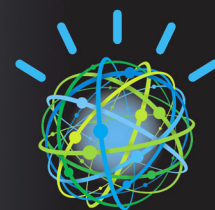


Эпоха когнитивных систем: Принцип построения и работы IBM Watson



IBM WATSON™

Публикации
Redguide

Роб Хай (Rob High)



- Знакомство с когнитивными системами типа IBM Watson, способными изменять подходы, принципы принятия решений и работы организаций
- Описание возможностей IBM Watson, в том числе возможностей обработки естественных языков
- Изучение вопроса о том, каким образом основанные на фактах меры помогают улучшить общие результаты



Краткий обзор

IBM® Watson™ является первым шагом на пути к созданию когнитивных систем, знаменующих новую эру вычислений. В системе Watson использованы современные достижения в области программирования, но она существенно отличается от существующих решений. Сочетание ряда функций и возможностей делают систему Watson уникальной:

- ▶ *Обработка естественного языка* – позволяет разобраться в сложностях неструктурированных данных, которые составляют 80% данных, имеющихся в мире на сегодняшний день.
- ▶ *Построение и оценка гипотез* – применяет средства расширенного анализа для сравнения и оценки группы ответов на основе только существенных фактов.
- ▶ *Динамическое обучение* – позволяет улучшить обучение на основе получаемых результатов, чтобы система становилась умнее с каждой итерацией и взаимодействием.

Ни одна из этих возможностей Watson сама по себе не является уникальной, но когда они объединяются в одном решении, такое решение становится мощным средством, позволяющим:

- ▶ раздвинуть рамки ограничений, сложившихся в области программирования;
- ▶ не ограничиваться только структурированными данными, хранящимися в локальных системах, а начать пользоваться огромным массивом глобальных неструктурированных данных;
- ▶ заменить детерминированные приложения, управляемые деревом решений, на вероятностные системы, способные развиваться вместе со своими пользователями;
- ▶ вместо поиска по ключевым словам, предлагающего список источников, где потенциально можно найти ответ, использовать интуитивно понятный диалоговый инструмент поиска ответов с оценкой по степени уверенности.

Когнитивные системы типа IBM Watson способны изменить подходы, принципы принятия решений и работы организаций в будущем. В этой публикации IBM Redguide™ рассматривается, каким образом сочетание средств обработки естественного языка, динамического обучения, построения и оценки гипотез позволяет Watson выдавать прямые ответы с учетом степени достоверности.

Какой язык трудно понимать компьютеру и почему

Язык служит средством выражения мысли. Он используется для передачи суждений при общении людей. С его помощью мы передаем опасения, надежды, сведения о прошлом и установки на будущее. Есть мнение, что именно язык позволяет нам думать, строить предположения и формировать представления. Он лежит в основе нашего сознания, нашей способности понимать окружающий нас мир или, во всяком случае, нашей способности использовать такое понимание и обмениваться им.

И этот язык крайне неточен.

Наш язык – это бесконечная игра слов, масса специфических особенностей, множество идиоматических выражений, многозначность и омонимия. Горло у нас может быть луженое, а нога – широкая. Как могут быть *высокий голос* и *тонкий голос* одинаковыми по смыслу, а *умник* и *умница* иметь противоположное значение? Как можно *гореть нетерпением* и *сгореть от стыда*? Почему мы *выписываем счет*, когда *вписываем соответствующие данные*?

Но при всем этом наш язык на удивление точный.

Мы умудряемся передавать столько смысла и добиваться такого понимания и участия, несмотря на все трудности языка! Каким-то образом мы понимаем недосказанности, не останавливаемся на несоответствиях и противоречиях, не обращаем внимания на неупорядоченность и отсутствие ясности – и во многом правильно понимаем друг друга.

Вот эта разница между точностью и правильностью имеет большое значение. *Точность* – это механическая или научная точность совпадения, которое можно найти во фрагменте текста. Наличие или отсутствие отдельного слова во фрагменте можно определить с высокой степенью точности. *Правильность* – это степень, с которой один фрагмент позволяет разумным людям сделать вывод о том, что второй фрагмент следует считать верным.

Что, если, говоря “2 + 2,” мы имели в виду конфигурацию автомобиля: два передних сиденья и два задних?

В примере “2 + 2” ответ равен точно 4. Этот факт известен нам из курса математики. Из него же мы узнали, что, сколько бы нулей мы ни добавили после точки, чтобы повысить точность вычисления, в ответе мы все равно получим 4. А что, если говоря “2 + 2,” мы имели в виду не буквальное математическое выражение, а подразумевали конфигурацию автомобиля: *два передних сиденья и два задних*? Или, допустим, психолог под выражением “2 + 2” понимает семью из *двух родителей и двух детей*? Вот в таких других контекстах ответ *четыре* будет неправильной интерпретацией того, что мы пытаемся передать средствами языка.

На самом деле, для того чтобы правильно ответить на вопрос, во многих случаях необходимо учитывать имеющийся контекст. При отсутствии достаточной фактической информации трудно правильно ответить на вопрос, даже если вы можете найти точный ответ на элементы вопроса в буквальном смысле.

