

Міністерство освіти і науки України  
Національний технічний університет України «Київський політехнічний інститут  
імені Ігоря Сікорського»  
Інститут прикладного системного аналізу  
Кафедра математичних методів системного аналізу

Дипломна робота на тему:

Порівняння моделей на основі нейронних мереж для  
задачі машинного розуміння тексту

Виконав студент групи КА-33

Ходак А.Р.

Науковий керівник:

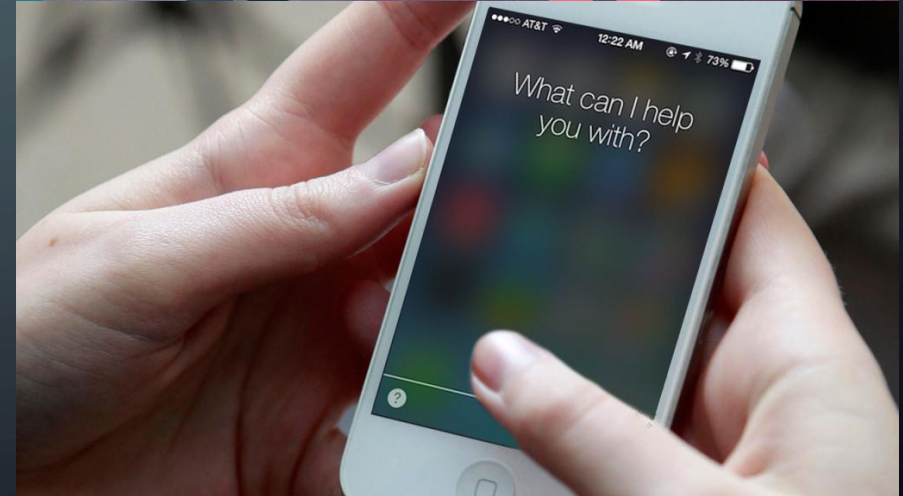
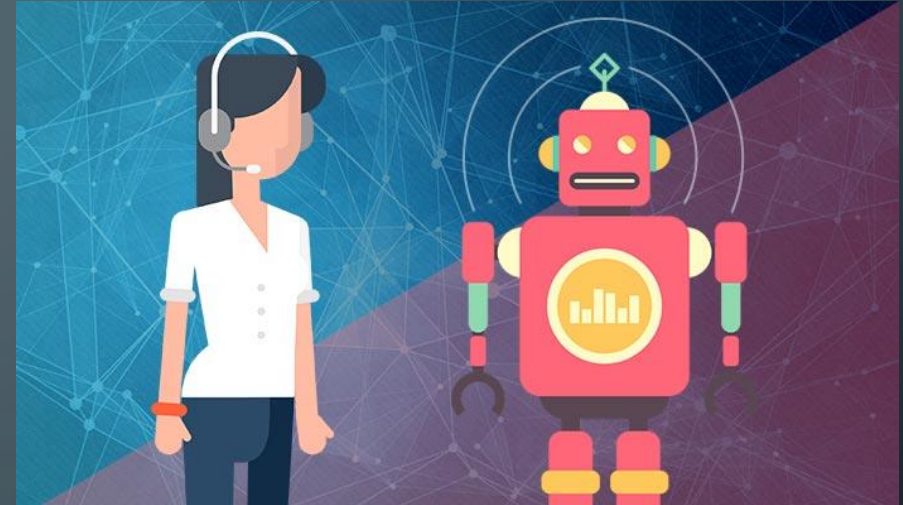
к.т.н, доцент Тимошенко Ю.О.

# ПОСТАНОВКА ЗАДАЧІ

- Маючи дві відкриті імплементації моделей нейронних мереж, порівняти результати їх роботи на двох нових наборах даних питання-відповідь SQuAD (2015 р.) та MS MARCO (грудень 2016 р.)

