

Попов І. В., Колбасін В. О., Сердюк І. В.
*Національний технічний університет
«Харківський політехнічний інститут», м. Харків*

Спілкування людини зі штучним інтелектом є досить складним процесом, який в наш час, на жаль, не є досконалим. Досить амбітною метою є розробка штучного інтелекту, який може вільно брати участь у розмові. Основним етапом для досягнення цієї мети є вивчення природної мови використовуючи сенсорні дані як, наприклад, зображення (grounded language learning).

Одним з варіантів вирішення цієї проблеми можуть бути так звані “дискримінаційні запитання” (en. Discriminative Questions). Штучний інтелект потрібно навчити задавати такі дискримінаційні запитання, які допомагають розрізняти візуальні екземпляри. А саме: комп'ютеру на вхід надаються 2 візуальних приклади. Йому потрібно згенерувати таке запитання, яке б дозволило точно розрізнити їх між собою.

Для цього було розглянуто такі методи, як:

- Підпис до зображення природною мовою.
- Візуальні відповіді на запитання (en. VQA).
- Візуальна генерація запитань (en. VQG).

Усі ці методи базуються на таких штучних нейронних мережах, як згорткові нейронні мережі (en. CNN) та мережі довгої короткострокової пам'яті (en. LSTM).

Було використано датасет Visual Genome, в якому представлені анотації об'єктів, атрибутів та зв'язків на великій кількості зображень [1].

Для оцінювання згенерованого питання використовувалася метрика ΔBLEU, яка дуже часто використовується в подібних роботах [2].

В результаті роботи були проаналізовані підходи, та розроблені методи, які дозволяють розрізнити зображення за допомогою генерування дискримінаційного питання.

Література:

1. R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations // IJCV. – 2017 123(1) : 32-73.
2. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a method for automatic evaluation of machine translation // Association for Computational Linguistics. – 2002. – P. 311-318.