Supervised Machine Learning Approach Discovers Protective Sequence for Avoiding Sexual Victimization in Criminal Suit Documents

Abstract

Effective self-protective behaviors, such as victim's physical resistance for avoiding sexual victimization have been studied. However, effective self-protective behavioral sequences, such as offender's physical violence followed by victim's physical resistance, have not been studied often. Our study aims to clarify these sequences through supervised machine learning approach. The samples consisted of 88 official documents on sexual crimes regarding women committed by male offenders incarcerated in a Japanese local prison. The crimes were classified as completed or attempted cases based on judges' evaluation. All phrases in each crime description were also partitioned and coded according to the Japanese Penal Code. The Support Vector Machine learned the most likely sequences of behaviors to predict completed and attempted cases. Around 90% of cases were correctly predicted through the identification of sequences of behaviors. The sequence involving the offender's violence followed by victim's physical resistance predicted attempted sexual crime. However, the sequence involving victim's general resistance followed by the offender's violence predicted completed sexual crime. Timing of victim's resistance and offender's violence could affect potential avoidance of sexual victimization.

Keywords: Criminal Suit Documents; Supervised Machine Learning; Protective Action; Rape; Sexual Coercion.

Introduction

Sexual crime violates victim's human rights and needs to be prevented before it occurs. To prevent the crime, several protective actions were proposed for potential victims (Ullman, 2007). Among the protective actions, the most convincing strategy is physical resistance, namely physical action against offenders such as fighting, fleeing, guarding one's body with one's arm, and struggling(Clay-Warner, 2002; Sarnquist et al., 2014; Senn et al., 2015; Tark & Kleck, 2014). The second effective strategy is forceful verbal resistance, which refers to a verbal response leaving no room for the offender to talk, such as screaming, yelling, and swearing at the offender (Clay-Warner, 2002; Tark & Kleck, 2014; Ullman, 2007; Zoucha-Jensen & Coyne, 1993). The third strategy is non-forceful verbal resistance, which is a verbal response leaving some room for the offender to talk, such as reasoning, arguing, persuading, or appeasing the offender (Fisher, Daigle, Cullen, & Santana, 2007). University women who received a training regarding the first and second strategies reduced the risk of sexual victimization than those who did not (Senn et al., 2015). The third non-forceful verbal resistance was especially effective for child victims (Leclerc, Wortley, & Smallbone, 2011b) and sexual crime without offender's physical violence (Fisher et al., 2007).

Although these protective actions were well reported (Senn et al., 2015), behaviors before and after the protective actions were still unclear. On the one hand, victim's protective actions paired to offender's behavior were reportedly effective to decrease the risk of sexual victimization (Fisher et al., 2007; Ullman, 1998): Victims' physical

resistance after the offender's physical violence was effective to reduce the risk of sexual victimization. Similarly, victims' forceful verbal resistance after the offender's verbal coercion was effective to reduce the risk. On the other hand, other studies suggested that offender's physical violence after the victim's resistance increase the risk of sexual victimization, because offender's violence stops victim's resistance (Balemba, Beauregard, & Mieczkowski, 2012; Jordan, 2005). Hence antecedent offender's violence and consequent victim's physical resistance might reduce the risk of sexual victimization, whereas antecedent victim's resistance and consequent offender's violence might increase the risk of sexual victimization. Still, direct comparison of these behavioral sequences was rare so that behavioral sequences of protective action were still unclear.

Our study aims to clarify the behavioral sequences of protective actions. Our research question is what behavioral sequence predicts completed and attempted (but not completed) sexual crimes. To clarify the sequence, we focused behavioral interactions between a victim and an offender during a sexual crime. Specific interaction which predicts attempted sexual crime is regarded as a protective behavioral sequence for avoiding victimization. Another interaction which predicts completed sexual crime is regarded as predictive behavioral sequence for victimization. Both protective and predictive sequences clarify the knowledge regarding sequences of protective action and are beneficial for protective action training (Senn et al., 2013).

The present study sampled women-victim cases and excluded child-victim cases, because victims' protective action, offenders' behavior, and effects of protective actions

were different between women and child victims. Child victims more received gifts from offenders(Leclerc & Wortley, 2015; Leclerc, Wortley, & Smallbone, 2011a), more used non-forceful verbal resistance(Leclerc, Wortley, & Smallbone, 2010), and less protected efficiently(David Finkelhor, Asdigian, & Dziuba-Leatherman, 1995b, 1995a) than women victims. We regarded those less than 13 years old as children according to Japanese law (Maeda, 2015) and excluded cases including these child victims, although definitions of children were different among countries and eras (David Finkelhor et al., 1995a; Leclerc & Wortley, 2015). In sum, to eliminate ambiguity of sample, we excluded child-victim cases and analyzed cases where victims were more than 13 years old.

Further, to label the sexual crime as completed and attempted case, we utilized official suit documents on sexual crime in Japan. Attempted crime has a less severe penalty than completed crime in Japan (Yamashita & Yamaguchi, 2016), so the term for these attempts is clearly described in the documents. Furthermore, the documents also describe behavioral chains between an offender and a victim during the crime. The described interaction was useful to clarify behavioral sequences at the crime.

Based on the label of crime (completed or attempted) and behavioral sequences in the documents, we tested four hypotheses: To confirm previous findings of protective action (Leclerc et al., 2011b; Senn et al., 2015), victim's physical resistance, forceful verbal resistance, and non-forceful verbal resistance would predict attempted sexual crime (Hypothesis 1). According to the parity effects of protective action (Fisher et al., 2007; Ullman, 1998), the offender's antecedent physical violence and victim's

consequent physical resistance would predict attempted sexual crime (Hypothesis 2). Similarly, the offender's antecedent verbal coercion and victim's consequent forceful verbal resistance would predict attempted sexual crime (Hypothesis 3). According to the effect of offender's physical violence on victim's resistance (Balemba et al., 2012; Jordan, 2005), antecedent victim's resistance and consequent offender's physical violence would predict sexual victimization (Hypothesis 4).

