Использование рекурентных нейронных сетей для ранжирования списка гипотез в системах распознавания речи

Кудинов М.С.

В статье представлены предварительные результаты использования рекурентных нейронных сетей для языкового моделирования на материале русского языка. Решалась задача ранжирования равновероятных гипотез распознавания. Для уменьшения разреженности данных модели оценивались на лемматизированном новостном корпусе. Также для предсказаний использовалась морфологическая информация. Для финальной сортировки был использован метод опорных векторов для ранжирования. В статье показано, что комбинация нейронных сетей и морфологической модели дает лучшие результаты, чем 5-граммная модель со сглаживанием Кнессера-Нея.

• рекурентная нейронная сеть • ранжирование гипотез

The paper demonstrates preliminary results of the experiments on equiprobable hypothesis re-scoring with recurrent neural networks (RNN). RNNs proved to be successful for language modelling on various tasks for English including speech recognition and phrase completion but their applicability to inflictive languages is not well studied yet. However, for now we trained the model only on lemmas with additional morphological information to decrease data sparseness. We demonstrate that the model performs better than the popular 5-gram model with Knesser-Ney smoothing.

• recurrent neural network • hypotheses ranking

1. ВВЕДЕНИЕ

Известно, что проблема статистического моделирования флективных языков представляет большую сложность, чем для английского языка [1]. Основные проблемы возникают вследствие большого количества морфологических форм слов (лемм) и более свободного порядка слов [2]. Обе проблемы в результате усиливают разреженность данных и снижают эффективность *n*-граммных моделей.

В то время как использование *p*-граммных моделей на первых стадиях распознавания сегодня является стандартной практикой [3], возможности для последующей обработки в рамках алгоритма распознавания, осуществляющего несколько проходов по входным данным, гораздо шире. Например, для переранжирования гипотез, возвращаемых процедурой лучевого поис-

ка Витерби, может быть использована морфологическая, синтаксическая и семантическая информация. В последнем случае значения слов представляются посредством вложения слов в некоторое векторное пространство. К методам, осуществляющим такие вложения, относятся: латентносемантический анализ [4], вероятностное тематическое моделирование [5] или нейронные сети [6]. В 2010 году была представлена языковая модель на рекурентной нейронной сети (RNNLM) [7]. Использование данной модели позволило улучшить предыдущие результаты на стандартных наборах данных как в перплексии, так и в пословной ошибке в экспериментах по распознаванию речи. Несмотря на то что модель была предложена для английского языка, в [8] были приведены обнадеживающие результаты, полученные на небольшом наборе данных для чешского языка. Сходство чешского и русского языков общеизвестно, а значит, перспективы применения рекурентных нейронных сетей к русскому материалу выглядят многообещающе.

Тем не менее проблема обучения рекурентной нейронной сети для языков с богатой морфологией является более сложной, по крайней мере, если использовать оригинальный подход из [7]. В дополнение к уже упомянутым трудностям, связанным с разреженностью данных, обучение модели, использующей словник, содержащий все допустимые словоформы, потребовало бы слишком много времени. Поэтому в данной работе было решено поставить лишь предварительные эксперименты и решить более простую задачу, а именно произвести переранжирование гипотез распознавания исходя из оценок отдельной лексической модели, основанной на рекурентной нейронной сети, и морфологической модели, основанной на условных случайных полях.

Статья организована следующим образом. В разделе 2 приводится общая информация о рекурентных нейронных сетях. В разделе 3 обсуждается применимость оригинальной архитектуры рекурентной нейронной сети к статистическому моделированию флективных языков и сопутствующим проблемам. Наконец, в разделе 4 приводятся данные экспериментов с комментариями и выводами.

2. РЕКУРЕНТНАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ ЯЗЫКА

Рекурентные нейронные сети впервые были рассмотрены в [9] Элманом в 1990 году. В данном исследовании также была высказана идея о применимости рекурентной нейронной сети для моделирования языка. Тем не менее вследствие значительной вычислительной сложности и отсутствия доступных лингвистических корпусов достаточного объема на тот момент метод не получил широкого распространения.

