



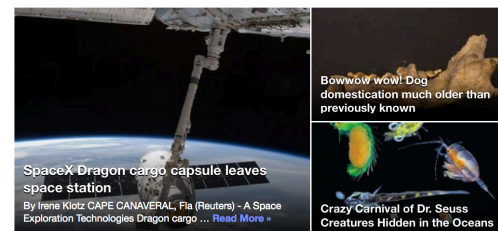
Hierarchical Neural Language Models for Joint Representation of Streaming Documents and their Content

Nemanja Djuric*, Hao Wu*, Vladan Radosavljevic, Mihajlo Grbovic, Narayan Bhamidipati
Florence, May 22nd, 2015

Introduction

- Neural language models have recently gained in popularity
 - Efficiently learn low-dimensional representations of words (word2vec)
 - A number of applications outside NLP (graph mining, smartphone apps, ...)
- We consider learning distributed representations for documents
 - Prior work using neural models by Le and Mikolov (paragraph2vec)
- What if relationships between documents exist?
 - Our motivation: Model click-through logs on Yahoo News
 - Answers in a common thread, web pages visited by a user
- We propose an efficient, highly flexible neural language model that can capture such information

Science



Paracetamol/Tylenol in pregnancy may lower testosterone in boys

By Kate Kelland LONDON (Reuters) - Pregnant women who take the painkiller paracetamol regularly for long periods may put their unborn sons' testosterone levels at risk, leading to possible reproductive
[Reuters](#)



The sci-fi idea that we'll soon be growing babies in artificial wombs has 3 major problems

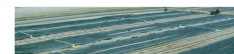
Imagine a future in which babies are born not from women, but from machines that replicate the...
[Business Insider](#)

Depression Linked with Parkinson's Disease Risk

People who have been diagnosed with depression may have an increased risk of developing Parkinson's disease later on, a new study suggests. In the study, researchers looked at about 140,000 people
[LiveScience.com](#)

Ancient Bacteria Sweat Fuel Under New Mexico's Desert Sun

May 20 ... Under the desert sun of New Mexico, scien...



Word and document embeddings

- Find appropriate representation to get meaningful insights and to be able to apply machine learning methods on such data
- Many methods used in practice
 - Bag-of-words, probabilistic LSA, Latent Dirichlet Allocation
- Neural language models
 - Probabilistic predictions of a word given the surrounding ones
 - Efficient training using recently proposed word2vec
 - Extended to document representation using paragraph2vec
- For our use case the methods not fully applicable
 - e.g., assume temporally close streaming documents are similar
 - How to add such constraints?

Neural language models

- User specifies neighborhood size and dimensionality of a word
- Continuous bag-of-words

$$\mathcal{L} = \sum_{t=1}^T \log \mathbb{P}(w_t | w_{t-c} : w_{t+c})$$

$$\mathbb{P}(w_t | w_{t-c} : w_{t+c}) = \frac{\exp(\bar{\mathbf{v}}^\top \mathbf{v}'_{w_t})}{\sum_{w=1}^W \exp(\bar{\mathbf{v}}^\top \mathbf{v}'_w)}$$

