

ОЦЕНКА ТРУДНОСТИ ПОНИМАНИЯ УЧЕБНЫХ ТЕКСТОВ ДЛЯ ВЫСШЕЙ ШКОЛЫ

М.А. Зильберглейт, М.М. Невдах, Ю.Ф. Шпаковский
Белорусский государственный технологический университет, Минск

Приведены результаты экспериментов с использованием различных методик (дополнения, экспертных оценок и парных сравнений) для получения объективных критериев относительно трудности текстов. Выделены и вычислены значения 49 параметров учебных текстов. Осуществлено снижение признакового пространства методами многомерного статистического анализа. С помощью дискриминантного анализа разработано решающее правило для оценивания трудности понимания учебных текстов для высшей школы.

Статистические методики анализа данных с поддержкой компьютерных технологий обладают огромным потенциалом в разрешении многих практических задач обработки текстовых массивов. Одной из областей анализа текста с точки зрения его доступности для читателя является читабельность.

В настоящее время отдельные аспекты читабельности с использованием современных информационных технологий привлекают специалистов из различных областей знаний, так как трудность учебных текстов волнует не только педагогов и родителей, но и самих учащихся. Исследования, посвященные автоматизированной оценке трудности учебных текстов (на примере белорусских изданий), не проводились.

Целью доклада является описание методики автоматизированной оценки трудности учебных текстов по философии и экономической теории для высшей школы. В решении данной задачи можно выделить три этапа: 1) нахождение и реализация методов для определения трудности понимания различных текстов данной группой лиц; 2) выбор формальных характеристик текста (и только тех, которые поддаются точному измерению); 3) создание автоматизированной системы, которая бы на основе ответов испытуемых, полученных экспериментальным путем, предсказывала понятность текста для будущих читателей.

Экспериментальным материалом послужили издания для вузов по философии и экономической теории [1–6]. Всего было отобрано 48 отрывков длиной 1 800 – 2 000 печатных знаков. Выбор данной величины обусловлен тем, что, начиная с объема в 1 800 знаков, статистические характеристики текста относительно постоянны [7].

На *первом этапе* проведены эксперименты с использованием различных методик. С этой целью были проанализированы основные методы определения трудности понимания текста: постановка вопросов к тексту, сводка основного содержания текста, составление плана или схем текста, интонирование, пересказ и др. [8]. В исследовании применялись наиболее надежные методы – методика дополнения и метод экспертных оценок. Впервые для оценки трудности понимания учебного материала использовался метод парных сравнений.

Методика дополнения – это заполнение пропусков в тексте, в котором слова через определенный интервал заменены точками. Плюсы данной методики состоят в том, что пропускается всегда только одно слово и не по усмотрению исследователя, а по строгому правилу. В текстах на основе результатов предварительного эксперимента пропускалось каждое седьмое слово.

Суть *экспертных оценок* трудности текста заключалась в следующем: после прочтения отрывка испытуемым предлагалось оценить его трудность по семибалльной шкале: 1 – сверхлегкий текст, 2 – очень легкий, 3 – легкий, 4 – текст со средней

трудностью, 5 – трудный, 6 – очень трудный, 7 – сверхтрудный текст. Для того чтобы исключить поверхностное знакомство испытуемых с текстом и возможное искажение результатов при оценке его трудности, студентам перед оценкой трудности понимания текста по шкале предлагалось выписать несколько ключевых слов и выразить основное содержание отрывка одним предложением. При проведении методики дополнения и экспертных оценок фиксировалось также время работы с текстом.

Оценка трудности учебных текстов проводилась среди студентов (75 человек) старших курсов Белорусского государственного технологического университета.

Суть *метода парных сравнений* заключалась в том, что каждому испытуемому предлагался набор текстов, размещенных парами, и после прочтения студент должен был указать, какой из отрывков обладает заданным признаком (в нашем случае – какой отрывок воспринимается легче). Оценка каждого текста производилась путем сравнения с каждым другим текстом того же набора. Так как у нас в наборе имелось 24 отрывка по философии и столько же по экономической теории, следовательно, по одному предмету было составлено 276 пар. За один этап эксперимента студенту предъявлялось восемь пар текстов. Такое количество не вызывало утомления у испытуемого.

