумя аспектами компьютерной фонетики и фонологии являются распознавание речи и синтез речи. Распознавание речи – это процесс анализа речевого сигнала на его составляющие звуки и фонемы и, по сути, создание фонетической транскрипции речи. Далее, обработка конвертирует транскрипцию в обычный текст и выводит его на экран, или для дальнейшей обработки, например, приложением по пониманию речи.

Примечание: распознавание речи и понимание речи не является одним и тем же процессом, как это принято считать. Распознавание речи является необходимым предшествующим этапом гораздо более сложного процесса понимания.

Синтез речи – это процесс создания электронных сигналов, имитирующих звуки и просодические особенности речи, и затем воссоздания их в слова и фразы для воспроизведения электронным динамиком, или для последующей обработки в приложении по формированию текста.

Когда вы слушаете иностранную речь, вы замечаете, что она прерывается только для вдоха воздуха, поэтому довольно сложно сегментировать речь на звуки и слова. Это можно назвать проблемой сегментации, о которой было упомянуто в предыдущей лекции. Компьютер также сталкивается с данной проблемой при попытке распознавания речи.

Первые устройства распознавания речи были разработаны не для того, чтобы «слышать» отдельные звуки. Компьютеры были запрограммированы для хранения в памяти акустических моделей целого слова или даже фразы, а затем для поиска подобных моделей в любой последующей речи, которую они должны были распознать. В компьютеры был встроен маленький фиксированный словарь. Более того, намного лучше распознавалась речь одного и того же человека, которая содержала фразы, имеющиеся в этом словаре. С «пониманием» речи другого говорящего уже возникали проблемы. Если использовались слова, которых не было в словаре, то эти программы были бесполезны. Также, если слова произносились без пауз, то точность распознавания ухудшалась. Или если слово не было произнесено полностью, например «missipi» вместо «Mississippi», то распознавание также не было точным.

Эффекты коартикуляции также усложняли ситуацию. Компьютеры могли распознать слово «his» [hɪz], если оно было произнесено отдельно. Но распознать «his» в последовательности слов «hissoap», которая произносится как [hɪssɒp], где последний звук произносится как глухой «s», уже было проблематично. Кроме того, словари состояли в основном из слов, не особо схожих фонетически, для того, чтобы избежать путаницы между такими словами как, например, «pirate» и «pilot».

Сегодня многие телефонные системы, оснащенные функцией интерактивного взаимодействия с пользователем, имеют приложения распознавания речи, которые предлагают вам «нажать 1 или сказать да; нажать 2 или сказать нет» или, например, сделать заказ из меню, произнеся одно или несколько слов. Современные мобильные устройства позволяют запрограммировать целые фразы, такие как, например, «позвонить в офис» или «открыть календарь». В такую систему встроен небольшой запас слов, и поэтому они могут искать голосовые сигналы, схожие со звуковыми моделями ключевых слов и, в основном, воспринимают их верно.

В наиболее современных устройствах распознавания речи, которые приобретаются для использования на персональных компьютерах, имеются наиболее обширные словари. Для



большей точности они должны адаптироваться к голосу определенного человека и определять отдельные звуки в речевых сигналах, даже если они довольно короткие, так как это может привести к значительной разнице в значении.

Адаптация происходит следующим образом: пользователь произносит несколько фраз, заранее заложенных в компьютер, который в свою очередь извлекает акустические модели из каждого звука, характерного для этого пользователя. Позже компьютер использует эти модели для процесса распознавания. Это называется распознавание речи говорящего.

Так как не существует двух одинаковых фраз, и поскольку в сигналах обычно присутствуют шумы, процесс сопоставления, лежащий в основе распознавания речи, является статистическим. На фонетическом уровне вычисления могут классифицировать входящий звук как «l» с 65% достоверностью и как «r» с 35% достоверностью. Также компьютер может использовать другие факторы для принятия решения. Например, если компьютер «уверен», что слово начинается со звука «d», то наиболее вероятно, что после него следует «r», потому что в английском языке нет слов, начинающихся с «dl». Система использует уже запрограммированные знания о фоноактических ограничениях. Но если звук находится в начале слова, необходима дополнительная информация, чтобы определить, какая это фонема – «l» или «r». Если последующими звуками являются «ur», то «l» – правильный вариант, потому что «loor» – это слово, а «goor» словом не является. Если же компьютер не может принять решение, то может быть предложен перечень слов на выбор. Например, «late» или «rate», чтобы человек, использующий систему, сам принял решение. Наиболее современные устройства распознавания речи, которые используются в наши дни, имеют доступ к огромному корпусу текстов, состоящего из нескольких миллиардов слов. Они могут использовать предыдущую частоту встречаемости, чтобы принять решение. Если встает выбор между «look» или «rook», то предпочтительнее будет использование наиболее часто встречающегося слова «look».