Our study utilized supervised machine learning models as a statistical model. This is because the number of behavioral sequences increases exponentially the number of variables and destroys the premise of psychological statistical analysis: The 0, 1, and 2 behavioral sequences in our study require 18, 324, 5832 variables. The 324 and 5832 independent variables did not fit well with regression analysis for the prediction of a binary dependent data (completed or attempted). In contrast, Support Vector Machine in the supervised machine learning is robust against the increased number of variables (Bishop, 2006), so we used the Support Vector Machine like other studies(Costa, Fonseca, Santana, de Araújo, & Rego, 2017).

Methods

Sample

We identified the 128 sexual offence cases consisted of 72 male inmates who were imprisoned in April 20XX in a local Japanese prison as repeat offenders. Among them, 12 cases were inaccessible, because of offenders' transportation; furthermore, the 28

cases involved child victims (aged under 12 years). Thus, these cases were excluded from the analysis. Finally, we analyzed 88 sexual offence cases. Of these, the 35 involved teen victims (aged between 13 and 19 years) and 52 involved adult victims (aged over 20 years). One case included a charge of public lewdness; therefore, the victim's age was unknown.

Measures

Categories of sexual crime. Table 1 shows four categories of sexual crime in our study: completed rape, attempted rape, completed sexual coercion, and attempted sexual coercion. Although the definition of rape and sexual coercion differs slightly in previous studies (Clay-Warner, 2002; Fisher et al., 2007; Ullman & Knight, 1992), we utilized the Japanese Penal Code to fit with the finalized criminal suit documents in Japan. Completed rape is an offender's realization of penile-vaginal penetration achieved by either or both of illegal physical force and verbal coercion (Maeda, 2015; Yamashita & Yamaguchi, 2016). Attempted rape did not involve realization of penile-vaginal penetration, but include offender's intent of penile-vaginal penetration. For instance, in a case that offender exposed his private parts to a victim and penetrated her vagina with his finger in her private room, the Japanese judges regarded the offender has intent of penile-vaginal penetration and wrote "rape" in the section on charged offence and "with intention to rape" in the criminal behavior description section.

Completed sexual coercion involves any sexual behaviors other than penile-vaginal penetration achieved by either or both of illegal physical force and verbal coercion (Maeda, 2015; Yamashita & Yamaguchi, 2016). The completed sexual coercion did not

involve offender's intent of penile-vaginal penetration (Table 1). In a case where offender touched victim's breast in a public train with many passengers, the Japanese judges did not regard the offender has intent of penile-vaginal penetration so the judges never write the term of "rape" in the documents. Attempted sexual coercion did not involve realization of any sexual behavior, but included offender's intent of the sexual behavior. For instance, in a case that offender prepared spy camera in his bathroom and forced his victim to take shower, but she noticed the camera before taking shower, the judges regarded the offender has intent of sexual behavior but did not realize his behavior. Hence, they wrote "attempted" in the section on the charged offence and "failed to accomplish one's purpose" in the criminal behavior description section. Based on these descriptions, we categorized cases as completed rape (n = 24), attempted rape (n = 13), completed sexual coercion (n = 49), and attempted sexual coercion (n = 2).

Code of Behaviors. All phrases in the criminal description were partitioned. In total, 560 phrases were coded according to the following definitions.

Victim's Resistance. Physical resistance is physical action against an attacker (Clay-Warner, 2002). Forceful verbal resistance refers to a verbal response leaving no room for the offender to talk (Ullman, 2007). Non-forceful verbal resistance refers to a verbal response leaving some room for the offender to talk (Fisher et al., 2007). Several phrases included "resist" (n = 5) or "fierce resistance" (n = 1) only; these phrases cannot be regarded as specific type of resistance, so they were coded as general resistance. Table 2 shows details of victims' resistant behaviors.

Offender's Behavior. Sexual behavior is a behavior that "unnecessarily stimulates

and excites sexual desires," "harms the grace of a citizen," and "is against sexual morality" (Maeda, 2015), as defined in the sections on Rape, Forcible Indecency, and Public Indecency in the Japanese penal code (Yamashita & Yamaguchi, 2016). Physical violence is defined as the illegal use of physical force, regardless of physical contact (Maeda, 2015) in the Assault section of the Japanese penal code (Yamashita & Yamaguchi, 2016). Verbal Coercion is defined as "intimidating another through a threat to another's life, body, freedom, reputation, or property" in the Intimidation section (Yamashita & Yamaguchi, 2016), and "causes the other to perform an act which the other person has no obligation to perform, or hinders the other from exercising his or her rights" in the Compulsion section (Yamashita & Yamaguchi, 2016). Persuasion (non-forceful verbal behaviors) is verbal communication without threat and compulsion. Table 2 shows details of offenders' behavior at the crime.

The transfer of possessions is defined as transferring others' property against their will (Maeda, 2015) in the Theft and Robbery sections (Yamashita & Yamaguchi, 2016). Although there are various types of property (Maeda, 2015), we focused on the transfer of money only to clarify mercenary motives. Here, offenders obtained the victim's cash (n = 5), cash card (n = 1), and credit card (n = 1).

Crime Location. The location of the encounter was categorized according to indoor/outdoor and private/semi-public/public criteria (Beauregard, Proulx, Rossmo, Leclerc, & Allaire, 2007). Private refers to a privately owned site not open to the public. Semi-public refers to a privately owned site open to the public, especially for business purposes. Public is a publicly owned site. An indoor private location includes the

177 victim's house (n = 31), hotel room (n = 9), victim and offender's houses (n = 9),
178 offender's house (n = 3), and someone else's house (n = 3). Indoor semi-public locations
179 include the elevator (n = 2), plastic greenhouse (n = 2), restaurant (n = 2), trash area (n
180 = 2), bar (n = 1), cafe (n = 1), and toilets in an apartment (n = 1). Indoor public locations
181 include toilets in the park (n = 2), car on the road (n = 3), and train (n = 2). Outdoor
182 private locations include the building area of someone's house (n = 4) and a school (n =
183 1). Outdoor semi-public locations include parking lots (n = 5), a station (n = 2), a field
184 (n = 2), a corridor in an apartment (n = 2) and a building (n = 2). Entrance in an
185 apartment (n=1), escalator in a building (n=1), and stairs in a building (1) are also
186 included. Outdoor public locations include roads (n=12) only.