- Skip-gram

$$\mathcal{L} = \sum_{t=1}^T \log \mathbb{P}(w_{t-c} : w_{t+c} | w_t)$$

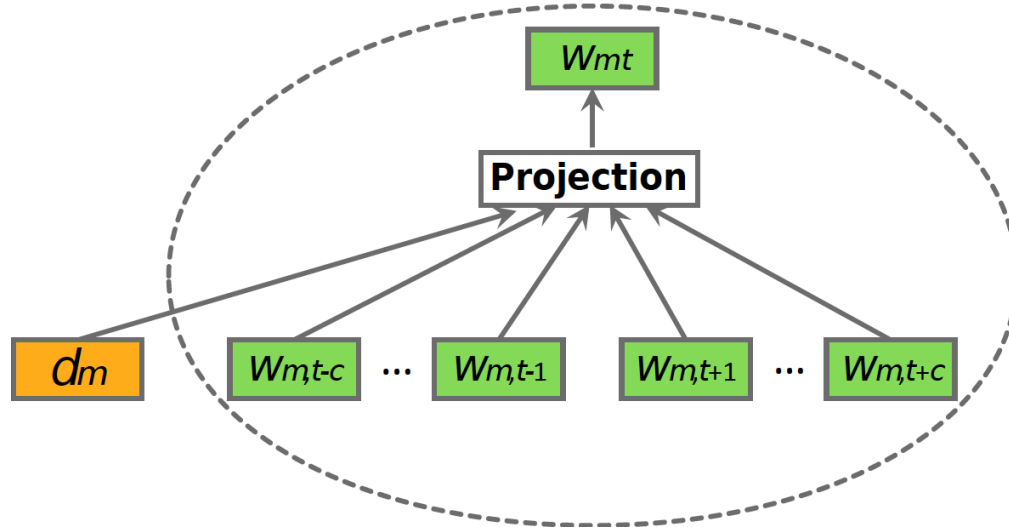
$$\mathbb{P}(w_{t+j} | w_t) = \frac{\exp(\mathbf{v}_{w_t}^\top \mathbf{v}'_{w_{t+j}})}{\sum_{w=1}^W \exp(\mathbf{v}_{w_t}^\top \mathbf{v}'_w)}$$

- Trained efficiently using hierarchical soft-max or negative sampling (Mikolov et al., NIPS 2013)

Neural language models for document representation

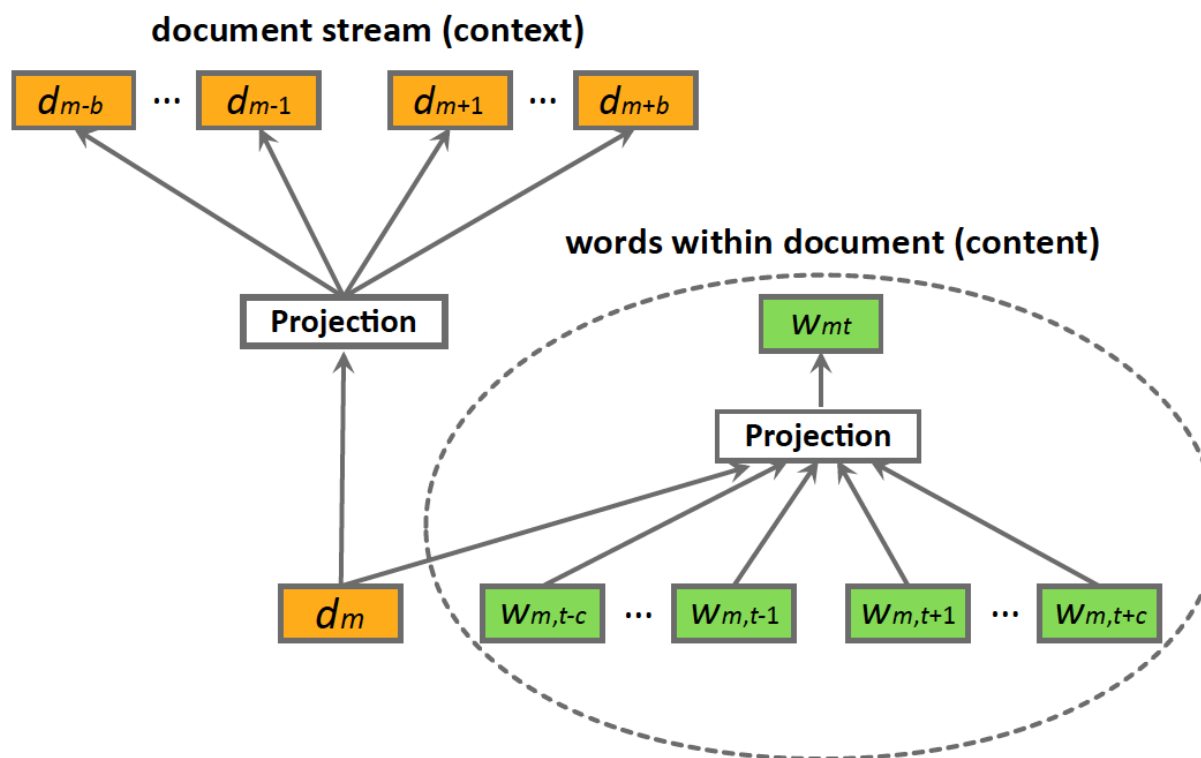
- *paragraph2vec* assumes additional “paragraph vector”, always in neighborhood of each word (Le and Mikolov, ICML 2014)
- As we scan input text during training, this vector is always updated