Передовые устройства распознавания речи также могут использовать синтаксические правила для устранения неоднозначности высказывания.

Если в синтаксическом контексте «It's too _____» необходимо выбрать между словами «late» и «rate», то программа выберет слово «late», так как после «too» может следовать прилагательное, но не существительное или глагол. Также может быть использована статистическая неоднозначность, основанная на корпусе английского языка. «It's too late» встречается гораздо чаще, нежели «It's too rate». Статистическая модель может быть построена на основе таких фактов, благодаря которым устройство выбрало бы слово «late», а не «rate» в определенном контексте.

В 1779 Христиан Готлиб Кратценштейн получил награду за создание подобной машины. Это был «инструмент, сооруженный наподобие труб одного из регистров органа, который в точности воспроизводит гласные звуки». Создание этой машины также помогло ответить на вопрос, поставленный Императорской академией в Санкт-Петербурге: «Какова природа и характер гласных звуков «а», «е», «i», «о», «u», которая делает их отличными друг от друга?».

Кратценштейн соорудил набор «акустических резонаторов», схожих по форме с человеческим ртом. Эти гласные звуки артикулируются и резонируют с помощью вибрирующей пластины, порождающей импульсы воздуха. Это похоже на импульсы, поступающие из легких через вибрирующие голосовые связки.

Почти сто лет спустя молодой ученый Александр Грейам Белл, восхищенный речью и процессом ее создания, изготовил «говорящую голову» из слепка человеческого черепа. Он использовал различные материалы для создания оболочки, неба, зубов, губ, языка, щек и т. д. Также он установил металлическую гортань с голосовыми связками, сделанными путем растягивания куска резины с прорезями. Все части контролировались кнопочной системой управления при помощи сложного набора рычагов. Эта гениальная машина воспроизводила гласные звуки, некоторые носовые звуки и даже кое-какие короткие комбинации звуков.

Благодаря достижениям в области акустической теории о речеобразовании и благодаря технологическому прогрессу, машинное производство звуков речи значительно шагнуло вперед. Нам больше не нужно создавать физические модели механизмов создания речи; теперь мы можем имитировать этот процесс, создавая физические сигналы в электронном виде. Звуки речи могут быть сокращены на небольшое количество акустических компонентов.



Один из способов создания синтетической речи – это смешать эти компоненты в правильном соотношении, в зависимости от речевых звуков, которые нужно сымитировать. Данный метод создания синтетической речи предусмотрен следующим рецептом:

- Начните с тона той же частоты, что и вибрирующие голосовые связки, с тона выше, если синтезируется голос ребенка или женщины, с тона ниже, если синтезируется мужской голос.
- Сделайте акцент на гармонические колебания, соответствующие формантам, необходимым для конкретного гласного, плавного или назального звука.
- Добавьте шепелявость и гудение для фрикативных согласных.
- Добавьте назальную звучность назальным звукам.
- Вырежьте звук на определенное время, чтобы сделать паузу или создать аффрикативный звук. И так далее...

Все эти ингредиенты смешиваются электронным способом при помощи компьютера, чтобы создать как можно более понятную и естественно звучащую речь. Поскольку пункт (2) является ключевым для процесса, этот метод синтеза речи называется формантным синтезом.

Большинство синтетических речей по-прежнему имеют механический тембр или произношение. Это вызвано небольшими неточностями в имитации и тем, что суперсегментные факторы, как, например, изменение интонации и система акцентных типов еще не полностью понятны. Такие факторы могут привести к большей путанице, чем неверно произнесенные фонемы. В настоящее время главная область исследований синтеза речи связана именно с выявлением и программированием правил ритма и своевременности, которые используют носители языка. Тем не менее понять сегодняшние устройства синтеза речи несколько не сложнее, чем человека, который говорит на диалекте, слегка отличающегося от вашего собственного. Также практически не возникает проблем с пониманием информации, где контекст довольно узкий, как, например, в прогнозе погоды, который читает синтетический голос.

Альтернативным подходом к формантному синтезу является конкатенативный синтез. Задачей конкатенативного синтеза является достижение плавности человеческой речи. Это требует хорошей электронной настройки речевой просодии: то есть продолжительности, интонации, тональности и громкости. Именно на ней основывается естественность. Сейчас, во втором десятилетии двадцать первого века, естественность почти достигнута лучшими устройствами синтеза речи. Тем не менее отдельные единицы не всегда идеально сочетаются друг с другом, и совершенство просодического эффекта остается недостижимым, поэтому люди часто слышат «акцент».