Другой важной вехой в развитии нейросетевых языковых моделей является работа И. Бенджио 2003 года, в которой предлагается метод предсказания последующего слова по левому контексту длины $n-1$, таким образом формируя своего рода n-граммную нейросетевую модель n-го порядка. Однако в отличие от n-граммной модели в данном случае предсказание осуществляется на основании вложений слов в векторное пространство R^M. Каждое входное слово (допустим, с индексом l) в словаре объемом $|V|$ слов представляется в виде $|L|$-мерного вектора $w = <0_1,...,1_l,0_{l+1},...,0_{|L|}>$ с единственной ненулевой координатой $w_l = 1$. На вектор слева умножается матрица U, что эквивалентно выборке l-го столбца U. Другими словами, U действует как словарная таблица, осуществляющее однозначное отображение слов на их векторные представления.

Аналогичная техника была использована Т. Миколовым в [7], который использовал рекурентную сеть Элмана для предсказания слов по контексту. Результирующая модель описывалась следующими уравнениями:

$$P(w_t \mid w_{t-1}, h_{t-1}) = y_{w_t} \quad (1)$$

$$y_t = s(U \cdot h_t) \quad (2)$$

$$h_t = W \cdot x + V \cdot h_{t-1}, \quad (3)$$

где $\sigma(x) = \frac{1}{1+e^{-x}}$ логистическая функция активации,

а $s(y_k) = \frac{e^{y_k}}{\sum_i e^{y_i}}$ софтмакс-функция, h_t – рекурентный слой; y – выходной слой,

где каждому k-му элементу соответствует вероятность $P(w_t \mid w_{t-1}, h_{t-1})$, $V_{H\times H}$ – матрица весов рекурентного слоя, $W_{H\times|L|}$ словарная таблица, отображающая слова в векторные представления, $U_{|L|\times H}$ – матрица весов выходного слоя; H – количество нейронов скрытого слоя.

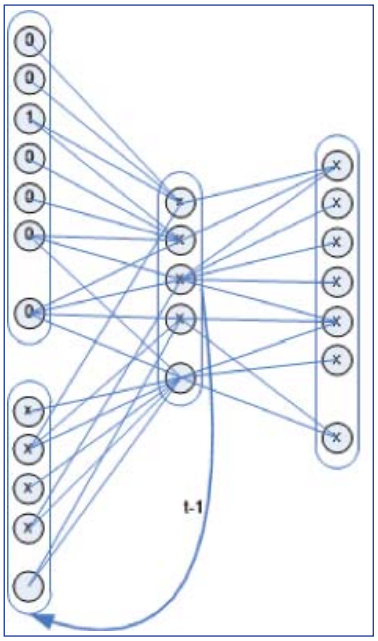

Рис.1. Рекурентная нейронная сеть для статистического моделирования языка

Поскольку h_t потенциально сохраняет в себе весь левый контекст, данная модель выглядит более мощной, чем *n*-граммная нейросетевая модель.

К сожалению, в действительности последнее утверждение не совсем верно, поскольку норма градиента $\frac{\partial h_t}{\partial h_k}$, k <t, отражающего влияние предыдущих значений на скрытом слое на последующие, стремится к нулю (или к бесконечности) с экспоненциальной скоростью по (t–k) [10], [11]:

$$\frac{\partial h_t}{\partial h_k} = \prod_{k<i\le t} \frac{\partial h_i}{\partial h_{i-1}} = \prod_{k<i\le t} V^T diag(\sigma'(h_{i-1}))$$

Стремление градиента к нулю или к бесконечности определяется наибольшим собственным значением матрицы V, причем поведение градиента обязательно демонстрирует один из этих типов ([6]).

Хотя за прошедшие 20 лет с момента обоснования данной проблемы было предложено немало способов её решения [6], [12], в [7] утверждается, что данная проблема не является существенной для моделирования языка. Таким образом, в данной работе будет рассмотрен случай стандартной архитектуры Элмана с алгоритмом распространения ошибки обратно по времени (backpropagation through time).

