

Глубокие рекуррентные нейронные сети для аспектно- ориентированного анализа тональности отзывов пользователей на различных услугах

Тарасов Д. С. (dtarasov3@gmail.com)

Интернет-портал reviewdot.ru, Казань, Россия

Ключевые слова: рекуррентные нейронные сети, анализ тональности, извлечение
аспектных терминов, унифицированный подход

Глубокие рекуррентные нейронные сети для нескольких языковых аспектов Анализ настроений по отзывам пользователей

Тарасов Д.С. (dtarasov3@gmail.com)

Reviewdot research, Казань, Российская Федерация

Глубокие рекуррентные нейронные сети (RNN) — это мощные модели последовательностей, применимые для моделирования естественного языка. В этой работе мы изучаем применимость различных архитектур RNN, включая одно- и двунаправленные модели Элмана и модели долгой краткосрочной памяти (LSTM), для анализа тональности на основе аспектов, который включает в себя извлечение терминов аспекта и задачи прогнозирования полярности терминов аспекта. Мы показываем, что единая архитектура RNN без ручной разработки признаков может быть обучена выполнять все эти подзадачи на английских и русских наборах данных. Для подзадачи извлечения аспектов наша система превосходит сильные базовые линии усложненного случайного поля (CRF) и достигает самой современной производительности на российском наборе данных. Для предсказания полярности аспектных терминов наши результаты выступают самым эффективным методом, но все еще хороши для многих практических приложений.

Ключевые слова: рекуррентные нейронные сети, полярность настроений, извлечение
аспектных терминов, единый подход

1. Введение

Во многих практических системах обработки естественного языка (NLP) желательно иметь единую архитектуру, которую можно было бы адаптировать к различным задачам и языкам без необходимости разработки новых наборов функций. Недавний успех глубоких нейронных сетей в целом и глубоких RNN в частности дает надежду на то, что эта цель теперь достижима. RNN были применены к ряду задач английских когнитивных NLP, продемонстрировав свои превосходные возможности в задаче заполнения слогов [Mesnil et al, 2013] и анализе мнений [Irsoy and Cardie, 2014].

Хотя эти результаты являются многообещающими, до сих пор неясно, можно ли теперь использовать RNN для замены других моделей в практической многоцелевой системе NLP. И может ли единая архитектура RNN эффективно выполнять множество различных задач.

В нашей работе оценивается ряд архитектур RNN на трех разных наборах данных: наборе данных ABSA Restaurants (на английском языке) из SemEval-2014 [Pontiki et al, 2014] и двух российских наборов данных (рестораны и автомобили) из SentiRuEval-2015.

Мы показываем, что производительность RNN по извлечению терминов аспекта близка к современному уровню техники и результатов по прогнозированию настроений, но при этом значительно отстает от лучших систем формирования, превосходят сильные базовые показатели и обеспечивает достаточную производительность для использования в практических приложениях. Мы обсуждаем факторы, влияющие на результаты RNN, и предлагаем возможные направления дальнейшего улучшения их производительности в этих задачах.

2. Связанная работа

Анализ настроений или добыча мнений — это компьютерное исследование отношения людей к объектам. При анализе отзывов пользователей две основные задачи — это извлечение аспектных терминов и предсказание полярности аспектных настроений.

Методы извлечения терминов аспекта можно грубо разделить на контролируемые и неконтролируемые подходы. В контролируемом подходе извлечение аспектов обычно рассматривается как проблема маркировки последовательностей и часто решается с использованием вариантов методов условного случайного поля (CRF) [Gan et al, 2009; Breck and Cardie, 2007], включая системы полу-CRF, которые работают на уровне фразы и, таким образом, позволяют включать функции уровня фразы [Choi and Cardie, 2010]. Такие системы в настоящее время содержат самые современные результаты по извлечению терминов из отзывов пользователей [Pontiki et al, 2014]. Однако успех подходов CRF и полу-CRF зависит от доступа к богатым наборам функций, таким как деревья синтаксического анализа зависимостей, теги именованных сущностей и другие компоненты предварительной обработки, которые часто недоступны на языках с ограниченными ресурсами, таких как русский. Неконтролируемые подходы к извлечению терминов пытаются сократить затраты и усилия, связанные с ручным выбором признаков и аннотированием обучающих данных. Эти подходы обычно используют тематические модели, такие как скрытое распределение Дирихле, для изучения терминов аспектов [Brody and Elhadad, 2010]. Однако их производительность ниже, чем у контролируемых систем, обученных на данных домена.