Поверхностная обработка естественного языка

Во многих системах обработки естественного языка предпринимались попытки акцентировать внимание на точности в рамках особых правильно построенных правил. Например, при анализе тональности текста часто выполняется поиск определенной группы слов и их синонимов на сайтах социальных сетей. Никакой дополнительной оценки контекста, в котором эти слова употребляются, такие системы не выполняют; они просто подсчитывают, сколько раз эти слова встречаются рядом с определенным брендом в одной фразе. Например, при рассмотрении фразы

“... Остановился у IBM Donut Store, чтобы выпить кофе сегодня утром, очень порадовало ...” система делает утверждение, что находящиеся рядом название бренда и выражение “очень порадовало” указывают на позитивные эмоции. А теперь посмотрим, что было бы, если бы эта фраза имела продолжение “... очень порадовало сообщение о том, что на днях открывается кофейня Fictional Coffee Shop, и меня не будет соблазнять аппетитный вид пончиков каждое утро”. В таком случае система выпускает из внимания, что эмоции связаны не с IBM Donut Store. Такой подход мы называем *поверхностной обработкой естественного языка (NLP)*, поскольку, обладая достаточно высокой точностью в более узких целях, она не совсем правильная.

В то же время, важно понимать, что поверхностная обработка на практике играет важную роль во многих системах. Если перед вами стоит задача составить статистически обоснованную оценку тенденций в изменении эмоций на больших массивах информации, то отсутствие правильности в каком-то отдельном примере, по-видимому, не будет иметь особого значения. Если предположить, что число ошибочно-позитивных результатов примерно равно числу ошибочно-негативных результатов, то они компенсируют друг друга. И если общее число исключаемых из подсчета случаев останется примерно постоянным в массивах с течением времени, то остальные данные (за вычетом исключаемых) представляют статистически обоснованную информацию о трендах. В такой ситуации дополнительные расходы на обработку, которая нужна для обеспечения правильности в каждом отдельном случае, были бы необоснованными.

Поверхностная обработка текстов на естественном языке может быть довольно точной в более узких целях, но она не совсем правильная.

Но в ситуациях, когда значение имеют все без исключения случаи, системы, предназначенные для обеспечения высокого уровня точности без учета степеней правильности, обычно не справляются. То есть, они дают хорошие результаты в узких рамках своего предназначения, но при изменении параметров они не могут обеспечить хорошие результаты. Эти системы можно сравнить со способами кирпичной кладки в строительстве. Кирпичи являются прочным материалом, из них можно довольно легко строить конструкции. Десятилетиями и веками люди улучшали способы кирпичной кладки, чтобы они были достаточно точными. Они научились строить довольно большие, красивые и долговечные конструкции. Но, несмотря на высокую прочность при нагрузке, кирпичные здания имеют низкую прочность при растяжении. Они легко разрушаются при землетрясениях и не поддерживают больших пролетов. А после определенного момента и прочность их тоже ухудшается.

Такие ограничения можно наблюдать сегодня и в некоторых потребительских товарах. Например, можно включить свой любимый персональный цифровой помощник с речевым управлением и сказать ему: “Найди мне пиццу”. В ответ вы получите список близлежащих пиццерий - именно то, что вы искали. А теперь скажите: “*Не ищи мне пиццу*”. И снова вы получите список близлежащих пиццерий – не совсем то, что вы просили. И если вы скажете: “Найди мне пиццу *поближе*” или “Найди мне пиццу *подальше*”, в ответ вы получите все те же списки местных заведений. Дело в том, что эти системы разработаны в соответствии с определенным набором правил и выполняют поиск определенного сочетания ключевых слов, чтобы определить ответ. Таким системам неизвестно, как следует проводить различие между вариантами, не описанными в правилах. Они могут быть точными, но не обязательно правильными.

Глубокая обработка естественного языка

Для того чтобы преодолеть ограничения кирпичных строений, стали строить большие здания из стали и железобетона. Аналогия просматривается и в языке: когда требуется не столько точность в узких параметрах, сколько правильность, методы обработки естественного языка изменяются. При оценке вопроса эти методы намного шире используют контекст. Такой подход мы называем *глубокой обработкой естественного языка*, или, если речь идет об ответе на вопросы на естественном

языке, – *вопросно-ответной системой контентной аналитики (Deep Question-Answering, DeepQA).*

Когда требуется большая точность, методы обработки естественного языка изменяются.

IBM Watson относится к системам глубокой обработки естественного языка. Для того чтобы ответ был правильным, система стремится оценить как можно более широкой контекст. Это не только контекст, который содержится в самом вопросе, но и контекст из базы знаний (так называемого *корпуса*), доступной для поиска ответов.

При подготовке к телешоу JEOPARDY! системе Watson предложили ответить на следующий вопрос (подсказка) из категории "Блоги Линкольна":

"Treasury Secy. Chase just submitted this to me for the third time – guess what pal, this time I'm accepting it (Только что министр финансов Чейз подал мне ее в третий раз – представь себе, старик, на этот раз я ее принимаю)".

Прежде всего, обратите внимание на сокращение, "Secy.", которое должно было означать *секретарь*. Далее обратите внимание, что *секретарь* здесь означает не человека, который записывает поручения начальника и ведет дневник деловых встреч. Значение имеет сочетание слов *Treasury Secretary (министр финансов США)* как существительное, определяющее должность. Таким образом, для ответа на этот вопрос системе Watson нужно было найти фразу, в которой бы шла речь о подаче и принятии чего-то между министром финансов Чейзом и Линкольном (категория подсказки). Но тут следует обратить внимание еще и на то, что в категории не сказано именно "президент Линкольн". Правильным оказался ответ: "Что такое отставка?".

Когда спустя некоторое время после телетрансляции игры JEOPARDY!, в которой участвовала система IBM Watson, в начальной школе рассматривали этот пример, ученик пятого класса предложил "Что такое заявка в друзья?" в качестве возможного варианта ответа.

Без контекста нам оставалось бы только гадать.

