Scrutinize Analysis of Crime Investigation – Assigned Weighted Approach
Thiagarajan Kittappa¹*, S. Mrithul Snehal², J. Amrutha³, J. John Osborne⁴, A. Nishanth⁵ and Anish Anburaja⁶

Abstract
This study focuses on trust rating between police records and observed current case’s data, employing a weighted graph approach to make informed decisions in different criminal investigations. Four distinct cases, ranging from theft to homicide, were examined, each with its own assigned trust levels. The outcomes of these cases varied, with trust ratings playing a significant role in influencing the final decisions made by law enforcement agencies. The results underscore the importance of considering specific evidence and circumstances in the determination of suspects guilt or innocence in AI based outputs.

Keywords: Agent, Clients, Weighted Signed Graph.

Introduction
In today's world, there are a lot of criminal offenses happening around the world. And these criminal offense data are useful for investigating the crime, and the availability of data pertaining to these criminal incidents plays a pivotal role in the process of crime investigation and the apprehension of offenders. By correlate the similarities between the data in the police record and the data collected by the crime scene investigators to determine if there was an AI based pattern between these criminal acts. If there is data available that supports this hypothesis. Department of Investigators can conclude that these acts were done by a proposed criminal entity.

The objective of the research is to enhance the operational efficiency and facilitate more accurate identification of potential suspects. This proposed algorithm discusses the trust level rating through a weighted sign approach to improve efficiency. The outcomes of this study hold significant promise for law enforcement agencies, enabling them to expedite and streamline the investigation process, ultimately contributing to the pursuit of justice in an optimized manner.

Definitions

Simple Match Coefficient (SMC)
The term "simple match coefficient" is often used in the context of data matching and record linkage. It's a measure of how well two data records match or correspond. The simple match
Scrutinize Analysis of Crime Investigation – Assigned Weighted Approach

coefficient is typically a binary measure, indicating whether two records match (1) or do not match (0) based on certain criteria or attributes. It's a straightforward way to assess similarity or agreement between data records.

**Rao's Coefficient (RC)**

Rao's coefficient, also known as Rao's quadratic entropy, is a statistical measure used to assess the dissimilarity or distance between two probability distributions. It is named after the Indian-American statistician C. Radha krishna Rao.

Rao's coefficient is a mathematical way to quantify how two probability distributions differ from each other. It's often used in fields like statistics, information theory, and ecology to compare the similarity or dissimilarity between datasets. The formula for Rao's coefficient involves the use of probabilities from both distributions, and the result can range from 0 (indicating identical distributions) to a positive value (indicating increasing dissimilarity).

**Jaccard Coefficient (JC)**

The Jaccard coefficient produces a value between 0 and 1, where 0 indicates no similarity (the sets have no elements in common), and 1 indicates complete similarity (the sets are identical). It's a useful metric for comparing the similarity of data samples or sets in various applications, such as text analysis, recommendation systems, and biology.

- **SMC**: \( S(i, j) = \frac{P+S}{P+Q+R+S} \)
- **RC**: \( S(i,j) = \frac{P}{P+Q+R+S} \)
- **JC**: \( S(i,j) = \frac{P}{P+Q+R} \)

Table 1.1: Consider some examples for table 1.1

<table>
<thead>
<tr>
<th>PA</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>++</td>
<td>Past Record Matches with Present</td>
</tr>
<tr>
<td>+ -</td>
<td>Past Record Matches but The Newly Collected Data Has Some Percentage of Variations.</td>
</tr>
<tr>
<td>- +</td>
<td>First Time the Record Is Being</td>
</tr>
<tr>
<td>- -</td>
<td>Past Record Doesn’t Match with Present Collected</td>
</tr>
</tbody>
</table>

Flow Diagram of Proposed System
Examples

Case (I)

Necessary records for this case,
1) Past Record (PR)
2) Newly Collected Record (NCR)

Type of Crime is Theft

Assume that, the police database has case records on various types of crimes which are committed by an individual or a group of people. In this scenario consider theft to be the crime.