При синтезаторах, работающих на основе слогов, программа преобразования текста в речь примет предложение «The number is 5557766» в качестве входных данных и выдаст [θə] [nɪm] [bɜː] [ɪz] [faɪv] [faɪv] [faɪv] [sɛv] [ɛn] [sɛv] [ɛn] [sɪks] [sɪks] в качестве исходящих данных.

Компьютерная программа «Синтезатор» будет искать в памяти различные слоги и объединять их для дальнейшей электронной обработки с целью создания реалистичной интонации и сглаживания границ слога. Телефоны и мобильные устройства, которые преобразовывают текст в речь и сообщают о звонящем, которого нет в ваших контактах, как, например, «неизвестный номер», часто произносят эту фразу неясно. Ее довольно трудно понять. Существует много трудностей преобразования текста в речь. Мы рассмотрим две.

Первая – это проблема слов, которые пишутся одинаково, но произносятся по-разному. «Read» может произноситься как [rɛd] в «She has read the book», и как [ri:d] в «She will read the book».

Как система преобразования текста в речь определит, что есть что? Можете не сомневаться в ответе: машина должна знать структуры предложений так же, как и человек, для того, чтобы сделать правильный выбор. Заурядных, неструктурированных знаний недостаточно. Например, мы можем запрограммировать систему преобразования текста в речь произносить «read» как [rɛd], когда предыдущее слово является формой глагола «have», но данный подход неприменим в нескольких случаях.

Во-первых, «have» может «управлять» произношением, находясь на определенном расстоянии справа или слева от нужного слова, как например, в предложениях «Has the girl with the flaxen hair read the book?» и «Oh, read a lot of books, has she!».



Необходимо знать структуру, лежащую в основе. Речь идет о том, что «has» является вспомогательным глаголом для основного глагола «read». Если мы попробуем стратегию «произносить «read» как [red] всякий раз, когда глагол «have» в непосредственной близости к нужному слову» мы можем совершить ошибку в таком предложении, как «The teacher said to have the girl read the book by tomorrow», где данный глагол следует произносить как [ri:d]. Если вы знаете английский язык, вы знаете и синтаксические структуры в английском. Только благодаря структурным знаниям можно эффективно решить проблему омографов.

Вторая трудность – непоследовательное правописание, пример которого можно увидеть в первых двух строках данного стихотворения:

«I take it you already know
Of tough and bough and cough and dough»

Все эти слова фонетически разные, но трудно найти правила, которые определяют, когда «gh» в словах, например «ough», произносится как [f], а когда вовсе не произносится или произносится, как «ou». Современные компьютеры имеют достаточную емкость для хранения записанного произношения каждого слова, а также других возможных произношений, которые могут быть определены посредством широкого статистического анализа. В этот перечень могут входить: акронимы, аббревиатуры, иностранные слова, имена собственные, числа, включая дроби, а также такие символы, как (#, &, *, %) и т. д. Такой список достаточно полезен – он похож на запоминание, а не на определение произношения – и охватывает большую долю областей, в том числе и слова, заканчивающиеся на «ough». Это основа конкатенативного синтеза на уровне слов. Однако данный список никогда не будет окончательным. Новые слова, иностранные слова, имена собственные, аббревиатуры и акронимы постоянно пополняют английский язык, что не позволяет предугадать их правильное произношение. Система преобразования текста в речь требует правила для преобразования элементов, не входящих в ее словарь, которые должны выводиться формантным синтезатором или конкатенативным синтезатором на основе единиц, которые по форме короче, чем само слово.

Например, исследователи из Университета штата Северная Каролина смогли разработать систему коммуникации для человека, который из-за тяжелой формы заболевания, не мог произнести ни слова, и лишь кивал головой. Благодаря движению головы в качестве утверждения или отрицания, этот человек смог выбирать слова, отображаемые на экране компьютера, и составлять предложения для выражения своих мыслей, которые в последствии произносились синтезатором. Если мы хотим, чтобы компьютеры понимали грамматику английского и говорили грамотно, мы должны научить их морфологии.

У нас не должно быть машин, которые говорят: «The cat is sit on the mat» или «My five horse be in the barn». Поэтому, чтобы компьютеры понимали английский, им следует знать, что слово «sitting» содержит две морфемы «sit» + «-ing», тогда как в слове «spring» – одна морфема. Также в слове «reinvent» две морфемы – «re» + «invent».

