RESEARCH ARTICLE   OPEN ACCESS

Word Problem Solver for Probability

Priyanka Dalvi[1], Sanjana Oulkar[2], Siraj Cotecha[3], Dr. Abhijit R. Joshi[4], Prof. Harshaldalvi[5]

[1](B.E. Student, Department of I.T., DJSCOE, University of Mumbai, Mumbai, India)
[2](B.E. Student, Department of I.T., DJSCOE, University of Mumbai, Mumbai, India)
[3](B.E. Student, Department of I.T., DJSCOE, University of Mumbai, Mumbai, India)
[4](Vice Principal (academic), Head of Department-IT, DJSCOE, University of Mumbai, Mumbai, India)
[5](Assistant Professor, Department of I.T., DJSCOE, University of Mumbai, Mumbai, India)

ABSTRACT
Word Problems are designed to help students to learn the application of mathematical concepts, algebraic identities and formulae in the real world. Variables are assigned the values of 'real-world' entities and a logical approach in solving them is established. They help the students to bridge the gap between theoretical knowledge and the real world application of it by giving them hypothetical situations about the same. Probability is a measure or estimation of how likely it is that a particular event will happen. Probability concepts need to be properly understood before attempting to solve any problem related to it. In view of this a survey was conducted. Students from various schools and coaching classes were approached for the same. The study shows that majority of the students experience difficulties in identifying and understanding what exactly the word problem signifies and what approach it demands. Also, the process of learning Probability needs to be specialized given the different understanding levels of each and every student in contrast to the generalized education techniques that are being used in traditional classrooms. Keeping in mind these issues, Word Problem Solver for Probability is implemented, which caters to the learning needs of each and every student individually by providing a step-by-step solution to all problems from the Probability domain.
Keywords - Events, Probability, Problem Solver, Sample Space, Teaching Aid Word.

I. INTRODUCTION

Probability is one of the most important topics of Mathematics. It encompasses a variety of problem types and their variations. It is observed that despite knowing the formulae and calculations, students cannot solve probability word problems correctly. Reasons for this can be traced to lack of basic understanding of the concepts and a correct approach to look at them. Once these basics are understood, students will be able to tackle any problem without difficulty. Word Problem Solver for Probability will help students in this area. It will guide them about Probability, help them understand the questions and provide them with an accurate step-by-step solution. Also, students can use it for verifying their answers and identifying mistakes, if any, in their solutions. Word Problem Solver for Probability will allow for a personalized way of teaching taking into account every single student. In order to enhance the traditional monotonous approach of classroom teaching, Word Problem Solver for Probability is developed. It not only serves as a learning tool for students but also as a teaching aid for the teachers.

The rest of the paper is organized as - Section 2, covers the literature survey giving overview of the existing systems used for teaching concepts of probability. Section 3 talks about the Proposed Approach that is followed to develop the system. Section 4 discusses Results obtained along with

sample outputs. The paper ends with conclusion and the future scope.

II. LITERATURE SURVEY

Probability provides a quantitative description of the likely occurrence of a particular event. Probability is conventionally expressed on a scale from 0 to 1; a rare event has a probability close to 0, a very common event has a probability close to 1, [4]. Probability is used widely in areas of study such as mathematics, statistics, finance, gambling, science, artificial intelligence, machine learning and philosophy too. Word problems related to Probability are of three types [3]. They are:
1. Specific events are described and probability is asked.
2. A set of variables is given. The probability of specific combinations and permutations (arrangements) is expected.
3. The probability of certain events is given and conditional probability is expected.

Various methods are implemented for teaching Probability currently. Majorly the existing methods are confined to classroom teaching and to some

extent students refer to material available online for the same. The approach of these is however, mainly syllabus centric and does not focus on overall understanding of the topic and development of student.

But, learning is defined to be a goal-directed act. The involvement of the student in learning is as important and fundamental as the teacher. Correct knowledge can be passed on in a correct manner only with the mutual coordination and cooperation of both; the teacher and student.

In this section it is intended to find out the reasons and coin conclusions about why a simple topic like Probability is considered to be tough to understand and grasp by the students. In doing so, all the existing systems involved in this area are evaluated.

2.1. TRADITIONAL APPROACH

The traditional method involves classroom teaching, normally associated as the chalk and board method, where teachers teach the students manually. The general method observed is explaining the concepts first followed by making the students solve a few examples in class and the remaining as homework.

