Peer Reviewed
Received 11 March 2017
Received in revised form 15 May 2017
Accepted 15 May 2017
Available online 2 June 2018

Forensic Science Seminar
ISSN 2157-118X
Volume 8 Number 2
30 June 2018

Multi-Expert Estimations of Burglars' Risk Exposure and Level of Pre-crime Preparation Based on Crime Scene Data

Martin Boldt [a], **Veselka Boeva** [a], **Anton Borg** [a]

a. Department of Computer Science and Engineering Blekinge Institute of Technology, Sweden.

*CORRESPONDING AUTHOR. E-mail: {martin.boldt, veselka.boeva, anton.borg}@bth.se

ABSTRACT Law enforcement agencies strive to link crimes perpetrated by the same offenders into crime series in order to improve investigation efficiency. Such crime linkage can be done using both physical traces (e.g., DNA or fingerprints) or "soft evidence" in the form of offenders' modus operandi (MO), i.e. their behaviors during crimes. However, physical traces are only present for a fraction of crimes, unlike behavioral evidence. This position paper presents a method for aggregating multiple criminal profilers' ratings of offenders' behavioral characteristics based on feature-rich crime scene descriptions. The method calculates consensus ratings from individual experts' ratings, which then are used as a basis for classification algorithms. The classification algorithms can automatically generalize offenders' behavioral characteristics from cues in the crime scene data. Models trained on the consensus rating are evaluated against models trained on individual profiler's ratings. Thus, whether the consensus model shows improved performance over individual models.

KEY WORDS Multi-expert decision making, Classification, Crime linkage, Offender profiling, Forensic science

I. INTRODUCTION

For crime categories that involve serial offenders, i.e. where the same offender commits two or more crimes in the same crime category (e.g. burglaries or robberies), law enforcement agencies strive to link those crimes into crime series [9, 14]. The linking of crimes enables investigators to get a more complete understanding, based on the combined knowledge and evidence collected from the different crime scenes, compared to investigating each crime in isolation [16]. In addition, such linking also enable more efficient use of the police force's scarce resources compared to investigating each crime individually [14].

The linking of crimes into series can be done based on physical evidence, e.g., DNA or fingerprints. However, such evidence is only available in a fraction of all crime scenes and the processing of physical evidence is also costly and time- consuming [8]. Thus, it is many-times difficult for law enforcement agencies to handle large amounts of physical evidence from high-volume crime categories, e.g., burglaries [15]. However, "soft evidence" in the form of offender's modus operandi (MO), i.e. habits, techniques and peculiarities of behavior when committing an offense [12], could be used for linking crimes committed in a similar fashion [1, 10]. This is done by 1) registering behavioral patterns at the crime scenes, 2) interpreting the patterns using criminal and geographical profiling, 3) comparing the behaviors of (unknown) offenders at different crimes, and 4) consider linking similar crimes into series [2, 19]. The process of linking crimes based on soft evidence is known as crime linkage and it rests on two key assumptions. First, the offender consistency hypothesis [6] stating that offenders display similar behaviors across time and place. Second, the offender specificity hypothesis [13] describing that offenders have an approach that deviate or is distinct from other offenders' approaches. These assumptions apply for offender behavior in residential burglaries [25]. The accumulated evidence published during the past 20 some years provide a base for conducting crime linkage based on behavioral consistency and distinctiveness for some offenders [20], some of the time, in various crime categories and types, e.g. commercial and residential burglary [2, 5, 11, 24].

A drawback with crime linkage is that crime analysts and offender profilers are required to manage a substantial amount of information that brings along a heavy cognitive load [17]. Law enforcement agencies have created computerized databases enclosing a large number of reported crimes that analysts can search for similarities [26]. Various data science methods and algorithms can be used in combination with such databases, e.g. in order to calculate similarity scores based on pair-wise comparisons

of crimes. Further, learning algorithms from the area of computer science and the field of machine learning can be used for grouping crimes with similar characteristics together, e.g., by using cluster algorithms [4]. Such intelligent models can be packaged in decision-support systems that assist crime analysts in the crime linkage process.

