Convert Probability Network to Artificial Neural Network based on Position, Time and Speed of Events

Imam Cholissodin1* and Marji1

1Faculty of Computer Science, Brawijaya University, Malang, Indonesia

*Corresponding Author: imamcs@ub.ac.id

Abstract. Building a probability function to get the best answer between events in many review problems is a very complex one. In this study, several events were given as a case study to find a solution, after which an analysis will be carried out when and where the event contained position and time features. So that the distance factor, travel time factor and speed factor for each event can be calculated. Then a concept comparison between Probability Network from the Bayes Theorem versus Artificial Neural Network (ANN) was made, accompanied by a review of the Propositional Logic (PL) or First Order Logic (FOL) concept compared to using the Uncertainty concept in calculating the probability of an event that has often been ignored because it is considered to have no influence and the probability is small whereas it can be the opposite. In fact, sometimes, the pseudo event is the main factor determining the real event and the probability is much greater than expected. In addition, we also give an Engineering Events Theorem that can intentionally set event-by-event syntheses to increase the probability of an event which is initially smaller than the probability of another event. After being given an injection/dummy/synthesis event, however, then the initially small probability of the event becomes very high and the time of occurrence can be accelerated or decelerated, or even minimized so that it does not occur.

1. Introduction

Probability is the possibility of any event that is calculated based on the most frequent occurrence to the number of trials. Probability is related to ideas or concepts of opportunity or possibility. This concept has been more rigorously formulated in mathematics, and is then used more widely not only in mathematics or statistics, but also in science and technology.

In 1763, an Englishman named Thomas Bayes was a very prominent figure in the field of probability. Bayes theorem is a theory used to solve conditional or partitioned opportunities which are an approach to an uncertainty measured by probability. The Bayes theorem is used on an event where in an experiment that produces 2 possible events, namely event A and event B, under the condition that both events are dependent on each other, the occurrence of event A will affect the probability of occurrence of event B. In accordance with subjective probabilities, if someone observes event B and has the confidence that there is a possibility that B will appear, then probability B is called the prior probability. After there is additional information that for example event A has appeared, there may be a change in the original estimate of the possibility that B will appear. The probability for B now is a conditional probability due to A and referred to as posterior probability. The Bayes theorem is a mechanism for renewing probabilities from priors to posterior probabilities. To date, there have been many applications...
of the Bayes theorem, for example for decision-makers in the presidential election, the development of insurance companies to target market share, diagnosis of a disease and data classification [1][2][3][4].

In this study, we are going to build a probability function to obtain answers to what factors actually affect events in the review of the above problems. First, we will provide some real sample events that actually take place in life that can be used as case studies to find the solution. Second, we will analyze when and where the event has time, speed and distance. Third, we will calculate the velocity factor (VF), time factor (TF) and distance factor (DF) for each event. And fourth, we will make a comparative study of the calculation of the probability of occurrence between the Artificial Neural Network (ANN) and the Bayes Theorem.

From the results of this study, it is expected that we will later build a more complex study of opportunities that can contribute to science which is actually much avoided because it is as if only using the existing theorems. In the future, we will calculate the probability of an event (V-T-D-F) in the event that has DF = 0 and DF <0, which is calculating the probability of a pseudo-event that has often been ignored because it is considered to have no influence and the probability is small. On the other hand, we think otherwise that the pseudo-event can actually be the main factor determining the real event and the chances are much greater than expected. Furthermore, we will also provide an Engineering Events Theorem that can be deliberately fabricated to increase the probability of an initial event having a smaller probability value among other events, but after an inject event/dummy event/synthesis event, then the small probability becomes very big.

2. Method

2.1. “Probability Network” Versus “ANN based ELM”

Probability algorithms that consider natural weighting factors for each event behind the events are based on a mutual exclusive model or another model to obtain probability values through weighting factors on each event behind both unexpected and invisible events, and modeling opportunities for events behind events. Thus, it can provide a solution to the complex probability calculation method in every event behind events in detail based on the mutual exclusive model and others. In any event, if the probability value is calculated, then the formula is always the number of occurrences of the event being searched divided by the number of events. The question, however, is how to solve or prove the equation.

