



# Question pairs Identification

Cuiting Zhong  
Kirti Pande

Jui Gupta  
Sagar Chadha

# Problem statement

*“Identify if a pair of questions are similar in content”*

**Why?**



Re-use content



Simpler data storage

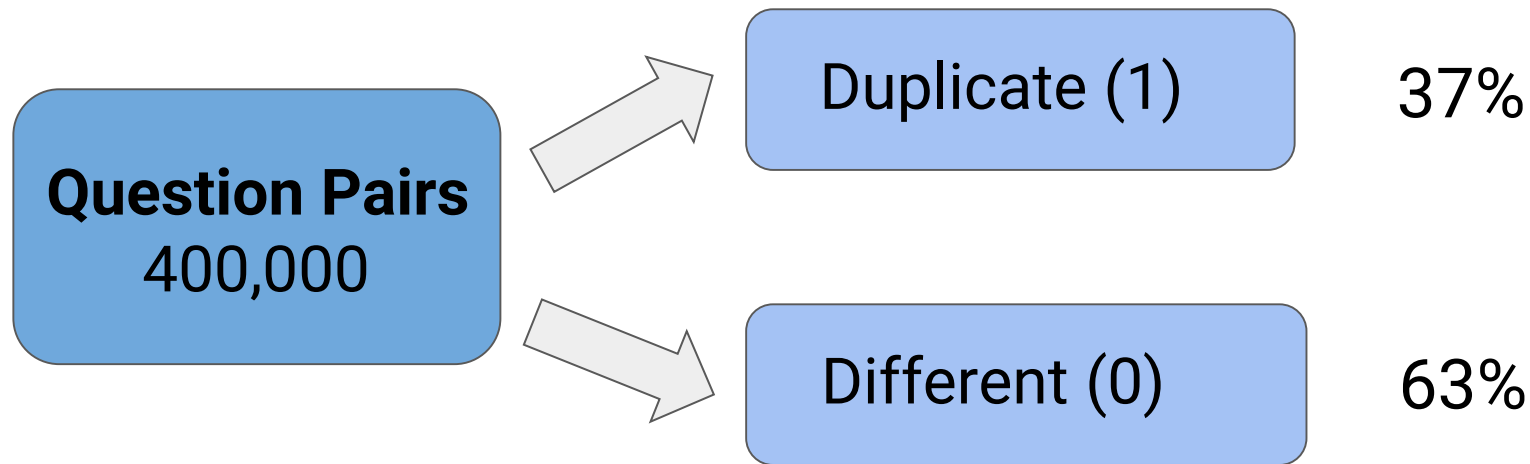


**\$\$**

Customer Experience

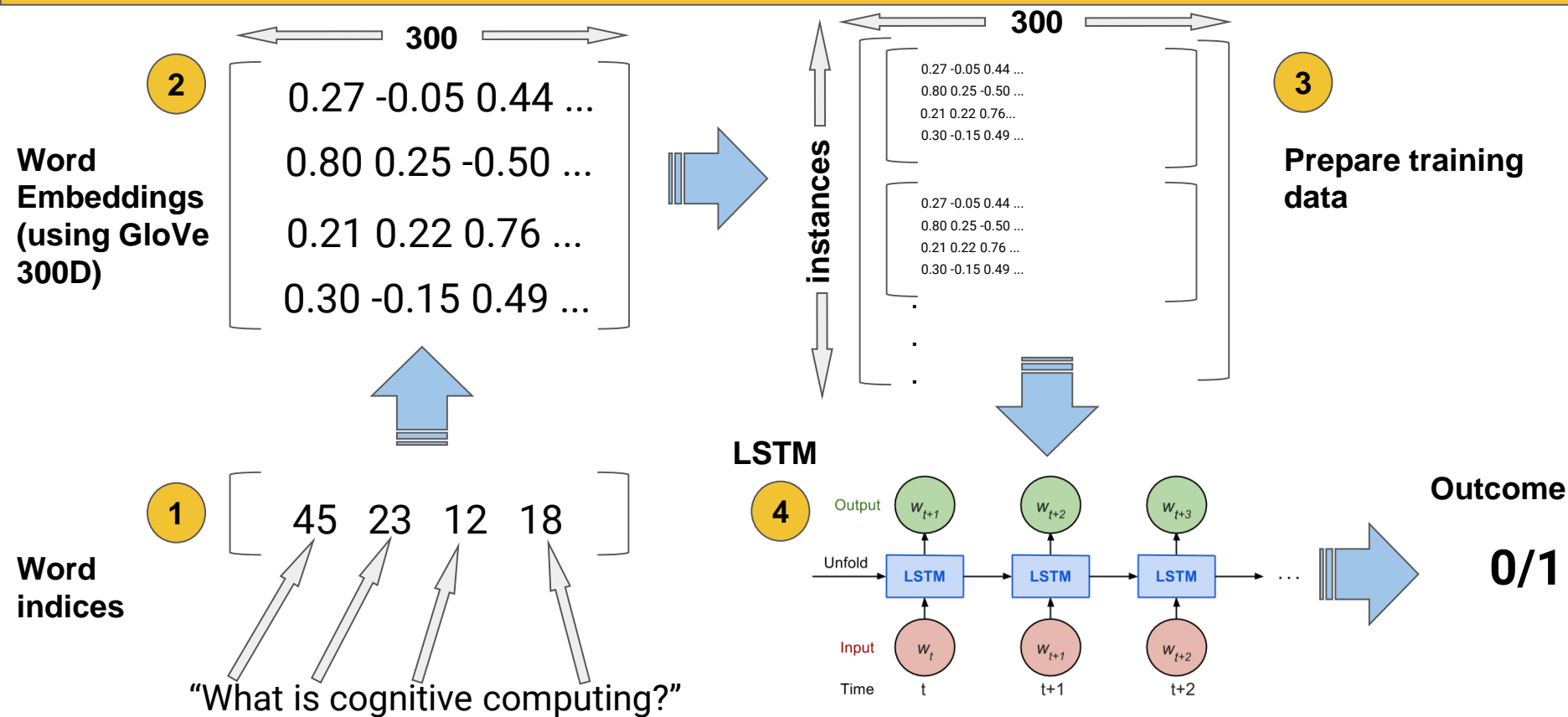
Data storage costs

