

A Neural Network Approach to Quote Recommendation in Writings

PRESENTED BY: WU, ZHIQIANG

Outline

- Introduction
- Background
- Context Encoder
- Quote Encoder
- Comparison
- Result
- Conclusion

Introduction

- Quote is a language phenomenon of transcribing the saying of someone else.
- Proper usage of quote can usually make the statement more elegant and convincing.
- A task to automatically recommend the quotes one may need while writing by user's typing.



“You have to be burning with an idea, or a problem, or a wrong that you want to right. If you're not passionate enough from the start, you'll never stick it out.”

— Steve Jobs

Like

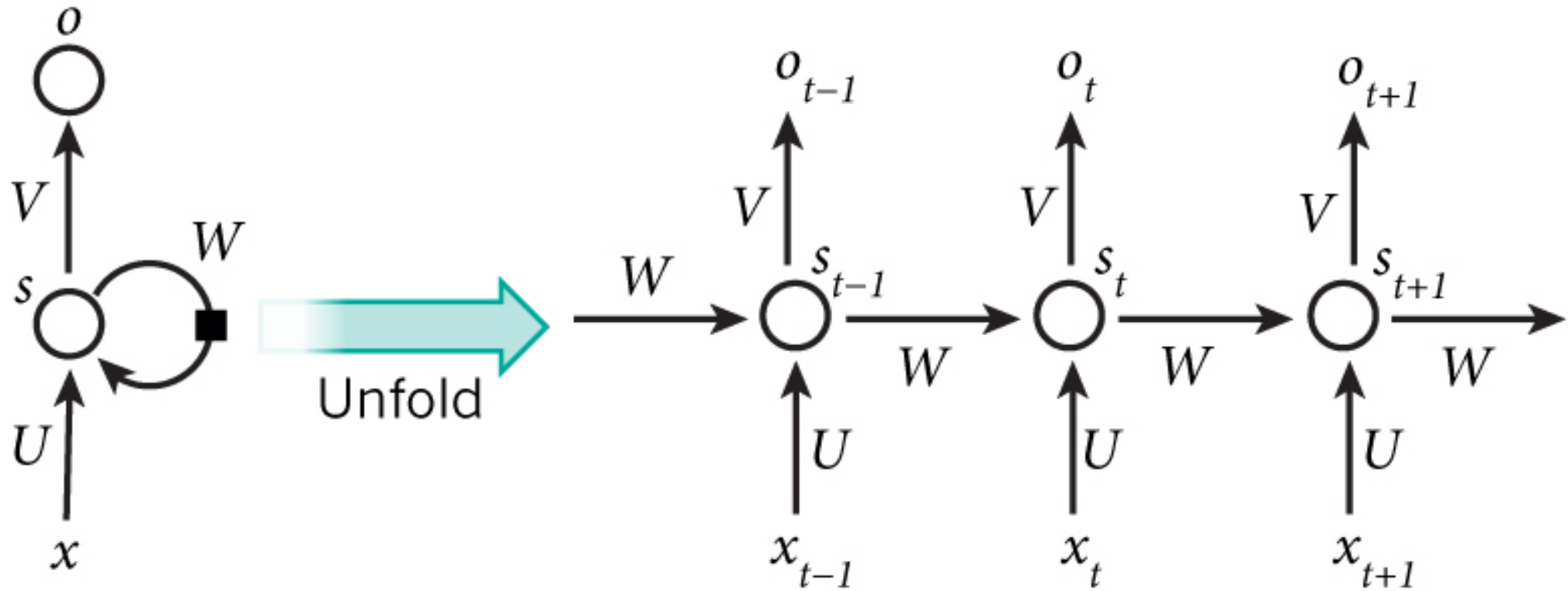
tags: [advice](#), [apple](#), [business](#), [creativity](#), [leadership](#), [leadership](#), [mac](#), [macintosh](#)

663 likes

What is the difference?