ANN is very different from Probability Network because when viewed visually, on probabilities, the connecting path is only limited to showing the preceding event and the next event, and the occurrence value must be taken from the frequency of occurrence of the event that has been compared with all the frequency of events and the learning processes really require extra preparation from the user. For ANN, on the other hand, the pathway between events can be stated, one of which becomes the weighting of the occurrence value which allows the system to be able to carry out the learning process independently.

![Figure 1. “Probability Network” Vs “ANN based ELM”](image)
to continue to perform weight improvements automatically; hence, it has minimal human intervention as a user or developer. The network architecture of both can be compared clearly. At Fig. 1, it can be observed that the ANN architecture is very structured with layers that are detailed and neatly arranged that are very similar to the network in the human brain while Probability Network is very random and irregular, and looks chaotic.

2.2. Proposed Method: Convert Probability Paradigm to ANN Based on Event Decomposition

No matter how high the probability is when \( P((x),(t)) \sim 1 \) or even \( = 1 \), if the event does not occur, then \( P((x),(t+1)) = 0 \), and whatever the small value of \( P((x),(t)) \sim 0 \) or \( = 0 \), if the event occurs, however, then \( P((x),(t+1)) = 1 \), where \( x \) represents occurrence (an event) and \( t \) represents time. If the probability value is equalized as one of the features of an outcome of an event, then this probability value is considered to be very less contributing to the results of the decision; thus, the concept of probability itself needs to be improvised where in this case we propose a required technique or an alternative for the probability concept which has been used so far, which is the conversion process from probability paradigm to artificial neural network based on decomposition of cause events and decision events, and between the two, pseudo layers, namely hidden layer, is given and we also add distance, time, and speed of events features. In this study, the term convert is interpreted as a way of transforming beyond the limit of concept probability replaced with learning models by changing “time series probability” or a form of probability network into an artificial neural network. The following Figure 2 illustrates 2 events, for example A and B using conditional probability and a mutual exclusive model. In addition, part (iii) has the time of events in the form of time series.

![Figure 2](https://via.placeholder.com/150)

**Figure 2.** Illustration of Probabilities A and B (conditional, mutual exclusive, time, distance).

If using Fig. 2 (i), then Equation (1) which indicates that the two events above have a non mutual exclusive can be formulated. Meanwhile, if based on Fig. 2 (ii), it can be formulated as in Equation (3) which states that two events are not interrelated or one event is not attached to another event (between mutually exclusive events, including mutually exclusive).

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}
\]

(1)

Based on Equation (1), it can be described as Equation (2), non mutual exclusive.
\[ P(B|A) = \frac{P(A \cap B)}{P(A)} \]

\[ P(B|A) = \frac{P(A|B)P(B)}{P(B|A)P(A) + P(A|B)P(B)} \quad (2) \]

\[ P(A \lor B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (3) \]

Based on Equation (3), it can be broken down into Equations (4 & 5), mutual exclusive.

\[ P(A \cap B) = 0 \quad (4) \]

\[ P(A \lor B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \lor B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

Based on Equation (3), it can be broken down into Equations (4 & 5), mutual exclusive.

\[ P(A \cap B) = 0 \quad (4) \]

\[ P(A \lor B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \lor B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]

The concept of probability can be explored more deeply by describing each event with its decomposition events. For example, Based Fig. 3, when analyzed from \( P(A \to B) = P(B \mid A) \), then we get an event set from A and B. Set A = sub Feature Set of A = \( \{a_1, a_2, \ldots, a_n\} \), Set B = sub Target Set of B = \( \{b_1, b_2, \ldots, b_n\} \).

\[ P(A \to B) \neq P(B|A) \]

\[ P(A \to B) = P(B \mid A) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P(A \to B) = P(A \cup B) = P(A) + P(B) \quad (5) \]
In Equation 6 there are things that seem to ignore event A. This is the first problem. Then the second one is related to the delta value or the difference between Conditional Probability and Probability of implication \( |\Delta P_{ic}| \) shown in Equation 7.