Ответ этого ученика в какой-то мере интересен: он многое говорит о том, насколько глубоко укоренились социальные сети в жизни следующего поколения общества. Но это и поучительный пример: он также может быть принят как вполне разумный ответ на вопрос. Но мы знаем, что это неправильный ответ, потому что у нас есть исторический контекст. Нам известно, что в конце девятнадцатого века сети Facebook еще не было. Заметьте, что только контекст позволил нам повысить правильность ответа, выданного системой. Без контекста нам оставалось бы только гадать.

Следует отметить, что у нас, людей, нет особых трудностей в обработке своего языка, несмотря на то что иногда мы все-таки путаемся. Но в целом мы гораздо лучше понимаем смысл написанной нами информации, чем это обычно делают компьютеры.

У нас есть врожденные качества, подсказывающие нам, как нужно разрешать неоднозначность языка, – мы бы хотели понять их суть и использовать в вычислительных системах. Такое направление было главной задачей сообщества специалистов в области искусственного интеллекта в течение последних четырех десятилетий. И в огромной мере нам удалось повысить точность обработки языка. Но только после создания системы Watson мы получили возможность наконец добиться того уровня правильности, который необходим для успешного функционирования информационных систем в реальном мире с широким использованием естественного языка.

Решение задачи глубокой обработки естественного языка помогло бы справиться и с другой проблемой. Мы наблюдаем стремительный рост объема данных. Десяносто процентов всех имеющихся в мире данных появились за последние два года. Эта

тенденция будет только усиливаться по мере превращения мира во все более взаимосвязанный и оснащенный. А 80 процентов всей информации в мире является неструктурированной – это текстовая информация, к которой относятся произведения литературы, отчеты, статьи, материалы исследований, диссертации, электронные письма, блоги, сообщения Twitter, чаты и текстовые сообщения. Нам нужны компьютеры, способные понимать этот поток информации, чтобы эффективнее использовать его.

IBM Watson понимает естественный язык

Для того чтобы вести эффективный поиск в потоках неструктурированной информации, нужны совершенно другие вычислительные возможности – их называют *когнитивными системами*. Примером когнитивных систем служит IBM Watson. Эта система способна разделять высказывания на обычном языке и находить взаимосвязи между фразами текста так же правильно, как человек, но намного быстрее и на гораздо больших объемах, чем это может сделать любой человек без помощи технических средств. Система может с высоким уровнем безошибочности определить, какой ответ на вопрос будет правильным.

Но Watson понимает не отдельные слова языка. Он понимает особенности языка, на котором общаются люди. По этим особенностям он может определить, подразумевает ли одна фраза из текста (которая называется *вопрос*) другую фразу (которая называется *ответ*), с высоким уровнем точности в изменяющихся условиях.

В телешоу JEOPARDY! системе Watson нужно было определить, подразумевает ли вопрос “Джоди Фостер получила его (дословно: забрала домой) за роль в ‘Молчании ягнят’” ответ “Джоди Фостер получила Оскар за роль в ‘Молчании ягнят’”. В данном случае *получить (забрать домой)* означает *выиграть премию Оскар*, но это справедливо не всегда. Иногда *получить (забрать домой)* подразумевает простуду, продукты, да мало ли что еще. И наоборот, не всегда можно забрать домой то, что выигрываешь. Например, можно выиграть у конкурентов контракт на работу, но его невозможно забрать домой.

Важен контекст. Важны временные и пространственные ограничения. Все они вместе позволяют когнитивной системе проявлять в работе характеристики, свойственные человеку. Если вспомнить, о чем мы говорили раньше, то методу на основе правил понадобилось бы близкое к бесконечному число правил, чтобы описать все случаи, которые могут встретиться в языке.

Для того чтобы вести эффективный поиск в потоках неструктурированной информации, нужны совершенно другие вычислительные возможности - их называют когнитивными системами.

Watson разбивает на части вопросы и возможные ответы из корпуса текстов, а затем сотнями разных способов анализирует их и контекст высказывания. Используя полученные результаты, Watson устанавливает степень достоверности своей интерпретации вопроса и возможных ответов.

Но мы немного забежали вперед. Каким образом Watson получает свои ответы на вопросы? Система работает в таком порядке:

1. Получив вопрос, Watson выполняет его синтаксический анализ, чтобы выделить основные особенности вопроса.
2. Система генерирует ряд гипотез, просматривая корпус в поисках фраз, которые с некоторой долей вероятности могут содержать необходимый ответ.

3. Система выполняет глубокое сравнение языка вопроса и языка каждого из возможных вариантов ответа, применяя различные алгоритмы логического вывода.
 Это трудный этап. Существуют сотни алгоритмов логического вывода, и все они выполняют разные сравнения. Например, одни выполняют поиск совпадающих терминов и синонимов, вторые рассматривают временные и пространственные особенности, тогда как третьи анализируют подходящие источники контекстуальной информации.
4. Каждый алгоритм логического вывода выставляет одну или несколько оценок, показывающих, в какой степени возможный ответ следует из вопроса, в той области, которая рассматривается данным алгоритмом.
5. Каждой полученной оценке затем присваивается весовой коэффициент по статистической модели, которая фиксирует, насколько успешно справился алгоритм с выявлением логических связей между двумя аналогичными фразами из этой области в “период обучения” Watson. Эта статистическая модель может быть использована впоследствии для определения общего уровня уверенности системы Watson в том, что возможный вариант ответа следует из вопроса.
6. Watson повторяет процесс для каждого возможного варианта ответа до тех пор, пока не найдет ответы, которые будут иметь больше шансов оказаться правильными, чем остальные.