3. ПРИМЕНИМОСТЬ ПОДХОДА К МОДЕЛИРОВАНИЮ ФЛЕКТИВНЫХ ЯЗЫКОВ

При наличии словаря существенного объема статистическое моделирование флективных языков составляет дополнительную техническую проблему для нейросетевого подхода. Большое количество различных словоформ приводит к пропорционально большему размеру выходного слоя, а из (3) видно, что сложность алгоритма обучения линейна по объему выходного слоя. Чтобы обойти эту проблему, можно было бы использования схему на рис. 2. Каждое входное слово предварительно лемматизируется внешним морфологическим анализатором. Леммы используются для предсказания последующих лемм. Далее предсказанной леммы запускается линейный классификатор (например, логистическая регрессия), предсказывающий словоформу по лемме и морфологическим признакам контекста. Данный подход позволяет миновать проблему разрастания словаря. Другой подход мог бы состоять в том, чтобы разделить выходной слой на два вектора – словарный (леммы) и морфологический (морфологические признаки). Ошибка предсказания в данном случае получалась бы суммированием ошибок на двух векторах.

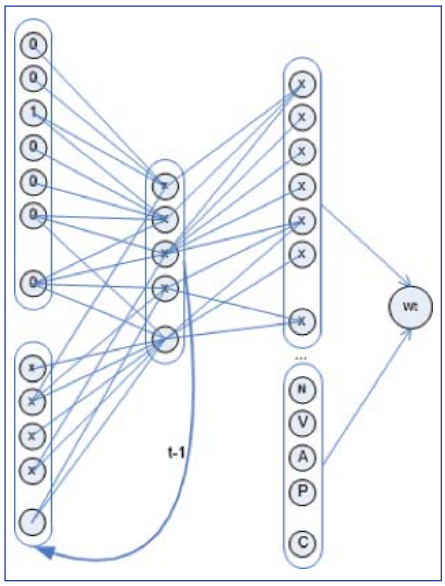

Рис.2. Рекурентная нейронная сеть с внешним классификатором

Тем не менее в данной статье речь пойдет о предварительном эксперименте, целью которого является проверка гипотезы о том, что комбинация нейронных сетей, обученных на леммах, дает лучший результат, чем комбинация n-граммных моделей с дисконтированием Кнессера-Нея.

3.1. Эксперименты

Модель была натренирована на новостном корпусе объемом приблизительно в $2 \cdot 10^6$ токенов. Примерно 10% данных было выделено для валидации. Каждый текст был обработан морфологическим аннализаторо/лемматизатором для русского языка [13] со встроенным словарем примерно в $2 \cdot 10^6$ словоформ. Выходом анализатора являлся текст, в котором все известные словоформы были заменены соответствующими леммами, а неизвестные – специальным токеном «UNK». Второй сгенерированный текст был получен только заменами неизвестных токенов на «UNK». Таким образом, было получено 2 пары тренировочной и тестовой выборок для лемм и словоформ соответственно. На этих корпусах проводились обучение и эксперименты по определению перплексии.

Для эксперимента по ранжированию гипотез использовались списки гипотез, полученные от внешней системы распознавания фирмы Nuance. Использовался русскоязычный корпус предложений со студийным качеством записи и транскрипциями. Аудиофайлы подавались на вход системе распознавания. На выходе получалось до 10 гипотез. В результате была получена коллекция неотсортированных списков гипотез. Как правило, список не содержал полностью правильной гипотезы, и она добавлялась вручную. Далее каждая гипотеза обрабатывалась теми же инструментами, которые использовались при подготовке корпусов: т.е. были проведены лемматизация и замены неизвестных слов. Полученные корпуса были обработаны обученными на предыдущем этапе моделями. В результате для каждой из гипотез были получены списки откликов от каждой модели – *n*-граммной со сглаживанием Кнессера-Нея и рекурентных нейронных сетей с различными размерами скрытого слоя. Всего в обучающем корпусе для ранжирования было 1300 фраз со средним значением 5 гипотез на фразу. В тестовом корпусе было 300 фраз.