Совсем недавно модели рекуррентных нейронных сетей были предложены для решения задач маркировки последовательностей, в том числе аналогичной задачи интеллектуального анализа мнений [Irsoy, Cardie, 2014], демонстрируя результаты, превосходящие все предыдущие системы. Важно отметить, что эти результаты были получены с использованием только векторов слов в качестве признаков, что устраняет необходимость в сложных проектировании признаков.

Глубокие рекуррентные нейронные сети для анализа настроений на основе нескольких языковых аспектов

Точно так же задача прогнозирования полярности настроений решается в рамках контролируемого и неконтролируемого обучения. Самая современная производительность при обнаружении полярности терминов в настоящее время достигается за счет использования машин опорных векторов (SVM) с богатыми наборами функций, которые включают деревья синтаксического анализа и большие словари мнений, а также предварительную обработку для разрешения отрицания [Pontiki et al, 2014].]. Неконтролируемые методы анализа настроений обычно используются на построении словарей полярностей, для которых в настоящее время существует ряд подходов [Brody and Elhadad, 2010], а затем применяют эвристики для определения полярности терминов.

Недавно были разработаны методы на основе нейронных сетей для определения тональности на уровне документов и на уровне фраз, включая автокодировщики на основе дерева [Socher et al, 2011; 2013] и сверточные нейронные сети [dos Santos and Gatti, 2014; Blunsom et al, 2014] и RNN типа Элмана применялись для анализа настроений на уровне предложений с многообещающими результатами [Weng et al, 2014].

3. Методология

3.1. Наборы данных

Набор данных SemEval-2014 ABSA Restaurants [Pontiki et al, 2014] был загружен через MetaShare (<http://metashare.ilsr.gr:8080/>). Этот набор данных является подмножеством набора данных (Ganu et al, 2009). Он содержит утверждения на английском языке из обзоров ресторанов (3041 в обучении и 800 предложений в тестовом наборе), аннотированные терминами аспектов, встречающимися в предложениях, полярности митерминов аспектов и полярности микатегорий аспектов.

Набор данных «Рестораны России» и соответствующий набор данных «Автомобили», предоставленные организаторами SentiRuEval-2015 участникам, состоят из аналогично аннотированных обзоров на русском языке с рядом важных отличий. Эти наборы данных содержат целые обзоры, а не отдельные предложения, и аннотированы тремя категориями аспектных терминов: «явные» (примерно эквивалентные понятию аспектных терминов SemEval-2014), «имплицитные» и так называемые «полярные факты» — утверждения, которые не содержат явных суждений, но тем не менее говорят о том, что-то хорошее или плохое о рассматриваемом аспекте.

Вспомогательный набор данных для обучения русских неконтролируемых векторов слов был построен на основе конкатенации неаннотированных отзывов об автомобилях и ресторанах, предоставленных организаторами SentiRuEval-2015, и 300 000 отзывов пользователей о различных потребительских товарах из базы данных reviewdot.ru (полученной в результате сканирования более 200 интернет-магазинов и каталогов).

3.2. Оценка человеческого несогласия

В рамках этой работы мы решили оценить человеческие разногласия в наборе данных SentiRuEval-2015 Restaurants, потому что нашли много примеров, которые показались неоднозначными. Для этого мы разделили набор данных на две части (70/30) и назначили двух судей-людей. Судьям-людям получили «рекомендации по аннотации», отправленные организаторами SentiRuEval, и 70% аннотированного набора данных. Затем их попросили аннотировать оставшиеся 30% аспектными терминами (явные, неявные и полярные факты), а результаты с равнили с их одной аннотацией с использованием показателей оценки, описанных в разделе «Показатели».

3.3. Рекуррентные нейронные сети

Рекуррентная нейронная сеть [Elman, 1990] — это тип нейронной сети с рекуррентными связями. Это делает их применимыми для задач последовательного прогнозирования, включая задачи НЛП. В этой работе мы рассмотрим простые сети типа Элмана и архитектуры с долговременной кратковременной памятью.