\[
|\Delta P_{ic}| = \frac{P(\neg A)P(\neg B)}{P(A)P(\neg A) + P(A)}
|\Delta P_{ic}| = \frac{P(\neg A)P(\neg B)}{P(A)}
\]  

(7)

If the concept of Probability of implication from \( P(A \rightarrow B) \), the truth table approach is used, where calculated \( P(A \rightarrow B) \) is True. Previously we need to identify when the implications are true (True), or when they are false (False). If taken a simple one then \( A \rightarrow B \) is False, when A is True and B is False. Then Equation 5.3 is obtained, which is calculating \( P(A \rightarrow B) \) is True from the negation \( P(A \rightarrow B) \) is False. To find out the detailed description of Equation 5.3, it can be explained by the truth table by looking at the concept of congruence or equality between columns \( (A \rightarrow B) \) and \( (\neg A \lor B) \).

\[
P(A \rightarrow B) = 1 - (P(A = True)P(B = False))
P(A \rightarrow B) = 1 - (P(A)P(\neg B))
\]  

(8)

In Equation 8 there are things that do not seem to ignore the occurrence of A, and this is what distinguishes it from Equation 6. Then related to the delta value or the difference between the Conditional Probability and the Probability of implication base truth table \( |\Delta P_{itc}| \) is shown in Equation 9.

\[
|\Delta P_{itc}| = \frac{P(A)(1 - P(A)P(\neg B)) - (P(A) - P(\neg B))}{P(A)}
\]  

(9)

**Figure 4.** Conditional Probability Vs Bayesian Network.

**Table 1.** Bayesian Network

|      | (i)                                      | (ii)                                      |
|------|-----------------------------------------|------------------------------------------|
| A=True| A=1−P(A)                               | B=True P(B|A)=\frac{A \cap B}{A} \quad 1−P(B|A) |
| A=1−P(A)   | B=False P(B|\neg A)=\frac{\neg A \cap B}{\neg A} \quad 1−P(B|\neg A) |

\[ P(A) \]
Based on Figure 4 (ii) as a form of P (A → B) or approached by P (B | A), which is about Bayesian Network, it can be mapped the value of opportunities in the form of Table 1, as a value that has been previously known. From Table 1, the values of P (B) and P (B → A) can be calculated or approached by looking for the value of P (A | B). Equation 10 is a description to get the value of P (B), while Equation (1) to calculate P (B → A) and P (A | B).

\[
P(B = T) = \sum_{i=0}^{2} (P(A = i) P(B = T | A = i)); i = 0 = False(F), i = 1 = True(T)
\]

\[
P(B = T) = (P(A = F) P(B = T | A = F)) + (P(A = T) P(B = T | A = T))
\]

Equation (10)

Based on Figure 5 (ii) as a form of P (A → B) or approached by P (B | A), which is about Bayesian Network, it can be mapped the value of opportunities in the form of Table 1, as a value that has been previously known. From Table 1, the values of P (B) and P (B → A) can be calculated or approached by looking for the value of P (A | B). Equation 10 is a description to get the value of P (B), while Equation (1) to calculate P (B → A) and P (A | B).

\[
P(B = T) = \sum_{i=0}^{1} \sum_{k=0}^{1} \sum_{j=0}^{1} \left( \prod_{j=1}^{n} P(A_j = t_j) \prod_{j=1}^{n} P(B = T | A_j = t_j) \right)
\]

Equation (11)

In Equation 11 is a description of "Bayesian Network with n-Events of A". But in most cases, the output (Event of B) can also be "m-Events of B" which if visualized will be similar to that in Figure 5 (i). So that it can also be modeled in the form of Equation 12 which is a description of "Bayesian Network with Events of A & m-Events of B", which is y = \{1, 2, ..., m\}, k = 0 = False (F), k = 1 = True (T), and \(t_1, t_2, ..., t_n\) are identical to \(k\).

\[
P(B_y = T) = \sum_{i=0}^{1} \sum_{k=0}^{1} \sum_{j=0}^{1} \left( \prod_{j=1}^{n} P(A_j = t_j) P(B_y = T | A_j = t_j) \prod_{k=0}^{1} \prod_{x \in \{1, 2, ..., m\} \setminus y} P(B_x = k | A_j = i_j) \right)
\]