187 The approach to the crime location was coded as "Invade" and "Go with." "Invade"
188 means that the offender approached the victim's private place alone (Leclerc, Chiu, Cale,
189 & Cook, 2016), invading the space through an open door (n=8), through an open
190 window (n = 8), through a window (n = 4), through the door (n = 3), or through the vent
191 (n=1). In addition to these numbers, six offenders invaded the victim's home, but their
192 invasion methods are unknown. "Go with" means that the offender moved to the crime
193 location with the victim (Leclerc et al., 2016), bringing the victim (n =14) or moving the
194 victim by his car (n = 1) and taxi (n = 1). In addition to these numbers, two offenders
195 moved with the victim, but their transportation is unknown (n = 2).

196 *Bystander*. A bystander is an individual present, who is not the victim or offender: "a
197 third person detected the crime (n = 2)," and "a third person (n = 1) and the victim's
198 sibling (n = 1) came to the situation."

Coding Process

The following case is a dummy attempted rape case: "The offender invaded the victim's house through an open window, saying, "I will kill you if you make a noise." The offender then touched the victim's private parts, and tried to conduct sexual intercourse with her; however, she fled, meaning that he failed to accomplish his purpose." When we code this case, the code can be "offender's invade→ a victim encounters the offender at private indoor setting → offender's verbal coercion → offender's sexual behavior → offender's sexual behavior → victim's physical resistance → offender's failure to achieve goal."

Sequence 1 (continuous two behaviors) includes "Invade→ Private Indoor," "Private Indoor → Verbal Coercion," …, and "Physical Resistance → Failure to achieve goal." Here, the sequence with "Failure to achieve goal" is excluded from the analysis, because this is the classification criterion of attempted case. The selected sequences were linked with the attempted class, and these sequences were weighted to predict the attempted class. Similarly, all cases were used and the Support Vector Machine learned the weights of sequences. The final weights of these sequences show the most predictive sequences.

Plan of Analysis

To show the probability of behavioral sequence, conditional probability was applied. Furthermore, to predict attempted and completed cases through a behavioral sequence, the Linear Support Vector Classifier was used in scikit-learn 0.18.1. The results of prediction have four categories: A true positive (TP) indicates that both judge and

classifier supported the completed sexual crime, while a false positive (FP) indicates that the classifier supported the completed sexual crime but the judge did not support. Furthermore, a false negative (FN) indicates that the judge supported the completed sexual crime but the classifier did not support, while a true negative (TN) indicates that neither the judge nor the classifier supported the completed sexual crime. To evaluate the results of prediction, we utilized index of accuracy: accuracy is (TP+FN) / (TP+TN+FP+FN). For the validation of the accuracy, the 10 cross-validation is utilized: Total sample (N = 88) is randomly partitioned into 10 equal-sized subsamples (n = 8 or 9). A single subsample is retained as test data, whereas the other subsamples are used as training data (9 subgroups, n = 79 or 80). With training data, the predictive model (weights of sequence) is estimated. The model analyzes retaining test data as a test and provides accuracy. Next, another single subsample is selected as test data, the other subsamples are training data, and the model provides accuracy. Similarly, we can test 10 models and provide 10 accuracies. The average of 10 accuracies indicates robust accuracy of the total sample.

Results

Comparison of rape and sexual coercion cases

Table 3 shows several significant differences between the rape and sexual coercion cases. Victims in sexual coercion cases were attacked by unknown strangers more frequently than those in rape cases. The rate of completed sexual coercion cases is also

higher than the rate in completed rape cases. In contrast, victims used physical resistance and general resistance in rape cases more frequently than those in sexual coercion cases did. Furthermore, the rape cases occurred in indoor private settings more frequently than sexual coercion cases. Except for these indexes, rape and sexual coercion were not differed in other indexes such as victims' and offenders' age.

Interconnections of victim's protective action and offender's failure of sexual crime

Table 4 shows the conditional and unconditional probabilities of offenders' behavior and victim's protective action. The probabilities in rape and sexual coercion cases were quite similar; therefore, Table 4 shows the combined probabilities only. Table 4 shows that the chance of consequent rape (sexual coercion) avoidance is predicted by the victim's antecedent physical resistance (38%), forceful verbal resistance (33 %), non-forceful verbal resistance (11 %), general resistance (83 %), and bystander's intervention (75 %). The unconditional chance of consequent rape (sexual coercion) avoidance is 3%, meaning that these victims' antecedent resistant behaviors and bystander's intervention increased the chance of successfully thwarting rape (or sexual coercion) completion.

Furthermore, victim's resistance behavior and bystander's intervention were connected with each other. Figure 1 shows the interconnections between victim's protective action and offender's failure of sexual crime. Victim's physical resistance increased the chance of victim's forceful-verbal resistance. The victim's forceful-verbal resistance increased the probabilities of victim's non-forceful-verbal resistance and bystander's intervention. Further, the bystander's intervention increased the

probabilities of victim's physical resistance. All of victim's resistance and bystander's intervention increased the probabilities of offender's failure of sexual crime. Figure 1 indicated the protective actions were connected with each other and had both direct and indirect effects on increasing the probabilities of offender's failure of sexual crime.

Prediction Accuracy of attempted and completed sexual crime with Behavioral Sequence

We used 0 (single behavior), 1 (two continuous behaviors), 2 sequences (three continuous behaviors) as sequence units and built models to predict completed and attempted cases. Table 5 shows the prediction accuracies of the models. All accuracies were over 80%. Especially, models in rape cases show over 88%. Taking into account random chance (64.9 %, Table 3), the sequence of continuous behavior predicted rape avoidance well.