3.3.1. Простая рекуррентная нейронная сеть

В сети типа Элмана (рис. 1а) активации скрытого слоя $h(t)$ на временном шаге t вычисляются путем преобразования текущих входных $x(t)$ и предыдущего скрытого $h(t-1)$. Выход $y(t)$ вычисляется из скрытого $h(t)$.

Более формально, учитывая последовательность векторов $\{x(t)\}$, где $t = 1..T$, RNN типа Элмана вычисляет память и выходные последовательности:

$$h(t) = f(Wx(t) + Vh(t-1) + b) \tag{1}$$

$$y(t) = g(Uh(t) + c) \tag{2}$$

где f — нелинейная функция, такая как сигмовидная или гиперболическая тангенсная функция, а g — выходная функция. W и V — матрицы весов между входными и скрытыми слоями, а также между скрытыми единицами. U — матрица выходных весов, b и c — векторы смещения, x — входные данные с открытыми и выходными единицами. $h(0)$ в уравнении (1) можно установить на постоянное значение, которое выбирается произвольно или обучается обратным распространением.

Глубокие RNN могут быть определены многими возможными способами [Pascanu et al, 2013], но для целей этой работы глубокие RNN были получены путем наложения нескольких повторяющихся слоев друг на друга.

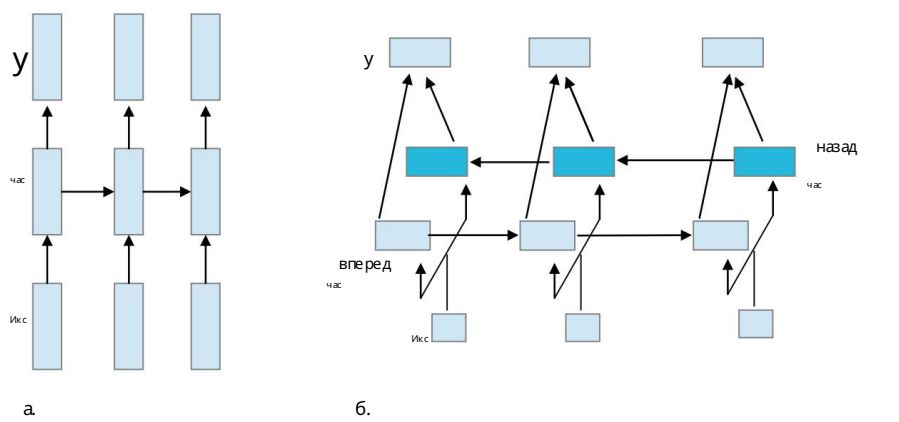


Рис. 1. Рекуррентные нейронные сети, развернутые во времени в три шага а. Простая рекуррентная нейронная сеть б. Двухнаправленная рекуррентная нейронная сеть

Глубокие рекуррентные нейронные сети для анализа настроений на основе нескольких языковых аспектов

3.3.2. Долгая краткосрочная память

Структура LSTM [Hochreiter and Schmidhuber, 1997] позволяет тренироваться на задачах с долгосрочными зависимостями. В LSTM простая функция активации сверху заменена составной функцией активации LSTM. Каждая скрытая единица LSTM дополняется переменной состояния $s(t)$. Активации скрытого слоя соответствуют «ячейкам памяти», масштабируемым по активациям «выходных вентилях» и вычисляемым следующим образом:

$$h(t) = o(t) * f(c(t)) \quad (3)$$

$$c(t) = d(t) * (c(t-1) + i(t)) * f(Wx(t) + Vh(t-1) + b) \quad (4)$$

где $*$ обозначает поэлементное умножение, $d(t)$ — динамическая функция активации, которая масштабирует состояние с помощью «ворот забывания», $i(t)$ — активация входного элемента.