Equation (12)

Based on Equation 12, the results of P(B₁), P(B₂), and P(Bₘ) are obtained. Whereas if you use the ANN-based ELM with n-Events of A (as features) & m-Events of B (as target), you can see the training process in Equation 19, while the testing process is in Equation 20 based on Fig. 1 (ii), which is X ≈ n-Events of A, while Y ≈ m-Events of B which will be included in Equation 13. The following is a description of the derivative with the least square estimator (LSE) method to obtain a beta estimator from ELM based ANN, start from Equation (13-18) [7].

\[
H = \frac{1}{\left(1 + \exp\left(-X_{train} W^T + \text{ones}(N_{train}, 1) b \right) \right)}
\]

Equation (13)

\[
H \beta = Y
\]

Equation (14)

\[
H \beta + \epsilon = Y
\]

Equation (15)

\[
\epsilon = Y - H \beta
\]
\[
\min_{\beta} (\varepsilon^T \varepsilon) = \min_{\beta} (Y - H\beta)^T (Y - H\beta) \\
= \min_{\beta} (Y^T Y - Y^T H\beta - (H\beta)^T Y + (H\beta)^T (H\beta)) \\
= \min_{\beta} (Y^T Y - Y^T H\beta - \beta^T H^T Y + \beta^T H^T (H\beta)) \\
= \min_{\beta} (Y^T (Y - H\beta) - \beta^T H^T (Y - H\beta)) \\
\frac{\partial}{\partial \beta} (\varepsilon^T \varepsilon) = \frac{\partial}{\partial \beta} (Y^T (Y - H\beta) - \beta^T H^T (Y - H\beta)) = 0 \\
= \frac{\partial}{\partial \beta} (Y^T Y - Y^T H\beta - \beta^T H^T Y - \beta^T H^T H\beta) = 0 \\
= \frac{\partial}{\partial \beta} (Y^T Y) - \frac{\partial}{\partial \beta} Y^T H\beta - \frac{\partial}{\partial \beta} \beta^T H^T Y + \frac{\partial}{\partial \beta} \beta^T H^T H\beta = 0 \\
= 0 - (Y^T H)^T - (Y^T H)^T + 2H^T H\beta = 0
\]

(16)

Derivative results from \( \beta \) based Equation 16 can be further more simplified. Where are the variables from \( \beta \) changed into \( \hat{\beta} \) as the value of the estimator which can directly be calculated by only operating the matrix \( H \) with \( Y \) as in Equation 17.

\[
-H^T Y - H^T Y + 2H^T \hat{H} = 0 \\
-2H^T Y + 2H^T \hat{H} = 0 \\
-H^T Y + H^T \hat{H} = 0 \\
H^T \hat{H} = H^T Y \\
(H^T H)^{-1} (H^T H) \hat{\beta} = (H^T H)^{-1} H^T Y \\
\hat{\beta} = (H^T H)^{-1} H^T Y \\
\]

(17)

If in ELM, Equation 17 can be written as Equation 18.

\[
\hat{\beta} = H^+ Y
\]

(18)

If Equation 14 is reverse the side arrangement becomes \( Y = H\beta \) and “ \( \beta \) changed into \( \hat{\beta} \) ” along “ \( Y \) changed into \( \hat{Y} \) ”. Then we get Equations (19) and (20).

\[
\hat{Y} = \left( H = \frac{1}{\left[1 + \exp\left(- \left( X_{\text{train}} W^T + \text{ones}(N_{\text{train}}, 1) b \right) \right) \right]} \right)^{\left[(H^T (H))^{-1} H^T \right]} \left[Y\right] \\
\hat{Y} = \left( \frac{1}{1 + \exp\left(- \left( X_{\text{test}} W^T + \text{ones}(N_{\text{test}}, 1) b \right) \right) \right)^{\left[(H^T (H))^{-1} H^T \right]} \left[Y\right]
\]

(19)

(20)
If based on Figure 3, then the Feature Set = \{A, B, C\}, while the Target Set = \{B, C, D, E\}. There are members of the set included in both parts, namely the event node as a feature and also as a target. This is a kind of gradual learning process, for example, the target results in the first stage will be a feature for the next stage although there are some with only one stage. Besides the learning process, there will definitely be a testing process for every related event or even an unrecognized relationship, meaning that from the two events at Fig. 3, there is an unrecognized relationship, but there is actually a relationship between the two events. When taken to ANN, then, the direction of the arrow from the causal event (as a feature) that shifts to the resultant event (as a target) will be reversed for different learning processes.