# АКТУАЛЬНІСТЬ РОБОТИ

- Поліпшення персональних асистентів
- Наукові асистенти
- Автоматизація служб підтримки
- Удосконалення чат-ботів



# ЗАДАЧА ВІДПОВІДІ НА ЗАПИТАННЯ НА ОСНОВІ ПРОЧИТАНОГО (RCQA)

- Набір даних представлений множиною трійок  $(Q, P, A)$ , де
  - $Q = (q_1, \dots, q_i, \dots, q_M)$  - питання довжиною  $M$  слів,
  - $P$  – уривок довжиною  $N$  слів,
  - $A = (a_b, a_c)$  - відрізок для відповіді,  $1 \leq a_b \leq a_c \leq N$ .
- Задача RCQA - оцінка умовної імовірності  $Pr(A|Q, P)$  та передбачення відповідей для тестових елементів

$$A^* = \mathop{arg \max}_{A \in A(P)} Pr(A|Q, P)$$

, де  $A(P)$  – множина відповідей-кандидатів в  $P$ .

# SQUAD (STANFORD QUESTION-ANSWERING DATASET)

- 2015 р.
- Перший набір даних такого розміру зі справжніми запитаннями-відповідями
- 100 тис. зразків
- Джерело: 536 статей англійської Вікіпедії
- 80% навчальна, 10% розробницька, 10% тестова вибірки
- Людська точність - F1: 77.0%, exact match: 86.8%

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?

**gravity**

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

**graupel**

Where do water droplets collide with ice crystals to form precipitation?

**within a cloud**

**Figure 1:** Question-answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage.

# MS MARCO

- 2016р., Microsoft Research
- 100 тис. пар питання-відповідь
- 1 млн контекстних абзаців
- Усі запитання – реальні запити до пошукової мережі Bing
- Відповіді створені вручну об'єднанням інформації з топ 10 релевантних результатів пошукової видачі
- Відповідь – не обов'язково відрізок тексту з контекстного уривку

Запит: **bulimia nervosa age range**

Тип запиту: numeric

Відповіді: [**'9 to 40 years old'**]

Абзаци: [ {"вибрано" : 1,

“текст абзацу”: Bulimia Nervosa Overview. Bulimia nervosa (bulimia) is characterized by repeatedly eating abnormally large amounts of food in a short period of time and then trying to compensate for the binge by purging (i.e., engaging in self-induced vomiting, using laxatives or diuretics), exercising excessively, or fasting. **The average age of onset for bulimia is 18 years, but it occurs in children as young as 9 and adults into their mid-40s and older.** Most bulimics go through a period of prolonged dieting or restricted eating before the cyclic binging-purging episodes begin ]]

# МЕТРИКИ

- SQuAD:

