# Predictive Analysis: Assigning Weightage and Difficulty Level of Question using Data Mining

Kamal Vora K. J. Somaiya College of Engineering. Sumeet Jain K. J. Somaiya College of Engineering.

## ABSTRACT

Generation of question papers is a task that requires high expertise and flair. The design and formation of the test depends on the questions. The test should generate the quiz such that all different levels of questions are inserted so as to make the question paper more competitive. This research work aims at assigning weightage to a questions in the quiz based on the difficulty level. Assigning the weightage to question is a tedious task. This paper aims to discuss various parameters involved in assigning weightage to the questions in the test paper and decide the difficulty level of the question.

#### **General Terms**

Assignment of weightage, deciding the difficulty level.

#### **Keywords**

Weightage

#### **1. INTRODUCTION**

Aptitude tests are very important to identify the knowledge of person. They can prove helpful in deciding the right career path. The better the test, accurate would be the prediction result. For the test to be better, correct weightage has to be assigned to the question. Assigning correct weightage to the question implies giving correct importance to the question and thereby with the degree of importance a person can decide the level of difficulty to the question. Accurate assignment of weightage would help in giving accurate results. Basically there are three levels of difficulty to the question.<sup>[2]</sup>

- High(difficult questions)
- Moderate(manageable questions)
- Low(easy question)

To group the questions into these sections, the system needs to know the total weightage that is to be assigned to the question. To assign the weightage a predefined scale needs to be set. Reaching a specify measure in scale means to have achieved a certain level of difficulty.

Doing this would provide three main benefits.

- Question paper containing questions of difficulty level can be set.
- Marks can be allocated efficiently to each module of the course/syllabus/field.
- Can be helpful to generate multiple question papers with same difficulty level so as to be as fair as possible when multiple people appearing test for same goal.

Param Mehta K. J. Somaiya College of Engineering. Smita Sankhe Assistant Professor K. J. Somaiya College of Engineering.

## 2. ASSIGNING WEIGHTAGE

Assigning weightage to a question is tedious task. Various parameters are defined which are required to evaluate a question. Each parameter gives the measure of evaluation for the question. Satisfying the parameter gives the level of weightage.

Parameter		Weight(range)
Level of IQ (sense)		2-10
Length of question		2-10
Pattern		
a.	Repetition of	2-8
h	keyword	0-2
D.	image	
Type of question		
a.	true/false type	2
b.	simple MCQ	4
с.	calculated MCQ	6
d.	check box(multiple	8
	correct answers)	
e.	text box	10

The evaluation measure for each parameter is given as:

- Level of IQ (sense): The more sense the question makes the more weightage to be assigned to the question. This also means question is related to the subject on which test is taken. Thus more the question makes sense and is more related to test subject more weightage is given to the question.
- Length of question: In a time bound exams where timing is less the length of question plays a crucial role. Smaller the question, less time will it take easier and faster would the user will be able to solve. Thus longer the question more is the weightage.
- Pattern:
- Repletion of keyword: The more time the keyword of the question is repeated implies more information in the question is given about key terms making question easy to answer thus lower weightage to be assigned.
- Image: Visual perception to a thing gives better idea. Thus providing the image in question makes it easy to solve a question as compared to a question where image is not given thus question with image to be assigned lesser weightage.

- Type of question:
- True/false: true false question has only two options thus probability of giving correct answer is high i.e. 0.5 thus making question easy and less weightage to be assigned.
- Simple MCQ: simple MCQ question has lesser probability of giving correct answers. Suppose for question with 4 options the probability of right answer is 0.25 thus a slight more weightage to be assigned.
- Calculated MCQ: The only difference from simple MCQ type of question is that in simple MCQ type of question the answers can be given directly whereas MCQ type questions require complex calculations.
- Check box (Multiple answers): Multiple answers to a question decrease the probability of giving absolute correct answer. For example a question having 4 options 4! (4 factorial i.e. 24) possible combinations can be applied. That means probability of correct answer is 0.04167 which means difficult is the question. Hence higher weightage to be assigned
- Text Box: For the textual answers the number of possible answers can be infinitely large thus making probability of getting right answers very less so highest weightage to be assigned to this type of question.

The total weightage that is to be assigned to the question should be the cumulative sum of weightage from all the parameters. Adding the weightage from each parameter will give the final total weightage to the question.

The minimum possible weight to a question is 10 meaning the easiest possible question and maximum possible weightage is 40 meaning the most difficult question. The weightage to general question can be anything between this range.

Now this weightage can be further used to determine the level of difficulty to the question.