Next is how to make the possibility of D events occur faster based on Figure 3 (ii). In this case, several alternative solutions can be selected. All alternatives can be mapped as follows:

- Select P(A→C→D), or
- Select P(A→B→D)

At the first point, namely “Select P(A→C→D)”, there are 2 stages of the learning process. The first stage is (A→C) and the second stage is (C→D). Similarly, in the second point, it can also be translated into 2 stages. The question is whether it is possible to do by-pass or shortcut that can connect A directly to D (A→D) or from D itself, which if mapped, will bring up a pseudo path that is actually invisible or nonexistent at the following point:

- Select P(A→D), or
- Select P(D(t_i)→D(t_j)), where \( t \) is the time of the event, and \( i, j \) denotes the time index \( (i<j) \).

Based on the 2 points above, if visualized in the form of a graph, the pseudo path represented by a dashed line will look like in Figure 6. The graph formed states a part of the alternative path that can be used as a liaison between each node, as well as nested on the node itself. If between each node, then one or several nodes become the feature nodes and one or several other nodes function as the target nodes whereas if using nested conditions, then one node has a dual role, as a feature node as well as the target node, using the concept of events at previous times as in the second point.

|     | A | B | C | D | E |
|-----|---|---|---|---|---|
| A   | 0 | 2 | 4 | ad|   |
| B   | 0 | 3 |   |   |   |
| C   | 0 | 2 | 1 |   |   |
| D   |   | 0 |   |   |   |
| E   |   |   |   |   | 0 |

(i)

(ii)

![Figure 6. Example Table and Network of Events (Part 2).](image)

3. Results and Discussion

3.1. Mapping any type feature as “ANN based ELM”

The technique used in this mapping is to change the learning paradigm (learning system model) from the possibility, which the system will be set to learn from reality. The term "possibility" is closer to the concept of probability, while "reality" is closer to ANN. So that the probability value will be replaced by the intensity of the event as a predicted or predicted target which includes the speed factor (VF), time factor (TF) and distance (DF) and other factors each incident. So in the mapping features can be divided into several sections as in Figure 7 which shows in detail the distribution of features used in ANN. It can also be used for mapping targets.
| Feature 1st | ... | ... | Feature kth |
|------------|-----|-----|------------|
| Position of Event | dd-mm-yy- | hh-minute- | ... |
| Event start | ... | ... | Feature end |
| (TF1) | hh-minute- | ... | (TF2) |

**Figure 7.** Mapping features of ANN based ELM.

Why must use ANN as an alternative solution to some problems in probability concepts, it can be understood that when we predict B from A \( (A \rightarrow B) \). If using the concept of probability, then A and B must know the value of the intensity of their respective events. This is very different from ANN, where there is a learning and testing process, indeed when the training process the intensity values A and B must be known, but when the testing process, only the intensity value A is known. While the intensity value B will be obtained from ANN_{in}(A).

### 3.2. Result Convert Probability Paradigm to ANN

In this section a simple transformation process is given as an example of the Convert Probability Paradigm to ANN. Look again Fig. 2, for example there are 2 events, namely A and B using conditional probability and a mutually exclusive model. If calculating the result of \( P(B \mid A) \), it can be described as follows.

\[
P(B \mid A) \approx P(A \rightarrow B) \\
\approx P(\text{if } A \text{ then } B) \\
\approx P(\neg A \lor B)
\]

(21)