Обработка морфем компьютерами называется вычислительной морфологией. Компьютер должен знать состав слова для того, чтобы понимать слова и использовать их грамматически верно. Для обработки слов компьютер запрограммирован на поиск корней и аффиксов. В некоторых случаях этот процесс прост. Слово «Books» легко разобрать по составу: «book» + «-s», так же как и слова «walking» – «walk» + «-ing», «fondness» – «fond» + «-ness» и «unhappy» – «un-» + «happy». Эти случаи являются простыми, так как их правописание понятно, а морфологические процессы являются общими. Есть слова намного сложнее. Например: «profundity» = «profound» + «-ity», «galactic» = «galaxy» + «-ic» и «democracy» = «democrat» + «-y».

Один из способов состоит в том, чтобы поместить морфологические формы всех слов в словарь компьютера. Несмотря на то, что современные компьютеры могут справляться с высокой вычислительной нагрузкой, миллионы морфологических форм будут вызывать проблемы из-за обобщенности процессов. Как только новое слово возникает в языке, появляются однокоренные слова. Например, «blog» = «blogs», «blog's», «blogging», «blogged», «reblog», «bloggable» и т. д. Словарь будет постоянно устаревать. Более того, не все варианты можно предугадать. Хотя слово «heaten» не является словарным словом, компьютер должен знать, что под данным выражением имеется в виду «нагревать». Таким же образом в языке



существует много сложных слов, и невозможно предугадать все возможные их вариации. Когда слово «podcast» произошло от двух слов «pod» + «cast», этого слова не было в словаре ни одного компьютера.

Чтобы понять предложение, вы должны знать его синтаксическую структуру. Если вы не знаете структуру предложения «dogs that chase cats chase birds», то вы не поймете, кто охотится на птиц – кошки или собаки. Точно также компьютеры, которые знают язык, должны определять синтаксическую структуру. Парсер – это компьютерная программа, которая пытается освоить знания носителей языка. Как и мы, парсер использует грамматику для определения структуры слов. В предложении «The little orange rabbit hopped» программа может ошибочно предположить, что «orange» – это существительное. Позже, когда выявляется ошибка, парсер заново анализирует предложение и определяет, что «orange» – это прилагательное. Такой метод используется для работы со сложными по структуре предложениями, которые могут ввести людей в заблуждение. Например, «The old man the boats» или «The Russian women loved died».

Другой способ справиться с такими неоднозначными случаями для компьютера – это рассмотреть каждый синтаксический анализ в параллели, если это позволяет грамматика. Только законченный анализ может являться верным. По данной стратегии, два анализа «The Russian women loved died» будут проведены параллельно. В первом «Russian» – будет являться прилагательным, во втором – существительным. Синтаксический анализ в первом случае дойдет до «The Russian women loved», но после этого выполнить его будет невозможно, так как слово «died» не может находиться в данной позиции глагольной конструкции. Анализ не может быть завершен, так как слово «died» осталось без разбора. В следующем синтаксическом анализе, при рассмотрении двух существительных «Russian women» вместе, обнаруживается наличие определительного придаточного предложения, что было бы очевидно, если бы предлог «that» стоял перед словом «women». Затем программа определяет наличие словосочетания в предложении «The Russian women loved». Предложение заканчивается глаголом «died», который может образовывать глагольную конструкцию. Таким образом, анализ успешно завершен.

В некоторых случаях возможности компьютеров превосходят людей, так как их семантические возможности намного шире. Например, попробуйте определить все возможные значения выражения «Time flies like an arrow». Компьютерная программа справляется с этим с легкостью. Несколько слов могут принадлежать к нескольким синтаксическим категориям. В итоге получается, что значений более 5. Мы не только хотим, чтобы компьютеры понимали язык. Мы также хотим, чтобы они могли составлять новые предложения, которые не были предварительно сохранены. И это также требует знания синтаксических правил грамматики. В некоторых случаях программирование может быть очень простым. Например, компьютерная программа для создания оскорблений в стиле Шекспира представляет три столбца слов, где первый столбец – список простых прилагательных, второй – список сложных прилагательных, а третий – список существительных:

Простые прилагательные:	Сложные прилагательные:	Существительные:
– bawdy;	– beetle-headed;	– baggage;
– churlish;	– clay-brained;	– bladder;
– goatish.	– fly-bitten.	– codpiece.