Обработка и анализ результатов экспериментов позволили выявить информацию относительно трудности понимания учебного материала. На основании полученных данных найдено пять объективных критериев, определяющих трудность текста: процент правильно заполненных пропусков Y_1 ; относительное время работы с текстом Y_2 – с использованием методики дополнения; средняя оценка трудности восприятия текста Y_3 ; относительное время работы с текстом Y_4 – с использованием экспертных оценок; ранг текста Y_5 . Результаты экспериментов были сведены в табл. 1.

Таблица 1

Критерии трудности учебных текстов для высшей школы

Номер теста	Методика дополнения в процентах правильно заполненных пропусков	Относительное время работы с текстом (по методике дополнения)	Экспертные оценки испытуемых	Относительное время работы с текстом (по экспертным оценкам испытуемых)	Метод парных сравнений (ранг)
1	78,82	38,55	4,21	0,017	6
2	77,11	30,74	3,51	0,014	22
3	71,02	39,29	3,31	0,020	12
...
48	55,46	28,40	4,57	0,022	3

Для каждого показателя найдена середина диапазона всех полученных значений, в соответствии с которой производилось разбиение текстов на две группы: трудные (0), легкие (1). В итоге было получено разбиение текстов на группы по выделенным пяти показателям (табл. 2).

Таблица 2

Разбиение текстов на группы в соответствии с субъективной оценкой трудности текста

Номер теста	Методика дополнения в процентах правильно заполненных пропусков	Относительное время работы с текстом (по методике дополнения)	Экспертные оценки испытуемых	Относительное время работы с текстом (по экспертным оценкам испытуемых)	Метод парных сравнений (ранг)
1	1	0	0	1	0
2	1	1	1	1	1
3	1	0	1	0	0
...
48	0	1	0	0	0

Объективная трудность учебных текстов определялась путем анализа компонентов сложности текстов. Для этого на *втором этапе* были выделены и вычислены значения 49 параметров учебных текстов по философии и экономической теории: длина текста в абзацах, длина текста в словах, длина текста в буквах и др.

Очевидно, что использование большого количества параметров текста является неэффективным по ряду причин: сильная взаимосвязанность признаков, которая приводит к дублированию информации; неинформативность признаков, мало меняющихся при переходе от одного объекта к другому; возможность агрегирования по некоторым признакам. Однако ничем не оправданное уменьшение числа переменных может привести к потере точности экспериментов.

Для снижения признакового пространства были использованы кластерный и факторный анализ, метод корреляционных плеяд и вроцлавской таксономии, многомерное шкалирование. Для анализа данных и проведения статистического анализа использован программный пакет SPSS.

Кластерный анализ представляет собой многомерную статистическую процедуру, которая выполняет сбор данных, содержащих информацию о выборке объектов, и затем упорядочивает объекты в сравнительно однородные группы.

В исследовании для анализа данных в качестве критерия для определения подобия групп использовались следующие меры сходства: расстояние Евклида, квадрат расстояния Евклида, косинус угла, коэффициент корреляции, неравенство Чебышева, расстояние Минковского, манхэттенское расстояние.

Для кластеризации выделенных характеристик текста использовались основные методы кластерного анализа: простого среднего; группового среднего; ближнего соседа; дальнего соседа; невзвешенный центроидный; взвешенный центроидный; Варда. Количество кластеров по каждому методу варьировалось от трех до десяти. После выбора всех соответствующих параметров получена информация по формированию кластеров: порядок объединения кластеров, расстояние между ними, а также принадлежность характеристик текста к тому или иному кластеру.

Выводимые результаты для наглядности были представлены и в виде дендрограмм, которые позволяют не только перейти к любому признаку на любом уровне кластеризации, но и дают возможность судить о том, каково расстояние между кластерами или признаками на каждом из уровней.