3.3.3. Двухнаправленные RNN

В отличие от обычной RNN, которая может учитывать информацию только из прошлых состояний, двухнаправленная рекуррентная нейронная сеть (BRNN) [Schuster and Kuldip, 1997] может обучаться, используя все доступные входные данные в прошлом и будущем. В BRNN (рис. 16) состояния нейронов разделены на часть, отвечающую за положительное направление времени (состояния вперед), и часть за отрицательное направление времени (состояния назад):

$$h(t) \text{ вперед} = f(W \text{ вперед} x(t) + V \text{ вперед} h \text{ вперед}(t-1) + b \text{ вперед}) \quad (5)$$

$$h(t) \text{ назад} = f(W \text{ назад} x(t) + V \text{ назад} h \text{ назад}(t+1) + b \text{ назад}) \quad (6)$$

$$y(t) = g(U \text{ вперед} h \text{ вперед} + U \text{ назад} h \text{ назад} + c) \quad (7)$$

3.3.4. Обучение Все

Сети обучались с использованием алгоритма обратного распространения во времени (BPTT) [Werbos, 1990] с градиентным спуском мини-пакетов с одним предложением на мини-пакет, как предложено в [Mesnil et al, 2013]. Для задач маркировки последовательностей функция потерь оценивалась на каждом временном шаге, в то время как для задач классификации, таких как предсказание polarity терминов, функция потерь оценивалась только в позициях, соответствующих терминам, polarity которых прогнозировалась.

3.3.5. Регуляризация Для

предотвращения переобучения на входные сети был добавлен небольшой гауссовский шум. Большие сети также были упорядочены с помощью отсела [Hinton et al, 2012] — недавно предложенного метода, который пропускает определенную долю скрытых единиц для каждой обучающей выборки.

3.4. Вложения слов

Вещественнозначные векторы вложения для слов были получены путем неконтролируемого обучения языковой модели рекуррентной нейронной сети (RNNLM) [Миколов и др., 2010]. Были загружены английские вложения размером 80, обученные на 400M наборе данных Google News.

Тарасов ДС

с веб-сайта RNN Toolkit (<http://rnnlm.org/>). Русские эмбединги одинакового размера обучались на всём доступном наборе данных, описанном выше, той же методикой. Русский текст был предварительно обработан путём замены всех чисел на token #number, а все вхождения редких слов были заменены соответствующими словоформами.

3.5. Метрики оценки

Для задач извлечения терминов, где границы терминов трудно определить даже для человека, обычно рекомендуется использовать такие меры, как бинарное перекрытие, которое считается как среднее перекрытие между предсказанным и истинным выражением правильным [Breck et al, 2007], и пропорциональное перекрытие, которое вычисляет частичную правильность, пропорциональную количеству перекрытия [Johansson and Moschitti, 2010].

Из описания задачи SemEval-2014 следует, что использовалась точная версия F-меры (учитываются только точные совпадения), хотя организаторы отмечают, что «в ряде случаев аннотаторы не соглашались с точными границами многоаспектных терминов».

Для российских наборов данных SentiRuEval-2015 из-за нескольких подходов к аннотации довольно часто встречаются многоаспектные (4- и 5-аспектные термины) и довольно велики человеческие разногласия (как будет показано ниже). Организаторы SentiRuEval-2015 используют две метрики для извлечения терминов-аспектов — основную (на основе точного подсчёта) и дополнительную (на основе пропорционального совпадения).

В наборах данных SentiRuEval-2015 все термины помечены как «релевантные» (относящиеся к целевому объекту) или нерелевантные (относящиеся к чему-то другому), а официальные метрики считают правильными только идентификацию релевантных терминов. Мы считаем, что идентификация термина аспекта и классификация его как «релевантного» или нет — это две принципиально разные задачи, и их следует измерять отдельно. Из-за чрезвычайно низкого присутствия (менее 5%) нерелевантных терминов их исключение довольно ложно для алгоритма машинного обучения, и поиск алгоритмов, которые делают это хорошо, представляет собой проблему, представляющую значительный теоретический интерес. Такие системы не могут быть идентифицированы с помощью официальных метрик, так как вклад обнаружения «релевантности» в общее значение F1 довольно мал.