В тестах были использованы *n*-граммные модели со сглаживанием Кнессера-Нея, порядков 3, 4, 5, натренированные на леммах и на словоформах. Модели на основе рекурентных нейросетей различались размером скрытого слоя. Были протестированы модели с объёмами слоя 100,200,300, 400 и 500. Все рекурентные сети обучались на лемматизированном корпусе. Кроме этого использовалась оценка, возвращаемая морфологическим анализатором. В результате было получено 12 оценок. Для ранжирования использовалась модель ranking SVM, где в качестве признаков использовались оценки моделей. Результирующая модель обучалась ранжированию гипотез в списке на 2 категории – верная и неверная гипотеза. Фактически данный подход дает интерполяцию моделей. В качестве метрик для оценки в этом случае выбраны уровень пословной ошибки (word error rate, WER%) и процент случаев выбора правильной гипотезы (sentence error rate, SER%).

3.2. Результаты

Результаты экспериментов приводятся в таблицах 1 и 2. В таблице 1 приведены перплексии всех используемых моделей. В таблице 2 приведены результаты эксперимента по ранжированию – уровень пословной ошибки (WER%) и процент точность выбора правильной гипотезы (SER%).

Стоит отметить, что перплексии моделей, натренированных на лемматизиванном и нелемматизованном корпусе, строго говоря, несравнимы по перплексии, поскольку количество неизвестных токенов, а значит, и словарный состав корпусов, различны. Таким образом, важным обнадеживающим выводом,

который можно сделать из приведенной таблицы, является то, что модели на рекурентных нейронных сетях демонстрируют существенно лучшие показатели в эксперименте, чем 5-граммная модель со сглаживанием Кнесера-Нея.

Таблица 1

Перплексии моделей на тестовой выборке

Model	Perplexity	Model	Perplexity
KN3lem	272,8	RNN100	240,13
KN4lem	272,2	RNN200	230,45
KN5lem	273	RNN300	231
KN3tok	128,72	RNN400	231,87
KN4tok	130,76	RNN500	231,21
KN5tok	132		

Рассмотрим теперь результаты эксперимента по ранжированию. Стоит сделать следующие замечания. Первое из них состоит в заметном превосходстве рекурентных нейронных сетей над сглаженными n-граммами. Второй заметный факт – это противоречивое влияние морфологической модели на конечный результат: улучшение пословной ошибки при явной тенденции к голосованию за неверную гипотезу предложения. Это можно объяснить тем фактом, что оценка, возвращаемая морфологическим анализатором, пропорциональна вероятности лучшего разбора $P\left(tag_1^T \mid word_1^T\right)$. По этой причине данная оценка имеет тенденцию к выбору гипотез с наименьшей энтропией разбора. Стоит признать, что данная оценка не вполне подходит к решаемой нами задаче. Третий заметный факт состоит в несколько хаотичном характере результатов рекурентных моделей: некоторые из них демонстрируют достаточно скромные результаты, однако их интерполяции обеспечивают наилучшие результаты.

Таблица 2

Результаты моделей в эксперименте по ранжированию

Model	WER%	SER%	Model	WER%	SER%
KN5lem	16,62	40,8	RNN100	17,55	43,67
KN5tok	18,09	42,72	RNN200	15,35	40,5
KN5lem + morph	15,58	43.98	RNN300	17,09	43,98
KNlem all	17,05	40,82	RNN400	16,58	41,77
KNlem all + morph	15,74	43,67	RNN500	17,43	43,67
KNlem+tok all	15,74	39,24	RNN all	15,35	38,29
KNlem+tok all + morph	15,89	43,35	RNN all + morph	14,58	41,45
all models	14,78	40,5			

Эксперименты по ранжированию в целом демонстрируют превосходство рекурентных моделей. Наилучшая комбинация задействует оценку, возвращаемую морфологическим анализатором и оценки, полученные от рекурентных моделей. Таким образом, обеспечивается комбинирование морфологической и словарной информации. Данный результат свидетельствует о том, что результаты в данном направлении могут быть продолжены.