# Data

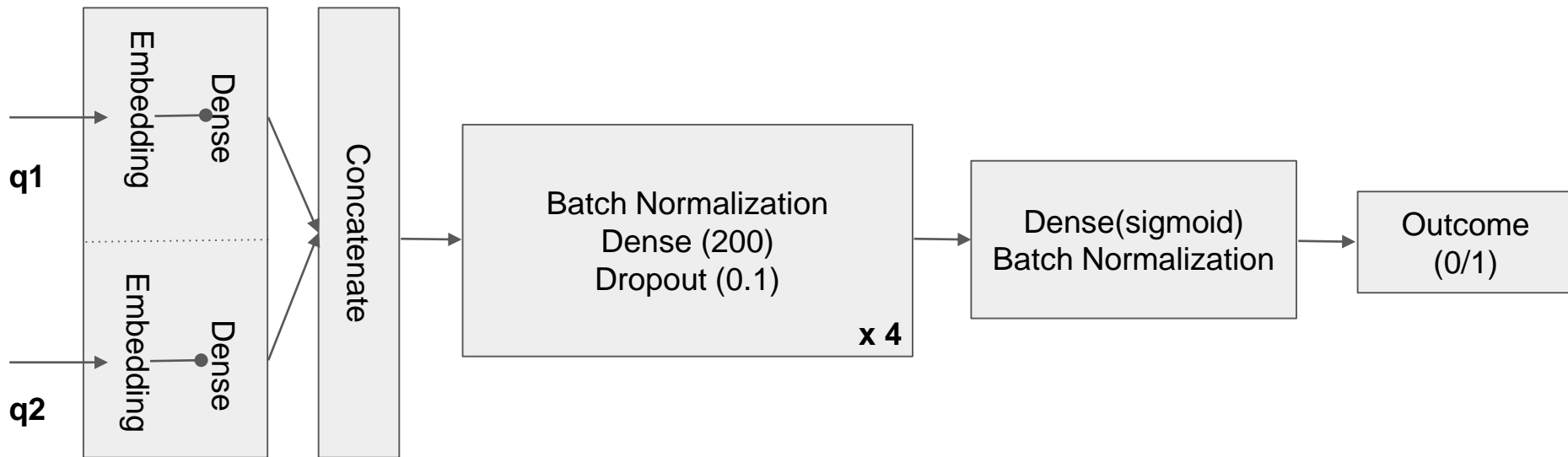


Question 1	Question 2	Is_duplicate (target)
Which one dissolves in water quickly sugar, salt, methane or carbon dioxide?	Which fish would survive in salt water?	0
Astrology: I am a Capricorn Sun Cap moon and cap rising...what does that say about me?	I'm a triple capricorn (Sun, moon and ascendant in Capricorn). What does this say about me?	1