Для целей данной статьи, если не указано иное, мы применяем F-меру, основанную на пропорциональном перекрытии, чтобы облегчить сравнение результатов, полученных на разных наборах данных. Для английских ресторанов набор данных ABSA F-мера вычисляется на тестовом наборе данных из 800 предложений (который не использовался при разработке моделей). Для российских наборов данных, поскольку тестовые данные не были доступны вовремя этой работы, мы выделяем набор для разработки из 5000 слов и используем 7-кратную перекрёстную проверку на оставшихся данных, аналогично подходу [Isroy and Cardie, 2014]. Поскольку мы участвовали в ряде треков SentiRuEval-2015, для сравнения и обсуждения также приведены официальные результаты по метрикам SentiRuEval-2015.

Для задач классификации, таких как определение полярности настроений и категоризация аспектов, макросреднее F-меры не может быть использовано из-за того, что некоторые категории (например, полярность «конфликт», называемая «оба» в российском наборе данных) встречаются крайне редко (рус. Набор данных Sian Restaurant содержит менее 80 экземпляров полярности «обеих» на 3000 экземпляров аспектных терминов). F-мера для таких категорий подвержена огромной ошибке выборки, а также может быть неопределённой (с нулевой прецедентной и отзывом), что делает макросреднее

Глубокие рекуррентные нейронные сети для анализа настроений на основе нескольких языковых аспектов

среднее значение также не определено. Чтобы предотвратить возникновение этой проблемы, SemEval-2014 использует Точность вместо F-меры. Организаторы SentiRuEval-2015 помимо макросреднего используют микросреднее F1. В этой статье для задач классификации мы показываем общую точность, вычисляя макросреднее как дополнительную меру, где это возможно.

3.6. Базовые показатели

Для задачи извлечения терминов мы рассматриваем несколько базовых систем: простой многослойный перцептрон (MLP) с прямой связью, MLP на уровне кадра (MLP с прямой связью с входными данными только функций встраивания слов в окне контекста слова), логистическая регрессия. Функции встраивания слов, а также CRF с использованием в качестве функций словососновой и POS-тегов.

4. Результаты и обсуждение

4.1. Задача извлечения термина аспекта

Таблицы 1–3 суммируют наши результаты по извлечению аспектных терминов. Первоначально для набора данных «Русский ресторан» нам было очень трудно улучшить простую базовую линию CRF.

Ручная проверка аннотации выявила ряд досогласованных решений в представленных обучающих данных, например, в одном месте термин «официантка Любовь» («с лужанка Любовь») был помечен целиком, а в другом аналогичном случае имя «служит» не было помечено как часть слова. Это привело нас к оценке человеческого несогласия, которая оказалась очень близкой к базовым результатам, что сделало выделение терминов очень сложной задачей.

Тем не менее, мы обнаружили, что расширенная прямая RNN превосходит базовую линию CRF при извлечении аспектов, а глубокая модель LSTM превосходит базовые линии CRF и Frame-NN во всех подзадачах, в то время как простая BRNN, хотя и обеспечивает достаточное хорошее результаты, не могла улучшить эти базовые показатели по сравнению с Английский набор данных.

Мы считаем, что непоследовательная аннотация в обучающем наборе приводит к переобучению простых BRNN, потому что сложные локальные модели изучаются достаточно, как могут быть обнаружены длительные зависимости в данных.

В целом, как показано в Таблице 2, наша система дает наилучший результат при извлечении всех терминов аспектов в соответствии с пропорциональным измерением и лучший результат при извлечении всех терминов аспектов в наборе данных автомобилей в соответствии с точной мерой, с охранным при этом второй лучший результат в наборе данных ресторанов. Однако эти хорошие результаты следует интерпретировать с осторожностью из-за относительно небольшого числа участников, общего отсутствия сильных конкурентов и низкого качества данных (по крайней мере, в области ресторанов).

Поэтому, чтобы лучше понять возможности системы, мы оценили нашу систему на английском наборе данных SemEval-2014. Преимущество этого набора данных в том, что он тщательно очищен от ошибок, а также результаты с современных систем легко доступны для сравнения. Таблица 3 показывает, что в этом наборе данных наша система не достигла лучших результатов. Тем не менее, производительность LSTM довольно хороша (эквивалентно 6-му лучшему результату из 28 участников).