Protective Sequence for Avoiding Sexual Victimization (Hypothesis 1, 2, and 3)

Table 6 shows the protective sequence for avoiding sexual victimization. As hypothesized (1), attempted sexual crime was predicted by victim's general resistance (0 sequence 1st place w = -2.00), physical resistance (0 sequence 3rd place w = -1.54), forceful verbal resistance (0 sequence 2nd place w = -1.76), and non-forceful verbal resistance (0 sequence 7th place w = -0.17). Moreover, as expected (2), the sequence of offender's antecedent violence and victim's consequent physical resistance was also protective for avoiding sexual victimization (1 sequence 6th place: w = -1.00, 2 sequence 4th place: w = -0.82). Similarly, the sequence of offender's antecedent verbal coercion and victim's consequent forceful verbal resistance was also protective for avoiding

sexual victimization (1sequence 3rd place: w = -1.20, 2sequence 3rd place: w = -1.18) [hypothesis 3]. Further, victim's general resistance after the offender's sexual behavior is also protective for avoiding sexual victimization (1sequence 1st place: w = -2.09, 2sequence 1st place: w = -2.11)

Predictive Sequence for Sexual Victimization (Hypothesis 4)

Table 7 shows the predictive sequence for sexual victimization. As hypothesized (4), the sequence of victim's antecedent general resistance and offender's consequent violence was predictive for sexual victimization (1 sequence 2nd place: w = 0.76, 2 sequence 8th place: w = 0.26). Further, offender's antecedent violence and offender's consequent sexual behavior was predictive for sexual victimization (1 sequence 1st place: w = 0.88, 2 sequence 1st place w = 0.40). Table 4 also shows indoor public setting is predictive for sexual victimization (0 sequence 1st place w = 1.09). These findings suggest that a victim's physical resistance in response to an offender's antecedent physical contact was protective in avoiding sexual victimization. However, an offender's physical contact in response to a victim's antecedent resistance was predictive for sexual victimization.

Discussion

Protective Action for Avoiding Sexual Victimization (Hypothesis 1)

Our study confirmed the effects of protective action for avoiding sexual victimization. In line with environmental criminology theory (Braga, 2005; Clarke,

1997; Cornish & Clarke, 2014; Felson & Clarke, 1998; Guerette & Santana, 2010), we confirmed that physical resistance was the effective protective action for avoiding sexual victimization. Physical resistance requires that offenders expend additional labor such as catching the victim again, and pose additional risk such as injury to the offender (Guerette & Santana, 2010). This labor and risk might be effective in reducing the potential of sexual victimization. Effects of physical resistance were mainly reported in North America (Clay-Warner, 2002; Fisher et al., 2007; Senn et al., 2015; Tark & Kleck, 2014; Ullman, 2007) with a few exceptions (Sarnquist et al., 2014). Our findings with a Japanese sample confirmed generalizability of previous findings into the Asian population. We also found that the effects of forceful verbal resistance were comparable to the effects of physical resistance, similar to previous studies (Clay-Warner, 2002; Zoucha-Jensen & Coyne, 1993). Interconnections between victim's protective action and offender's failure of sexual crime suggested indirect effects of forceful verbal resistance (Figure 1). Antecedent victim's forceful verbal resistance was linked to consequent bystander intervention and victim's non-forceful verbal resistance, both of which increased the chance of avoiding sexual victimization. Forceful verbal resistance adds the cost of crime, such as clear resistance from the potential victim, during the initial step, and might add other costs of crime, such as being caught by bystanders, in the second step. The two-step effects of forceful verbal resistance might make the total effect comparable to the effects of physical resistance. We also found that victim's non-forceful resistance was effective for avoiding sexual victimization, but the effect size of victim's non-forceful resistance was smaller than the effect size of victim's

physical resistance and forceful verbal resistance. One reason stems from sample differences. Our study did not include child-victim cases for whom the non-forceful verbal resistance was effective (Leclerc et al., 2011b), so that non-forceful resistance might not show the protective effects like previous study. Our study also include rape victims who preferred physical resistance(Fisher et al., 2007) so that the effects of physical resistance might be expanded, whereas the effects of non-forceful resistance might be diminished.

Parity between Victim's Protective Action and Offender's Criminal Behaviors predicted attempted sexual crime (Hypothesis 2 and 3)

As hypothesized (2), the sequence of offender's antecedent violence and victim's consequent physical resistance was effective for avoiding sexual victimization. The sequence of offender's antecedent verbal coercion and victim's consequent forceful physical resistance was effective for avoiding sexual victimization (hypothesis 3). Moreover, the sequence of offender's antecedent sexual behavior and victim's consequent physical resistance was effective for avoiding sexual victimization. These findings clarified the temporal order of the parity between an offender's antecedent physical contact and the victim's consequent physical resistance (Fisher et al., 2007; Nurius & Norris, 1996; Ullman, 1998). Victim's physical resistance responding to an offender's antecedent physical contact might prevent additional criminal behaviors by the offender and decrease the potential of sexual victimization. Similarly, victim's forceful verbal resistance responding to an offender's antecedent verbal coercion might prevent additional criminal behaviors by the offender and decrease the potential of

sexual victimization.

Predictive Sequence for Sexual Victimization (Hypothesis 4)

As hypothesized (4), the sequence of victim's antecedent general resistance and offender's consequent violence predicted sexual victimization (w = 0.76). The sequence of offender's antecedent violence and offender's consequent sexual behavior predicted sexual victimization. Taking into account that the small effect size of single violence (w = 0.17), offender's violence need to be interpreted with antecedent and consequent behaviors of his violence. The offender's violence followed by his sexual behavior on a victim could predict sexual victimization, because his violence could prevent additional resistance from the victim (Jordan, 2005). In contrast, the offender's violence followed by victim's physical resistance could predict avoidance of sexual victimization, because his violence cause counterattack from the victim and increase the cost of crime (Fisher et al., 2007).

Limitations

Our study has limitations regarding sample and behavioral coding. First, the number of sample is too small to generalize our findings(Pang, Lee, & Vaithyanathan, 2002; Tong & Koller, 2001), so our findings are preliminary and requires caution for interpretation. Moreover, our sample did not include child-victim cases so that protective action and sequence for avoiding sexual victimization might be biased. Previous study suggested that child-victims' physical resistance might have adverse effects on sexual victimization(Finkelhor et al., 1995a, 1995b) and their non-forceful verbal resistance could be effective to reduce the risk of sexual victimization(Leclerc et

al., 2011b). Future study needs large sample including child-case victims. Second, our behavioral coding was based on criminal suit documents; the documents focused on criminal behaviors, so several general behaviors might not have been described well, such as giving gifts and playing games (Leclerc et al., 2016). The documents were also written by individual judge. Description of crime situation could be changed by judges (Zaleski, Gundersen, Baes, Estupinian, & Vergara, 2016). Actually, several victim's resistant behavior was describe only "resistance" and cannot categorize specific resistant behavior. Individual differences of judges need to be controlled near the future.