In a classroom of minimum 35-40 students, it is difficult to ensure each and every student has understood the concept properly. Even a small doubt in understanding concepts of a critical topic like Probability can create various difficulties for the students in solving further problems. Moreover, students cannot always find the appropriate guidance when they get stuck in solving and practicing the exercises on their own. Thus, the absence of a teacher at that point of time proves to be a major hindrance.

2.2. ONLINE VIDEOS

'Learning Probability' videos available online have been successful in imparting distant education to students across the globe. They are easy to access and students can select a particular sub-topic they want to watch and follow the same. As of today, these videos receiving as many as 5,83,772 views clearly speaks out for their popularity and excellence in reaching out efficiently to students.

However, with a lot of content available online, it is difficult for a student to identify and choose for himself/herself what is apt and correct. Also, the learning pace requirements of each student are different at different levels. Precisely stating, school students need a step by step approach towards the final answer because every correct step and formula used carries marks which also surfaced in the survey. On the other hand, college students who want to excel in competitive examinations need a quick understanding of the problem and arrive to the answer in a stipulated amount of time. Online videos

fail to address these needs. Language barrier is also a major problem.

2.3. MATHWAY

Mathway is a free tutor online service which resolves users' mathematics problems based on users' stated problem. It is a web application. Mathway provides features like Basic Maths, Pre-Algebra, Algebra, Trigonometry, Precalculus, and Calculus to specify the category of the problem [5]. Choosing the category provides a set of commonly used symbols for those types of problems and makes solving them easier. However, systems like Mathway have a few drawbacks. They are:

1. Web applications like Mathway absolutely require compatible web browsers. If a browser vendor decides not to implement a certain feature, or abandons a particular platform or operating system version, this may affect a huge number of users.
2. User Interface is not interactive.
3. Mathway cannot be used to solve word problems.
4. A basic understanding of Probability is needed in order to use Mathway efficiently.
5. Applications like Mathway depend entirely on the availability of the server delivering the application. At times server is too busy to solve submitted problems.

To overcome these problems in existing system and also to address the learning requirement of students in online scenario, Word Problem Solver for Probability is developed.

III. THE PROPOSED APPROACH

The system is implemented using the principles of Artificial Intelligence because it has the ability to create never-ending thought process that can be used to solve problems.

3.1. APPROACH TO BUILD THE SYSTEM

The proposed system accepts a problem inputted by the user and aims to solve it. For doing this, it should have the knowledge of the events and conditions mentioned in the problem. It needs to generate sample space and select the sample points that satisfy the mentioned condition in the problem and generate a solution. The knowledge for doing all this can be acquired from previously solved problems. Thus, it requires knowledge restoring from sample spaces stored and probability problems solved and referring to it later. For doing all this, the system will require some internal memory. This internal memory stores the percept history, which is the salient feature of model-based reflex agent [1]. Looking at the approach to build the system, now, let us see the framework of the system.

3.2 FRAMEWORK OF PROPOSED SYSTEM

The system uses object oriented architecture style because an object oriented design views the system as a series of cooperating objects and these objects are discrete and independent. They communicate through interfaces by calling methods and hence provide for reusability, testability, extensibility and make the system highly understandable.

The architecture of the system comprises of five major modules. Figure 1 illustrates the architecture of the system. The interactions of the various modules and their dependence on each other are shown in Figure 1. Now, let us see the role played by each module of the system.
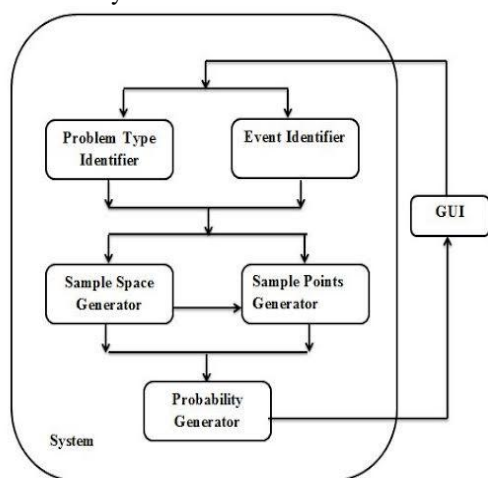

Fig. 1: Architecture of Proposed System

3.2.1 Problem Type Identifier:

As mentioned previously, word problems related to probability are of three types:
1. Specific events are described and probability is asked.
2. A set of variables is given. The probability of specific combinations and permutations (arrangements) is expected.
3. The probability of certain events is given and conditional probability is expected.