На рисунке Рис. 1 показано, каким образом Watson получает ответ на вопрос.

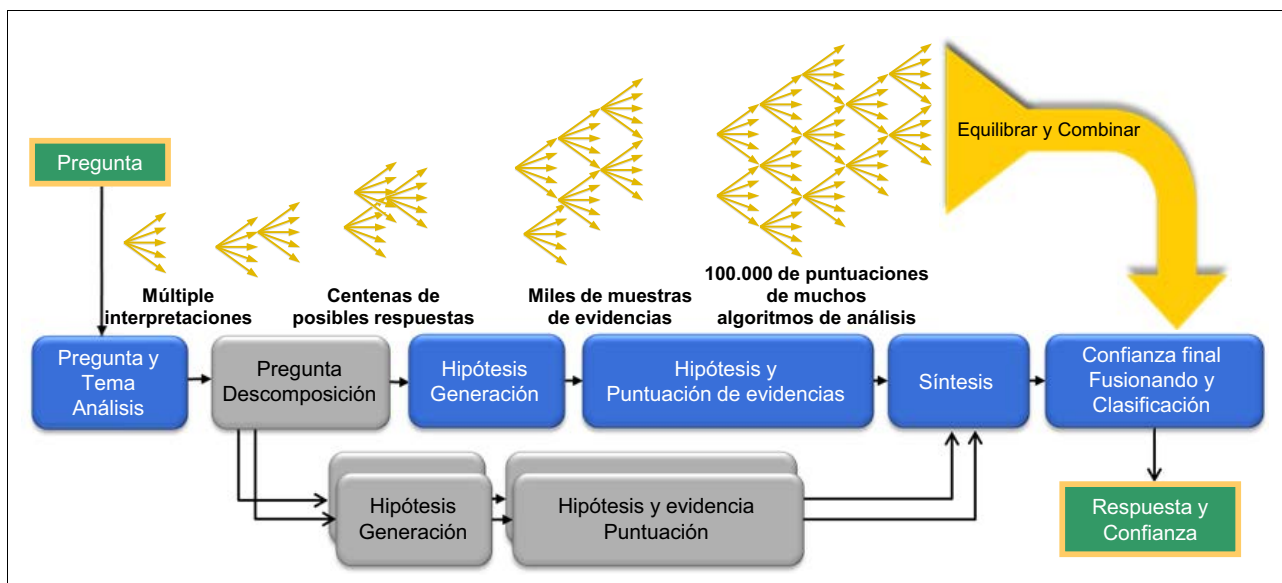


Рис. 1 Каким образом Watson получает ответ на вопрос

Особое значение для работы системы Watson имеет *корпус знаний*. Этот корпус включает в себя всевозможные виды неструктурированных знаний – учебники, методические рекомендации, практические руководства, часто задаваемые вопросы, социальные программы и новости. Watson обрабатывает корпус, преобразуя весь массив содержимого в более удобную для работы форму. В процессе обработки содержимое также подвергается проверке. То есть, система обращает внимание на то, соответствующий ли контент содержится в корпусе, и отсеивает статьи или страницы, которые устарели, стали бесполезными или были получены из потенциально ненадежных источников.

Некоторые алгоритмы логического вывода рассматривают пространственные и временные особенности фразы – это очень важно для разрешения неоднозначности во многих устных и письменных высказываниях. Когда мы говорим: “Найди мне пиццу”, само собой разумеется, что мы подразумеваем где-нибудь поблизости. Но понятие поблизости всегда относительно. В других примерах пространственное расположение можно понять по географическим маркерам; таким маркером может служить район города или штат (область) страны. Кроме того, в контексте большей части письменных источников присутствуют временные особенности. В словах “Будешь возвращаться – купи сыр в магазине” подразумевается определенный период времени. Предположительно, тот, кто писал их, и тот, кому они адресованы, имеют общее контекстуальное понимание того, когда они будут возвращаться домой.

Пространственная и временная оценка должна быть выполнена не только для вопроса, но и для возможного ответа.

В высказывании “В мае 1898 года Португалия отмечала 400-летие с момента прибытия этого исследователя в Индию” присутствуют одновременно пространственное и временное измерение. Празднование происходило в Португалии, но при этом отмечалось прибытие исследователя в Индию. Предполагается ли в высказывании, что исследователь совершил путь из Португалии в Индию? Был ли он вообще в Португалии? Обратите внимание, что торжества происходили в 1898 году, а само событие имело место на 400 лет раньше. То есть, это событие произошло в 1498 году. Ответ на вопрос содержался во фразе “27 мая 1498 года Васко да Гама высадился в городе Каликут (ныне Кожикоде)”. Пространственная и временная оценка должна быть выполнена не только для вопроса, но и для возможных ответов.

Контекст выводится из непосредственно полученной информации и знаний как таковых. Watson может получать оперативную информацию из заголовка документа, других фраз из него или из исходной базы данных, откуда этот документ взят. Контекстом также может быть общая история. Помните, нам было известно, что “Что такое заявка в друзья?” является, вероятнее всего, неверным ответом на вопрос из категории “Блоги Линкольна”. Просто мы знаем общий исторический контекст, в котором известно, когда произошли те или иные события по отношению к другим событиям. Нам известно, что сеть Facebook была создана не так давно, а Авраам Линкольн жил около 150 лет назад – задолго до того, как Facebook приобрел популярность. Контекст и построение логических выводов позволяют нам создать когнитивную основу для обработки естественного языка.

