ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

М.В. Акинин, А.И. Таганов, В.А. Балакин, В.В. Кузнецов

ВОПРОСЫ ПРИМЕНЕНИЯ АЛГОРИТМА ГЛУБОКОГО ОБУЧЕНИЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ВЕКТОРНОГО ПРЕДСТАВЛЕНИЯ СЛОЖНЫХ ОБЪЕКТОВ

(Рязанский государственный радиотехнический университет)

В последние годы методы, использующие глубокое обучение нейросетей, заняли ведущее положение в распознавании образов. Благодаря им планка качества методов компьютерного зрения значительно поднялась. В ту же сторону движется и распознавание речи.

Нейронные сети с одним скрытым слоем

Нейросеть со скрытым слоем универсальна: при достаточно большом количестве скрытых узлов она может построить приближение любой функции.

Для простоты понимания рассмотрим перцептрон. У перцептрона бинарные входы и бинарный выход (0 или 1). Количество вариантов значений на входе ограничено. Каждому из них можно сопоставить нейрон в скрытом слое, который срабатывает только для данного входа.

Разбор «условия» для каждого отдельного входа потребует 2^n скрытых нейронов (при n данных). На практике в большинстве случаев могут быть «условия», под которые подходят несколько входных значений, и могут быть «накладывающиеся друг на друга» «условия», которые достигают правильных входов на своём пересечении. Затем необходимо использовать связи между этим нейроном и нейронами на выходе, чтобы задать итоговое значение для конкретного случая [1].

Универсальностью обладают не только перцептроны. Сети с сигмоидами в нейронах (и другими функциями активации) также универсальны: при достаточном количестве скрытых нейронов, они могут построить сколь угодно точное приближение любой непрерывной функции. Продемонстрировать это значительно сложнее, так как нельзя просто изолировать входы друг от друга.Поэтому можно сделать вывод, что нейронные сети с одним скрытым слоем универсальны. Однако то, что модель может работать как справочная таблица, – не самый сильный аргумент в пользу нейросетей. Под универсальностью понимается только то, что сеть может подстроиться под любые выборки, но это вовсе не значит, что она в состоянии адекватно интерполировать решение для работы с новыми данными. Универсальность ещё не объясняет, почему нейросети так хорошо работают. Правильный ответ лежит несколько глубже.

Рассмотрим конкретный результат.

Векторные представления слов

Впервые векторные представления слов были предложены профессором Й. Бенгио более 10 лет назад. Сейчас это одна из перспективных тем для исследований в глубоком обучении. Также векторное представление слов – это одна из тех задач, с помощью которых лучше всего формируется интуитивное понимание, почему глубокое обучение так эффективно.

Векторное представление слова $W:words \rightarrow R^n$ – параметризованная функция, отображающая слова из некоторого естественного языка в векторы большой размерности (допустим, от 200 до 500 измерений). Например, это может выглядеть так:

$$W(\text{“cat”}) = (0.2, -0.4, 0.7,\ldots)$$
$$W(\text{“mat”}) = (0.0, 0.6, -0.1,\ldots)$$

Как правило, эта функция задаётся таблицей поиска, определяющейся матрицей θ, в которой каждому слову соответствует строка $W_\theta(w_n) = \theta_n$).

W инициализируется случайными векторами для каждого слова. Она будет обучаться, чтобы выдавать осмысленные значения для решения некоторой задачи.

Начнем обучать сеть определению, «корректна» ли 5-грамма (последовательность из пяти слов, например, 'cat sat on the mat'). 5-граммы можно «испортить», заменив в каждой какое-нибудь из слов на случайное (например, 'cat sat song the mat'), так как это почти всегда делает 5-грамму бессмысленной.