Thus before solving any word problem on probability, it is necessary to identify the type to which the problem belongs. The question entered by the user will be analyzed by the Problem Type Identifier Module. This module parses the word problem to identify specific keywords which would determine the type of the problem [2]. Keywords like "combination", "arrangement", "chosen", "selected" would indicate that the word problem is of the second type. Similarly if problem contains description of more than one event or if it contains words like "without replacement", "kept aside" then the module identifies the problem to be of the third type. If problems do not contain above keywords and only

describe one event, the problem is identified to be of the first type by the module.

Consider an example: Two die are thrown. What is the probability that the number on first dice is greater than the number on the second dice? [3].

For the above example, this module analyses the problem and identifies it to be of the "first type".

3.2.2 Event Identifier

After determining the problem type, the next task is to identify the event for which probability is to be determined and it is accomplished by Event Identifier module. For this, Event Identifier module analyses only a specific part of the problem statement. This specific part is that part of the word problem which follows the keywords like "probability", "chances", "odds". For the given example, the event is: "that the number on first dice is greater than the number on the second dice". The event thus identified is then passed to the next module which is the Sample Points Generator.

3.2.3 Sample Space Generator:

After identifying the problem type, the sample space for the mentioned problem will be generated. The module, i.e., Sample Space Generator will generate the sample spaces for problems of Type 1 and Type 2.The probability of any event is given by:

$$P(\text{Event A}) = \frac{n(\text{Sample Space of Event A})}{n(\text{Sample Space of Universal Set})} \qquad (1)$$

Therefore for finding the solution to the problem, first the sample space is calculated. Since the given example is of first type, hence, the sample space generator will generate the sample space which is as follows:

```
{
(1,1), (1,2), (1,3), (1,4), (1,5) ,(1,6)
(2,1) ,(2,2) ,(2,3) ,(2,4) ,(2,5) ,(2,6)
(3,1) ,(3,2) ,(3,3) ,(3,4) ,(3,5) ,(3,6)
(4,1) ,(4,2) ,(4,3) ,(4,4) ,(4,5) ,(4,6)
(5,1) ,(5,2) ,(5,3) ,(5,4) ,(5,5), (5,6)
(6,1) ,(6,2) ,(6,3) ,(6,4) ,(6,5) ,(6,6)
}
```

The number of sample points (for this example, 36) is passed to the module, Probability Generator. Since, the given example is the "first type" probability problem; so, the module will contain the sample sets of common events.

3.2.4 Sample Points Generator

This module takes as input the event identified by the Event Identifier module. This module will identify all the sample points, in the sample space generated by the second module, that satisfy the required condition. First step in identifying the sample points is to scan the event statement for keywords like 'even', 'odd', 'less than', 'greater than', 'together', 'exactly', 'atmost', 'at least', 'prime', 'perfect square'. After this, all the points in the sample space are tested for the condition specified by the keyword. If a set of points satisfy the condition, they are included in the sample points. For the given example, the event is getting a larger number on the first dice when two die are thrown.

```
{
(2,1)
(3,1) ,(3,2)
(4,1) ,(4,2) ,(4,3)
(5,1) ,(5,2) ,(5,3) ,(5,4)
(6,1) ,(6,2) ,(6,3) ,(6,4) ,(6,5)
}
```

The number of sample points (for this example, 15) is passed to the next module, i.e., Probability Generator.

3.2.5 Probability Generator:

After identifying the events and the sample points, now it is possible to calculate the required probability. The module Probability Generator takes the inputs from the 'Sample Space Generator' and the 'Sample Points Generator' modules and using the formula given below it calculates the probability of the given example.

$$P \text{ (Event A)} = \frac{n(\text{Sample Space of Event A})}{n(\text{Sample Space of Universal Set})}$$
$$= 15/36 \qquad (2)$$

In this way the probability problem is solved by generating a step-by-step solution.

IV. EVALUATION AND RESULTS

In this section we start with the walkthrough of the system as a user. The section ends with an evaluation study conducted to test the claim that the Word Problem Solver is a better system.

4.1 Working with the System

The system interface consists of two distinguished text areas and two command buttons. Figure 2 represents the initial screen which appears to the user. The text area named, "Enter the word problem" is reserved for the user to enter the question. The user can either type the question in the textbox or directly paste the question in that space.

The two command buttons, namely, Solve and Clear are also provided on the Graphical User Interface.

Fig. 2: User Screen

The user is expected to click "Solve" button after the question is entered. The solution will be generated in the Solution area only after "Solve" button is clicked.

"Clear" button is used to clear the screen if the user has entered the problem incorrectly. Also, this button can be used to enter a new problem once the student has understood the solution to the current problem.

Let us now analyze how the system works after the problem is entered. The problem can belong to any one of the three types that are mentioned in earlier sections.