С понимания языка все только начинается

Мы определяем когнитивные системы как применение для передачи и обработки мыслей тех способов, которые характерны для человека. В сочетании с мощными возможностями, заложенными в цифровых вычислениях, они позволяют решать задачи гораздо правильнее, с более высокой гибкостью и в широких масштабах огромных массивах информации.

Когнитивную систему можно представить в виде нескольких основных элементов (Рис. 2). В полях со средней интенсивностью окраски представлены текущие возможности когнитивных систем. В полях, выделенных более светлым тоном, представлены будущие возможности когнитивных систем.

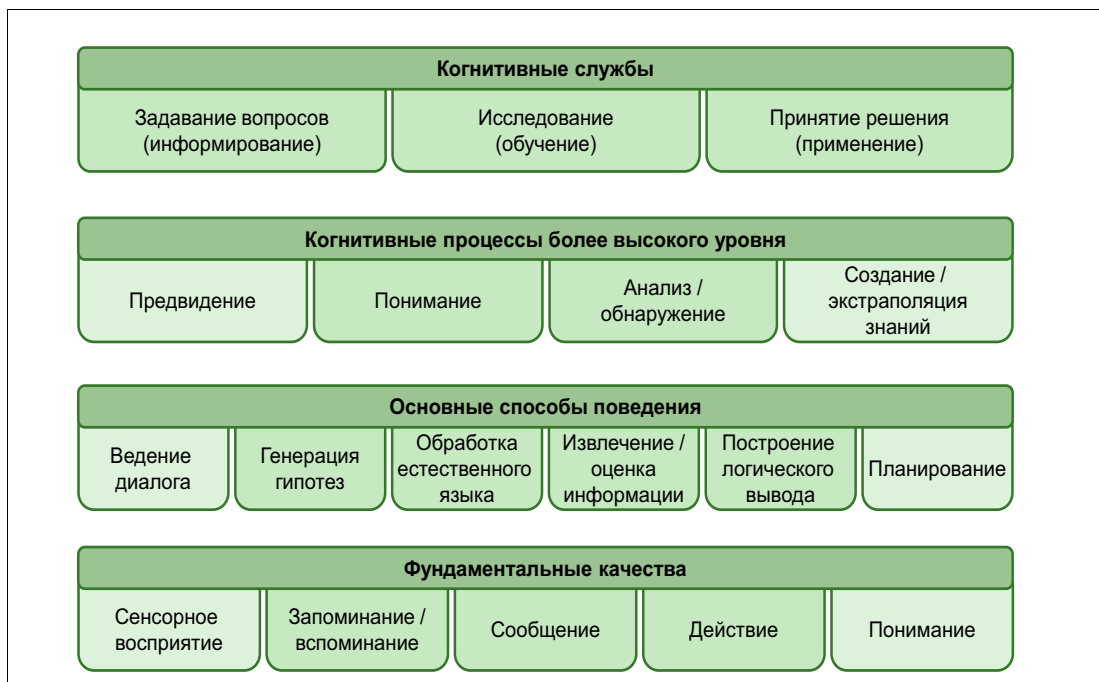


Рис.2 Элементы когнитивной системы

У когнитивных систем, как и у человека, имеется свой способ сбора, запоминания и извлечения информации, аналогичный воспоминаниям человека. Кроме того, когнитивные системы имеют основную способность передавать информацию и действовать. Эти способности разделены на определенные поведенческие конструкторы, как в следующих примерах:

- ▶ способность создавать и проверять гипотезы;
- ▶ способность разбивать на составляющие и строить логические выводы о языке;
- ▶ способность извлекать и оценивать полезную информацию (такую как даты, местоположения и характеристики).

Это самые главные способности, без которых ни компьютер, ни человек не сможет определить правильную взаимосвязь между вопросами и ответами.

Когнитивные процессы более высокого порядка могут достичь высокого уровня понимания, ориентируясь на основные способы поведения. Для того чтобы понять что-то, мы должны уметь разделить информацию на более мелкие элементы, которые достаточно хорошо упорядочены на рассматриваемом уровне. Физические процессы у человека протекают совсем не так, как процессы в космическом масштабе или на уровне элементарных частиц. Так же и когнитивные системы предназначены для работы на уровне человека, хотя они представляют огромное множество людей. В связи с этим понимание языка начинается с понимания более простых правил языка – не только формальной грамматики, но и неформальных соглашений, которые наблюдаются в повседневном использовании.

Когнитивные системы, как и человек, стремятся понять смысл, разделяя высказывания на составляющие элементы и связывая их с контекстом.

Однако когнитивные системы, как и человек, стремятся понять смысл, разделяя высказывание на составляющие единицы и связывая результат с контекстом и вероятностью того, что определенные термины будут использованы определенным образом. И, как и у человека, уверенность будет пропорциональна числу фактов, подтверждающих такую вероятность, и числу алгоритмов логического вывода, которые имеются у нас для проверки гипотезы.

Установив некоторый уровень понимания путем декомпозиции задачи в зависимости от ее возможной цели, когнитивные системы могут реконструировать элементы разными способами, каждый из которых можно проверить, чтобы представить себе новое целое. На основе этих комбинаций можно получать новые результаты и ценные знания, позволяющие находить ответы на вопросы и определять, какие вопросы могли вызвать те или иные ответы.

Эти способности можно использовать для решения задач, которые укладываются в определенные общие схемы. Мы можем задавать вопросы и получать на них ответы. Мы можем использовать систему, чтобы обнаруживать новые ценные знания и видеть подходы, которые не видели раньше. И мы можем использовать эти системы, чтобы обеспечивать принятие правильных решений или, во всяком случае, помогать другим принимать те решения, которые им необходимы.