В применении процедур кластерного анализа немаловажным аспектом выступает устойчивость структуры кластеров, отражающая реальную объективность классификации. В качестве методов проверки устойчивости могут быть использованы бутстреп-метод, предложенный Б. Эфроном в 1977 г., методы «складного ножа» и «скользящего контроля» [9]. Одним из наиболее простых и эффективных способов проверки устойчивости результатов является метод сравнения результатов, полученных для различных алгоритмов кластеризации, который и использовался в данной работе. Для этого все данные для наглядности были объединены в сводные таблицы, в которых четко прослеживаются особенности применения различных алгоритмов кластерного анализа, использующих разные меры сходства. Результаты формирования кластеров для текстов по философии согласуются практически по всем алгоритмам. Незначительно отличаются данные по методу Варда и центроидной кластеризации. Сравнение результатов с применением различных мер сходства показало, что наблюдаются заметные различия лишь в данных, полученных методами измерения близости и основанных на косинусах векторов значений и корреляции векторов значений.

Для текстов по экономической теории результаты формирования кластеров по различным алгоритмам отличаются только по методу Варда. При использовании раз-

личных мер сходства наблюдаются заметные различия лишь в данных, полученных методами измерения близости и основанных на корреляции векторов значений и манхэттенском расстоянии.

Проведенный кластерный анализ для учебных текстов показал, что целесообразно выделить следующие девять групп признаков: философия – 1, 4, 8, 13, 18, 22–25, 33, 34, 38, 40, 43, 45, 46; 2; 3; 5; 6, 7; 9, 14, 19, 30–32, 36, 37, 41, 42, 44; 10, 15, 27–29, 35, 47, 49; 11, 12, 16, 17; 20, 21, 26, 39, 48; экономика – 1, 4, 8, 13, 22–25, 40, 42–44, 46; 2, 9, 14, 18; 3, 39, 45, 48; 5–7; 10–12, 16, 17; 15, 19–21; 26–30, 35–37; 31–34, 38, 41; 47, 49. Для последующей обработки достаточно пользоваться одним признаком из каждой группы.

Снижение размерности набора переменных в методах *факторного анализа* базируется в основном на взаимной коррелированности исходных признаков. В связи с этим первый этап исследования заключался в вычислении корреляционной матрицы.

При изучении экспериментальных данных было установлено, что первые три фактора объясняют разброс дисперсии около 74 % для текстов по философии, около 64 % – для текстов по экономической теории. Так как факторный анализ является методом сокращения числа переменных, возникает вопрос, какие из факторов следует оставить для дальнейшей обработки. Исследователи рекомендуют руководствоваться здравым смыслом и оставлять только те факторы, которые имеют понятную или логическую интерпретацию. Однако установить заранее назначение каждого фактора не всегда представляется возможным, поэтому для начала были использованы формальные критерии: Кайзера и «каменистой осыпи» Р. Кэтелла.

Первый критерий, как правило, сохраняет слишком много факторов, в то время как второй – слишком мало, поэтому решение об оптимальном количестве факторов можно принять только после их вращения и интерпретации. Целью вращения факторов является получение простой структуры, которой соответствует большое значение нагрузки каждой переменной только по одному фактору и малое – по всем остальным. Нагрузка (значение лежит в пределах от –1 до 1) отражает связь между переменной и фактором. В работе использовались ортогональные методы вращения: варимакс, квартимакс и эквимакс. В результате были получены матрицы нагрузок для переменных.

Изучение результатов с использованием всех методов факторного анализа и методов вращения позволило выявить, как признаки распределились между факторами. Факторный анализ выделил следующие группы признаков: философия – 1, 44, 48; 2, 3, 5–17, 24, 47, 49; 4, 46; 18–21, 41; 22, 23, 25–40; 42, 43 (6 групп); экономика – 1, 4, 8, 13, 22–25, 40, 42–44, 46; 2, 9, 14, 18; 3, 39, 45, 48; 5–7; 10–12, 16, 17; 15, 19–21; 26–38, 41; 47, 49 (8 групп).

Для снижения признакового пространства использовался и метод *корреляционных плеяд*. Выделение корреляционных плеяд осуществлялось следующим образом: признаки упорядочивались, и рассматривались только те коэффициенты корреляции, которые соответствуют связям между элементами в упорядоченной системе.