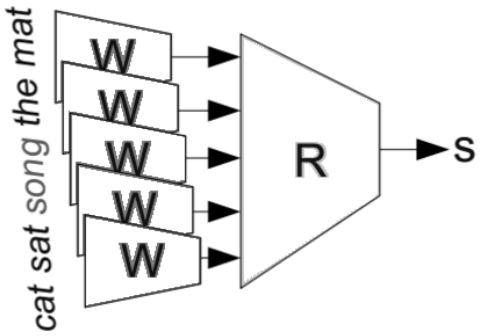

Рисунок 1 – Модульная сеть для определения корректности 5-граммы

Модель, которую мы обучаем, пропустит каждое слово из 5-граммы через W, получив на выходе их векторные представления, и подаст их на вход другому модулю, R, который попытается предсказать, «корректна» 5-грамма или нет. Нам нужно, чтобы было так:

$$R(W(\text{“cat”}), W(\text{“sat”}), W(\text{“on”}), W(\text{“the”}), W(\text{“mat”})) = 1$$
$$R(W(\text{“cat”}), W(\text{“sat”}), W(\text{“song”}), W(\text{“the”}), W(\text{“mat”})) = 0$$

Чтобы предсказывать эти значения точно, сети нужно хорошо подобрать параметры для W и R.

Однако, вероятно, что найденное решение этой задачи поможет находить в текстах только грамматические ошибки или что-то аналогичное. По-настоящему ценным может оказаться полученное значение *W*.

Изобразим, как устроено пространство векторных представлений с помощью хитрого метода визуализации данных высокой размерности – tSNE:

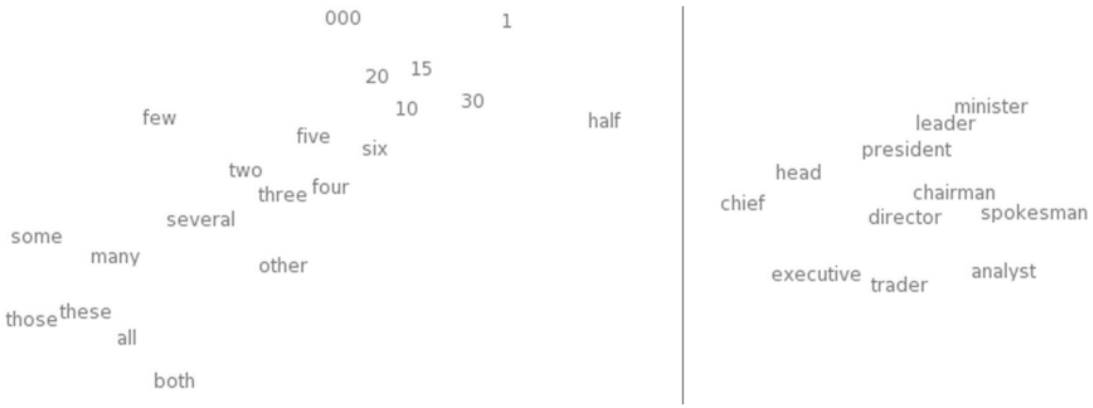

Рисунок 2 – Визуализация векторных представлений слов с помощью tSNE. Слева – «область чисел», справа – «область профессий»

Кажется естественным, что сеть сопоставит словам с похожими значениями близкие друг к другу векторы. Если заменить слово на синоним («некоторые хорошо поют» –> «немногие хорошо поют»), то «корректность» предложения не меняется. Казалось бы, предложения на входе отличаются значительно, но так как *W* «сдвигает» представления синонимов («некоторые» и «немногие») друг к другу, для *R* мало что меняется.

Это мощное средство. Число возможных 5-грамм огромно, в то время как размер обучающей выборки сравнительно мал. Сближение представлений похожих слов позволяет нам, взяв одно предложение, работать с целым классом «похожих» на него. Дело не ограничивается заменой синонимов, например, возможна подстановка слова из того же класса («стена голубая» –> «стена красная»). Более того, есть смысл и в одновременной замене нескольких слов («стена голубая» –> «потолок красный»). Число таких «похожих фраз» растёт по экспоненте от числа слов.

Очевидно, что это свойство *W* было бы очень полезным. Но как её обучают? Очень вероятно, что много раз *W* сталкивается с предложением «стена синяя» и распознаёт его как корректное перед тем, как увидеть предложение «стена красная». Сдвиг «красная» ближе к «синяя» улучшает работу сети.

Нам всё ещё надо иметь дело с примерами употреблений каждого слова, но аналогии позволяют обобщать на новые комбинации слов. Со всеми словами, значение которых мы понимаем, мы раньше сталкивались, но смысл предложения можно понять, никогда его до этого не слыша. То же умеют и нейронные сети.