In a previous initial study, we investigated the possibility of using intelligent models to automatically estimate both offenders' risk exposure and their level of pre-crime preparation for residential burglaries [3]. We did this using soft evidence in the form of MO characteristics that were recorded from crime scenes. The main motivation for an automated approach is because volume crimes (e.g. various forms of thefts, including burglaries) occur with such a high frequency that criminal profilers cannot analyze and rate each individual crime in- stance manually. Therefore, models that automatically estimate behavioral characteristics for unknown offenders using crime scene data, would be valuable for law enforcement agencies. Mainly since they would allow law enforcement agencies to use those behavioral characteristics in the crime linkage process, i.e. when linking sets of crimes to common offenders. Another benefit of such models would be that the behavioral characteristics that they calculate could be compared for crimes between different crime categories (e.g. burglaries and Diesel thefts) that otherwise more or less lack comparable MO behaviors. The previous initial study included an experimental evaluation of 16 learning algorithms and the models trained by the Naïve Bayes Multinomial algorithm both showed interesting results and was the most suitable candidates for the problem at hand. However, further research is required before such models could be used in an operational setting by law enforcement agencies.

In this work a planned extended study is described, which investigates whether multi-expert decision making concepts can improve the performance of intelligent models when estimating of- fender characteristics. The remainder of this position paper is structured as follows, in Section 2 we summarize the previous study. In Section 3 we elaborate on how we intend to extend the previous study. Finally, in Section 4 we present some conclusions and avenues for future work.

II. ESTIMATING OFFENDER CHARAC- TERISTICS

In the previous study, two criminal profilers from the national offender profiling group within the Swedish police manually rated offenders' risk exposure and level of pre-crime preparation for 50 burglaries each. However, there was an overlap of 25 burglaries to allow the analysis of inter-rater agreement, which turned out to be moderate. For each burglary a feature-rich and structured representation of the crime scene was available, containing some 137 features, e.g. type of residence and which entrance method the offender used. Based on the manual ratings of offender characteristics pro- vided by the profilers, a labeled dataset of 75 in- stances was created. Each instance has an offender risk exposure score as well as a pre-crime preparation score together with the feature-rich crime scene data. Both risk and plan scores were rated using the following scale: low, decreased, increased, high.

An experiment was executed in which 16 machine-learning algorithms were evaluated using a supervised-learning approach on the labeled dataset.Performance evaluation was done using stratified 10-times 10-fold cross-validation tests. Five performance metrics were used, with AUC as the primary metric. Statistical analysis of the AUC performance was done using the non-parametric Kruskal-Wallis test in combination with the Nemenyi post-hoc test. Models trained by the Naïve Bayes Multinomial algorithm outperformed more competing models than any of the other algorithms, and was therefore selected as the most suitable candidate for the problem at hand. The AUC measures were 0.79 (sd=0.15) and 0.77 (sd=0.16) for estimating offenders' risk and preparation scores respectively. The classification performance of the models were not excellent, but given that this was an initial study (with a quite limited dataset) the results indicate that models can pick up cues in the feature-rich crime scene data that are useful when generalizing offenders' risk and preparation scores.

Next, the models trained by Naïve Bayes Multinomial were then used for calculating both offenders' risk exposure scores as well as pre-crime preparation scores for 15,598 residential burglaries that all contained feature-rich and structured crime descriptions. For a subset of 153 burglaries (out of the 15,598) the police provided anonymized identifiers of the offenders, which allowed us to construct 41 linked crime series of linked burglaries. For each of the 41-crime series we calculated the variation of both the risk exposure and pre-crime preparation scores. We then compared those variations with randomly constructed "crime series" that contained equally many burglaries. Differences in score consistency between linked series and random series were studied using a Wilcoxon signed rank test, which showed that scores were significantly more consistent in linked series com- pared to random ones. Further analysis revealed that the scores also showed promising distinctiveness between linked series, as well as consistency for crimes within series compared to randomly sampled crimes. This indicates the usefulness of automatic models for estimating offenders' risk exposure and degree of pre-crime preparation. Further, that such behavioral scores could be used as a complement to traditional crime scene data in the crime linkage process when law enforcement officers try to link crimes together that most probably are committed by the same offender.