	Citation	Quotation
Length	Paper or document	Several sentences
Superficial Meaning	Same	May Different
Expression	Same	Different
Topics	Same	Similar
Ambiguity	No	Yes

Recurrent Neural Network (RNN)

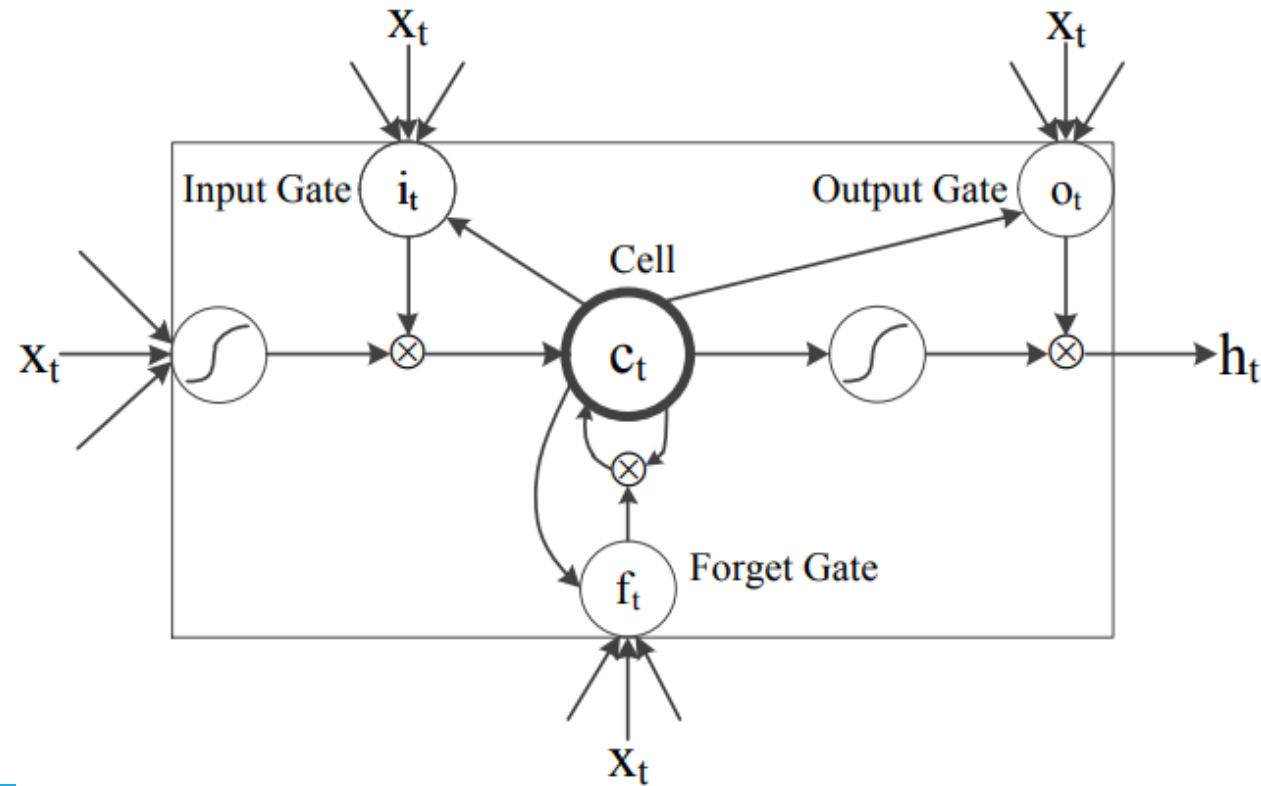


Recurrent Neural Network (RNN)

- $S_t = f(U_{x_t} + W_{S_{t-1}})$
- $o_t = f(V_{S_t})$
- RNN is able to capture the dependency of input sequence due to the recurrent structure.
- Drawbacks:
 - fails to capture the dependency
 - the problem of gradient vanishing and exploding

Long-Short Term Memory (LSTM)