Тарас ов ДС

Таблиц а 1. F-мера (пропорц иональное перекрытие) в наборе данных SentiRuEval, оц ененная с ис пользованием 7-кратной перекрест ной проверки

Мех од	SentiRuEval Набор данных о ресторанах				SentiRuEval Набор данных автомобилей			
	Явный	имплиц ит ный	мак рос с редний	я вный	имплиц ит ный	фак т	мак рос с редний	
Судья -человек 1	факт	58,7 33,0 53,6	— — — —					
Человеческий судья 2	69,1	62,3 27,0 51,4	— — — —	57,7 24,0 49,96	— — — —			
базовый уровень ИРК	65,0 68,2							
	54,0	43,0 30 33,3		70,1	75,4 15,2		53,6	
Лог истическая регресс ия								
МЛП	64,5	53,6 18,2 45,3	61,4 26,1	75,8 82,2 34,8 64,2	76,0 83,0 33,0			
Рамка-NN	67,9	51,8		64,0				
Прос той РНН	68,4 58,5 20,0 48,9			75,2	81,3 30,1		62,2	
Прос той RNN, дополненный одним будущим с ловом	68,9 60,0 25,3		51,4	75,8 82,0 31,4 63,1				
Прос той RNN, дополненный одним будущим с ловом + отс ев	71,1	56,0 20,1 49,06		76,0	82,1 24,3		60,8	
Двунаправленный РНН	69,8 61,2 19,1 50,3			76,1	81,5 32,1		63,2	
Двунаправленный ЛСТМ	73,5 64,3 23,5 53,76	77,0 82,5 36,3 65,3						

Таблиц а 2. F-мера набора данных SentiRuEval Test (по результатам SentiRuEval)

Метод	SentiRuEval Набор данных о рес тора нах				SentiRuEval Набор данных а втомоби лей			
	Пропорц иональные		Точные		Пропорц иональные		Точные	
	Явные	Все	Явные	Все	Явные	Все	67,2 52,2 57,5 64,5	
БРНН	71,7 70,4 61,7	59,9 71,9 60,0 62,6	66,8	— — — —				
ЛСТМ								
LSTM, г лубина 2	— — — —	74,8 71,4 65,1 63,0						
Друг ие с ис темы лучший результат	72,8 59,6 63,1 59,5		73,0 65,9 67,6 63,6					

Таблиц а 3. Результаты по анг лийс кому набору данных SemEval ABSA Restaurant (рас с читаны нами с ис пользованием офиц иальных метрик SemEval), эталонные результаты взя ты из [Pontiki et al, 2014]

Метод	Значение F1
ис х одный уровень	47,15
CRF с о с лова ми и POS-тег ами показывает б-й	75,20
лучший результат	79,60
Лучший результат	84,01
БРНН	76,20
ЛСТМ	79,80

Глубокие рекуррентные нейронные сети для анализа настроений на основе нескольких языковых аспектов

4.2. Задача прогнозирования полярности настроений

Таблицы 4–6 суммируют результаты полярности настроений. Здесь более сложные системы обычно дают лучшие результаты, чем более простые методологии.

Используя официальные метрики SentiRuEval-2015, мы получаем второй лучший результат в наборе данных ресторанов (к сожалению, результаты наших ведущих систем не были включены в официальные результаты из-за ошибок, которые мы допустили в формате данных. Эта ошибка стала очевидной только после выпуска тестовых наборов и поэтому не может быть исправлена). Кроме того, относительно плохие результаты частично объясняются тем, что наша система была оптимизирована для задачи прогнозирования полярности на всём сроке, что привело к неоптимальной производительности на задаче только с явным сроком (информация об официальных метриках была опубликована организаторами с опозданием, и мы были не в состоянии адаптировать все системы из-за нехватки времени и ресурсов). В наборе данных English ABSA Restaurant мы получаем точность 69,7, что значительно ниже лучших результатов, но все же

Несмотря на то, что наши результаты здесь ниже лучших систем, они достаточно хороши и имеют некоторую теоретическую ценность для демонстрации того, что одна и та же архитектура может использоваться как для задач маркировки последовательностей, так и для задач прогнозирования полярности. Также стоит отметить, что мы не использовали лексикона настроений, ни специальные этапы предварительной обработки для отрицания (мы обнаружили, что RNN при определенных условиях с помощью обучения отрицанию только на обучающих данных). Еще один важный вывод заключается в том, что использование скрытых активаций модели RNNLM в качестве признаков вместо векторов слов значительно повышает общую производительность системы. Наша гипотеза состоит в том, что задача предсказания следующего слова RNNLM включает в себя необходимость понимания зависимости слов — знание, которое оказалось решающим в задаче предсказания полярности терминов. Таким образом, эти знания из неконтролируемой модели могут быть использованы контролируемой RNN для повышения производительности.