## 3. DECIDING DIFICULTY LEVEL OF QUESTION

After finding out the exact weight of the question now it is time to decide the difficulty level of the question. Determining the difficulty level of question is very necessary as the weightage of the question may vary from 10 to 40 but the marks that are assigned to the question should be only discrete numbers. Hence deciding the level can help to assign the marks in the ratio of the level of difficulty.

The marks assigned to the question are assumed in the ratio 1:2:3. That is 1 mark for an easy question, 2marks for question with moderate difficulty level and 3 marks for question with the highest difficulty level.

The algorithm to calculate the ratio:

- $M \rightarrow$  Maximum Possible weight.
- $W \rightarrow$  Total weightage for that Question.
- $X \rightarrow$  Maximum Weight to that specific part.
- $F \rightarrow$  Assign Weightage to that specific part.
- $b \rightarrow$  Unused Weight.
- $S \rightarrow$  Decision Parameter
- M =∑ Xi
- $W = \sum Fi$
- $b = \sum (Xi Fi)$
- S = (M-b)/M

Since question are in ratio of 1:2:3 the total weight is 6 so question with easy difficulty will be given between range of 0 and 1/6 i.e. 0 to 0.1666 whereas for question with moderate difficulty will be in between 1/6 to 3/6 i.e. 0.1667 to 0.5000 and for difficult question it is upto 1i.e from .5001 to 1.

From above calculated value of S

If S lies between 0 to 0.1666 then Question type will be Easy

If S lies between 0.1667 to 0.5000 then Question type will be Medium

Else the Question type will be Hard.

Thus questions are classified into three difficulty level. For instance now the question with least difficulty level (easy) will get 1 mark while moderately difficult question will get 2 marks and question with highest difficulty level will get 3 marks.

#### 4. SYSTEM ARCHITECTURE

The question stored in the database will first be classified according to option given and then the weightage for each question will be calculated weight generator. Now this question along with the weightage will go into decider which will categorize question according to difficulty level.



Fig 1: System architecture

# 5. SYSTEM APPLICATIONS

This system can be used at many places.

- Automatic question paper generator: systems like automatic paper generator can use this algorithm as it provides proper classification of question and hence question of all category and varying range can be asked. This will also help to set a question paper of specified marks
- Aptitude test: since the weight assigned to the questions are generated by the system by applying pure statistical and probability knowledge the weightage assigned is accurate. Most accurate weightage assignment helps in generating better test that can be used to give better results.

Also in case of predictive analysis it will help to predict better future due to accuracy in assigning the weightage to question and classifying question in difficulty level.

# 6. SIMULATION AND RESULT

Consider a scenario where a question is entered into the database and the corresponding weightage is to be calculated. The general IQ level of this question amounts to 6 units. The length of the question is of 140 characters which accounts to a weightage of 8. The keyword present in the question is repeated 4 times, it adds up to the weightage of 4 units. Additional 2 units is assigned to the question as there is no image reference option to the question. The question type is simple MCQ, and hence according to the probabilistic model the weightage added is 4 units. Final weightage corresponding to the question adds up to 24 units.

To decide the difficulty level of the question, the above algorithm (Algorithm 1) is used:

The maximum possible weightage (M) for a question aggregates to 40.

While the total assigned weightage (X) to the question is 24.

The ratio (S) = 24/40 = 0.6

Thus according to the algorithm this ratio falls in the category "Hard". This means that this is the question of highest difficulty level and thus the maximum marks that could be given to individual question is allocated to this question.

Parameter	Values	Weightage
IQ	6	6
Length of question	140	8
Repetition	4	2
Image present	No	2
Type of question	Simple MCQ	4
То	24	

# 7. CONCLUSION

In the previous system, the questions are assigned marks that are uniform to all questions irrespective of the difficulty and various parameters that lead to flaws in prediction in the desired result. Our stated algorithm allows us to allocate marks depending on the weightage assigned to various parameters that cumulate to the total weightage of the question, which helps to make a more appropriate prediction.

The difficulty of a question is a matter of individual perspective and hence never be certain. Using probabilistic model, it tends to give a better result even in uncertain and unforeseeable circumstances which is implemented in above stated algorithm.

## 8. ACKNOWLEDGMENTS

The authors would like to thank Ms. Smita Sankhe of K .J. Somaiya College of Engineering for providing specifications about the application scenario.

#### 9. REFERENCES

- Xindong Wu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang ·Top 10 algorithms in data mining 2012
- [2] Dept. of Comp.Sc., DM's College, Mapusa, Goa. Dynamic Examination Question Paper Generation System (DEQPGS) 17th International Conference on Management of Data (COMAD 2011) 19th December 2011 December 19—21
- [3] (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010
- [4] PACO: A predictive analysis system for manufacturing test, AUTOTESTCON, 2010 IEEE, T. Jones; J. Engler, IEEE, 13-1,Sept. 2010