

Лингвистические регулярности в явных и нейросетевых моделях векторного представления слов

Константин Архипенко

Содержание

- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Содержание

- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Нейросетевые модели

- Dense векторы
- Каждое измерение – это скрытый признак
- Наиболее популярный программный инструмент – **word2vec**

$$Italy : (-0.093, 0.156, -0.107, \dots) \in \mathbb{R}^{100}$$

word2vec

- Continuous Bag-of-Words – предсказание слова по его контексту
- Skip-gram – предсказание по слову его контекста, разбивается на n отдельных предсказаний
- Дерево Хаффмана – лог-линейная сложность!

$$CBOW : O(D \log |V|); SG : O(nD \log |V|)$$

Явные модели

- Разреженные (sparse) векторы
- Каждое измерение – это явный контекст слова
- Под контекстом слова здесь понимаются определенное слово и его относительная позиция, а не множество слов
- Будем использовать метрику ассоциации **PPMI**

ω_1^{-2}	ω_2^{-2}	...	$\omega_{ V }^{-2}$	ω_1^{-1}	...	$\omega_{ V }^{+2}$
0.153	0.000	...	0.082	0.000	...	0.227

PPMI

- PMI – pointwise mutual information (поточечная взаимная информация)
- PPMI – positive PMI

$$PPMI(\omega, c) = \begin{cases} PMI(\omega, c) & \text{если } PMI(\omega, c) > 0 \\ 0 & \text{иначе} \end{cases}$$

$$PMI(\omega, c) = \log \frac{P(\omega, c)}{P(\omega)P(c)} = \log \frac{freq(\omega, c) | corpus |}{freq(\omega) freq(c)}$$

Содержание

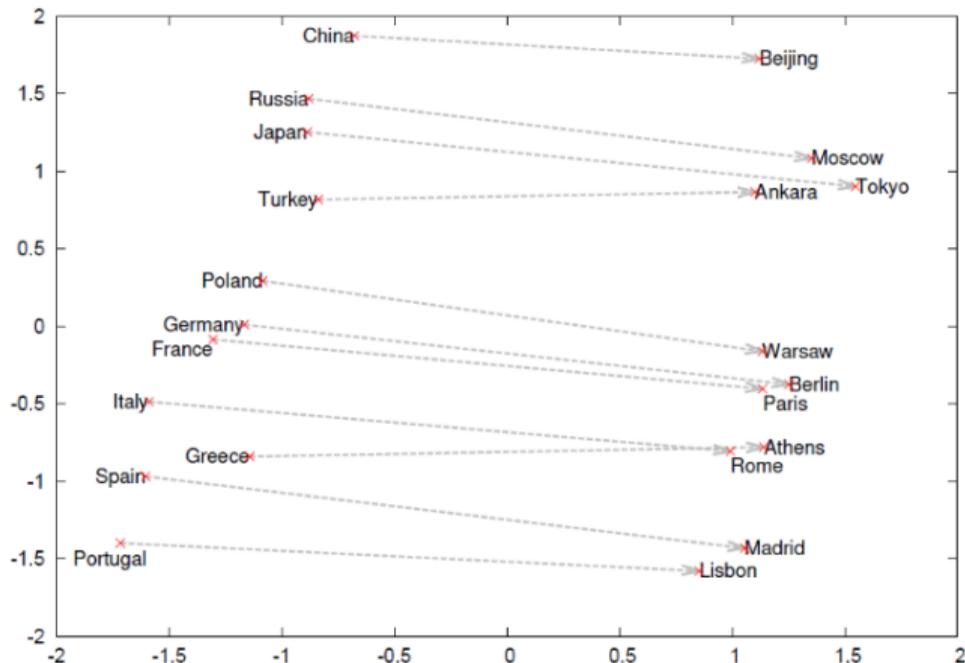
- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Mikolov et al. (2013)

- Векторные представления слов на основе нейронных сетей имеют интересную геометрию
- Они могут быть использованы для получения аналогий:

man относится к *woman* так же, как *king* относится к *queen*

Mikolov et al. (2013)



Mikolov et al. (2013)

man относится к *woman* так же, как *king* относится к *queen*

a относится к a^* так же, как b относится к b^*

$$a - a^* = b - b^*$$

$$b - a + a^* = b^*$$

$$\textit{king} - \textit{man} + \textit{woman} = \textit{queen}$$

Вопрос

Действительно ли получение аналогий является уникальной способностью нейронных сетей?

