Mathematics tools for digital forensics

M. Eugenia Cornejo Ivana

Sˇtajner-Papuga

Jesu´s Medina Andreja Tepavˇcevi´c

mariaeugenia.cornejo@uca.es ivana.stajner-papuga@dmi.uns.ac.rs jesus.medina@uca.es andreja@dmi.uns.ac.rs

Department of Mathematics Faculty of Sciences University of C´adiz University of Novi Sad

C´adiz, Spain Novi Sad, Serbia

The main aim of this research is to increase awareness among forensics experts of some new mathematical methods and techniques and their applicability. Nowa- days contemporary non-classical mathematical theories are producing powerful tools for modeling uncertainty and decision problems under uncer- tainty.

One of the current problems is a classiﬁcation or ranking of a big amount of forensics data. Aggregation operators are a mathematical tool that can extract a representative value from a pale of data, and, therefore, are useful in ranking diﬀerent data, e.g., diﬀerent recorded crimes in a certain area. Therefore, it is important to develop a methodology for the automatic prioritization of suspicious data or rank diﬀerent crimes.

One of the current problems is a classiﬁcation or ranking of a big amount of forensics data. Aggregation operators are a mathematical tool that can extract a representative value from a pale of data, and, therefore, are useful in ranking diﬀerent data, e.g., diﬀerent recorded crimes in a certain area. Therefore, it is important to develop a methodology for the automatic prioritization of suspicious data or rank diﬀerent crimes.

The proposed method will incorporate fuzzy sets, monotone set functions, and integral aggregation operators.

1. Fuzzy sets are used for modeling segments of data that needs to be evalu- ated.
2. Functions for all data that needs to be evaluated are formed (numerical values are extracted for membership functions of fuzzy set from the ﬁrst step).
3. Monotone set functions are used to describe importance and interaction of segments of data that needs to be evaluated.
4. Integral aggregation operator (Choquet integral) is used to evaluate each data.

Characteristic function ⇛ Membership function

*U* - universal set

*A* - fuzzy subset of *U*

*µA* - membership function that represents fuzzy set *A*

*µA* : *U* → [0*,* 1]

*U* - universal set - ﬁle size

*A* - fuzzy subset of *U* - at least around 6MB

1

0.66

0.5

0.33

Membership function for fuzzy set ”At least around 6MB”.

2MB

3MB

4MB

6MB

10MB

20MB

Definition 1 *Let X be a universe. A* monotone set function, i.e., a fuzzy measure, *on* P(*X*) *is a mapping ν* : P(*X*) −→ [0*,* 1] *satisfying the following properties:*

*(i) ν*(∅) = 0*,*

*(ii) if A* ⊆ *B then ν*(*A*) ≤ *ν*(*B*)*, for all A, B* ∈ P(*X*)*.*

Definition 2 *Let X be a* discrete universe*, f* : *X* → {*ω*1*, ω*2*, . . . , ωn*} *be an arbitrary* simple function *and ν* : P(*X*) −→ [0*,* 1] *a fuzzy measure on* P(*X*)*. The* Choquet integral *of f with respect to ν is given by:*

∫

(*C*)

*X*

*f dν* =

Σ*n*

*i*=1

(*ωi* − *ωi*−1) · *ν* (*Ai*) *,*

*where ω*0 = 0*,* 0 *< ω*1 ≤ *ω*2 ≤ *. . .* ≤ *ωn and Ai* = {*x* ∈ *X* | *f* (*x*) ≥ *ωi*}*.*

Let *X* = {*x*1*, x*2*, . . . , xn*} be the set of all relevant segments of the considered dataset and *f* : *X* → {*ω*1*, ω*2*, . . . , ωn*} be a simple function associated with a data with relevant segments from *X*, such that 0 *< ω*1 ≤ *ω*2 ≤ · · · ≤ *ωn*.

Without loss of generality, we can assume that the segments in *X* are ordered in order to satisfy:

*f* (*x*1) = *ω*1*, f* (*x*2) = *ω*2*, . . . , f* (*xn*) = *ωn.*

Fuzzy measure *ν* will be used for describing importance and interaction of segments.

The opinion of an expert, on the level of importance of each segment in *X* is

*ν*({*x*1})*, ν*({*x*2})*, . . . , ν*({*xn*})*.*

The joint importance of different segments, needs to be calculated pre- serving the main property of a fuzzy measure, that is, the monotonicity:

* the impact of two segments should be stronger than the impact of each segment on its own,
* the impact of three segments should be stronger than impact of two segments, and of a single segment...