- LSTM is a kind of recurrent neural network.
- Designed for the long term issue



Framework

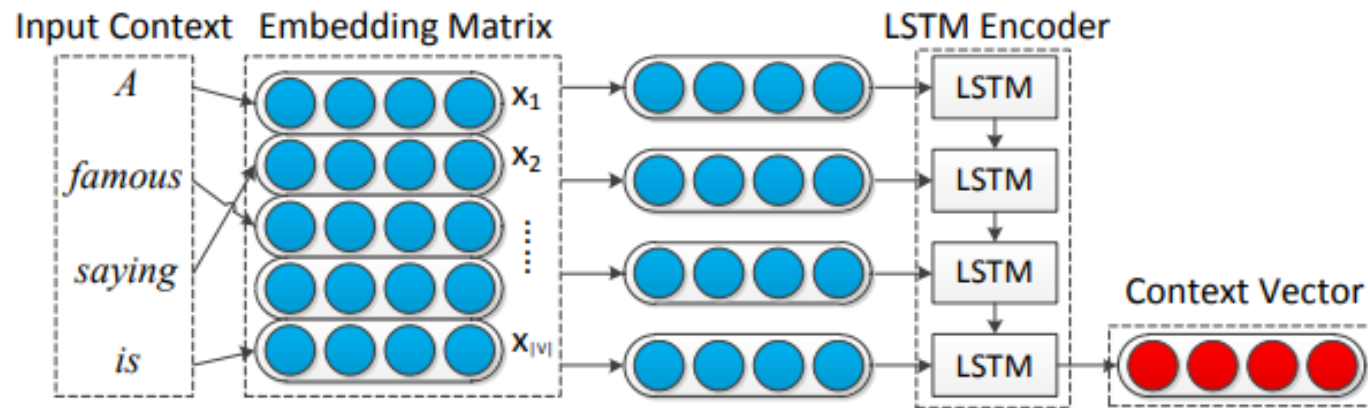
- Query context: V_c
- Quote vector: V_q
- Relevant scores $r_c = V_q V_c^T$ $r_c = [r_{cq_1} \quad \dots \quad r_{cq_{|Q|}}]$
- Candidate quotes are ranked according to the relevance scores as final recommendation list.
- **The same context will map to only one quote that called Gold Quote.**

Context Encoder

- The query context divided into two parts: left context and right context
- $V_c = V_l + V_r$
- To avoid the risk of too long input sequence

Laura said: -- “we are commanded to **[quote]** and the sun always shines at four oaks, for me.”

Context Encoder



Quote Encoder

- Difference:
 - Different expression from common writings
 - Different from typical meaning
 - Implicit meanings
 - Different writers