- **Эксперимент:** сравним нейросетевые векторные представления слов с явными

Содержание

- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Специальные датасеты для аналогий

- Четыре слова в аналогии: a относится к a^* так же, как b относится к b^*
- Даны три слова: a , a^* , b
- Необходимо найти наиболее подходящее слово b^* , отличное от первых трех:

$$b^* = \arg \max_{\omega} \text{sim}(\omega, b - a + a^*)$$
$$\text{sim}(x, y) = \frac{1 + \cos(x, y)}{2} = \frac{1}{2} + \frac{(x, y)}{2\|x\|\|y\|}$$

- **MSR**: 8000 синтаксических аналогий
- **Google**: 19000 синтаксических и семантических аналогий

Синтаксические и семантические аналогии

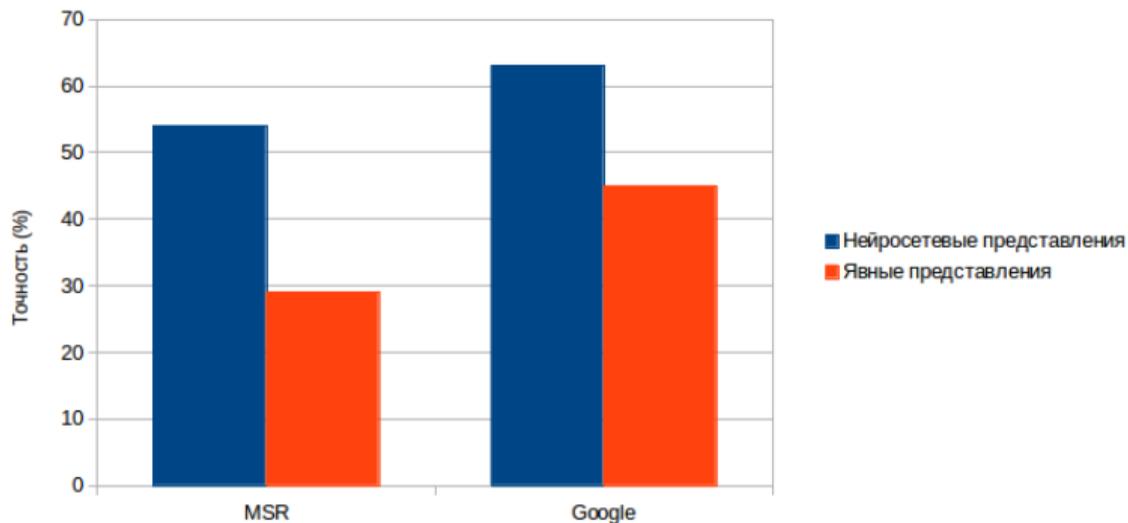
Синтаксические аналогии:

- Сравнительная и превосходная степени прилагательных: **good** относится к **best** так же, как **bad** относится к **?**
- Прилагательное → наречие: **bad** относится к **badly** так же, как **good** относится к **?**

Семантические аналогии:

- Столицы государств: **Russia** относится к **Moscow** так же, как **China** относится к **?**
- Мужской род → женский род: **husband** относится к **wife** так же, как **son** относится к **?**

Раунд 1



Нейронные сети оказались значительно лучше

Содержание

- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Внимательный взгляд на целевую функцию

- Мы хотели найти ближайшее слово b^* к $b - a + a^*$
- Следующим способом: $b^* = \arg \max_{\omega} \text{sim}(\omega, b - a + a^*)$
- Если все векторы, соответствующие словам, нормированы, то

$$\begin{aligned} b^* &= \arg \max_{\omega} \text{sim}(\omega, b - a + a^*) = \\ &= \arg \max_{\omega} (\text{sim}(\omega, b) - \text{sim}(\omega, a) + \text{sim}(\omega, a^*)) \end{aligned}$$

Что мы делаем не так

England – London + Baghdad =

Что мы делаем не так

$$\textit{England} - \textit{London} + \textit{Baghdad} = \textit{Mosul}$$

$$\text{sim}(\textit{Iraq}, \textit{England}) = 0.15; \text{sim}(\textit{Mosul}, \textit{England}) = 0.13$$

$$\text{sim}(\textit{Iraq}, \textit{London}) = 0.13; \text{sim}(\textit{Mosul}, \textit{London}) = 0.14$$

$$\text{sim}(\textit{Iraq}, \textit{Baghdad}) = 0.63; \text{sim}(\textit{Mosul}, \textit{Baghdad}) = 0.75$$

Что мы делаем не так

- Проблема доминирования одного слагаемого
- Гораздо сильнее проявляется в явных векторных представлениях
- Возможно, поэтому они плохо справились с задачей

Как сделать лучше?