Упорядочение производилось на основании принципа максимального корреляционного пути. Для удобства построения графа были составлены упорядоченные корреляционные матрицы. На основании упорядочения признаков были построены графы, которые представляют собой кратчайший незамкнутый путь. В графах соединены все исследуемые параметры текстов. Если задать определенное пороговое значение коэффициента корреляции (r_0), полученные графы можно разбить на подграфы (плеяды), проводя разрыв между признаками со значением сопряженности, меньшим r_0 . Используя прямое (z) и обратное (z^{-1}) преобразования Фишера, было определено r_0 для заданного объема выборки. Исходя из поставленной цели и анализа корреляционной матрицы ис-

следуемых характеристик текстов, был задан пороговый коэффициент корреляции $r \geq 0,9$, что позволило выявить наиболее связанные друг с другом признаки. Исходный граф распался на пять подграфов для текстов по философии и шесть подграфов для текстов по экономической теории. Признаки, не вошедшие в выделенные группы, требуют дальнейшего исследования. Использование метода корреляционных плеяд позволило выделить следующие группы близких параметров текста: философия – 1; 2; 3; 4; 5–7; 8; 9–12, 14–21; 22, 23, 25–38; 24; 39, 40; 41; 42; 43; 44; 45; 46; 47, 48; 49 (18 групп); экономика – 1; 2; 3; 4; 5–7; 8; 9, 11, 12, 14, 16–18, 20, 21; 10; 13; 15, 19; 22–25, 27–33, 35–37; 26; 34; 38; 39, 40; 41; 42; 43; 44; 45; 46; 47, 48; 49 (23 группы).

С помощью метода *вроцлавской таксономии* было получено нелинейное упорядочение изучаемых элементов текста. С целью построения дендрита были вычислены матрицы расстояний (на основе расстояния Евклида) между изучаемыми характеристиками учебных текстов. Далее из составленных матриц расстояний между признаками были выбраны единицы с близкими значениями. В результате для текстов по философии были получены пары признаков с близкими значениями (например, 1–22, 2–12, 3–7 и т. д.). Далее были найдены пары с общим признаком, которые затем объединялись друг с другом. Например, пары 43–22 и 22–1 образовали цепочку 43–22–1. В результате было получено 16 отдельных конструкций, называемые скоплениями 1-го порядка.

Полученные скопления не удовлетворяют основному условию дендрита, а именно, они не связаны в единое целое. Для достижения этой цели было выбрано наименьшее расстояние между единицами, входящими в различные скопления 1-го порядка. В результате были получены скопления 2-го порядка. Объединение признаков в скопления 3-го, 4-го, n -го порядков происходило до тех пор, пока любые две точки исследуемого множества параметров не оказались связанными друг с другом.

Исходя из поставленной цели и анализа дендрита, была определена максимальная величина расстояния между признаками, равная 0,08 для текстов по философии и 0,15 для текстов по экономической теории. В первом случае исходный дендрит распался на семь наиболее связанных групп признаков, во втором – на пять.

Использование метода *вроцлавской таксономии* позволило выделить следующие группы признаков: по философии – 1, 4, 8, 13, 18, 22, 23, 25, 33, 34, 38, 40, 43, 45, 46; 2; 3; 5; 6; 7; 9, 14; 10; 11, 12; 15; 16, 17; 19; 20, 21; 24; 26; 27, 35; 28; 29; 30; 31, 32, 37; 36; 39; 41; 42; 44; 47; 48; 49 (28 групп); по экономической теории – 1, 4, 8, 9, 13, 14, 18, 22–25, 34, 38, 40–43, 45, 46; 2; 3; 5; 6; 7; 10; 11; 12; 15, 19; 16; 17; 20; 21; 26, 39; 27, 28, 35; 29–33, 36, 37; 44; 47; 48; 49 (21 группа).

Задача *многомерного шкалирования* заключалась в преобразовании исходной матрицы 49×49 в гораздо более простую матрицу 49×2 и визуальным ее представлением.

После расположения точек в заданном пространстве для всей модели в многомерном шкалировании были вычислены стресс и коэффициент R^2 . Наилучшей моделью для текстов по философии ($\text{stress} = 0,210$, $R^2 = 0,856$) стала модель, полученная с использованием меры сходства, основанной на неравенстве Чебышева; по экономической теории ($\text{stress} = 0,230$, $R^2 = 0,763$) – модель, полученная с использованием квадрата расстояния Евклида. На их основе были получены следующие группы признаков: по философии – 1; 2, 4–14, 47, 49; 3, 15–18; 19–21, 45; 22, 23, 25–38; 24; 39, 40; 41; 42, 43; 44, 48; 46 (11 групп); по экономической теории – 1, 39, 40, 46; 2; 3, 4, 9–12, 14, 16, 17, 47, 49; 5–7, 18; 8, 13, 42, 44; 15, 19–21, 43; 22–38, 41, 45, 48 (7 групп).