This ampliﬁcation of strength continues with each additional segment. Each step should be more emphasized than the previous one. It is done in the following manner:

The joint importance of different segments, needs to be calculated pre- serving the main property of a fuzzy measure, that is, the monotonicity:

* the impact of two segments should be stronger than the impact of each segment on its own,
* the impact of three segments should be stronger than impact of two segments, and of a single segment...

This ampliﬁcation of strength continues with each additional segment. Each step should be more emphasized than the previous one. It is done in the following manner:

*ν*(Ai) = *ν*({xn*,* xn−1*, . . . ,* xi}) = @i(*ν*({xi})*, ν*({xn*,* xn−1*, . . . ,* xi+1}))*,*

where @*i* : [0*,* 1]2 → [0*,* 1] are aggregation operators satisfying the inequality *x* ≤

@*i*(*x, y*), for all *x, y* ∈ [0*,* 1], with *i* ∈ {1*,* 2*, . . . , n* − 1}.

Given the considered dataset *Y* = {*y*1*, y*2*, . . . , yd*} and the set of relevant seg- ments *X* = {*x*1*, x*2*, . . . , xn*} corresponding to dataset *Y* , the methodology for the automatic ranking of data is carried out following the next steps:

Given the considered dataset *Y* = {*y*1*, y*2*, . . . , yd*} and the set of relevant seg- ments *X* = {*x*1*, x*2*, . . . , xn*} corresponding to dataset *Y* , the methodology for the automatic ranking of data is carried out following the next steps:

Step 1: Each *xi* ∈ *X* is observed as a fuzzy set with its own membership func- tion. Each *xi* is given on its own universe *Ui*.

Given the considered dataset *Y* = {*y*1*, y*2*, . . . , yd*} and the set of relevant seg- ments *X* = {*x*1*, x*2*, . . . , xn*} corresponding to dataset *Y* , the methodology for the automatic ranking of data is carried out following the next steps:

Step 1: Each *xi* ∈ *X* is observed as a fuzzy set with its own membership func- tion. Each *xi* is given on its own universe *Ui*.

Step 2: For each data from *Y* , a simple function *fyj* : *X* → {*ω*1*,j, ω*2*,j, . . . , ωn,j*} is assigned, where *ωi,j* is the value of a membership function of the fuzzy set *xi* for the exact value from *Ui* that appears in the *j*-th data from *Y* .

Given the considered dataset *Y* = {*y*1*, y*2*, . . . , yd*} and the set of relevant seg- ments *X* = {*x*1*, x*2*, . . . , xn*} corresponding to dataset *Y* , the methodology for the automatic ranking of data is carried out following the next steps:

Step 1: Each *xi* ∈ *X* is observed as a fuzzy set with its own membership func- tion. Each *xi* is given on its own universe *Ui*.

Step 2: For each data from *Y* , a simple function *fyj* : *X* → {*ω*1*,j, ω*2*,j, . . . , ωn,j*} is assigned, where *ωi,j* is the value of a membership function of the fuzzy set *xi* for the exact value from *Ui* that appears in the *j*-th data from *Y* .

Step 3: The fuzzy measure *ν* : P(*X*) → [0*,* 1] describing the importance and interaction of segments of data that need to be evaluated is obtained from the procedure previouslz explained.

Given the considered dataset *Y* = {*y*1*, y*2*, . . . , yd*} and the set of relevant seg- ments *X* = {*x*1*, x*2*, . . . , xn*} corresponding to dataset *Y* , the methodology for the automatic ranking of data is carried out following the next steps:

Step 1: Each *xi* ∈ *X* is observed as a fuzzy set with its own membership func- tion. Each *xi* is given on its own universe *Ui*.

Step 2: For each data from *Y* , a simple function *fyj* : *X* → {*ω*1*,j, ω*2*,j, . . . , ωn,j*} is assigned, where *ωi,j* is the value of a membership function of the fuzzy set *xi* for the exact value from *Ui* that appears in the *j*-th data from *Y* .

Step 3: The fuzzy measure *ν* : P(*X*) → [0*,* 1] describing the importance and interaction of segments of data that need to be evaluated is obtained from the procedure previouslz explained.

Step 4: Each *yj* ∈ *Y* is e∫valuated by using the integral aggregation operator

(Choquet integral) (*C*) *fyjdν,* and ranked according to obtained values.