По мере развития функций когнитивных систем они должны приобрести возможность сенсорного восприятия.

В будущем по мере развития функций когнитивных систем мы ожидаем, что они приобретут возможность сенсорного восприятия. Мы рассчитываем, что они должны будут уметь не только читать текст, но и слышать, видеть и ощущать, получая таким образом элементарное представление о своей среде. Кроме того, мы надеемся, что эти системы смогут воспринимать информацию, например опознавать внешние формы и изменение условий, что будет дополнять их контекст новой информацией, расширяя их возможности строить логические выводы и делать заключения. Мы рассчитываем также, что они смогут усвоить способы поведения и когнитивные процессы более высокого порядка, например, научатся вести диалог, планировать различные стратегии решения задач, а также предвидеть ход развития событий и экстраполировать это предвидение на новые знания.

По сути, когнитивные системы усвоят многочисленные способы поведения, которые люди считают “естественными”, и будут применять их в массовом порядке, помогая людям решать задачи, которые сегодня во многих случаях лежат за гранью их понимания. Мы вступаем в новую эру. В эру, когда компьютеры будут не только эффективнее выполнять обычные процедурные задачи, но начнут применять когнитивность, свойственную человеку, чтобы люди стали разумнее во всем, чем бы они ни занимались.

Задачи могут быть разными

Продолжая работы по проекту IBM Watson, мы открываем другие сферы использования этой системы. В классическом примере “Спросите у Уотсона” пользователь задает системе Watson вопрос (или предлагает подсказку, медицинскую карту больного и т. д.), из которого Watson путем логических построений получает ответ. У Watson’a есть уверенность в том, что из вопроса логически вытекает ответ, и есть факты, которые подтверждают ответ. Watson находит применение в таких сферах, как диагностика онкологических заболеваний, управление использованием средств (то есть, предварительное утверждение страхового покрытия плановых медицинских процедур), кредитный анализ и фундаментальные исследования.

Система Watson полезна во всех случаях, когда специалистам требуется помощь в получении наиболее релевантной информации для своей предметной области.

Задавая другие более важные вопросы, вы станете думать о проблемах своего бизнеса совершенно иначе.

Одним из самых больших откровений, связанных с Watson, стало то, что при поиске ответов на вопросы можно понять, что вы задаете совсем не те вопросы. Когда Watson отвечает на ваши вопросы, даже если он дает правильный ответ, вы можете

осознать, что нужно было задать другие, более важные вопросы, которые позволили бы взглянуть на вашу бизнес-проблему совершенно иначе. Вы начинаете мыслить так, что вам становятся яснее угрозы со стороны конкурентов и возможности на рынке, которые до сих пор не приходили вам в голову.

Эти варианты применения, открывающие новые грани, продолжают совершенствоваться благодаря работам, которые ведутся в IBM в лабораториях IBM Research and Software Development. Последние достижения в построении цепочек логического вывода (которые определяют, что из *этого* следует *то*, а из него, в свою очередь, следует что-то еще, и так далее) углубляют понимание. Важно знать, что диабет ведет к повышению уровня сахара в крови. Но для лечения больного в комплексе гораздо важнее не ограничиваться этим знанием, а сделать следующий вывод: высокий уровень сахара в крови ведет к слепоте. Многоуровневые логические выводы такого типа можно изобразить в виде графа логического вывода, на котором можно наблюдать самые разные аспекты в нисходящем направлении. Более того, сходимость на графе является мощным средством получения более важных выводов - например, ответов, которые обнажают более глубокие ценные знания и их скрытые последствия. Мы можем агрегировать предыдущие значения степени достоверности и достигнуть более высокого уровня уверенности в том, что конкретный ответ является предпочтительным ответом на поставленный вопрос.

Мы можем делать обратные выводы – например, определять вопросы, которые могли породить те или иные ответы.

Мы можем делать обратные выводы - по сути, определять вопросы, которые могли породить те или иные ответы. Выяснив, что у больного с “невыразительным лицом” в анамнезе имеется тремор покоя, можно сделать вывод о том, что он страдает болезнью Паркинсона. Но если при этом выясняется, что больной также испытывает трудности при ходьбе, это может свидетельствовать о дегенерации черного вещества Земмеринга, которая осталась бы незамеченной без ответа на те вопросы, которые ранее не задавали.

IBM инвестирует ресурсы в существенное усовершенствование системы Watson, которые, на наш взгляд, приведут к другим важным достижениям в сфере здравоохранения, финансовом секторе, работе контакт-центров, органов государственного управления, химической отрасли и на пути к созданию разумной планеты. Достижения такого типа приближают эпоху когнитивных систем.

Во многих сферах система Watson применяется совместно с другими более традиционными формами вычислений, такими как статистический анализ, правила и бизнес-процессы, совместная работа и отчетность, для решения задач бизнеса. Представьте, например, как IBM объединяет другие виды статистического анализа со способностью системы Watson отвечать на вопросы о возможных событиях, которые могут свидетельствовать о рисках для инвестиций. IBM помогает нашим заказчикам улучшать процессы управления рисками и процессы оценки для финансовых учреждений. Аналогичным образом и ценные знания, которые мы получаем при анализе ответов заказчиков с использованием глубокой естественногоязыковой обработки, могут свидетельствовать об изменениях поведения при совершении покупок и использовании продуктов, которые без этой системы невозможно было бы обнаружить в структурированных данных. В области здравоохранения Watson помогает страховым компаниям с процессами предварительного утверждения курсов лечения управлять использованием средств.