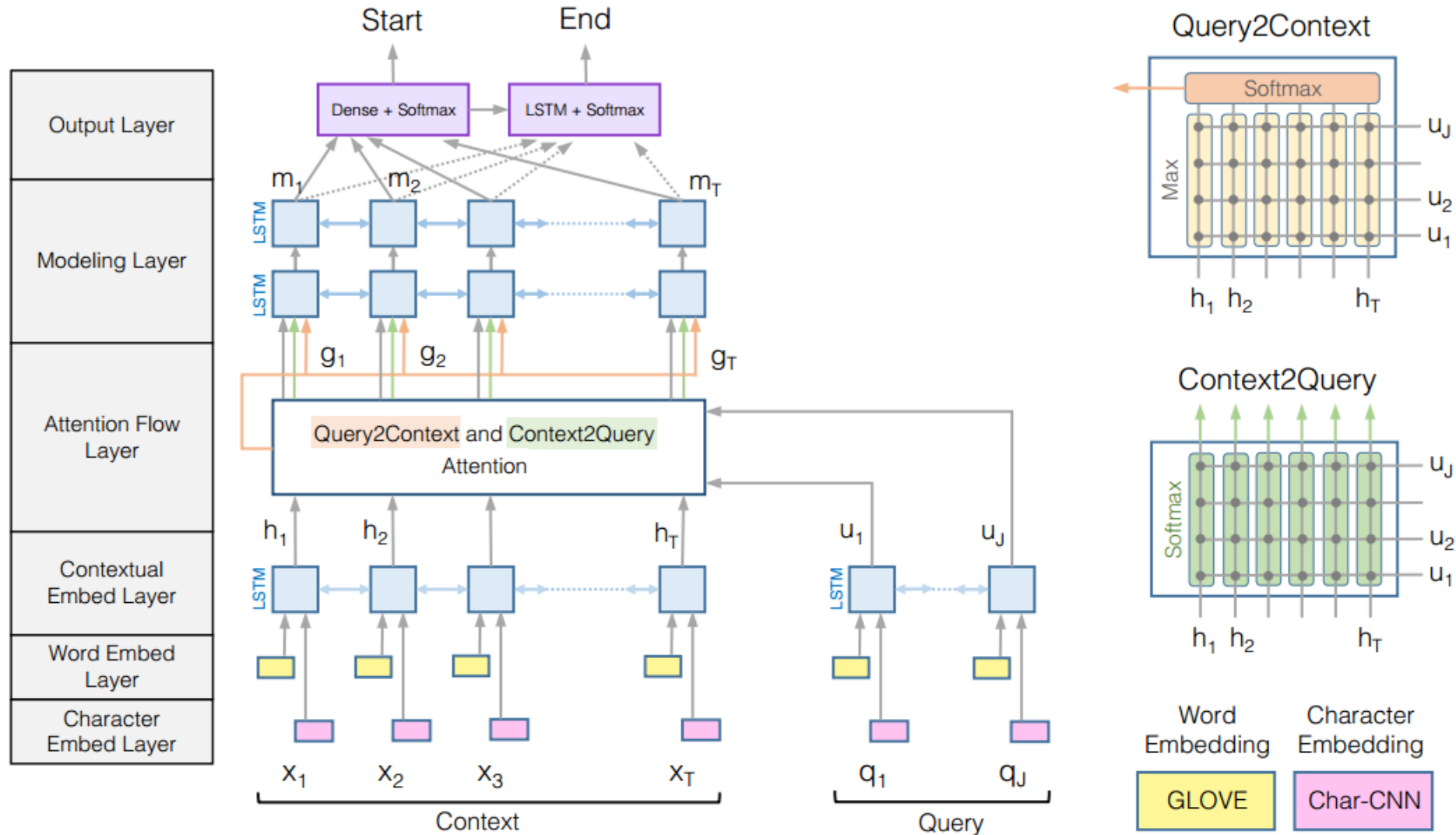
- f1:  $F1 = 2 * \frac{Precision * recall}{Precision + recall}$

- Exact match: відсоток відповідей, які точно відповідають одному з варіантів відповідей