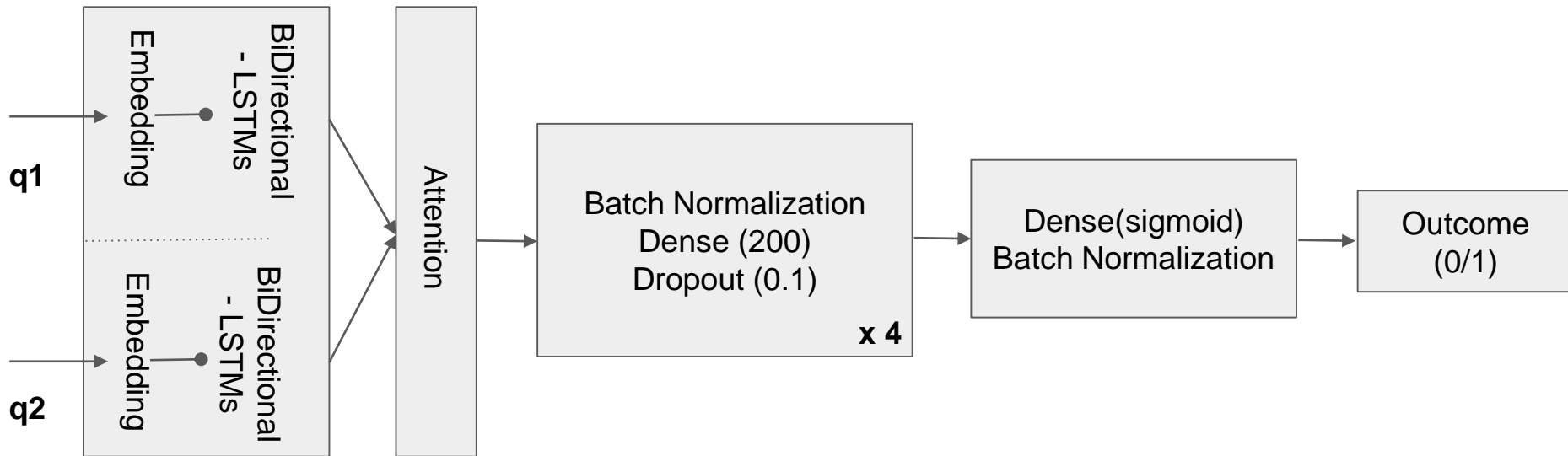
# Modelling Approach



# Modelling Approach: Bag of Embeddings



# Modelling Approach: Bi-LSTM with Attention



# Model Evaluation

**Loss function:** Binary cross entropy (Log Loss)

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

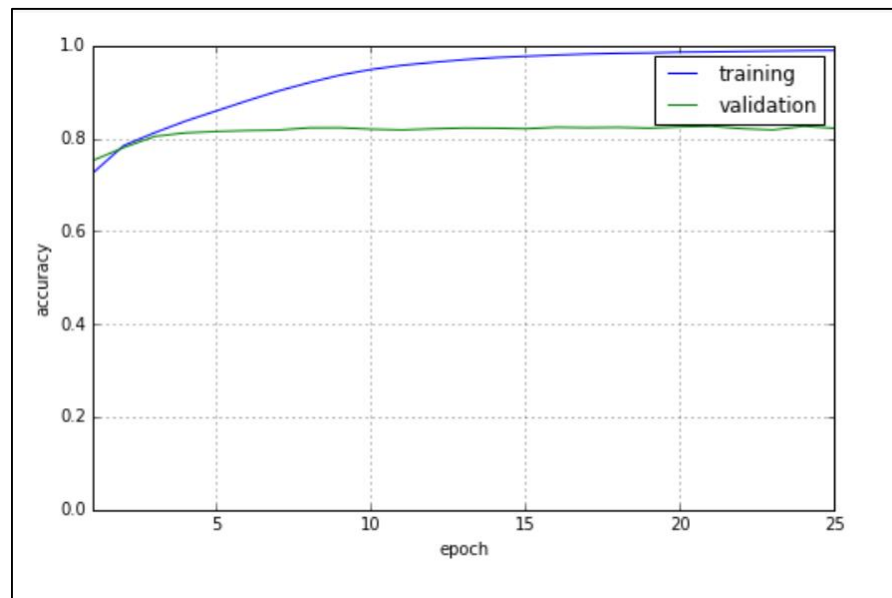
**Metric Scores -**

**Baseline accuracy: 63%**

**Validation accuracy: 82.63%**

**Test set accuracy: 82.43%**

**Learning Curve**



# Model Performance Comparison

**Baseline Accuracy: 63%**

(If all predictions are 0)