“Success is the prize for those who stand true to their ideas.” --

Josh S. Hinds

+ Attributions

489 Votes

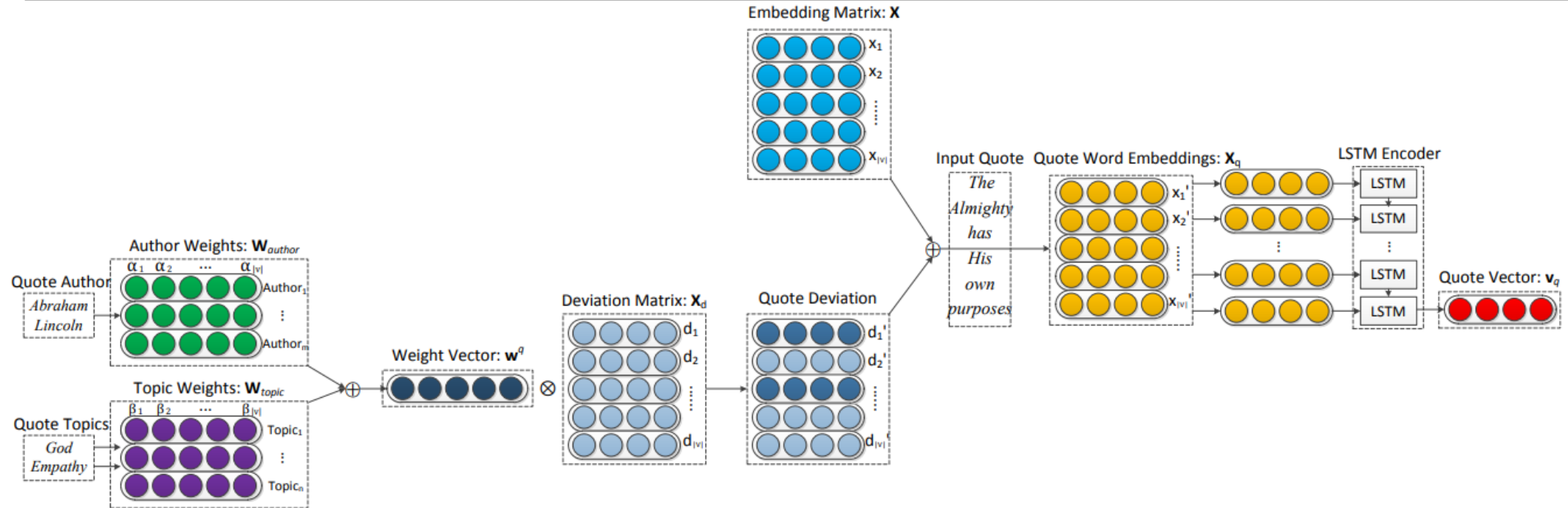
Life

Life Experiences

Success

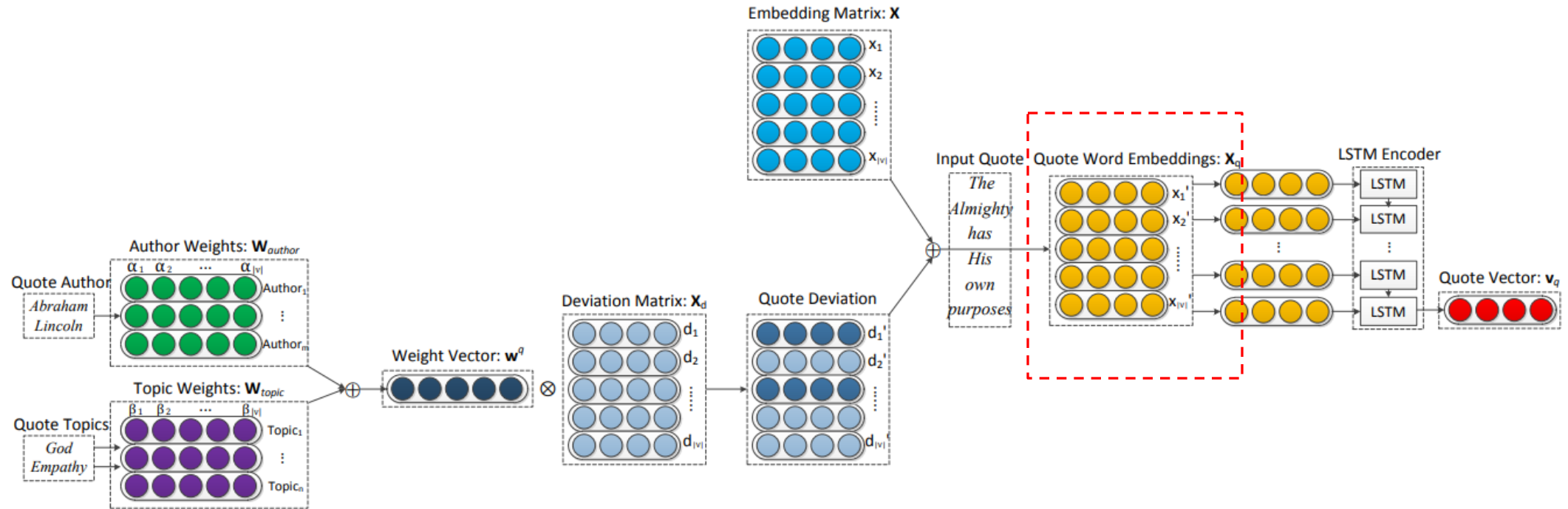
Truth

Quote Encoder

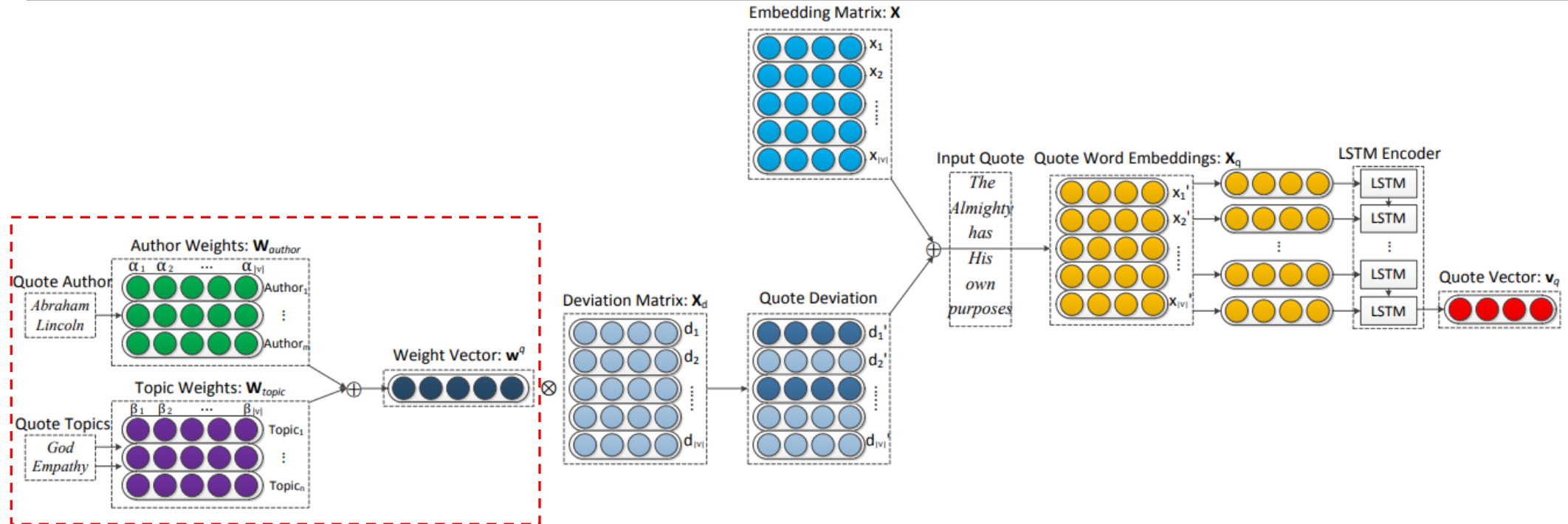


$$X_q = X + \text{diag}(w_{author}^q + w_{topic}^q) \times X_d$$

Quote Encoder

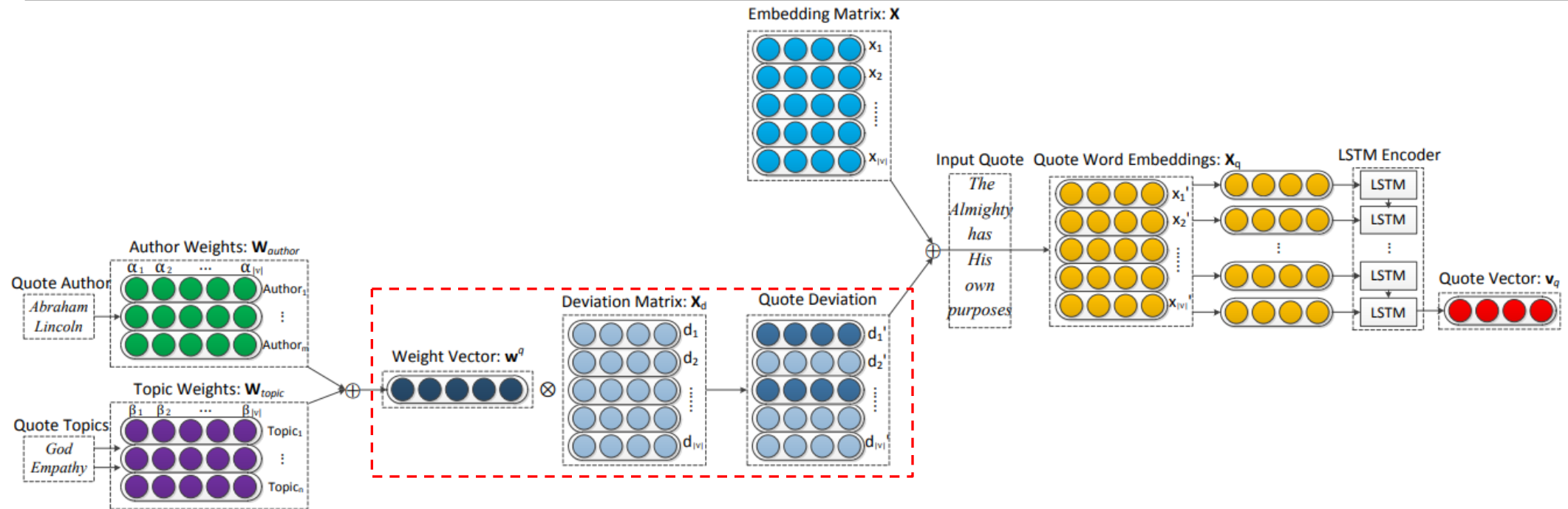