Правильность повышается в результате обобщения

IBM продолжает развивать и разрабатывать когнитивные системы этого типа. И теперь мы стоим перед классическим выбором, с которым на каждом шагу сталкивается человек, – специализация или генерализация. Мы можем сузить

технологии NLP, ограничив ее одной отдельной областью, – например, сосредоточить внимание только на лингвистических особенностях этой области. Весьма заманчивый подход – кстати, он даже может быть необходим на ранних этапах развития, чтобы убедиться в жизнеспособности технологии. Но он, скорее всего, отбросит нас назад, в то время, когда мы строили здания из кирпича. Если способность к адаптации с присущей человеку маневренностью выделяет когнитивные системы среди других, однозначно нужна генерализация. Нам необходимо опознавать и делать логические выводы на основе более широкой лингвистической вариативности, в более широком ряду обстоятельств, по мере изменения наших знаний, изменений контекста и изменений в современной лингвистике.

Применяя такой подход, мы сможем с большей легкостью адаптироваться для решения новых и более серьезных задач. Мы уже применяем Watson в сфере здравоохранения и индустрии финансовых услуг – и получаем существенные преимущества:

- ▶ мы можем использовать Watson в разных областях, в которых проблемы обходятся дорого;
- ▶ развиваются алгоритмы обработки языка, заложенные в Watson, и появляется возможность обработки более широкой лингвистической вариативности.

Такой подход облегчает адаптацию системы Watson для новых областей и расширяет возможности ее использования в уже существующих областях.

Мы стоим перед классическим выбором - специализация или генерализация.

Варианты применения в здравоохранении представляют особый интерес, так как здесь часто требуется точность и правильность. Правильность нужна для того, чтобы соответствующим образом интерпретировать текст в истории болезни и на его основании сделать логический вывод о состоянии больного. В рекомендациях Национальной всеобщей онкологической сети (NCCN) по лечению рака молочной железы подчеркивается, что следует опираться на наличие точных терминов в медицинской карте больного. Затем требуется дополнительная точность, чтобы найти факты, подтверждающие определенное лечение.

Мы вступаем в новую эру вычислений - менее четких, но более точных.

Когда мы сталкиваемся с языковой аномалией (явлением в языке, которое не встречалось раньше), мы принимаем решение о том, уникальна ли проблема для этой области или она типична для более широкого спектра языковых проблем. По возможности мы возвращаемся к базовым алгоритмам, чтобы определить, можно ли распространить алгоритм на новую ситуацию. Как и любой человек, при использовании этого подхода мы можем связывать свое понимание с новым опытом и таким образом наращивать контекстную базу системы.

В результате мы увеличиваем правильность, область и масштаб применения:

- ▶ правильность лингвистического вывода (получение верного ответа, с вескими основаниями и своевременно);
- ▶ область применения;
- ▶ расширение масштаба - использование для обработки больших массивов данных во многих других областях.

Мы рассчитываем, что система будет еще полезнее при нахождении “наилучшей альтернативы”, анализе социальных настроений, переработке нефти и во многих других сферах. Мы только вступаем в новую важную эру вычислений, которые ориентированы не столько на точность, сколько на правильность. Эру применения моделей человеческого поведения для решения масштабных вычислительных задач. Эру когнитивных систем.

Другие источники дополнительной информации

Дополнительная информация о системе Watson приведена на веб-сайте, посвященном IBM Watson:

ibm.com/innovation/us/watson/index.shtml

автор данного руководства

Данное руководство написано Робом Хаем в сотрудничестве с Международной организацией технической поддержки (ITSO).



Роб Хай (Rob High) – сотрудник IBM, вице-президент и главный инженер проекта IBM Watson Solutions, IBM Software Group. Отвечает за все вопросы разработки стратегии технического развития и концепции решений IBM Watson. Роб, как ведущий член группы проекта IBM Watson Solutions, работает в сотрудничестве с группами проектирования, исследования и разработки Watson. Опыт работы в области программирования составляет 37 лет; в течение последнего 21 года Роб занимается мониторами транзакций и объектно-ориентированными средствами на основе компонентов, в частности, SOMObject Server, Component Broker, а в последнее время – IBM WebSphere® Application Server. Ранее Роб был главным архитектором WebSphere Foundation, ответственным за архитектуру WebSphere Application Server и связанных продуктов, интегрированных на одном рабочем ядре с ним.

Теперь и вы можете стать автором публикаций!

Предлагаем вам возможность продемонстрировать свою квалификацию, продвинуться по карьерной лестнице и стать автором печатных трудов – в одно и то же время! Участвуйте в проекте ITSO – участвуйте в создании книги, делаясь опытом в своей предметной области и обогащая этот опыт на основе использования передовых технологий. Ваша работа поможет улучшить отношение к вашим продуктам и повысить степень удовлетворенности заказчиков, а вы сможете расширить сеть своих технических контактов. Вы можете участвовать в программе непосредственно на месте или дистанционно, продолжительность составляет от двух до шести недель.