Таблица 4. Результаты задачи прогнозирования полярности для всех терминов в наборе данных SentiRuEval (среднее значение макроканды F1 для положительных и отрицательных классов и общая точность по всем терминам)

Метод	Рестораны		Для всех аспектов	
	Макро Точность	F1 Макро Точность	F1	56,2 57,4 55,2
ТДННN=3	61,0			
РНН	63,1	59,2	57,1	57,1
БРНН	67,4	60,3	60,3	56,9
ЛСТМ	70,2	61,1	62,4	58,0
Возможности LSTM + RNNLM *	74,1	62,5	65,0	59,1

* Получение с помощью активации скрытого слоя RNNLM

Таблица 5. Результаты классификации полярности только явных терминов (согласно официальным результатам SentiRuEval-2015)

Метод	Рестораны	Для всех аспектов
БРНН	61,9	64,7
Функции LSTM + RNNLM	—	65,3
Лучший результат	82,4	74,2

Таблица 6. Результаты классификации полифонических терминов в наборе данных ABSA Restaurants SemEval-2014 (по нашим оценочным метрикам)

Метод	Точность
Базовый уровень	64,00
Лексиконастроений над графами зависимостей *	69,50
БРНН	65,10
ЛСТМ	69,70
Лучший результат	82,92

* Значение взято из [Wettendorf et al, 2015].

5. Выводы

В аспекте задачи извлечения терминов модели рекуррентных нейронных сетей демонстрируют отличную производительность. На протяжении набора данных SentiRuEval-2015 наша система показала лучший результат в извлечении всех терминов аспектов в соответствии с пропорциональным измерением и лучший результат в извлечении всех терминов аспектов в наборе данных автомобилей в соответствии с точной мерой, удерживая при этом второй лучший результат в наборе данных ресторанов. В наборе данных English SentEval-2014 мы получили достаточно хорошие результаты, эквивалентные 6-му лучшему известному результату в этом наборе данных. Из всех моделей RNN наилучшие результаты были получены с глубокой двунаправленной LSTM с двумя скрытыми слоями.

Для предсказания полифонических терминов аспекта мы получили второй лучший результат на наборе данных автомобилей SentiRuEval-2015 и третий лучший результат на наборе данных SentiRuEval-2015 автомобильных ресторанов. Мы также получили хорошие результаты по предсказанию полифонических терминов. Поскольку нам известно, это первый случай, когда модели LSTM применялись для предсказания полифонических терминов аспекта с достаточно хорошими результатами.

В целом наша работа показывает, что модели RNN полезны для анализа настроений на основе аспектов и могут использоваться для быстрого прототипирования и развертывания систем мониторинга на разных языках.

Благодарности

Автор благодарит Екатерину Изотову за помощь в преобразовании формата данных, анонимных рецензентов за полезные комментарии и организаторов SentiRuEval за подготовку и проведение оценки, благодаря чему эта работа стала возможной.

Глубокие рекуррентные нейронные сети для анализа настроений на основе нескольких языковых аспектов