Quote Encoder



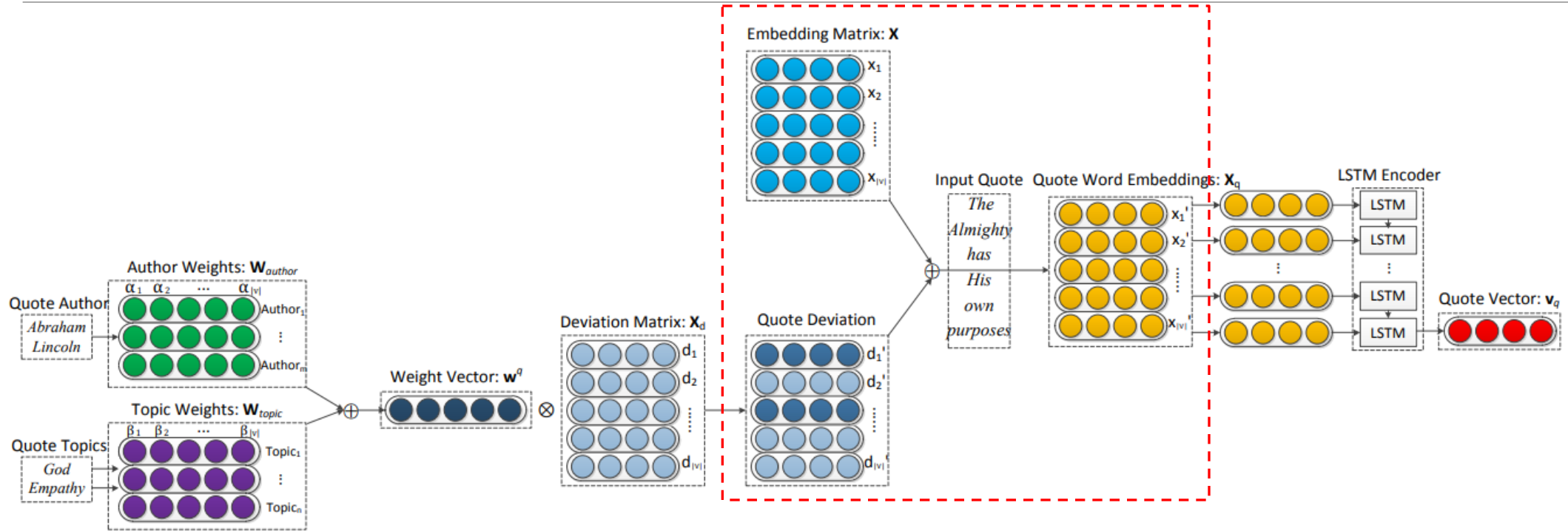
$$X_q = X + \text{diag} \left(w_{author}^q + w_{topic}^q \right) \times X_d$$

Quote Encoder



$$X_q = X + \text{diag}(w_{author}^q + w_{topic}^q) \times X_d$$

Quote Encoder



$$X_q = X + \text{diag}(w_{author}^q + w_{topic}^q) \times X_d$$

Evaluation Metrics

- Took 300 characters before and after each quote tag

- Training : Validation : Test = 8: 1 : 1

- Recall@k : $Recall@k = \frac{|G_k|}{|C|}$

- NDCG@k : $NDCG = \frac{1}{Z_n} \sum_{i=1}^n \frac{2^{r(i)}}{\log(i+1)}$

$$NDCG@k = \frac{1}{|C|} \sum_{c \in C} NDCG(c)$$

- MRR : $MRR = \frac{1}{|C|} \sum_{c \in C} \frac{1}{rank(c)}$

Dataset	C-Q pairs	Quotes	Authors	Topics
Training	51457	3061	748	814
Validation	6433	3090	754	816
Test	6433	3102	760	820
All	64323	3158	762	822