Conclusion

Despite these limitations, our supervised machine learning model including victim's and offender's behaviors during sexual crime clarified the protective sequence for avoiding sexual victimization. We summarize three points. First, the sequence of an offender's antecedent violence and a victim's consequent physical resistance was effective protective action, but the sequence of a victim's antecedent resistance and an offender's consequent violence was predictive for sexual victimization. Hence, protective training needs a lecture how to restrain an offender's counterattack. Second, forceful verbal resistance was especially effective after the offender's verbal coercion. Hence, offender's verbal coercion could be a sign to use forceful verbal resistance. Third, our model showed protective sequences avoiding for sexual victimization, which were not clarified by predominant methodology. Use of supervised machine learning

models in other official criminal documents, such as murder and robbery case, could discover protective sequences avoiding for these crimes. Protective sequence is fundamental in resistance training (Senn et al., 2013, 2015), and contribute to the improvement of resistance training (Senn et al., 2015).

Compliance with ethical standards

Funding

The present study was not funded by any foundation.

Conflict of interest

The first author declares that he has no conflict of interest.

Ethical approval

All procedures performed in the present study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

The present study abbreviated informed consent because of three reasons. First, participants' informed consent and researchers' will do not affect our sampling methods. This is because our criminal suit documents are based on daily activity logs in Japanese courts. Regardless of the participants and researchers' will, Japanese courts created and stored the documents as their professional tasks. Second, if we analyzed only those who could get informed consent in prison, the data could be biased strongly and cannot be a

representative data of sexual offenders in a Japanese prison. Third, analysis of criminal documents is the best method to clarify effective behavioral sequences for avoiding rape. The effective behavioral sequences for avoiding rape were essential to prevent sexual victimization.

Following these reasons, we abbreviated informed consent. Abbreviation of informed consent was frequent in epidemiological study (e.g., Information about influenza and Ebola virus was frequently used without informed consent from patients). The present study was also acknowledged by an ethical committee in a local university and a research committee in a local prison in Japan.

References

Balemba, S., Beauregard, E., & Mieczkowski, T. (2012). To resist or not to resist?: The effect of context and crime characteristics on sex offenders' reaction to victim resistance. *Crime & Delinquency*, *58*(4), 588–611. https://doi.org/10.1177/0011128712437914

Beauregard, E., Proulx, J., Rossmo, K., Leclerc, B., & Allaire, J.-F. (2007). Script analysis of the hunting process of serial sex offenders. *Criminal Justice and Behavior*, *34*(8), 1069–1084. https://doi.org/10.1177/0093854807300851

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0387310738

441 Braga, A. A. (2005). Hot spots policing and crime prevention: A systematic review of
442 randomized controlled trials. *Journal of Experimental Criminology*, *1*(3), 317–
443 342.
444 Clarke, R. V. G. (1997). *Situational Crime Prevention*. Criminal Justice Press Monsey,
445 NY. Retrieved from
446 http://www.popcenter.org/library/reading/pdfs/scp2_intro.pdf
447 Clay-Warner, J. (2002). Avoiding rape: The effects of protective actions and situational
448 factors on rape outcome. *Violence and Victims*, *17*(6), 691–705.
449 https://doi.org/10.1891/vivi.17.6.691.33723
450 Cornish, D. B., & Clarke, R. V. (2014). *The Reasoning Criminal: Rational Choice*
451 *Perspectives on Offending*. Transaction Publishers.
452 Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017).
453 Evaluating the effectiveness of educational data mining techniques for early
454 prediction of students' academic failure in introductory programming courses.
455 *Computers in Human Behavior*, *73*, 247–256.
456 https://doi.org/10.1016/j.chb.2017.01.047
457 Felson, M., & Clarke, R. V. (1998). Opportunity makes the thief. *Police Research Series,*
458 *Paper*, *98*, 1–36.
459 Finkelhor, D., Asdigian, N., & Dziuba-Leatherman, J. (1995a). The effectiveness of
460 victimization prevention instruction: An evaluation of children's responses to
461 actual threats and assaults. *Child Abuse & Neglect*, *19*(2), 141–153.
462 https://doi.org/10.1016/0145-2134(94)00112-8

463 Finkelhor, D., Asdigian, N., & Dziuba-Leatherman, J. (1995b). Victimization prevention
464 programs for children: a follow-up. *American Journal of Public Health*, *85*(12),
465 1684–1689. https://doi.org/10.2105/AJPH.85.12.1684
466 Fisher, B. S., Daigle, L. E., Cullen, F. T., & Santana, S. A. (2007). Assessing the efficacy
467 of the protective action–completion nexus for sexual victimizations. *Violence*
468 *and Victims*, *22*(1), 18–42. https://doi.org/10.1891/vv-v22i1a002
469 Guerette, R. T., & Santana, S. A. (2010). Explaining victim self-protective behavior
470 effects on crime incident outcomes: A test of opportunity theory. *Crime &*
471 *Delinquency*, *56*(2), 198–226. https://doi.org/10.1177/0011128707311644
472 Jordan, J. (2005). What would MacGyver do? The meaning(s) of resistance and
473 survival. *Violence Against Women*, *11*(4), 531–559.
474 https://doi.org/10.1177/1077801204273299
475 Leclerc, B., Chiu, Y.-N., Cale, J., & Cook, A. (2016). Sexual violence against women
476 through the lens of environmental criminology: Toward the accumulation of
477 evidence-based knowledge and crime prevention. *European Journal on Criminal*
478 *Policy and Research*, *22*(4), 593–617.
479 https://doi.org/10.1007/s10610-015-9300-z
480 Leclerc, B., & Wortley, R. (2015). Predictors of victim disclosure in child sexual abuse:
481 Additional evidence from a sample of incarcerated adult sex offenders. *Child*
482 *Abuse & Neglect*, *43*, 104–111. https://doi.org/10.1016/j.chiabu.2015.03.003
483 Leclerc, B., Wortley, R., & Smallbone, S. (2010). An Exploratory Study of Victim
484 Resistance in Child Sexual Abuse: Offender Modus Operandi and Victim