So if using ANN, it can be ported into 2 solutions, first by involving both components \( (A \rightarrow B) \) with a time factor, as in Equation 22. Second by just involving component A as a feature, as in Equation 23. Then Sub features of \( A = \{a_1, a_2, ..., a_n\} \) appear. So that at \( P(A \mid B) \) if an approach is made to \( \approx P(B \rightarrow A) \), it can also use the rules in Equations (21-25).

\[
\approx \text{ANN}_{in}[(A \rightarrow B)_{i \rightarrow t}] \rightarrow \text{ANN}_{out}[(A \rightarrow B)_{t}] \\
\approx \text{ANN}_{in}[(A)] \rightarrow \text{ANN}_{out}[(B)]
\]

(22)

(23)

If calculating the result of \( P(A) \), it can be translated into \( P(A_{i,t} \rightarrow A_t) \), where \( t \) is the time series when specified as the target, \( i = \{1, 2, ..., nFeatures\} \). So if it is changed using ANN, it can be seen as in Equation 24. If \( P(B) \), it can be described using ANN as in Equation 25.

\[
\approx \text{ANN}_{in}[(A)_{t \rightarrow i}] \rightarrow \text{ANN}_{out}[(A)_{i}] \\
\approx \text{ANN}_{in}[(B)_{i \rightarrow t}] \rightarrow \text{ANN}_{out}[(B)_{t}] \\
\]

(24)

(25)

### 4. Conclusion

Several illustrations have succeeded in showing a probability network that can be easily converted into an artificial neural network. Then the mapping of features used is based on changing the learning model by the possibility that is replaced by learning from the data reality. The results of this study do not assume that the paradigm of probability network conversion to ANN is the same as ANN Probability, but the difference is very far, where \( \text{ProbNet} \rightarrow \text{ANNProb} \). Then, the related techniques that can be used start from appearing pseudo paths between events (nodes to nodes (n2n)) to creating nested
paths for the event itself (single node to self (n2self)) or combining both, which is “n2n + n2self”. Since this study only focused on conceptual ideas, then the next study will implement the ANN results to be evaluated related to the effectiveness and efficiency with the probability network. This certainly will involve several cases to compare the results between the two algorithms although in theory, it is very clear that if the ANN approach is closer to or similar to the way the human brain works in solving any cases. Meanwhile, the probability of containing the concept of uncertainty is very deep and has many limitations when compared to ANN. Furthermore, in the future, every event will be carried out by mapping using Convolution Neural Network (CNN), which in turn makes ANN able to be directed to deeper learning (Deep Learning algorithm) for every event, both simple and very complex one.

References

[1] McGrayne, Sharon Bertsch. (2011). “The Theory That Would Not Die: How Bayes' Rule Cracked The Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy”. New Haven: Yale University Press. 13-ISBN 9780300169690/10-ISBN 0300169698; OCLC 670481486.

[2] Hajek, A. and Hartmann, S. (2010). "Bayesian Epistemology", in: Dancy, J., Sosa, E., Steup, M. (Eds.) (2001) A Companion to Epistemology, Wiley. ISBN 1-4051-3900-5 Preprint.

[3] Newberg, L. (2009). "Error statistics of hidden Markov model and hidden Boltzmann model results". BMC Bioinformatics 10: 212.

[4] Phillips, L.D.; Edwards, W. (October 2008). "Chapter 6: Conservatism in a simple probability inference task (Journal of Experimental Psychology (1966) 72: 346-354)". In Jie W. Weiss and David J. Weiss. A Science of Decision Making:The Legacy of Ward Edwards. Oxford University Press. pp. 536. ISBN 978-0-19-532298-9.

[5] Cholissodin, I., Sutrisno. 2018. “Prediction of Rainfall using Simplified Deep Learning based Extreme Learning Machines”. Journal of Information Technology and Computer Science (JITECS), accredited by number 21/E/KPT/2018 valid from July 9, 2018 to July 9, 2023.

[6] Russell, Stuart dan Norvig, Peter. 2003. Artificial Intelligence: A Modern Approach. International Edition, Edisi 2. Pearson Prentice-Hall Education International. New Jersey.

[7] Derivations of the LSE for Four Regression Models. http://facweb.cs.depaul.edu/sjost/csc423/documents/technical-details/lsreg.pdf