A typical dialogue between the user and Word Problem Solver is shown in Figure 3 where user enters the problem - "3 Physics and 4 Chemistry books are to be arranged on a shelf. Find the probability that no 2 physics books are arranged together".

To generate the solution of this question, first, The Problem Type Identifier module will identify the key word "arranged" in the question and mark this problem to be of "Type 2". Now, it is necessary to identify the event for which probability is asked. The Event Identifier Module will analyze the entered problem. For the question entered, the event is "that no 2 physics books are arranged together". Because the entered problem is identified to be of "Type 2" by the Problem Type Identifier, the Sample Space Generator module will generate the sample space for the problem. For the given problem, there are 7 elements in all (3 Physics and 4 Chemistry). The sample space will be generated by using permutation as the question is pertaining to 'arrangement'. Hence, the sample space is 5040 (7! = 5040). The Sample Point Generator module will then identify all the sample points that satisfy the event identified by the Event Identifier module. For the given problem, since no two physics books can be arranged together, the Chemistry books can be arranged in 4P4 ways. The solution area displays the possible arrangement of books. Also, the remaining 3 Physics books can be arranged in 5P3 ways. Hence, the sample points will be generated by taking a multiplication of the two arrangements (Physics and Chemistry) by using the

fundamental theory of multiplication. Thus, the sample points generated is 1440 (4P4 *5P3). The Probability Generator Module will take the above calculated values and generate the probability by using the formula:

$$P \ (Event) = \frac{(\text{Sample Space of Event })}{(\text{Sample Space of Univeral Set })} = \frac{1440}{5040} \qquad (3)$$

The individual calculations of each module are displayed in the solution area thus providing the user with a step wise detailed solution. In this way, the student can understand how to solve the probability problems.
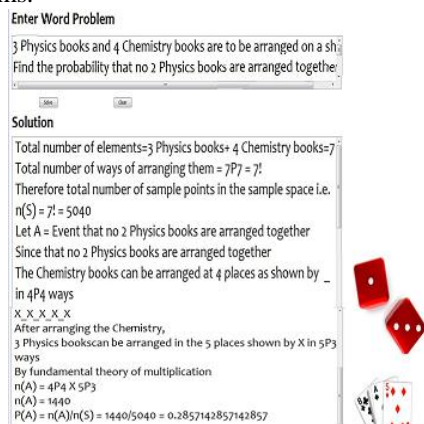

Fig. 3: Solving of Problem Type 2

Consider Case 2 as shown in Figure 4 where user enters the problem "1 dice is thrown and 3 coins are tossed. Find the probability of atmost 1 tail when you get an even number"

To generate the solution of this question, first, The Problem Type Identifier module will parse the word problem to identify the specific keywords. In this case, the problem consists of two keywords, namely, dice and coins. So, it recognizes that the question consists of more than one event and identifies this problem to be of "Type 3". After the type of problem is identified, the Event Identifier Module will analyze the entered problem. For the question entered, the event is "atmost 1 tail when you get an even number". Because the entered problem is identified to be of "Type 3" by the Problem Type Identifier, the Sample Space Generator module will generate the sample space for the problem. For the given problem, there are two different events and hence two different sample spaces are generated. The first event is 1 dice is thrown. The sample space for this event is {1, 2, 3, 4,5,6}. The second event is 3 coins are tossed. The sample space for this event is {HHH, HHT, HTH, HTT, THH, THT, TTH,TTT}. The Sample Point Generator module will identify all the sample points that satisfy the event identified by the Event Identifier module. For the given problem, the event of having atmost one tail is given by P(A) = {HHH, HHT, HTH, THH} and the event of getting an even number is P(B) = {2, 4, 6} The Probability

Generator Module will generate the probability by using the Conditional Probability formula:

$$P(A \cap B) = P(A) * P(B/A) \qquad (4)$$

For a problem involving conditional probability, the first step involved in finding the probability of the dice event are displayed. Then the calculations needed to find the probability of the 3 coins event are shown and the calculation of the final probability is presented. Thus, the student gets better understanding of the word problems from the probability domain.
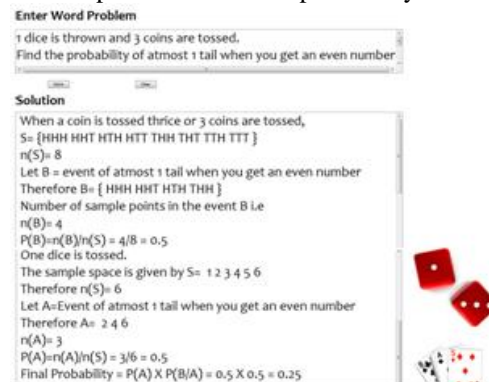

Fig. 4: Solving of Problem Type 3

4.2 Experimental Result and Analysis

A study was conducted so as to understand how the system benefits the users. And thus, a number of students were asked for hands on of the system and their feedback was recorded. Table 1 gives a detailed description of the parameters considered and their evaluation.