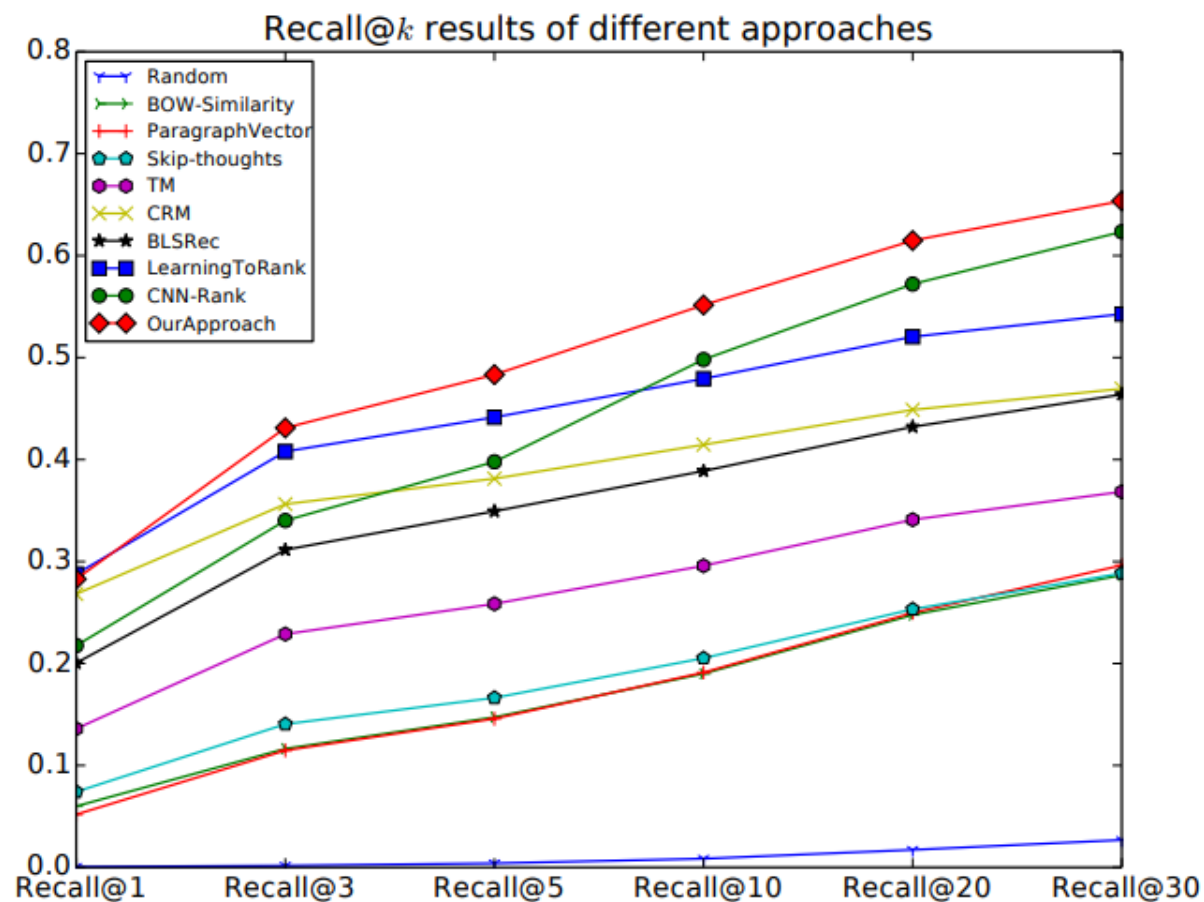
Comparison Approaches

Table 2: Results of different approaches.

Approach	MRR	NDCG@5
Random	0.006	0.002
BOW-Similarity	0.106	0.105
ParagraphVector	0.102	0.101
Skip-thoughts	0.123	0.123
TM	0.197	0.203
CRM	0.324	0.330
BLSRec	0.272	0.282
LearningToRank	0.360	0.372
CNN-Rank	0.315	0.313
LSTM	0.322	0.329
LSTM+	0.356	0.365
Our Approach	0.379*	0.392*

(* shows that the improvement over others are statistically significant with $p \ll 0.001$.)

Comparison Approaches



Results

to see the beauties of haying-time, with such voluntary labor; but I was too old and too much interested with my experiment to be cajoled by a lot of youngsters. I promised them a week of haying in each fifty-two, but that was all the concession I would make. Laura said: -- “we are commanded to **[quote]** and the sun always shines at four oaks, for me.” It was pretty of her to say that; but what else would one expect from Laura? The twelve acres from which the fodder oats had been cut were ploughed and fitted for sugar beets and turnips. I was not at all certain that the beets would do anything

[a] Make hay while the sun shines .

[b] New brooms sweep clean .

[c] We 're making progress .

[d] Least said soonest mended .

[e] There is a time for everything .

[f] No use crying over spilt milk

[g] Strike while the iron is hot .

[h] Delays are dangerous .

[i] Prevention is better than cure .

[j] Absence makes the heart grow fonder .

Conclusion

- Quote is hard to recommend due to the literary techniques and differ meanings.
- Query context divided into left and right parts.
- Make the relevance for author and topics of quotes
- create the weight vector for authors and topics of quotes
- The query context divided into left and right contexts for avoiding the risk of too long input
- Deviation embedding matrix is to learn different word embedding matrix for words in quotes
- The performance of their approach is better than existing methods. However, the recall@k is still lower when k is smaller.

Q&A

