

УДК 378(075.8)

**ОБ ИНФОРМАЦИОННОМ ПОДХОДЕ
К ФОРМИРОВАНИЮ ТЕСТОВЫХ СИСТЕМ**

Воробьев Г.Н.

Учреждение образования

«Могилёвский государственный университет продовольствия»

г. Могилёв, Республика Беларусь

Тестовую систему, содержащую вопросы и возможные ответы на них, можно рассматривать как систему классификации, основанную на множестве примеров, определяющих оценку результатов учебной деятельности обучающихся в данной предметной области. Такое представление позволяет проверить знание обучаемым определенных показателей учебы в пределах шкалы оценок, используя аналог метода построения дерева решений, предложенный в [1, с. 392]. Предполагается выделение определенных свойств, применяемых для проведения классификации, в данном случае – выставление

балла за ответы на поставленные вопросы. В указанной классификации важное значение имеет метод построения дерева решений, которое с наибольшей вероятностью позволяет корректно классифицировать все возможные ответы тестируемых. При этом выдвигается требование выполнения принципа простейшего ответа, соответствующего исходным данным, смысл которого состоит в определении уровня подготовки за минимальное количество шагов в тесте. Например, не следует продолжать тестирование, если на первых его шагах обнаружено, что уровень подготовки не соответствует требуемому минимуму. В этой связи, при построении дерева решений важное значение имеет порядок тестирования, определяемый некоторой функцией выбора, основанной на эвристическом поиске.

Предлагается эристика, основанная на количественной оценке информации, содержащейся в дереве решений и выборе свойств для его организации на основе максимума количества информации, содержащейся в тестовых примерах. Теория информации позволяет формализовать интуитивные рассуждения о важности первостепенного внесения в корень дерева классификации определенного свойства.

Пусть в тестовой системе создан набор V обучающих примеров мощности n , составляющих основу классификации. В соответствии со шкалой оценок, элементы множества V будут отнесены к различным подмножествам V_i , каждое соответственно мощностью n_i , где $i = 1, 2, \dots, m$, m – число градаций в шкале оценок, $\sum_{i=1}^m n_i = n$, $\bigcup_{i=1}^m V_i = V$, $V_i \cap V_j = \emptyset$, если $i \neq j$. При

составлении тестовой системы мы предполагаем, что все примеры появляются с равной вероятностью. Тогда информативность распределения V может быть определена по формуле Шеннона и составит $I(V) = -\sum_{i=1}^m p(V_i) \log_2 p(V_i)$ бит, где $p(V_i) = n_i / n$. Количество информации $I(C)$, связанное с некоторым свойством C , влияющим на оценку, необходимо увязать с требованиями образовательного стандарта, в смысле: "иметь представление", "знать", "уметь". С информационной точки зрения это количество информации может быть определено, как разность общего количества информации $I(V)$ в дереве и количества информации $E(C)$, необходимого для выставления корректной оценки при выборе свойства C . Так как обучающие примеры, связанные с указанным свойством C , учитывают квалификационные требования и тем самым образуют подмножества C_j , $j = 1, 2, \dots, k$, то значение $E(C)$ может быть вычислено по формуле $E(C) = \sum_{j=1}^k \frac{|C_j|}{n} I(C_j)$.

Такой анализ можно рекурсивно выполнить для каждого поддеревя до полного построения всего дерева квалификации.

Литература.

1. Люгер, Д. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем / Д. Ф. Люггер. – М.: Издательский дом "Вильямс, 2003. – 864 с.