Векторные представления обладают ещё одним примечательным свойством: отношения аналогии между словами определяются значением вектора разности между их представлениями. Например, вектор разности «мужских-женских» слов – постоянный:

$$W(\text{“woman”}) - W(\text{“man”}) \quad W(\text{“aunt”}) - W(\text{“uncle”})$$
$$W(\text{“woman”}) - W(\text{“man”}) \quad W(\text{“queen”}) - W(\text{“king”})$$

Это не должно удивлять: наличие местоимений, имеющих род, говорит о том, что замена слова «убивает» грамматическую правильность предложения. Мы пишем: «она – тётя», но «он – дядя». Аналогично, «он – король» и «она – королева». Если мы видим в тексте «она – дядя», скорее всего, это грамматическая ошибка. Если в половине случаев слова заменили случайным образом, то вот, должно быть, наш случай.

Оглядываясь на прошлый опыт можно предположить, что векторные представления сумеют представить пол и множественное/единственное число. Выясняется, что и более сложные отношения «закодированы» аналогично [2].

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Рисунок 3 – Пары отношений

Важно, что все эти свойства W – побочные эффекты. Мы не накладывали требований о том, что представления похожих слов должны быть близко друг к другу. Мы не пытались сами настраивать аналогии с помощью разностей векторов. Мы всего лишь попытались научиться проверять, «корректно» ли предложение, а свойства откуда-то взялись «сами собой» в процессе решения задачи оптимизации.

Великая сила нейронных сетей заключается в том, что они автоматически учатся строить «лучшие» представления данных. В свою очередь представление данных – существенная часть решения многих задач машинного обучения. А векторные представления слов – это один из наиболее удивительных примеров обучения представлений.

Литература

1. Харалик Р.М. Статистический и структурный подходы к описанию текстур [Текст] / Р.М. Харалик // ТИИЭР. – 1979. – №5.– С. 98-120.

2. T. Mikolov. Efficient Estimation of Word Representations in Vector Space [Текст]/ T. Mikolov, K. Chen, G. Corrado et al. // Google Inc., Mountain View, CA. – 2013.

А.К. Алимурадов, П.П. Чураков

МЕТОД ПОВЫШЕНИЯ ТОЧНОСТИ СЕГМЕНТАЦИИ СИГНАЛ/ПАУЗА НА ОСНОВЕ КОМПЛЕМЕНТАРНОЙ МНОЖЕСТВЕННОЙ ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ

(Пензенский государственный университет)

Сегментация на информативные участки и паузы является одной из важных задач при обработке речевых сигналов. Точное обнаружение границ сигнала не только повышает качество обработки, но и уменьшает количество вычислительных и расчетных операций. Поэтому исследование и разработка методов, повышающих точность сегментации сигнал/пауза, являются весьма актуальными.

На сегодняшний день существует много различных подходов к сегментации сигнал/пауза, которые успешно решают проблему эффективного обнаружения границ речевого сигнала. Среди наиболее известных методов сегментации можно выделить следующие:

– методы, основанные на использовании значений кратковременной энергии (*Short-time Energy, STE*) и количества переходов сигнала через нулевое значение в короткие промежутки времени (*Short-time Zero-crossing Rate, ZCR*) [1];

– методы, основанные на использовании значений информационной энтропии (*Information Entropy, IE*) [2];

– методы, основанные на использовании мел-частотных кепстральных коэффициентов (МЧКК, *Mel-frequency cepstrum coefficients, MFCC*) [3];

Проведенные авторские исследования [4] данных методов выявили низкую эффективность в условиях зашумленной обстановки. При отношении сигнал/шум (*Signal-to-Noise Ratio SNR*) 10 дБ коэффициент действительного обнаружения (*Detection rate, DR*) у метода на основе *STE + ZCR* равен всего лишь 72,1%, а у метода на основе на использовании МЧКК - 76,2%.

Основная причина больших погрешностей в сегментации связана с использованием неэффективных и неадаптивных методов обработки сложных нестационарных и зашумленных речевых сигналов. В данной статье авторами предлагается метод сегментации сигнал/пауза, с использованием: адаптивного анализа на основе комплементарной множественной декомпозиции на эмпирические моды (КМДЭМ) [5]; правила разграничения на основе физиологическо-