Вместо аддитивной целевой функции будем использовать мультипликативную!

$$\arg \max_{\omega} (\text{sim}(\omega, b) - \text{sim}(\omega, a) + \text{sim}(\omega, a^*)) \rightarrow$$

$$\rightarrow \arg \max_{\omega} \frac{\text{sim}(\omega, b)\text{sim}(\omega, a^*)}{\text{sim}(\omega, a)}$$

Как сделать лучше?

Вместо аддитивной целевой функции будем использовать мультипликативную!

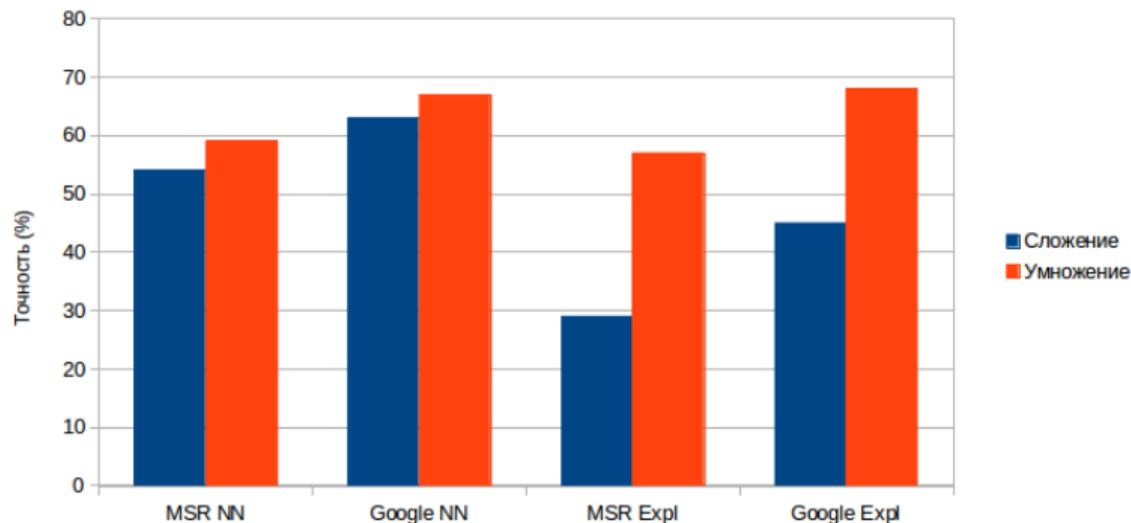
$$\arg \max_{\omega} (\text{sim}(\omega, b) - \text{sim}(\omega, a) + \text{sim}(\omega, a^*)) \rightarrow$$

$$\rightarrow \arg \max_{\omega} \frac{\text{sim}(\omega, b)\text{sim}(\omega, a^*)}{\text{sim}(\omega, a) + \varepsilon}$$

Содержание

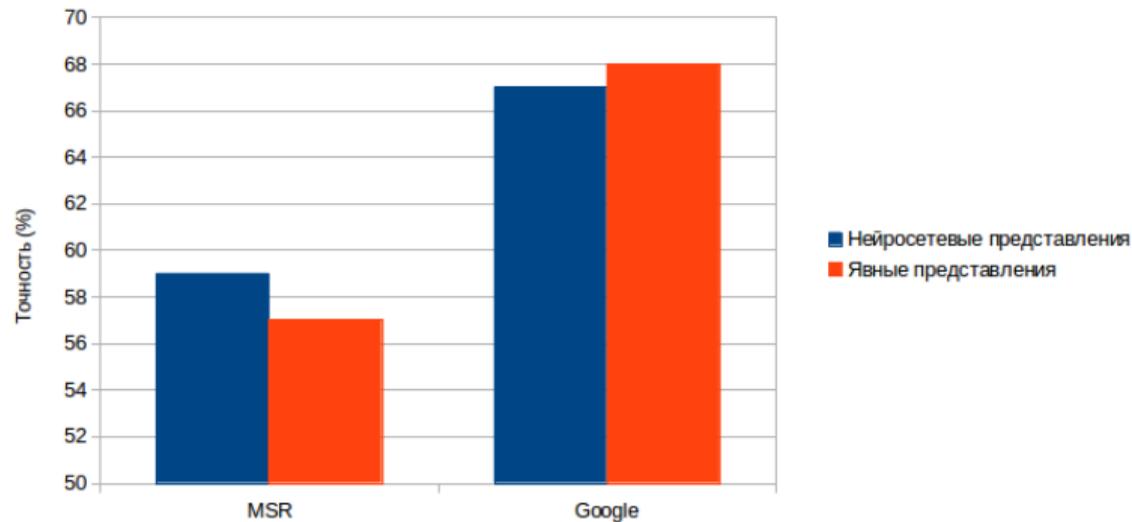
- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 **Неужели нейронные сети настолько лучше?**
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Сложение → умножение