<table>
<thead>
<tr>
<th>PR</th>
<th>NCR</th>
<th>TL</th>
<th>TR</th>
<th>DECISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>++</td>
<td>++</td>
<td>++</td>
<td>P1</td>
<td>COMPLETE TRUST LEVEL</td>
</tr>
<tr>
<td>++</td>
<td>+</td>
<td>-</td>
<td>P2</td>
<td></td>
</tr>
<tr>
<td>++</td>
<td>-</td>
<td>-</td>
<td>P3</td>
<td></td>
</tr>
<tr>
<td>++</td>
<td>--</td>
<td>--</td>
<td>P4</td>
<td></td>
</tr>
</tbody>
</table>

(P1+P2+P3+P4)/4 = COMPLETE TRUST LEVEL

Use cases

- (++); The police concluded in the report that the crime happened between 11pm-1am and there is data available about the crimes that happened between 11pm-1am in database and the data’s which are available in database conveys that theft is the major cause committed between 11pm-1am
- (+-); The police also reported that the crime happened in a nearby shop (specific). Even though the database has complete data about the crime happened in that area but the data about shop (specific) is not present in the database so the similarity decreases.
- (-+); The police reported that these are foot prints along the pathway and collected the data but in the past record there is no data which has the footprints of the criminals.
- (--); An outsider was found roaming in the area around the above-mentioned time but his shoe size is larger than the footprints found at the crime scene and the database has data about the shoe size of criminals but not above-mentioned size.

Cases (II)

Necessary records for this case,
1) Past Record (PR)
2) Current Case Record (CCR)

Type of Crime is Murder

Consider an example in which crime scenario is taken has murder. Assume that the police of particular Jurisdiction suspect a person for murder. The weapon used for murder was a knife. This knife is similar to that of knife which is used in meat shop and the person who is accused for the murder is a meat shop owner.
Scrutinize Analysis of Crime Investigation – Assigned Weighted Approach

<table>
<thead>
<tr>
<th>PR</th>
<th>CCR</th>
<th>TL</th>
<th>TR</th>
<th>DECISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>++</td>
<td>--</td>
<td>P2</td>
<td>GOOD TRUST LEVEL</td>
</tr>
<tr>
<td>+</td>
<td>+-</td>
<td>--</td>
<td>P3</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>-+</td>
<td>--</td>
<td>P5</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>--</td>
<td>--</td>
<td>P5</td>
<td></td>
</tr>
</tbody>
</table>

(P2+P3+P5+P5)/4 = GOOD TRUST LEVEL

Use Cases
- If the client data matches with the agent but the person is deceptively accused, so the trust level will be (+-)
- If the client data have a similarity to the victim and the police believe that the person is the one who committed the crime then the trust level will be (-+).
- If the client data doesn’t have a similarity to the agent but the police accuses the person to be the victim then the trust level will be (--).
- If the client data data doesn’t match with the person but the police believes that the person was the victim (--).

Case (III)
Type of Crime is Kidnapping
Assume, that there is a kidnapping case at some specific place which is being handled by police of the same Jurisdiction. They believed that the kidnapping case is related to Maruti Omni van. Hence the police suspected the driver as the criminal but he was innocent.

<table>
<thead>
<tr>
<th>PR</th>
<th>CCR</th>
<th>TL</th>
<th>TR</th>
<th>DECISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>++</td>
<td>--</td>
<td>P3</td>
<td>BAD TRUST LEVEL</td>
</tr>
<tr>
<td>+</td>
<td>+-</td>
<td>--</td>
<td>P5</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>-+</td>
<td>--</td>
<td>P5</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>--</td>
<td>--</td>
<td>P5</td>
<td></td>
</tr>
</tbody>
</table>

(P3+3P5)/4 = BAD TRUST LEVEL

Use Cases
- If the client data matches even though the police does not suspect the criminal, then the trust level will be (-+)
- If the client data have similarities with the criminal but not with the person who was accused by the police to be a criminal then the trust level will be (--).
- If the client data have similarities with the criminal accused by the police then the trust level will be (->) • If the client data doesn’t have similarities with the criminal then the trust level will be (--).