- MS MARCO

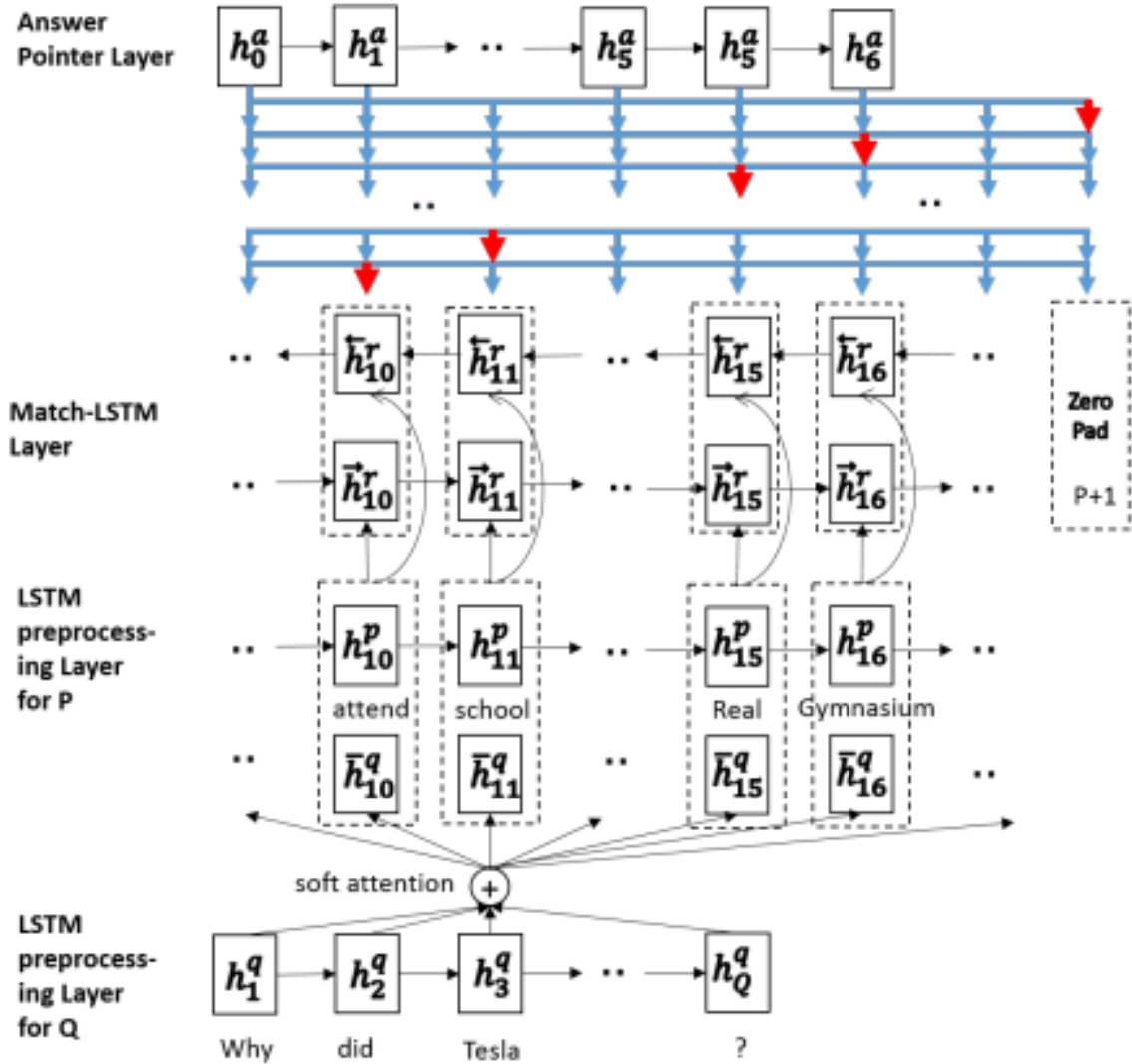
- Rouge-L: статистика, заснована на найдовшій спільній послідовності
  - BLEU: метрика правильності перекладу