III. MULTI-EXPERT DECISION MAKING

As crime profilers have different education, experience and domain knowledge it is interesting to consider the manual rating of

offender's crime characteristics using a multi-expert decision making approach. Such an approach enables both inter-rater agreements, as well as weighting of each individual rater's relative importance in order to reach a consensus decision regarding the scores. When linking crimes it is important that the decisions are as correct as possible, since the decisions made in the crime linkage process regarding which crimes to link into, or exclude from, series can have impact on the daily lives of several persons. It is therefore important that manual individual ratings of offender characteristics by profilers are aggregated in a systematic way into sound team evaluation ratings.

Although the work discussed in this paper investigates offender's risk exposure and planning for residential burglaries, the method is applicable on other crime types as well. Profilers, when manually rating offender's characteristics, use the same working method in more violent crimes as well, e.g., in rape, assault, and murders. As such, the possibility to have an initial automated estimate to compare the law enforcement officers' decision against is useful. By training models on both aggregated decisions (based on all profilers' ratings), as well as individual raters decisions allow for detection of outlier ratings. The detection of outlier ratings is interesting as an outlier may be more important than the aggregated decision. This might be in case the individual profiler possesses unusual and valuable domain knowledge otherwise lacking in the group, i.e. when an outlier is the correct decision, and which is not represented in the profiler weights. Such outlier ratings need to be manually investigated in more detail.

3.1. Proposed Approach

In this work, we propose a method that mimics a multi-expert estimation process of a team of criminal profilers who are involved in the rating of a given set of crimes. The defined process can be divided in the following two phases: (i) an individual estimation phase engaging individual profilers, and (ii) a consensus and decision phase involving the interaction between the team members. During the estimation phase each profiler is asked to rate/score, independently from the other profilers, both offender's risk exposure and level of pre-crime preparation for a set of crimes. In addition, each profiler gives his/her opinion about the expertise of the other profilers, by assigning weights to each one, including himself/herself. During the consensus and decision phases, the crime scores and pro- filer weights provided by each individual profiler are aggregated into team-based consensus scores. Then, learning algorithms are applied to generalize from these team-based consensus scores, using the feature-rich structured crime scene descriptions.

The proposed multi-expert decision making method is based on an aggregation approach developed by Tsiporkova and Boeva [22]. It resolves potential conflicts between the team members by mimicking a multi-step decision making process during which the decision makers have an opportunity to discuss and exchange views, ideas, in- formation, etc. At each decision step each expert (profiler) aggregates the outcome of the previous step according to the set of weights he/she has as- signed to himself/herself and the rest of the pro- filers. These weights express the relative degree of influence each profiler is inclined to accept from the rest of the team when forming a judgment for the different crimes.

The proposed multi-expert approach has a few advantages. First, that profilers are not allowed to completely ignore colleagues by using zero weights and in this way putting the decision process in a deadlock. This requirement is imposed in order to ensure that the applied recursive aggregation process is convergent [21]. Second, that the approach reflects different team interaction styles by incorporating mechanisms to deal with the reputation of the profilers [22, 23]. The goal is to minimize the total dissatisfaction of each crime rating by considering the profilers' particular views and implementing reputation-enhanced collaboration, but avoiding any kind of explicit public profilers' rating. The latter is crucial since it contributes to a positive collaboration atmosphere. Finally, the approach translates the chosen six level interval scale used by individual profilers, which spans low to high represented as [1-6], into a continuous scale, i.e. all values between 1.0 and 6.0 (inclusive). As it was discussed in the work by Sullivan et al. [18] responses can be rated or ranked in an ordinal scale, but the distance between responses is not measurable. In other words, one cannot assume that the difference between responses is equidistant even though the numbers assigned to those responses are. In contrast to this when we have interval data, the difference between responses can be calculated and the numbers can be used for further analysis and dis- crimination of the rating crimes.

We also plan to use the proposed multi-expert decision making model to simulate different team interaction styles and study how these affect the estimation performance of the applied learning models. The three team interaction styles described by Cooke and Szumal [7] will be investigated, that is the constructive, passive, and aggressive interaction styles. The constructive style is characterized by a balanced concern for personal and team outcomes. The passive style places greater emphasis on fulfillment of affiliation goals only, i.e. maintaining harmony in the team. While the aggressive style is characterized with that personal ambition is placed above concern for team outcome. Cooke and Szumal demonstrate that groups that predominantly use a constructive interaction style produce solutions that are: 1) superior in quality to those produced by passive groups, and 2) superior in acceptance (satisfaction) to those produced by either passive or aggressive groups.