Table 1: Test Cases and Results

Sr. No.	Evaluation Parameter	Evaluation Method	Evaluation Result
1	System Approach	Students were allowed to use the system to solve problems	Students found the step-by-step solutions very useful
2	System Design	Usability testing was performed with the help of few students	Students found the application easy-to-use and convenient
3	Syllabus Extent	Students were asked to solve various problems from syllabus	System correctly solved all the problems

To test the claim that Word Problem Solver for Probability is enhanced as compared to all the existing systems to learn, understand and practice Probability, a detailed study was conducted. Students from various board schools were approached. These

students were asked to attend Probability lectures held by teachers in classrooms. Then, the students were allowed to use Mathway to further practice and learn Probability. The same students were also introduced to Word Problem Solver for Probability. They were provided with the user manual of Word Problem Solver for Probability and allowed to use the system. The review and feedback of 75 students was recorded to compare the three approaches, i.e., classroom learning, Mathway and Word Problem Solver for Probability. As shown in Figure 5, the parameters used for the comparative analysis of the three systems are Ease of Use, Level of detail in Solution, Interactivity, and Syllabus Coverage. These parameters are the basic necessity of any teaching learning program and the success of a system in delivering correct knowledge depends on all of these.
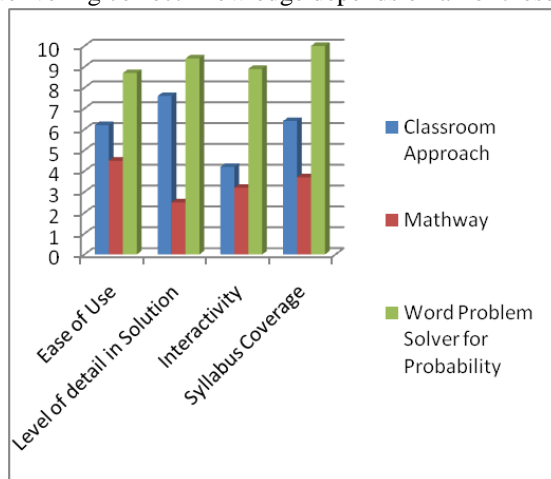

Fig. 5: Comparative Study

From the above analysis, it is evident that Word Problem Solver for Probability works more efficiently and is a better teaching learning model as compared to the existing systems.

V. CONCLUSION AND FUTURE SCOPE

By giving a step-by-step solution to various types of problems of Probability, Word Problem Solver for Probability can now serve as a powerful tool to help students to learn probability and the underlying concepts. Student can use it as a tool to solve his/her own doubts. By using Word Problem Solver student can identify the exact mistakes committed by him/her. The elaborate explanation of numerical can help the student to understand the mistake made by him/her and how to avoid it in future. Student can use the system repeatedly to solve the probability problem to master the concepts. Teacher can use it as teaching aid in classrooms for interactive teaching learning process.

The system can be enhanced further by incorporating audio features with the help of which the students can read aloud their question to the system instead of typing them in the input field. This will not only save the time of the students spent in typing long and elaborate problems but also add an interactive element to the learning process. Also, using an optical reader that can convert the question from an image into readable text can be integrated with the system. This will enable the students to click a picture of the question by a cell phone camera and use the same to obtain the detailed solution. The system can be extended to cover a large number of topics in Mathematics where students face difficulties. Systems like 'Word Problem Solver for Probability' can revolutionize the way mathematics is taught and learnt.

REFERENCES

[1] Stuart Russell, Peter Norvig, *Artificial Intelligence - A Modern Approach*, Second Edition, Pearson Publication, 2013.
[2] Wanintorn Supap, Kanlaya Naruedomkul, Nick Cercone, *Steps Towards Accurate Word Problem Translation*, IEEE, 2009.
[3] Algebra Standard Ninth Textbook, *Maharashtra State BoarFd of Secondary and Higher Secondary Education*, 2012.
[4] Statistics Glossary - Probability, "http://www.stats.gla.ac.uk/steps/glossary/probability.html", last accessed on 24-04-2014.
[5] Mathway, "http://www.mathway.com", last accessed on 15-10-2013.