Программа в случайном порядке выбирает слово из каждого столбца для составления оскорбительных словосочетаний. Мгновенно могут быть получены следующие оскорбления: «you goatish, rox-marked bladder;» «you lumpish, milk-livered hedge-pig». Не всегда ясно, какая часть языкового анализа или синтезирования языка является синтаксической, а какая часть – семантической. При рассмотрении фразы «The time flies like an arrow» должны ли все части синтаксического анализа рассматриваться семантическим анализатором? Или оба анализа должны происходить одновременно, так что анализ, в котором выражение «arrow» («стрела») обозначает время, что является семантически неправильным, непременно исключается? Или, как это сделали бы некоторые лингвисты, должен ли семантический анализ проводиться первым и «управлять» синтаксическим анализом? Эти вопросы активно обсуждаются в последних исследованиях психолингвистики и компьютерной лингвистики.



Вопрос о том, как сформулировать и выразить смысл, обсуждается в течение нескольких тысяч лет и продолжает порождать много исследований в области лингвистики, философии, психологии, когнитивной науки и информатики. Для простоты мы рассматриваем вычислительную семантику, как представление значений слов и морфем в компьютере, а также значений, полученных из их комбинаций.

В вычислительной семантике существуют две проблемы. Одной из них является создание семантического представления языка ввода (т. е. понимание речи). Другая проблема – взять семантическое представление и преобразовать его в естественный язык, который передаст значение и смысл (т. е. генерация речи). Например, рассмотрим отрывок из устной диалоговой системы под названием «The Circuit Fix-It Shop», разработанной, чтобы помочь техническому специалисту восстановить неисправную схему:

Компьютер: This is the Circuit Fix-It Shop. How may I help you?

Техник: I need to fix the RS111 circuit.

Компьютер: I am familiar with that circuit. The LED is supposed to be displaying alternately blinking one and seven.

Техник: Where is the LED?

Компьютер: In the middle right of the blue region is the LED.

Техник: No light.

Компьютер: What is the switch at when the LED is off?

Техник: The switch is down.

Компьютер: You put up the switch.

В процессе работы «The Circuit Fix-It Shop» компьютер должен создать семантическое представление ввода пользователя, воздействовать на него и создать другое семантическое представление, которое отображаются пользователю на обычном языке. Для составления предложений, компьютер пытается найти слова, которые соответствуют концептам, включенным в его семантическое представление. В системе «The Circuit Fix-It Shop» компьютер должен решить, что говорить о следующих словах: выключатель, пользователь, свет или что-либо другое. Ему нужно выбрать слова, соответствующие описанию состояния объекта в определенной области, спросить о состоянии объекта в представлении пользователя, сделать запрос пользователю, рассказать пользователю, что делать дальше и т. д. Если запрос обращен к пользователю, то следует использовать местоимение «ты». Если же речь идет о выключателе, используются следующие фразы: «the switch» или «a switch above the blue light». После того, как подходящие по смыслу выражения подобраны, компьютер начинает использовать синтаксические правила для составления грамматически верных предложений.

Чтобы достичь понимания речи, компьютер пытается найти в своем семантическом представлении концепты, которые соответствуют словам и структурам ввода. Когда технический специалист говорит «I need to fix RS111 circuit», система распознает, что «I» означает пользователя, который нуждается в чем-то, и компьютер должен это ему предоставить. Компьютер также знает, что если требуется исправить что-либо, то он должен предоставить информацию о том, как это работает. Он распознает «the RS111 circuit» как схему с определенными свойствами, которые содержатся в его семантическом представлении. Система сообщает, что работа этой конкретной схемы будет иметь решающее значение для последующего диалога. Когда предложение неоднозначно в структурном плане, как в случае с «He saw the boy with a bicycle» («use a bicycle to see the boy», либо «see a boy and a bicycle»), парсер произведет анализ каждой структуры. Семантическая обработка может исключать некоторые структуры, если они неправильные. Предложение «He saw the boy with a telescope» является семантически верным в двух анализах, а для понимания его подразумеваемого смысла требуются прагматические знания. Лингвисты-исследователи работают с компьютерными алгоритмами, которые будут распознавать язык жестов также, как они распознают речь. Люди, общающиеся на языке жестов, могут жестиковать перед камерой, как перед микрофоном. А компьютер попытается найти этот жест среди всех жестов, хранящихся в его памяти, посредством визуальной обработки, точно так же как он находит звуки из хранящихся в памяти звуковых волн через аудио обработку. Визуальная обработка в данном случае означает



определение жестов и положения рук, их траектории и направления. Эти алгоритмы довольно сложно построить, а успех в их выполнении до сих пор ограничен.