3.2. Study Design

The planned extended study consists of the following three phases. In the first phase the 4-6 profilers included in the study will participate in a work- shop that presents the study and the tasks that should be carried out. During the workshop the profilers should individually rate the risk and plan scores for 25 burglaries. Next, they should come to joint decisions (both for risk and plan scores) for each of the 25 burglaries. This workshop will apart from

describing the problem domain, allow participants to discuss the crime scene data avail- able and how to interpret and judge it by relating to concrete examples represented by the 25 burglaries. These discussions will also give the profilers a more detailed opinion about the other profilers' competence, which will be valuable in the next phase.

The second phase of the study includes two tasks: (i) that each rater individually rates the relative degree of influence he/she puts on each of the other profilers and him/herself, and (ii) that each profiler independently rates the risk and plan scores for another 125 burglaries. This work could be done little-by-little over the course of a couple of weeks when the profilers have some time to spare. Once all profilers have finished their tasks, the data is collected. Then it is possible to train individual models for estimation of both risk and plan scores based on the ratings from each individual profiler, i.e. 4-6 models (one per profiler) is created. The learning algorithm used for training the models will be Naïve Bayes Multinomial as it showed best performance in the previous study. Next, the consensus scores are calculated using the ratings from all profilers and the matrix of profiler weights according to the iterative method previously described in Section 3.1. An experiment is used to com- pare the estimation performance of the individual models against the consensus model over the full labeled dataset consisting of 150 burglaries. Evaluation is handled using stratified 10-times 10-fold cross-validation and suitable statistical tests. This is done twice, one time for the models that estimate offender risk exposure scores, and another time for the models that estimate offender pre-crime preparation scores.

The third phase of the study involves a simulation of the three team interaction styles according to the work by Cooke and Szumal, i.e. the constructive, passive, and aggressive styles. The matrix of profiler weights will be changed according to each of the three interaction styles. For each new weight matrix alternative consensus decisions for both the risk and plan scores will be calculated, and new models will be trained on these new scores. Finally, the estimation performance of each model will be compared to each other using a similar experimental setup as in phase two above.

IV. CONCLUSION AND FUTURE WORK

This position paper presents a method for aggregating multiple experts' ratings of offender behavioral characteristics using a feature-rich crime scene dataset. The method calculates consensus ratings from individual experts' ratings, which then are used as a basis for machine learning algorithms. The learning algorithms can be used to automatically eneralize offenders' behavioral characteristics from cues in the crime scene data. Models trained on the consensus rating are evaluated against models trained on individual profilers' ratings. Thus, investigating if there is any improvement in the performance of the consensus model compared to the individual models.

As future work it would be interesting to investigate if the use of consensus ratings as seeds for clustering algorithm could generate more informative clusters. This could further be extended as an interactive clustering approach.