485 Characteristics. *Sexual Abuse*, *22*(1), 25–41.
486 https://doi.org/10.1177/1079063209352093
487 Leclerc, B., Wortley, R., & Smallbone, S. (2011a). Getting into the script of adult child
488 sex offenders and mapping out situational prevention measures. *Journal of*
489 *Research in Crime and Delinquency*, *48*(2), 209–237.
490 https://doi.org/10.1177/0022427810391540
491 Leclerc, B., Wortley, R., & Smallbone, S. (2011b). Victim resistance in child sexual
492 abuse: A look into the efficacy of self-protection strategies based on the
493 offender's experience. *Journal of Interpersonal Violence*, *26*(9), 1868–1883.
494 https://doi.org/10.1177/0886260510372941
495 Maeda, M. (2015). *Detailed Explanation of Japanese Penal Code* (6th edition). Tokyo:
496 University of Tokyo Press.
497 Nurius, P. S., & Norris, J. (1996). A cognitive ecological model of women's response to
498 male sexual coercion in dating. *Journal of Psychology & Human Sexuality*, *8*(1–
499 2), 117–139. https://doi.org/10.1300/J056v08n01_09
500 Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification
501 using machine learning techniques. In *Proceedings of the ACL-02 conference on*
502 *Empirical methods in natural language processing-Volume 10* (pp. 79–86).
503 Association for Computational Linguistics. Retrieved from
504 http://dl.acm.org/citation.cfm?id=1118704
505 Sarnquist, C., Omondi, B., Sinclair, J., Gitau, C., Paiva, L., Mulinge, M., … Maldonado,
506 Y. (2014). Rape prevention through empowerment of adolescent girls. *Pediatrics*,

133(5), e1226–e1232. https://doi.org/10.1542/peds.2013-3414

Senn, C. Y., Eliasziw, M., Barata, P. C., Thurston, W. E., Newby-Clark, I. R., Radtke, H. L., & Hobden, K. L. (2013). Sexual assault resistance education for university women: study protocol for a randomized controlled trial (SARE trial). *BMC Women's Health*, *13*, 25. https://doi.org/10.1186/1472-6874-13-25

Senn, C. Y., Eliasziw, M., Barata, P. C., Thurston, W. E., Newby-Clark, I. R., Radtke, H. L., & Hobden, K. L. (2015). Efficacy of a sexual assault resistance program for university women. *New England Journal of Medicine*, *372*(24), 2326–2335. https://doi.org/10.1056/NEJMsa1411131

Tark, J., & Kleck, G. (2014). Resisting rape: The effects of victim self-protection on rape completion and injury. *Violence Against Women*, *20*(3), 270–292. https://doi.org/10.1177/1077801214526050

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, *2*(Nov), 45–66.

Ullman, S. E. (1998). Does offender violence escalate when rape victims fight back? *Journal of Interpersonal Violence*, *13*(2), 179–192. https://doi.org/10.1177/088626098013002001

Ullman, S. E. (2007). A 10-year update of "review and critique of empirical studies of rape avoidance." *Criminal Justice and Behavior*, *34*(3), 411–429. https://doi.org/10.1177/0093854806297117

Ullman, S. E., & Knight, R. A. (1992). Fighting back: Women's resistance to rape. *Journal of Interpersonal Violence*, *7*(1), 31–43.

https://doi.org/10.1177/088626092007001003

Yamashita, T., & Yamaguchi, A. (2016). *Statute Books* (Heise 28). Yubikaku. Retrieved from http://www.yuhikaku.co.jp/six_laws/detail/9784641104761

Zaleski, K. L., Gundersen, K. K., Baes, J., Estupinian, E., & Vergara, A. (2016). Exploring rape culture in social media forums. *Computers in Human Behavior*, *63*, 922–927. https://doi.org/10.1016/j.chb.2016.06.036

Zoucha-Jensen, J. M., & Coyne, A. (1993). The effects of resistance strategies on rape. *American Journal of Public Health*, *83*(11), 1633–1634. https://doi.org/10.2105/AJPH.83.11.1633

Table 1 Definition of rape and sexual coercion in Japan

	Use of illegal physical force or verbal coercion	Intent of sexual behavior	Realization of sexual behavior	Intent of penile-vaginal penetration	Realization of penile-vaginal penetration
Completed Rape	○	○	○	○	○
Attempted Rape	○	○	○	○	×
Completed Sexual Coercion	○	○	○	×	×
Attempted Sexual Coercion	○	○	×	×	×

Table 2 Code and example of victims' and offenders' behaviors

Subject	Code	Example	n
Victim			
	Physical Resistance	"flee"	6
		"(escaped from him and) step out onto a balcony"	1
		"overpower the offender"	1
	Forceful Verbal Resistance	"scream"	1
		"call the police with mobile phone"	1
		"shout"	1
		"alert police"	1
		"scream for someone to get help"	1
		"make a noise"	1.
	Non-Forceful Verbal Resistance	"demand accusingly"	1
		"She said 'I will do anything for you'"	1
		"She said 'I want to go back to my house'"	1
	General Resistance	"resist"	5
		"fierce resistance"	1
Offender			
	Sexual behavior	"sexual intercourse"	38
		"touch victim's private parts"	28
		"grab victim's breast"	25
		"tear victim's clothes off"	17
		"oral sex"	13
		"penetrate victim's vagina with offender's finger"	7
		"touch victim's bottom"	5
		"lick victim's cheek (n = 1), nipple (n = 1), breast (n = 1), and nipple and private parts (n = 1)"	4
		"expose offender's private parts"	4
		"kiss on the lips"	3
		"press oneself against victim"	3
		"hand job"	3
		"press offender's penis against victim's face (n = 2), bottom (n = 1)"	3
		"ejaculation"	3
		"record pornographic scene"	2
		"touch victim's thigh"	2