Значительно улучшились результаты для явных представлений

Раунд 2



Примерно одинаковые результаты

Выводы

- Плох был сам метод поиска аналогий, а не явные представления слов
- Никакой магии в нейросетях нет – с мультипликативной целевой функцией они показывают с явными представлениями схожие результаты

Содержание

- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Согласованность результатов

Датасет	NN и Expl	Никто	Только NN	Только Expl
MSR	43.97%	28.06%	15.12%	12.85%
Google	57.12%	22.17%	9.59%	11.12%
Суммарно	53.58%	23.76%	11.08%	11.59%

Результаты для разных типов аналогий

	Relation	Embedding	Explicit
GOOGLE	capital-common-countries	90.51%	99.41%
	capital-world	77.61%	92.73%
	city-in-state	56.95%	64.69%
	currency	14.55%	10.53%
	family (gender inflections)	76.48%	60.08%
	gram1-adjective-to-adverb	24.29%	14.01%
	gram2-opposite	37.07%	28.94%
	gram3-comparative	86.11%	77.85%
	gram4-superlative	56.72%	63.45%
	gram5-present-participle	63.35%	65.06%
	gram6-nationality-adjective	89.37%	90.56%
	gram7-past-tense	65.83%	48.85%
gram8-plural (nouns)	72.15%	76.05%	
gram9-plural-verbs	71.15%	55.75%	
MSR	adjectives	45.88%	56.46%
	nouns	56.96%	63.07%
	verbs	69.90%	52.97%

Результаты для разных типов аналогий

- Явные представления несколько лучше справляются с семантическими аналогиями, связанными с географией, а также с превосходной степенью прилагательных и существительными
- Нейросетевые представления выигрывают в аналогиях с глаголами, сравнительной степенью прилагательных, антонимами, а также в аналогиях "мужской род – женский род"

Содержание

- 1 Сравнение явных и нейросетевых моделей
 - Обзор моделей
 - Аддитивность векторных представлений слов
 - Экспериментальное сравнение: первый раунд
- 2 Неужели нейронные сети настолько лучше?
 - Мультипликативная целевая функция
 - Экспериментальное сравнение: второй раунд
- 3 Анализ ошибок нейросетевых и явных представлений
 - Результаты для разных типов аналогий
 - Проблема стандартного поведения

Проблема стандартного поведения

RELATION	WORD	EMB	EXP
gram7-past-tense	who	0	138
city-in-state	fresno	82	24
gram6-nationality-adjective	slovak	39	39
gram6-nationality-adjective	argentine	37	39
gram6-nationality-adjective	belarusian	37	39
gram8-plural (nouns)	colour	36	35
gram3-comparative	higher	34	35
city-in-state	smith	1	61
gram7-past-tense	and	0	49
gram1-adjective-to-adverb	be	0	47
family (gender inflections)	daughter	8	47
city-in-state	illinois	3	40
currency	currency	5	40
gram1-adjective-to-adverb	and	0	39
gram7-past-tense	enhance	39	20

Проблема стандартного поведения

- В качестве ответа на многие аналогии дается одно слово, вектор которого расположен близко к центру семантического типа этих аналогий
- Несколько сильнее проявляется в явных векторных представлениях слов

Заключение

- В задачах аналогии нейросетевые векторные представления слов не обладают никакой магией
- С этими задачами они справляются так же хорошо, как и явные представления слов
- Нейросетевые и явные представления по-разному справляются с разными аналогиями и не являются идеальными

Литература

-  O. Levy, Y. Goldberg.
Linguistic Regularities in Sparse and Explicit Word Representations.
CoNLL, 2014.
-  T. Mikolov et al.
Efficient Estimation of Word Representations in Vector Space
CoRR, 2013.