words within document (content)



The proposed method

- The existing approaches do not consider document interactions
- Solution: Hierarchical Document Vector (HDV) model



Methodology

- More specifically, we maximize the following function

$$\mathcal{L} = \sum_{s \in \mathcal{S}} \left(\alpha \sum_{d_m \in s} \sum_{w_{mt} \in d_m} \log \mathbb{P}(w_{mt} | w_{m,t-c} : w_{m,t+c}, d_m) + \right. \\ \left. \sum_{d_m \in s} \left(\alpha \log \mathbb{P}(d_m | w_{m1} : w_{mT}) + \sum_{-b \leq i \leq b, i \neq 0} \log \mathbb{P}(d_{m+i} | d_m) \right) \right)$$

- We model word probabilities given its document (in addition to surrounding words)
- We model document probability given surrounding documents
- Hyperparameter α trades off between optimization of document and word sequences
- The optimization problem is solved using stochastic gradient descent

Experiments – Movie genre categorization

- We combined MovieLens and IMDB data
 - MovieLens: ratings for 10,000 genre-specified movies generated by 70,000 users
 - IMDB: synopsis for each movie
- Filtered all movies that were rated 3 or below (on a scale 1 to 5), ordered them chronologically per user by rating's timestamp
 - Movie sequence treated as “document context”, synopsis as “document content”
- We learned movie representations using the following algorithms
 - LDA (content only)
 - paragraph2vec (content only)
 - word2vec (context only)
 - HDV (content + context)
- We set dimensionality to $D = 100$ and neighborhood size to $b = c = 5$

Experiments – Movie genre categorization

- Having learned the movie representations, we use them as an input to a classification algorithm
 - Each movie labeled as “horror”, “action”, “comedy”, ...
 - We used linear Support Vector Machine as a binary classifier

Algorithm	drama	comedy	thriller	romance	action	crime	adventure	horror
LDA	0.5544	0.5856	0.8158	0.8173	0.8745	0.8685	0.8765	0.9063
paragraph2vec	0.6367	0.6767	0.7958	0.7919	0.8193	0.8537	0.8524	0.8699
word2vec	0.7172	0.7449	0.8102	0.8204	0.8627	0.8692	0.8768	0.9231
HDV	0.7274	0.7487	0.8201	0.8233	0.8814	0.8728	0.8854	0.9872

- word2vec achieved higher accuracy than paragraph2vec
- HDV outperformed the competing approaches (due to advantage of both context and content modeling)

Experiments – Large-scale document representation

- We generated data set using 4 months of Yahoo News click logs
 - Nearly 200,000 distinct news stories
 - For a sample of Yahoo users, we sorted read news stories by the timestamp
 - More than 80 million news sequences generated in this way
- We illustrate a wide potential of this approach for online applications
 - Keyword suggestion
 - Document retrieval
 - Document tag recommendation
 - Document recommendation
 - News topic classification