	“open victim’s crotch”	1
Physical violence	“cover victim’s mouth with hand (n=24) or towel (n=1),” “cover victims’ eyes with adhesive tape (n=2) or unknown object (n=1),” “cover victim’s face with victim’s hood (n=1), flag (n=1) or unknown object (n=1)”	31
	“push down”	14
	“choke”	12
	“grip victim’s arm (n = 9), victim’s hair (n = 2)”	11
	“mount”	10
	“punch”	9
	“tie victim with banding band (n=1), belt (n=1), rope (n=1), towel (n=1), and unknown object (n=1)”	5
	“show knife (n=4) and imitation sword (n=1)”	5
	“pinion victim”	3
	“pull victim”	3
	“press knife against victim’s body (n=2) and private parts (n=1)”	3
	“push victim”	1
	“press scissors against victim’s body”	1
	“press pen against victim’s face”	1
	“kick victim's face”	1
	“lift offender’s hand against victim”	1
	“press a burning cigarette bottom against victim”	1
	“slap victim”	1
	“slash victim with knife”	1
Verbal Coercion	Threats to a victim’s life include: “I kill you if you make a noise” (n=16), “I will kill you” (n=4), “I will kill you if you move” (n=2), “Choose to be killed or have sex” (n=1), “I will kill you if you flee” (n=1), “I will kill you if you refuse my touch” (n=1), “I will not kill you if you do not make a noise” (n = 1), and ”Shall we die together?” (n=1).	27
	Threats to a victim’s body include: “Choose to choke or have sex” (n = 1), “Do what I tell you if you do not want to get punched” (n = 1), “Do you want to be beaten up?” (n = 1), “I will punch you” (n = 1), “I will shoot you if you open your eyes” (n = 1), “I will smash you if you raise your voice” (n =	8

	1), “Let me slash you with this knife” (n = 1), and “You can go back to your house if we can have sex together” (n = 1).	
	Threats to a victim’s reputation include: “Take off your clothes” (n=2), “I filmed you secretly. You do not want the film to be exposed on the Internet” (n=1), and “Let us go to the police” (n=1).	4
	Threats to a victim’s property include: “Pay X yen or be my girlfriend” (n=1) and “You can go back to your house if you pay money” (n=1).	2
	Threats to a victim’s freedom include: “I will take you away if you make a noise” (n=1).	1
	Threats to something else include: “Anything can happen if I get angry” (n=1), “Be quiet. You know what will happen if you make a noise” (n=1), “Be quiet. Your children are at risk” (n=1), ”Shout angrily” (n=1), “I am a mafia member” (n=1), “I have another collaborator” (n=1), “I will not do anything” (n=1), “You are being monitored by the gang” (n=1), “You are a target of the mafia” (n=1), and “You exposed our secret” (n=1).	10
	Orders to hinder victims from exercising their rights include: “Be quiet” (n=14), “Do not move” (n = 7), “Be quiet and do not move (n = 1)”“Do not look at my face” (n=1), “I will grab your breasts (Do not refuse)” (n=1), and “I will penetrate you (Do not refuse)” (n=1).	25
	Orders to perform an act include: “Suck” (n = 2) and “Lower your eyes” (n=1).	3
Persuasion	Offenders pretended to be a company manager and talked to the victim as her boss (n=2), pretended to be a security guard and talked about the victim’s shoplifting (n=1).	3
	They also frequently communicated with victims via telephone and e-mail (n=1), and offered kindness to them, such as “May I help you?” (n=1) and “Rest in my car” (n=1).	3
	They also made fake contracts with night service victims, such as “I will give you X yen for your service” (n =2).	2
	They also used real identities such as shop managers and telephoned the victim as a customer (n=1).	1

Table 3 Comparison of rape and sexual coercion cases

	Total		Rape		Sexual Coercion			*d.f.*	*p.*
	N = 88		*n* = 37		*n* = 51				
Age and Sex	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*		
Female Victim's Age	22.0[a]	6.3[a]	22.5	6.6	21.7[a]	6.1[a]	0.61	85.00	0.54
Male Offender's Age	42.3	8.4	43.4	9.9	41.5	7.2	0.98	62.77	0.33
Relationships	*n*	*%*	*n*	*%*	*n*	*%*			
Parent-Child	6	6.8	4	10.8	2	3.9			0.20
Romantic	1	1.1	1	2.7	0	0.0			0.42
Non-romantic	6	6.8	4	10.8	2	3.9			0.20
Unknown	75	85.2	28	75.7	47	92.2			0.03 *
Complete cases	73	83.0	24	64.9	49	96.1			0.00**
Alcohol Use									
Alcohol-induced Drunkenness	2	2.3	1	2.7	1	2.0			1
Resistance									
Physical	6	6.8	5	13.5	1	2.0			0.03 *
Forceful Verbal	6	6.8	4	10.8	2	3.9			0.16
Non-forceful Verbal	3	3.4	3	8.1	0	0.0			0.06
general	6	6.8	5	13.5	1	2.0			0.03 *
Bystanders									
Bystanders Intervention	4	4.5	2	5.4	2	3.9			0.56
Setting [b]									
In. Private	49	55.7	26	70.3	23	45.1			0.02 *
In. Semi-public	10	11.4	6	16.2	4	7.8			0.19
In. Public	7	8.0	3	8.1	4	7.8			0.63
Out. Private	5	5.7	2	5.4	3	5.9			0.65
Out.Semi-public	16	18.2	6	16.2	10	19.6			0.45
Out. Public	11	12.5	5	13.5	6	11.8			0.53

[a]: one case is charged with public lewdness, so the victim's age and sex are

unknown. [b]: several cases used multiple locations, so the percentage for settings is more than 100%. In.: Indoor, Out.: Outdoor, *:$p < .05$, **: $p < .01$.