REFERENCE

1. Bennell, C. and Canter, D.V., Linking com- mercial burglaries by modus operandi: tests using regression and ROC analysis, Science & Justice, 42, 2002, pp. 153-164.
2. Bennell, C. and Jones, N.J., Between a ROC and a hard place: A method for linking serial burglaries by modus operandi, Journal of Investigative Psychology and Offender Profiling, 2, 2005, pp. 23–41.
3. Boldt, M., Borg, A., Svensson, M., Hildeby, J., Predicting burglars' risk exposure and level of pre-crime preparation using crime scene data, (to appear in) Journal of Intelligent Data Analysis, 2018, 22(1).
4. Borg, A. and Boldt, M., Clustering Residential Burglaries Using Modus Operandi and Spatiotemporal Information, International Journal of Information Technology & Decision Making, 2016, 15(1), pp. 23-41.
5. Bouhana, N., Johnson, S.D. and Porter, M., Consistency and specificity in burglars who commit prolific residential burglary, Legal and Criminological Psychology, 2014, 21(1), pp.77-94.
6. Canter, D., Psychology of offender profiling, Handbook of psychology in legal contexts, Chichester, UK, John Wiley and Sons, 2000.
7. Cooke, R.A. and Szumal, J.L. The impact of group interaction styles on problem-solving effectiveness, Journal of Applied Behavioral Science, 1994, 30, pp. 415-437.
8. Davies, A., The use of DNA profiling and behavioural science in the investigation of sexual offences, Medicine, Science and the Law, 1991, 31, pp. 95-101.
9. Grubin, D., Kelly, P. and Bruns- don, C., Linking serious sexual assults through behaviour [Internet], 2001 [cited 2017-03-30], Available from: http://webarchive.nationalarchives.gov.uk/20110314171826
10. Hazelwood, R.R. and Warren, J.I., Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime, Aggression and Violent Behavior, 8, 2003, pp. 587-598.
11. Markson, L., Woodhams, J., and Bond, J.W., Linking serial residential burglary: comparing the utility of modus operandi behaviours, geographical proximity, and temporal proximity, Journal of Investigative Psychology and Of- fender Profiling, 2010, 7(2), pp. 91-107.
12. O'Hara, C.E. and O'Hara, G.L., Fundamentals of Criminal Investigation (7th ed.), Spring- field, IL, Charles C Thomas Publisher Ltd., 1956.
13. Pervin, L.A., Current Controversies and Is- sues in Personality, New York, John Wiley and Sons, 2002.
14. Reich, B.J. and Porter, M.D., Partially super- vised spatiotemporal clustering for burglary crime series identification, Journal of Royal Statistical Society, Series A (Statistical Society), 178, 2015, pp. 465-480.
15. Roman J., Reid, S., Reid, J., Chalfin, A., Adams, W. and Knight, C., The DNA Field Experiment: Cost-effectiveness Analysis of the Use of DNA in the Investigation of High-volume Crimes [Internet], 2008 [cited 2017-03-30], Available from: http://www.urban.org/UploadedPDF/411697 dna field experiment.pdf
16. Rossmo, K., Geographic Profiling, Boca Ra- ton, CRC Press, 1999.
17. Santtila, P., Korpela, S. and Häkkänen, H., Expertise and decision making in the linking of car crime series, Psychology, Crime & Law, 2004, 10(2), pp. 97-112.
18. Sullivan, G.M. and Artino, A.R., Analyzing and Interpreting Data from Likert-Type Scales, Journal of Graduate Medical Education, 2013, pp. 541-542.
19. Tonkin, M., Woodhams, J., Bull, R. and Bond, J.W., Behavioural case linkage with solved and unsolved crimes, Forensic Science International, 212, 2012, pp. 146-153.
20. Tonkin, M. and Woodhams, J., The feasibility of using crime scene behaviour to detect versatile serial offenders: An empirical test of behavioral consistency, distinctiveness, and discrimination accuracy,

Legal and Criminological Psychology, 2015, 22(1), pp. 99-115.
21. Tsiporkova, E. and V. Boeva, V. Nonparametric recursive aggregation process, Kybernetika, The Journal of the Czech Society for Cybernetics and Inf. Sciences 2004, 40(1), pp. 51 - 70.
22. Tsiporkova, E. and Boeva, V., Multi-step ranking of alternatives in a multi-criteria and multi-expert decision making environment, Information Sciences, 2006, 176(18), pp. 2673–2697.
23. Tsiporkova, E., Tourw`e, T. and Boeva, V. A Collaborative Decision Support Platform for Product Release Definition. The Fifth International Conference on Internet and Web Ap- plications and Services ICIW 2010, 2010, pp. 351-356.
24. Woodhams, J., Hollin, C.R. and Bull, R., The psychology of linking crimes: A review of the evidence, Legal and Criminological Psychology, 2010, 12(2), pp. 233-249.
25. Wright, R.T. and Decker, S.H., Burglars on the job, Boston, MA, Northeastern University Press, 1994.
26. Yokota, K. and Watanabe, S., Computer- Based Retrieval of Suspects Using Similarity of Modus Operandi, International Journal of Police Science & Management, 2002, 4(1), pp. 5-15.

POSTSCRIPT

Position paper.

First reported at the 30th Annual Workshop of the Swedish Artificial Intelligence Society SAIS 2017, May 15–16, 2017, Karlskrona, Sweden