

Оценка релевантности ответов пользователя на вопросы экзамена с использованием грамматик зависимостей

Виноградова Т. Б.

(Университет ИТМО)

Научный руководитель – к. ф.-м. н., Рыбин С.В.

(Университет ИТМО)

Исследования выполнены за счет стартового финансирования университета ИТМО в рамках НИР № 619423 «Разработка виртуального диалогового помощника для поддержки проведения дистанционного экзамена на основе аргументационного подхода и глубокого машинного обучения».

При разработке виртуального диалогового помощника для поддержки проведения экзамена важным этапом является разработка методов проверки ответов студентов на вопросы экзамена. Рассматривается задача оценки корректности и релевантности ответа пользователя на вопросы виртуального экзаменатора. В данной работе предлагается метод оценки ответов пользователя, представленных в виде деревьев грамматик зависимостей.

Методы неточного сравнения грамматик зависимостей часто используются для оценки релевантности ответов в вопросно-ответных системах. Существуют различные подходы к такому сравнению, например, подсчет пересечений множества отношений, сопоставление предикатов, неточное сравнение поиском в глубину.

Подсчет пересечений множеств отношений — статистический метод анализа словосочетаний в тексте. Он основан на анализе того, как много синтаксических конструкций входят одновременно и в текст вопроса, и в текст ответа. Сопоставление предикатов — метод, развивающий идею подсчета схожих словосочетаний вопроса и ответа. В этом методе для оценки пересечения используется нечеткое сравнение предикатов с использованием множеств контекстных слов из словаря WordNet, то есть учитывается контекст, в котором обычно употребляется то или иное слово.

Неточное сравнение поиском в глубину — метод, использующий представление грамматик зависимостей в виде деревьев. Для такого сравнения обычно применяется алгоритм сравнения деревьев, использующий простой способ сравнения токенов текста (совпадение лемм слов или вхождение одного слова в другое в качестве подстроки).

Метод, предлагающийся в работе, логически продолжает описанный выше алгоритм подобно тому, как сопоставление предикатов продолжает идею подсчета пересечений множества отношений. При оценке близости токенов в вершинах дерева зависимостей используются векторные представления токенов, полученные с использованием методов глубокого машинного обучения, учитывающие как само слово, так и окружающий его контекст.

Предложенный метод может быть использован при разработке системы автоматического оценивания ответов пользователя на вопросы виртуального экзаменатора.