Experiments – Large-scale document representation

- Given a query keyword, suggest related keywords

movies	batman	woods	hijack	tennis
characters	superman	tiger	hijacked	singles
films	superhero	masters	whereabouts	masters
studio	gotham	holes	transponders	djokovic
audiences	comics	golf	autopilot	nadal
actors	trilogy	hole	radars	federer
feature	avenger	pga	hijackers	celebration
pictures	sci	classic	turnback	sharapova
drama	sequel	par	hijacking	atp
comedy	marvel	doral	decompression	slam
audience	prequel	mcilroy	baffling	roger

- Given a query keyword, suggest related news articles

movies	tennis
'American Hustle,' 'Wolf of Wall Street' Lead Nominations	Tennis-Venus through to third round, Li handed walkover
3 Reasons 'Jurassic World' Is Headed in the Right Direction	Nadal rips Hewitt, Serena and Sharapova survive at Miami
Irish Film and TV academy lines up stellar guest list for awards	Williams battles on at Sony Open in front of empty seats
10 things the Academy Awards won't say	Serena, Sharapova again on Miami collision course
I saw Veronica Mars, thanks to \$35 donation, 2 apps & \$8 ticket	Wawrinka survives bumpy start to Sony Open
boxing	hijack
Yushin Okami's Debut for WSOF 9 in Las Vegas	Thai radar might have tracked missing plane
UFC Champ Jon Jones Denies Daniel Cormier Title Shot Request	Criminal probe under way in Malaysia plane drama
UFC contender Alex Gustafsson staring at a no-win situation	Live: Malaysia asks India to join the expanding search
Alvarez back as Molina, Santa Cruz defend boxing titles	Malaysia dramatically expands search for missing jet
Anthony Birchak Creates MFC Championship Ring	Malaysia widening search for missing plane, says minister

Experiments – Large-scale document representation

- Given a news article, suggest related keywords

News articles	Related keywords
Serena beats Li for seventh Miami crown	hardcourts biscayne sharapova nadal nishikori aces unforced walkover angelique threeset
This year's best buy ISAs	isas pensioners savers oft annuity isa pots taxfree nomakeupselfie allowance
Galaxy S5 will get off to a painfully slow start in Samsung's home market	mwc quadcore snapdragon oneplus ghz appleinsider samsung lumia handset android
'Star Wars Episode VII': Actors Battle for Lead Role (EXCLUSIVE)	reboot mutants anthology sequels prequel liv helmer vfx villains terminator
Western U.S. Republicans to urge appeals court to back gay marriage	lesbian primaries rowse beshear legislatures schuette heterosexual gubernatorial stockman lgbt

- Given a news article, suggest related news articles

Galaxy S5 will get off to a slow start in Samsung's home market
New specs revealed for one of 2014's most intriguing Android phones
LG G Pro 2 review: the evolutionary process
[video] HTC wins smartphone of the year
Samsung apparently still has a major role in Apple's iPhone 6
Samsung's new launch, the Galaxy S5, lacks innovative features
This year's best buy ISAs
Savings rates 'could rise' as NS&I launch new products
How to use an Isa to invest in property
Pensions: now you can have a blank canvas - not an annuity
Ed Balls' Budget Response Long on Jokes, a Bit Short on Analysis
Half a million borrowers to be repaid interest and charges

Experiments – Large-scale document representation

- We used linear SVM to classify news articles into one of 19 first-level topic tags from Yahoo's internal interest taxonomy
 - e.g., labels such as “home & garden” or “science”
 - Results after 5-fold cross validation
- The comparison with the competing approaches is given below

Algorithm	Avg. accuracy improvement
LDA	0.00%
paragraph2vec	0.27%
word2vec	2.26%
HDV	4.39%

- HDV outperformed the competing approaches

Conclusion

- We presented a flexible, efficient unsupervised learning framework to model the latent structure of streaming documents
 - Jointly modeling document context and document content
- Straightforward to extend to additional layers of hierarchy
 - e.g., we can model user relationships and learn their representations
- The method was evaluated on public and proprietary data
 - On the movie data we verified the usefulness of learned representations
 - On the news data we showcased vast application potential of the method

Thank you!

- Questions?

