

ОЦЕНКА КАЧЕСТВА УЧЕБНОГО МАТЕРИАЛА ПОСРЕДСТВОМ МЕТОДИК АНАЛИЗА ДАННЫХ

Саскевич А.В.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Стержанов М.В. – канд. тех. наук

В данной работе приведены результаты исследований по работе над оценкой качества учебного материала на примере обработки данных для интерактивного курса по изучению иностранного языка. В качестве основных инструментов применены скрапинг-системы и модели обработки натурального языка на языке Python.

При создании учебного материала применяется как ручной способ сбора и обработки материала, так и активно развивающиеся методики генерации учебного материала посредством инструментов машинного обучения и обработки больших объемов информации. В обоих случаях оценка качества может производиться как вручную, так и автоматически. Например, можно оценить сложность примера для изучающего курс высшей математики как количество математических операций, которые необходимо выполнить для решения задачи. С другой стороны, посредством разнообразных механизмов возможно в автоматическом режиме, с учетом каких-то скрытых и недоступных наблюдателю факторов оценить материал в достаточно крупном объеме. Подобные подходы уже применяются, например, в системах, позволяющих адаптивно формировать учебный материал [1].

В качестве данной работы проводилась разработка метрик, посредством которых возможна оценка учебного материала, сгенерированного автоматизированной системой для интерактивного курса иностранного языка. Условно процесс можно разделить на несколько этапов:

1. Скрапинг данных посредством любой доступной скрапинг-утилиты, например, библиотека Scrapy;
2. Извлечение текстовых и/или графических данных с целью формирования базы данных учебного материала;
3. Формирование атомарных учебных заданий (например, перевод фразы или перевод слова), которое может быть выполнено посредством любого модуля обработки естественного языка;
4. Расчет метрик для атомарных учебных заданий по некоторой методике [2].

Для расчета метрик для учебного языкового курса произведена следующая процедура:

1. Разделение текста на токены – отдельные слова и знаки препинания;
2. Подсчет частоты вхождения подобных токенов и их тип (существительные, глаголы и т.д.);
3. Соотнесение частоты и типа с некоторым взвешенным рейтингом, например, по формуле взвешенного среднего арифметического:

$$\frac{\sum x_i w_i}{\sum w_i} \quad (1),$$

где x_i – категоризированное значение признака, w_i – частота вхождения данного признака в выборку.

Затем задания упорядочены в порядке увеличения частоты, вычисленной на основе слов и фраз, входящих в их состав. Следующим этапом является формирование графа зависимостей, который призван показать, какие слова и фразы зависят от предварительного изучения других слов и фраз. Пример такого графа представлен на рисунке 1.

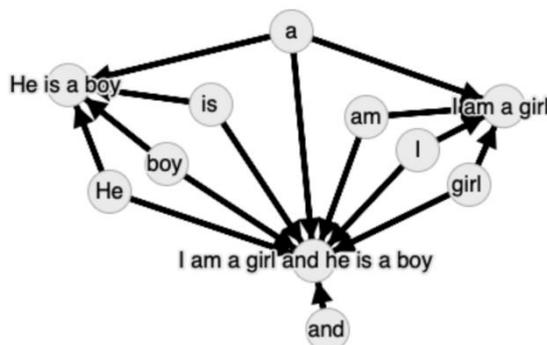


Рисунок 1 – Граф зависимостей слов при изучении

На такой граф могут быть перенесены различные варианты алгоритмов на графах, которые

могут помочь либо улучшить и дополнить метрики качества, описанные на собранных данных, либо заменить их альтернативным вариантом. В качестве применения подойдет множество различных классических алгоритмов на графах [3].

Топологическая сортировка позволяет выявить последовательность материала для изучения. Адаптация задачи под транспортную задачу с максимальным потоком позволяет определить точки входа для учащегося. Например, вершины, в которых отсутствуют входящие ребра, являются начальной базой для изучения.

На следующем этапе возможно соотнести полученные вершины графа с рассчитанными ранее значениями среднего взвешенного арифметического, позволяя упорядочить и выбрать в качестве стартовых элементов не просто слова, которые содержат в построенном графе исходящие ребра, но и которые являются, например, более популярными в своей категории – например, существительные и местоимения.

Для создания определенной сложности учебный материал можно наполнять словами, которые не были изучены ранее и отдельно, повышая рейтинг таких заданий. В качестве примера был собран набор слов и фраз для французского языка. На рисунке 2 представлены графики, на которых представлено, что формирование учебного материала в огромной степени зависит от включения материала, который будет предложен к изучению в контексте – например, внедрение неизвестных одного-двух слов в предложение, в котором учащийся уже знает остальные слова. Данный подход позволяет охватить до 80% доступного учебного материала, опираясь только на 20% изучаемого материала:

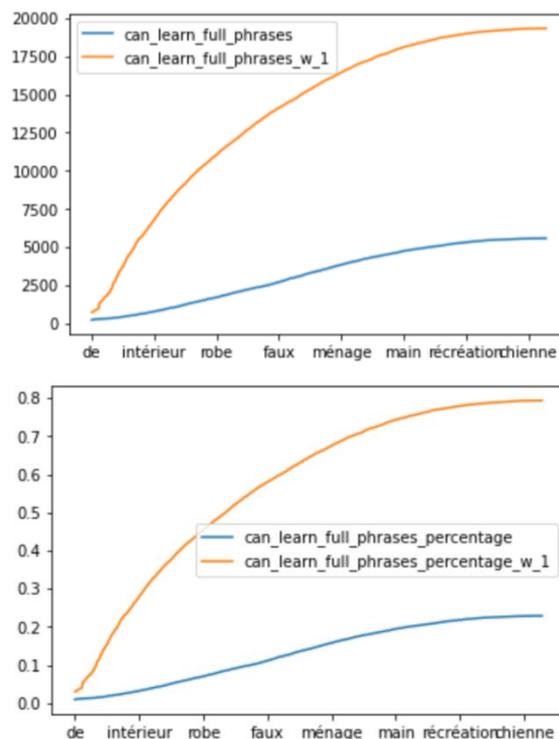


Рисунок 2 – Зависимость охвата учебного материала от количества изучаемого материала

Таким образом, применяя и комбинируя разнообразные статистические метрики, можно выполнить предварительную оценку учебного материала, в частности, сгенерированного посредством автоматизированных информационных систем. Предварительная оценка позволяет определить материал достаточного уровня, который в дальнейшем может быть протестирован на реальных учащихся.

Список использованных источников:

1. Тархов С. В. Адаптивное электронное обучение и оценка его эффективности // Открытое образование. 2005. №5. – С. 37-47.
2. Кузнецова Ирина Александровна Оценка качества систем дистанционного обучения // Вестник евразийской науки. 2011. №2 (7).
3. Неверов П. А. Использование графов для структурирования и дальнейшего совершенствования учебного материала // Известия РГПУ им. А.И. Герцена. 2008. №63-2.