Больше узнать об этой программе и подать заявку на участие вы можете на сайте:

ibm.com/redbooks/residencies.html

Следите за новостями о технических публикациях IBM Redbooks

- ▶ В сети Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ В сети Twitter:
<http://twitter.com/ibmredbooks>
- ▶ В сети LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Узнавайте о новых публикациях IBM Redbooks® и семинарах из еженедельной рассылки IBM Redbooks:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Получайте информацию о последних публикациях Redbooks по каналам RSS:
<http://www.redbooks.ibm.com/rss.html>

Примечания

Данная информация относится к продуктам и услугам, предлагаемым в США

IBM может не предоставлять в других странах продукты, услуги или компоненты, обсуждаемые в данном документе. Информацию о продуктах и услугах, распространяемых в вашей стране, можно получить в местном представительстве компании IBM. Любая ссылка на продукт, программу или услугу IBM не предполагает, что продукты, программы или услуги других компаний неприменимы. Вместо них могут использоваться любые функционально эквивалентные продукты, программы или услуги, не нарушающие прав компании IBM на интеллектуальную собственность. Однако ответственность за проверку работоспособности любых продуктов, программ и услуг других компаний лежит на пользователе.

Компания IBM может обладать заявками на патенты или патентами на предметы, обсуждаемые в данном документе. Обладание данным документом не предоставляет вам лицензии на эти патенты. Запросы на получение лицензии можно отправлять в письменном виде по адресу:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

Следующий абзац не относится к Великобритании и другим странам, где такие условия противоречат местному законодательству: INTERNATIONAL BUSINESS MACHINES CORPORATION ПРЕДОСТАВЛЯЕТ ДАННОЕ ИЗДАНИЕ НА УСЛОВИЯХ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ, В ЧАСТНОСТИ, ПОДРАЗУМЕВАЕМЫЕ ГАРАНТИИ НЕНАРУШЕНИЯ ПРАВ, ВОЗМОЖНОСТИ КОММЕРЧЕСКОГО ИСПОЛЬЗОВАНИЯ ИЛИ ПРИГОДНОСТИ ДЛЯ ПРИМЕНЕНИЯ В ОПРЕДЕЛЕННЫХ ЦЕЛЯХ. В некоторых государствах не допускается освобождение от явных или подразумеваемых гарантий в отношении некоторых транзакций, поэтому это заявление может к вам не относиться.

Эта информация может содержать технические неточности или типографские опечатки. В информацию периодически вносятся изменения, которые будут учтены во всех последующих изданиях этой публикации. Компания IBM может вносить усовершенствования и изменения в продукты и программы, описанные в данной публикации, в любое время без уведомления.

Все ссылки в данной информации на веб-сайты сторонних поставщиков приведены исключительно для удобства и никоим образом не служат рекламой таких веб-сайтов. Материалы на таких веб-сайтах не являются составной частью материалов для данного продукта IBM, и они могут использоваться только на страх и риск пользователей.

IBM может использовать и распространять любую предоставленную вами информацию по собственному усмотрению без каких-либо обязательств перед вами.

Приводимые здесь данные о производительности получены в контролируемой среде. Результаты, полученные в других средах, могут значительно отличаться от них. Некоторые показатели были получены в системах разработки, и нет никаких гарантий, что в общедоступных системах эти показатели будут такими же. Более того, некоторые результаты могли быть получены путем экстраполяции. Фактические результаты могут отличаться от заявленных. Пользователи должны проверить данные для своих конкретных сред.

Информация о продуктах других компаний была получена от поставщиков этих продуктов, из их опубликованных материалов или других общедоступных источников. Компания IBM не проверяла такие продукты и не может подтвердить правильность их работы, совместимость или другие заявленные характеристики продуктов других компаний. По вопросам о возможностях продуктов других компаний следует обращаться к поставщикам этих продуктов.

Эта информация содержит примеры данных и отчеты, применяемые в повседневных деловых операциях. Для большего сходства с реальностью примеры содержат имена людей, названия компаний, товарных знаков и продуктов. Все эти имена и названия являются вымышленными, и любое сходство с реальными именами и адресами случайно.

ЛИЦЕНЗИЯ НА ОБЪЕКТ АВТОРСКОГО ПРАВА:

Эта информация содержит примеры прикладных программ на исходном языке, иллюстрирующие приемы программирования на различных операционных платформах. Вы можете бесплатно

копировать, изменять и распространять эти примеры в любой форме в целях разработки, использования, маркетинга и распространения прикладных программ, соответствующих интерфейсу прикладных программ для той операционной платформы, для которой предназначены эти примеры. Эти примеры не проходили тщательного и всестороннего тестирования. По этой причине IBM не может предоставить явных или подразумеваемых гарантий надежности, удобства обслуживания и правильного функционирования этих программ.

Настоящий документ REDP-4955-00 создан/обновлен 5 апреля 2013 года.



Redbooks®


Товарные знаки

IBM, логотип IBM и ibm.com – товарные знаки или зарегистрированные товарные знаки International Business Machines Corporation в США и/или других странах. Эти и другие названия товарных знаков IBM при первом упоминании в данном документе отмечены символом товарного знака (® или ™); эти символы указывают на то, что это зарегистрированные в США или охраняемые общим правом товарные знаки, принадлежащие IBM на момент публикации данной информации. Такие товарные знаки могут быть также зарегистрированы или охраняемы общим правом в других странах. Действительный в настоящее время список товарных знаков IBM можно найти на веб-сайте ibm.com/legal/copytrade.shtml

Следующие термины являются товарными знаками International Business Machines Corporation в США и/или других странах:

IBM Watson™
IBM®

Redbooks®
Redguide™

Redbooks (логотип) ®
WebSphere®

Следующие термины являются товарными знаками других компаний:

JEOPARDY! – зарегистрированный товарный знак Jeopardy Productions, Inc. Все права защищены.

Другие названия компаний, продуктов или услуг могут являться товарными знаками или знаками обслуживания соответствующих компаний.