# МОДЕЛЬ ДВОНАПРЯМНОГО ПОТОКУ УВАГИ

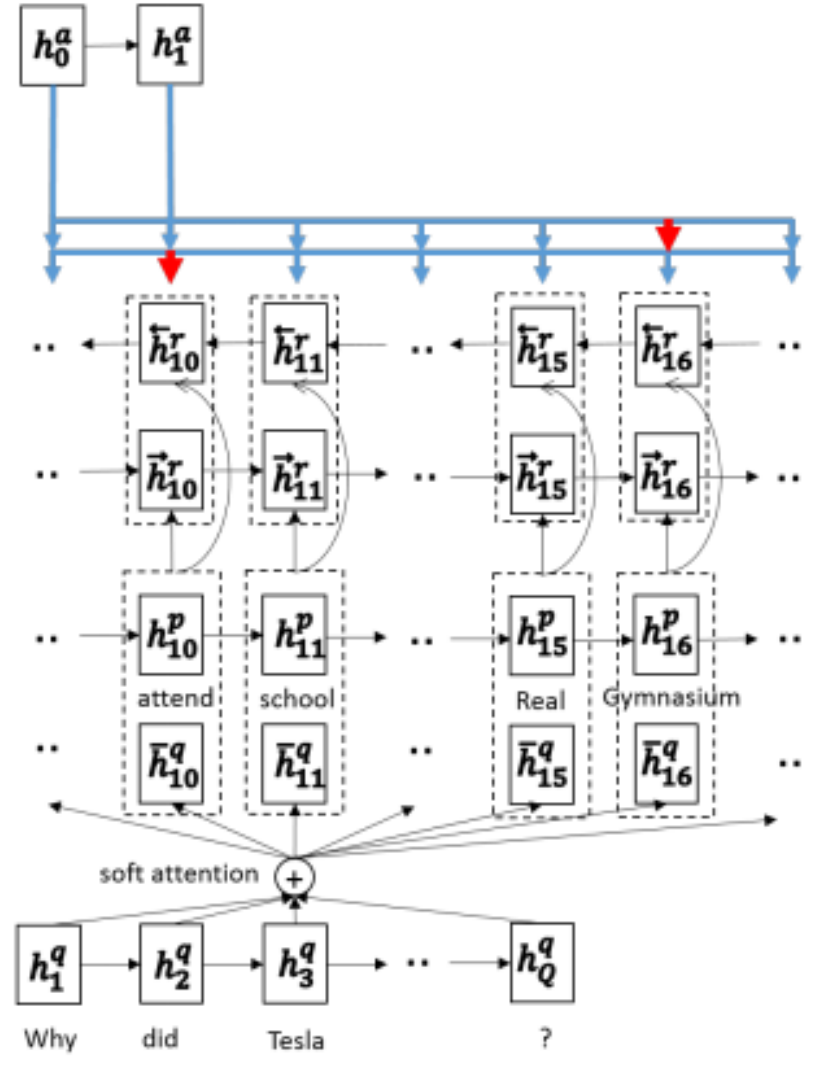




# МОДЕЛІ МАТЧ-LSTM ТА ВКАЗІВНИКА ВІДПОВІДІ



(a) Sequence Model



(b) Boundary Model

# ЕТАП ОБРОБКИ ДАНИХ

1. Завантаження даних, корпусів NLTK та GloVe.
2. Передобробка сирих даних: приведення набору даних до структури та формату .json
3. Для набору даних MS MARCO: у контекстних абзацах – визначення початку фрази-відповіді: обраховуються значення Rouge-L для всіх пар фраза з абзацу-відповідь, обирається фраза з найбільшим значенням Rouge-L.
4. Приведення обох наборів даних до однакової до структури.
5. Видалення пунктуації й токенизація контекстних абзаців, питань та відповідей за допомогою корпусу NLTK.
6. Отримання векторного представлення токенів за допомогою корпусу GloVe.
7. Індивідуальна

# РЕЗУЛЬТАТИ РОБОТИ МОДЕЛЕЙ НА НАБОРАХ ДАНИХ

Набір даних	SQuAD		MS MARCO	
Метрика	F1	Exact match	Rouge-L	Bleu
Мережа двонапрямого потоку уваги	73.3%	81.1%	22.23	18.83
Мережа двонапрямого потоку уваги (оригінальні результати)	73.3%	81.1%	-	-
Мережа Match-LSTM та вказівника відповіді	50.9%	60.02%	18.46	14.2
Мережа Match-LSTM та вказівника відповіді (оригінальні результати)	67.9%	77.02%	-	-
Найкращі результати серед усіх моделей на червень 2017 р.	r-net (ансамбль) 76.922%	r-net (ансамбль) 84.006%	r-net 42.2	r-net 42.9
Точність відповідей людини	77.0%	86.8%	46	47

# АНАЛІЗ РЕЗУЛЬТАТІВ

- Для мережі BiDAF підтверджені заявлені у статті результати
- На наборі даних MS MARCO обидві моделі показали результати, близькі до задовільних (остання з шести моделей на червень 2017 року в дошці змагання MS MARCO мають точності 19.20 у метриці ROUGE-L та 14.83 у метриці BLEU)
- Мережа BiDAF показала себе краще як на наборі даних SQuAD, так і на наборі даних MS MARCO.
- Мережа BiDAF навчається значно довше, ніж мережа Match-LSTM

Час навчання моделей	SQuAD	MS MARCO
BiDAF	20 год	27 год
Match-LSTM	6 год	10 год

- Набір даних MS MARCO є більш складним для поточних моделей, аніж набір SQuAD, і має менше тому науковцям варто зосередитися на досягненні високої точності саме на ньому.

# ВИСНОВКИ

## Новизна

- Отримані результати роботи двох моделей нейронних мереж на найновішому наборі даних для розуміння тексту MS MARCO

## Цінність:

- Розроблено модифікацію моделей BiDAF та Match-LSTM для роботи з набором даних MS MARCO
- Підтверджено результати моделі BiDAF на наборі даних SQuAD

# ПОДАЛЬШИЙ НАПРЯМОК РОЗВИТКУ

- Адаптація моделей BiDAF та Match-LSTM до роботи з наборами даних CNN/Daily Mail, Children's Book Test, LAMBADA
- Адаптація інших моделей до роботи з набором даних MS MARCO
- Розробка нових моделей, покликаних перевищити людську точність відповідей