Для дальнейшего изучения характеристик текста важнейшей задачей является выделение наиболее информативного признака из каждой полученной группы. В данной работе для оценки информативности признаков в качестве информационной использовалась мера $J(1, 2)$ расхождения между статистическими распределениями 1 и 2.

В итоге были вычислены информационные меры каждого из 49 признаков, а затем отобраны те из них, которые обладают наибольшей информативностью среди признаков своей группы. В результате число признаков было сокращено до возможного минимума.

Для дальнейшего исследования характеристик текста и их влияния на понятность учебного материала использовался дискриминантный анализ. При анализе дискриминантных функций в учет принимались только функции, у которых процент точности классификации – максимальный, а количество переменных при этом минимальное.

В текстах по философии чаще других фигурировали следующие признаки (в порядке убывания): 3, 24, 9, 40, 1, 43, 48; в текстах по экономической теории: 39, 47, 48, 15, 42, 41. Выделенные признаки позволяют сделать важный вывод относительно факторов трудности текста. Они связаны, прежде всего, с объемом текста (признак 3), с длиной слов и предложений (признаки 9, 15, 24), со сложностью организации текста (признаки 1, 47, 48), богатство словаря и абстрактностью изложения материала (признаки 39, 40–43).

Для автоматизированной оценки рукописи на основе полученных дискриминантных функций создана программа Readability analysis, предназначенная для автоматизации оценки трудности учебных текстов для студентов вузов. Новизна разработанного алгоритма программы заключается в том, что при оценке трудности текста он учитывает только статистически значимые (наиболее информативные) параметры. Кроме того, в основе алгоритма лежит не уравнение регрессии, которое используется в большинстве известных алгоритмов, а решающее правило в виде дискриминантных функций, которое позволяет объективно оценить трудность учебного текста.

Практическая значимость программы Readability analysis связана с тем, что она может быть использована в редакционно-издательской деятельности при подготовке учебной литературы для высшей школы. Анализ трудности текста на стадии его подготовки и дальнейшее усовершенствование материала позволят привести уровень сложности учебного текста в соответствии со способностями читателей.

Список литературы

1. Волчек, Е.З. Философия : учеб. пособие с хрестоматийными извлечениями / Е.З. Волчек. – Минск : Интерпрессервис, Экоперспектива, 2003. – 544 с.
2. Философия : учеб. пособие для студентов высших учебных заведений / В.С. Степин [и др.]; под общ. ред. Я.С. Яскевич. – Минск : РИВШ, 2006. – 624 с.
3. Философия : учеб. пособие для студентов высших учебных заведений / Ю.А. Харин [и др.]; под общ. ред. Ю.А. Харина. – Минск : ТетраСистемс, 2006. – 448 с.
4. Экономическая теория : учебник / Н.И. Базылев [и др.]; под общ. ред. Н.И. Базылева, С.П. Гурко. – Минск : Экоперспектива, 1997. – 368 с.
5. Давыденко, Л.Н. Экономическая теория : учебное пособие / Л.Н. Давыденко [и др.]; под общ. ред. Л.Н. Давыденко. – Минск : Вышэйшая школа, 2002. – 366 с.
6. Кажуро, Н.Я. Основы экономической теории / Н.Я. Кажуро. – Минск: ФАУинформ, 2001. – 672 с.
7. Косова, М.М. Описательная статистика учебных текстов по физике / М.М. Косова, М.А. Зильберглейт // Труды БГТУ. – 2006. – Сер. VI. Физ.-мат. науки и информатика. – Вып. XIV. – С. 167–170.
8. Невдах, М.М. Новая классификация методов определения понимания текста / М.М. Невдах, Ю.Ф. Шпаковский // Труды Белорусс. гос. технол. ун-та. – 2007. – Сер. IX. Издательское дело и полиграфия. – Вып. XV. – С. 100–104.
9. Ковальченко, И.Д. Количественные методы в исторических исследованиях / И.Д. Ковальченко. – М. : Высшая школа, 1984. – 384 с.