Рекомендации

1. Бланшом П., Грефенштетт Э. и Калхбреннер Н. (2014). Сверхточная нейронная сеть для моделирования предложений. В материалах 52-го ежегодного собрания Ассоциации компьютерной лингвистики.
2. Брек Э., Чой Ю, Карди К. (2007). Определение выражения мнения в контексте. В IJCAI, с. 2683–2688.
3. Броди С., Эльхадд Н. (2010). Неконтролируемая модель оценки аспектов для онлайн-обзоров. В Proceedings of NAACL, с. 804–812, Лос-Анджелес, Калифорния. Choi Y., Cardie C. (2010). Иерархическое последовательное обучение для извлечения мнений и их атрибутов. В материалах конференции ACL 2010 Short Papers, с. 269–274.
5. Душ Сантуш К.Н., и Гатти, М. (2014). Глубокие сверхточные нейронные сети для анализа тональности коротких текстов. В материалах 25-й Международной конференции по компьютерной лингвистике (COLING), Дублин, Ирландия.
6. Элман Дж. (1990). Нахождение структуры во времени. Когнитивная наука, 14(2):179–211.
7. Гану Г., Эльхадд Н. и Мариан А. (июнь 2009 г.). Beyond the Stars: улучшение прогнозов рейтингов с помощью текстового контента обзора. В WebDB (том 9, с. 1–6).
8. Хинтон Г.Э., Шивас тава Н., Крижевский А., Суцкевер И., Салах утдинов Р. (2012). Улучшение нейронных сетей за счет предотвращения совместной адаптации детекторов признаков. Препринт arXiv arXiv:1207.0580
9. Hochreiter, S., & Schmidhuber, J. (1997). Длинная кратковременная память. Нейронная связь, 9 (8), 1735–1780.
10. Иррой О., Карди К. Анализ мнений с помощью глубоких рекуррентных нейронных сетей (2014). EMNLP, Доха, Катар. с. 720–728 11.
- Йоханссон Р., Москитти А. (2010). Синтаксическая и семантическая структура для обнаружения выражения мнения. В материалах Четырнадцатой конференции по компьютерному изучению естественного языка, с. 67–76. Ассоциация компьютерной лингвистики.
12. Менил Г., Хе Х., Л. Денг и Юбенжо (2013). Исследование архитектур рекуррентных нейронных сетей и методов обучения пониманию устной речи. BINTERSPEECH, с. 3771–3775: ISCA.
13. Миколов Т., Карафиат М., Бургет Л., Чернокки Дж., Худангур С. (2010). Языковая модель на основе рекуррентной нейронной сети. BINTERSPEECH, с. 1045–1048.
14. Паскану, Р., Гульере, К., Чо К., и Бенжо, Ю (2013). Как построить глубокие рекуррентные нейронные сети. Препринт arXiv arXiv: 1312.6026.
15. Понтики М., Папагеоргиу Х., Галанис Д., Андруцопулос И., Павлопулос Дж. и Манандхар С. (2014). Семеваль-2014, задание 4: Аспектный анализ тональности. В материалах 8-го Международного семинара по семантической оценке (SemEval 2014) (с. 27–35).
16. Шустер М., Кулди К.П. (1997). Двухнаправленные рекуррентные нейронные сети. IEEE Сделки по обработке сигналов, 45 (11): 2673–2681.
17. Сочер, Р., Пеннингтон, Дж., Хуанг, Э.Х., Нг, А.И. и Мэннинг, К.Д. (2011, июль). Полуправляемые рекурсивные автоэнкодеры для прогнозирования распределения настроений. В материалах конференции по эмпирическим методам обработки естественного языка (с. 151–161). Ассоциация компьютерной лингвистики.

18. Сочер, Р., Перелыгин, А., Ву, Дж. Ю., Чанг, Дж., Мэннинг, К. Д., Нг, А. И. и Поттс, К. (2013, октябрь). Рекурсивные глубокие модели для семантической композиционной банки деревьев настраиваний. В материалах конференции по эмпирическим методам обработки естественного языка (EMNLP) (том 1631, с. 1642).
19. Венг, Р., Баолин, П., Юань, О., Чжоу, Чжан С. (2004) Структурная информационная глубокая рекуррентная нейронная сеть с полуправлением для анализа настраиваний. *Frontiers of Computer Science*, с. 1–14, <http://dx.doi.org/10.1007/s11704-014-4085-7>.
20. Вербос, П. Дж. (1990). Обратное распространение во времени: что оно делает и как с этим делать. *Труды IEEE*, 78 (10), 1550–1560.
21. Веттендорф К., Джеган Р., Корнер А., Зерхедж (2014) SNAP: многоступенчатый XML-конвейер для аналитичности на основе аспектов в материалах 8-го Международного семинара по семантической оценке (SemEval 2014), с. 578–584