Case (IV)
Necessary records for this case,
1) Police Doctrine (PD)
2) Behavior Of Suspection (BS)

Type of Crime is Homicide
Consider, that the police had collected the evidence of a murderer who engaged in the unlawful killing, but the evidence which they extracted has no reliability to the suspected personality, but there is either or possible offender who deliberate the suspicion with the previous or present actions.
Use Cases

- The murder has been done by the client, but the police doesn’t suspect him, the trust level will be (++)
- The murder has done in previous cases but not at present, yet the police doesn’t suspect him, the trust level will be (--).
- The murder has not done in previous cases but at present it had been done, yet the police doesn’t suspect him, the trust level will be (--).
- The murder has not been done by the client, but the police doesn’t suspect him, so the trust level will be (--).

<table>
<thead>
<tr>
<th>Respective variable case</th>
<th>Data in police record</th>
<th>Collected current case data</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Q</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>R</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>S</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Experimental Results and Comparison with Existed Records

For Case (I)

Observed data:

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>75</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>SMC</td>
<td>RC</td>
<td>JC</td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>2/5</td>
<td>4/9</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION: BAD TRUST LEVEL

For Case (II)

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>75</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>SMC</td>
<td>RC</td>
<td>JC</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION: BAD TRUST LEVEL

For Case (III)

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>SMC</td>
<td>RC</td>
<td>JC</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION: BAD TRUST LEVEL

For Case (IV)

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>SMC</td>
<td>RC</td>
<td>JC</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION: BAD TRUST LEVEL

Conclusion

Case I

The final decision may be influenced by the trust ratings of the agents involved. In this case,
person 1 complete trust and the high trust level of person 2 would hold more weight, making the conclusion about the time of the crime and the absence of shop-specific data relatively more reliable. However, conclusions related to footprints and the outsider's shoe size may be less reliable due to the lower trust levels. The average trust rating would depend on how much weight is assigned to each agent's input.

**Case II**

In the murder case, the trust levels vary depending on the situation, with the highest trust level (+ +) indicating a strong belief in the accused person's guilt and the lowest trust level (- -) suggesting a lack of evidence and similarity between the accused and the victim. The final decision would depend on the specific evidence and circumstances surrounding the murder case.

**Case III**

In the kidnapping case, trust levels vary, but they tend to be low, due to the possibility of wrongful accusations and the need for further investigation to determine the actual perpetrator. The final decision should consider concrete evidence and facts in order to establish the innocence or guilt of the accused driver.

**Case IV**

In this homicide case, trust levels vary, but they generally tend to be low, as the police do not suspect the client because the collected evidence is not reliable to the suspected personality. The final decision should consider the existing evidence in the given scenario and facts to determine the true culprit, to analyze whether the suspected person is involved or not.

**Nomenclature**

PA: Precedence of Acceptance  
PR: Past Record  
NCR: Newly Collected Record  
CCR: Current Case Record  
PD: Police Doctrine  
BS: Behavior of Suspection  
SMC: Simple Match Coefficient  
RC: Rao’s Coefficient  
JC: Jaccard’s Coefficient  
TR: Trust Rating  
TL: Trust Level

**Acknowledgement**

The authors would like to thank Prof. Ponnammal Natarajan worked as Director – Research, Anna University - Chennai and Dr. Sundar R Currently Director, Rajalakshmi Institute of Technology, Chennai, for their intuitive ideas and fruitful discussions with respect to the paper’s contribution.

**References**

Mohammad Reza Keyvanpour, Mostafa Javideh, Mohammad Reza Ebrahimi “Detecting and investigating crime by means of data mining: a general crime matching framework”. Department of Computer Engineering, Alzahra University, Tehran, Iran Islamic Azad University, Qazvin Branch, Qazvin Iran.  
K. Thiagarajan, A. Ragunathan, Ponnammal Natarajan, G. Poonkuzhali and Prashant Ranjan”
Weighted Graph Approach for Trust Reputation Management”, World academy of science, Engineering and Technology 56 2009.


M.R. Keyvanpour et al. / Procedia Computer Science 3(2011) 872–880 879 Mohammad Reza Keyvanpour / Procedia Computer Science 00 (2010) 000–000