79.24%

Max, SGD,  
Sigmoid

80.83%

Max, Adam,  
Sigmoid

80.95%

Mean, Adam,  
Sigmoid



82.63%

Attention, Adam,  
Sigmoid



# Next steps

- Use different pre-trained embeddings for the model.  
e.g. Word2Vec, fasttext
- Try different similarity measure in embedding concatenation.  
e.g. Manhattan distance
- Extract and combine more other NLP features.  
e.g. number/proportion of common words
- Another interesting problem that utilizes the same concept is that of question answering using a context passage. We can attempt that.



**Thank you!**  
**Any 'duplicate' questions?**

# Appendix

# Approach

Data preprocessing

Generated tokenized word indices

Pre-trained GloVe word embeddings

Create word embedding matrices

Prepare training data and label tensors



# Applications

Improve efficiency of question answer websites ( Quora, Stack Overflow, Reddit)

ChatBots

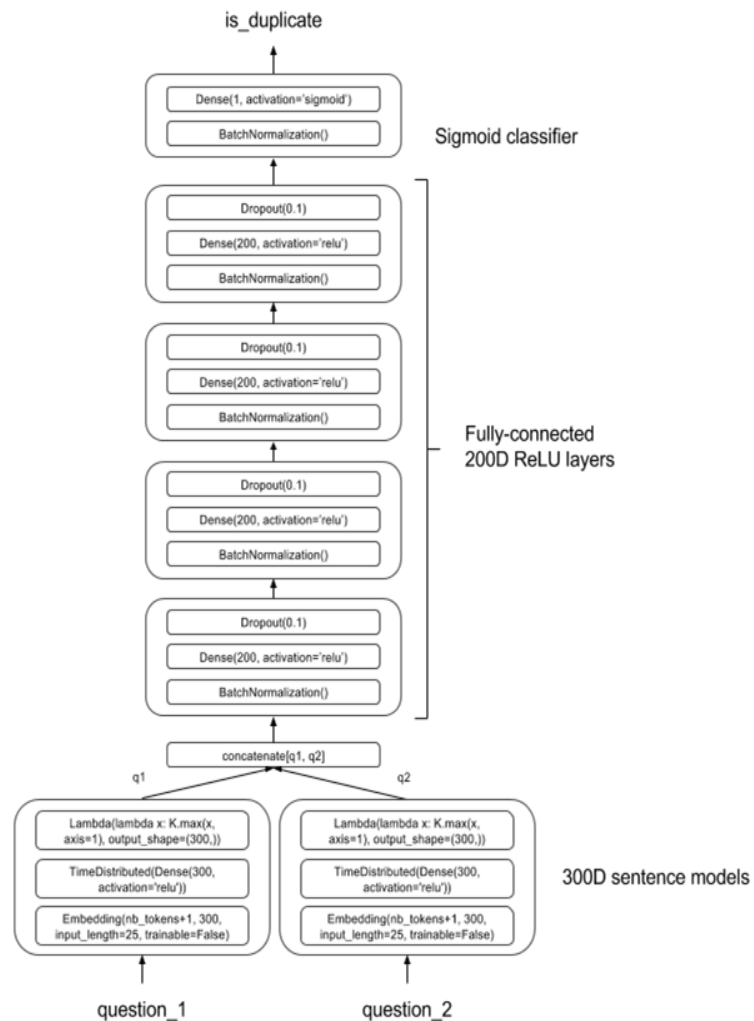
Enhance customer services and experience

Question generation systems for competitive exams

# Examples

Which one dissolve in water quikly sugar, salt, methane and carbon di oxide?	Which fish would survive in salt water?	0
Astrology: I am a Capricorn Sun Cap moon and cap rising...what does that say about me?	I'm a triple Capricorn (Sun, Moon and ascendant in Capricorn) What does this say about me?	1

# Approach : Model building





# Approach : Model with attention