3.3. Discussion

В статье был предложен простой эксперимент для проверки применимости рекурентных нейронных сетей с внешним классификатором грамматических форм к русскому языку. В ходе эксперимента комбинировались отклики различных языковых моделей с целью ранжирования списка гипотез, возвращенных системой распознавания речи. Результаты указывают на то, что

языковые модели на рекурентных нейронных сетях превосходят результаты сглаженных *n*-граммных моделей как по перплексии, так и по уровню пословной ошибки. Тем не менее эксперименты должны быть продолжены в двух направлениях: проверка воспроизводимости результатов при наличии большей обучающей выборки; конструирование языковой модели на рекурентной нейронной сети для предсказания словоформ русского языка.

СПИСОК ЛИТЕРАТУРЫ

1. *I. Oparin.* Language Models for Automatic Speech Recognition of Inflectional Languages. PhD thesis, University of West Bohemia, Pilsen, 2008.
2. *E.W.D. Whittaker.* Statistical Language Modeling for Au– tomatic Speech Recognition of Russian and English. PhD Thesis, Cambridge University, 2000.
3. *A. Deoras, T. Mikolov, S. Kombrik.* Approximate inference: A sampling based modeling technique to capture complex dependencies in a language model. Speech Communication, 2012
4. *J. Bellegarda.* Exploiting latent semantic information in statistical language modeling. Proc. IEEE. 88, 2000
5. *D. Gildea, T. Hoffman.* Topic-Based Language Models Using EM. In Proceedings of EUROSPEECH, 1999
6. *Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin.* A Neural Probabilistic Language Model. Journal of machine learning research, 2003
7. *T. Mikolov, M. Karafia´t, L. Burget, J. ˇCernocky´, S.Khudanpur.* Recurrent neural network based language model, In: Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), Makuhari, Chiba, JP
8. *T. Mikolov.* Statistical Language Models based on Neural Networks. PhD thesis, Brno University of Technology, 2012.
9. *J. Elman.* Finding Structure in Time. Cognitive Science, 14, 179-211, 1990.
10. [10] *Y. Bengio, P. Simard, P. Frasconi.* Learning Long-Term Dependencies with Gradient Descent is Difficult, IEEE Transactions on neural networks, 1994
11. *R. Pascanu, T. Mikolov, Y. Bengio.* On the difficulty of training Recurrent Neural Networks, CoRR, 2012 [12] Hochreiter, S. and Schmidhuber, J. (1996). Bridging long time lags by weight guessing and Long Short-Term Memory. In F.Silva, J.Principe, L.Almeida, Spatiotemporal models in biological and artificial systems
12. *S. Muzychka, A. Romanenko, I. Piontkovskaja.* Conditional Random Field for morphological disambiguation in Russian., Conference Dialog-2014, Bekasovo, 2014
13. *T. Joachims.* Optimizing Search Engines using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining, 2003

Сведения об авторе:

Кудинов Михаил Сергеевич,
родился в 1990 году в городе Усолье-Сибирское Иркутской области. В 2012 году закончил отделение теоретической и прикладной лингвистики филологического факультета МГУ и поступил в аспирантуру Вычислительного центра РАН им. Дородницына. В настоящий момент является сотрудником исследовательского центра «Самсунг» и аспирантом Вычислительного центра им. А.А. Дородницына Российской академии наук. В область научных интересов входят задачи, связанные с обработкой естественного языка – как текста, так и речи: анализ и извлечение информации из текста, компьютерная лингвистика, вопросно-ответные системы, в сфере речевых технологий особый интерес представляют языковые модели, основанные на нейронных сетях. E-mail: mikhailkudinov@gmail.com