Table 4 Conditional and unconditional probabilities of offender's and victim's behaviors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. In. Private	0	0	0	0	0	0	.02	0	.04	.22	.38	.32	0	0	0	0	0	0	0.02	.11
2. In. Semi-public	0	0	0	0	0	0	0	0	.18	0	.45	.36	0	0	0	0	0	0	0	.02
3. In. Public	0	0	0	0	0	0	.29	0	0	.14	0	.57	0	0	0	0	0	0	0	.01
4. Out. Private	0	0	0	0	0	0	0	0	0	0	.80	.20	0	0	0	0	0	0	0	.01
5. Out. Semi-public	0	0	0	0	0	0	.13	0	0	.06	.56	.25	0	0	0	0	0	0	0	.03
6. Out. Public	0	0	0	0	0	0	0	0	0	0	.82	.18	0	0	0	0	0	0	0	.02
7. Go with	.22	.17	.11	0	.06	0	0	0	0	.11	.17	.17	0	0	0	0	0	0	0	.04
8. Invade	.93	0	0	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.06
9. Persuasion	.11	0	.22	0	0	0	.22	0	0	.22	0	.22	0	0	0	0	0	0	0	.02
10. Ver. Coercion	.01	0	0	0	.01	0	.05	0	.01	.27	.16	.36	0	.01	.02	.02	0	.01	.05	.17
11. Violence	0	0	0	0	.01	0	.03	.02	0	.32	.32	.29	0	.01	.01	0	0	0	0	.24
12. Sexual Behav.	0	0	0	0	0	0	.02	.01	0	.04	.08	.65	.01	.05	.02	0	.07	.02	.02	.19
13. Failure of goal	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	0
14. Phy. Resistance	0	0	0	0	0	0	0	0	0	0	.50	0	.38	0	.13	0	0	0	0	.02
15. Forc. Ver. Resi.	0	0	0	0	0	0	0	0	0	.17	.17	0	.33	0	0	.17	0	.17	0	.01
16. Non-Forc. Ver. Resi.	0	0	0	0	0	0	.11	0	0	.11	0	0	.11	0	0	0	0	0	0	.01
17. Gen. Resi.	0	0	0	0	0	0	0	0	0	0	.17	0	.83	0	0	0	0	0	0	.01
18. Bystander	0	0	0	0	0	0	0	0	0	0	0	0	.75	.25	0	0	0	0	0	.01
19. Money	0	0	0	0	0	0	0	0	0	0.4	.60	0	0	0	0	0	0	0	0	.01
20. Unconditional	.07	.01	.01	0	.01	0	.04	.01	.01	.17	.24	.34	.03	.02	.01	.01	.01	.01	.01	nul

Note. N = 472[a]. The rows show antecedent behavior, and the columns show consequent behavior. The final row and column represent unconditional antecedent and consequent behaviors respectively. [a]: The number of total behaviors is 560, but the initial and final behavior in a case cannot be consequent and antecedent behaviors, so these ends of behaviors were excluded from consequent and antecedent data analysis. In.: Inside, Out.: outside.Ver.:Verbal, Behav.: Behavior,Phy.:Physical, Forc.:Forceful, Gene.: General, Resi.:Resistance

Table 5 Tenfold-cross-validated accuracy of complete/attempted sexual crimes with behavioral sequences

	Total	Rape	Sexual Coercion
0 sequence (one behavior)	0.872	0.933	0.963
0+1 sequence (one behavior + two continuous behaviors)	0.908	0.883	0.963
0+1+2 sequence (one behavior + two continuous behaviors + three continuous behaviors)	0.962	0.883	0.963

Table 6 Protective sequence for avoidance of sexual victimization

	0 sequence	*w*	1 sequence	*w*	2 sequence	*w*
1	V's general resistance	-2.00	O's sexual behavior => V's general resistance	-2.09	O's sexual behavior => O's sexual behavior => V's general resistance	-2.11
2	V's verbal resistance	-1.76	O's sexual behavior => V's physical resistance	-1.47	O's sexual behavior => O's sexual behavior => V's physical resistance	-1.60
3	V's physical resistance'	-1.54	O's Verbal coercion => V's verbal resistance	-1.20	O's persuasion => O's verbal coercion => V's verbal resistance	-1.18
4	Bystander Intervention	-0.84	O's violence => V goes with O	-1.08	O's verbal coercion => O's violence => V's physical resistance	-0.82
5	V goes with O	-0.64	O's sexual behavior => V's verbal resistance	-1.08	V's physical resistance => O's violence => O's violence	-0.73
6	O's verbal coercion	-0.18	O's violence => V's physical resistance	-1.00	O's violence => O's verbal coercion => O's violence	-0.69
7	V's non forceful verbal resistance	-0.17	V goes with O => O's Verbal coercion	-0.77	O's sexual behavior => O's sexual behavior => V's verbal resistance	-0.69
8	O invades	-0.06	V's verbal resistance => V's non forceful verbal resistance	-0.77	O's sexual behavior => V's verbal resistance => V's non forceful verbal resistance	-0.69
9	V encounters O at outdoor public setting	-0.06	V's verbal resistance => O's verbal coercion	-0.72	V encounters O at indoor semipublic setting => O's violence => V goes with O	-0.68
10	O's violence	0.17	O's verbal coercion => V's physical resistance	-0.69	O's violence => V goes with O => V's verbal coercion	-0.68

Note. Negative score indicates the negative predictive value on sexual victimization. O: Offender, V: Victim

Table 7 Predictive sequence for sexual victimization

	0 sequence	*w*	1 sequence	*w*	2 sequence	*w*
1	V encounters O at indoor public setting	1.09	O's violence => O's sexual behavior	0.88	O's violence => O's sexual behavior => O's violence	0.40
2	O's sexual behavior	0.75	V's general resistance => O's violence	0.76	O's verbal coercion => O's verbal coercion=> O's violence	0.38
3	O robbed V's money	0.50	V encounters O at outdoor semipublic setting => O's sexual behavior	0.65	O's verbal coercion => O's violence => O's sexual behavior	0.34
4	V encounters O at outdoor semipublic setting	0.36	V encounters O at indoor private setting => O's sexual behavior'	0.64	V encounters O at indoor semipublic setting => O's violence => O's violence	0.33
5	O's persuasion	0.35	V encounters O at indoor public setting => O's sexual behavior	0.45	V encounters O at indoor semipublic setting => O's violence => O's verbal coercion	0.31
6	V encounters O at indoor private setting	0.33	V encounters O at indoor semipublic setting => O's sexual behavior	0.39	O invades => V encounters O at indoor private setting => O's sexual behavior	0.29
7	V encounters O at indoor semipublic setting	0.26	V encounters O at outdoor public setting => O's sexual behavior	0.33	O's verbal coercion => O's verbal coercion => O's sexual behavior	0.27
8	V encounters O at outdoor private setting	0.22	O's verbal coercion => V goes with O	0.32	V's general resistance => O's violence => O's sexual behavior	0.26
9	O's violence	0.17	O robbed V's money => O's sexual behavior	0.32	O's sexual behavior => V's general resistance => O's violence	0.26
10	V encounters O at outdoor public setting	-0.06	O's sexual behavior => O's verbal coercion	0.30	V encounters O at outdoor semipublic setting=> O's sexual behavior => V's physical resistance	0.26

Note. Positive score indicates the positive predictive value on sexual victimization. O: Offender, V: Victim