This section applies the proposed model to a part of Bilbao crime dataset concerning the physical and/or psychical gender violence, which is displayed in the ﬁrst four columns of Table 1. Values in Table 2, and subsequently in Table 3, are based on membership functions of discrete fuzzy sets ”safe place”, ”safe day in the week” and ”safe neighborhood”. All those sets are formed for the purpose of this example, and in the future will be corrected by experts from the relevant ﬁeld.

The considered dataset is

*Y* = {crime*j* | *j* ∈ {1*, . . . ,* 18}} and the set of relevant segments is

*X* = {Place (P)*,* Work/not work day (W)*,* Neigbourhood (N)}*.*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Place (P) | Work/not work day (W) | Neigbourhood (N) |
| crime1 | Flat | Working day | MATIKO |
| crime2 | Flat | Weekend | OTXARKOAGA |
| crime3 | Flat | Weekend | ERREKALDEBERRI |
| crime4 | Flat | Weekend | SAN FRANCISCO |
| crime5 | Flat | Working day | BASURTU |
| crime6 | Flat | Public Holiday | URIBARRI |
| crime7 | No classiﬁcation | Working day | ABANDO |
| crime8 | No classiﬁcation | Working day | ATXURI |
| crime9 | No classiﬁcation | Working day | IRALABARRI |
| crime10 | Party Room/Disco | Working day | AMETZOLA |
| crime11 | Teaching center | Working day | ATXURI |
| crime12 | Train Station | Weekend | BASURTU |
| crime13 | Urban Public Road | Weekend | ABANDO |
| crime14 | Urban Public Road | Weekend | ABANDO |
| crime15 | Urban Public Road | Weekend | INDAUTXU |
| criem16 | Urban Public Road | Weekend | INDAUTXU |
| criem17 | Urban Public Road | Weekend | SOLOKOETXE |
| crime18 | Urban Public Road | Working day | CASCO VIEJO |

Table 1: Bilbao crime dataset related to physical and/or psychical gender violence.

Each element from the columns Place, Work/Not Work Day and Neighbourhood has its numerical representation extracted from the membership functions assigned to the fuzzy sets

* Dangerous Place (*µDP* ),
* Dangerous Work/Not Work Day (*µDW* ),
* Dangerous Neighbourhood (*µDN*).

These membership functions are discrete mappings and their values are based on an expert opinion.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Place (P) | Work/not work day (W) | Neigbourhood (N) | *µDP* | *µDW* | *µDN* |
| crime1 | Flat | Working day | MATIKO | 1 | 0.7 | 0.8 |
| crime2 | Flat | Weekend | OTXARKOAGA | 1 | 1 | 0.9 |
| crime3 | Flat | Weekend | ERREKALDEBERRI | 1 | 1 | 0.8 |
| crime4 | Flat | Weekend | SAN FRANCISCO | 1 | 1 | 0.9 |
| crime5 | Flat | Working day | BASURTU | 1 | 0.7 | 0.4 |
| crime6 | Flat | Public Holiday | URIBARRI | 1 | 0.9 | 0.8 |
| crime7 | No classiﬁcation | Working day | ABANDO | 0.8 | 0.7 | 0.9 |
| crime8 | No classiﬁcation | Working day | ATXURI | 0.8 | 0.7 | 0.8 |
| crime9 | No classiﬁcation | Working day | IRALABARRI | 0.8 | 0.7 | 0.7 |
| crime10 | Party Room/Disco | Working day | AMETZOLA | 0.7 | 0.7 | 0.5 |
| crime11 | Teaching center | Working day | ATXURI | 0.5 | 0.7 | 0.8 |
| crime12 | Train Station | Weekend | BASURTU | 0.7 | 1 | 0.4 |
| crime13 | Urban Public Road | Weekend | ABANDO | 0.9 | 1 | 0.9 |
| crime14 | Urban Public Road | Weekend | ABANDO | 0.9 | 1 | 0.9 |
| crime15 | Urban Public Road | Weekend | INDAUTXU | 0.9 | 1 | 0.8 |
| criem16 | Urban Public Road | Weekend | INDAUTXU | 0.9 | 1 | 0.8 |
| criem17 | Urban Public Road | Weekend | SOLOKOETXE | 0.9 | 1 | 0.7 |
| crime18 | Urban Public Road | Working day | CASCO VIEJO | 0.9 | 0.7 | 0.5 |

Table 2: Bilbao crime dataset related to physical and/or psychical gender violence.

For crime1, a simple function *f*1 : *X* → {*ω*1*,*1*, ω*2*,*1*, ω*3*,*1} is deﬁned as

*f*1(P) = 1*, f*1(W) = 0*.*7*,* and *f*1(N) = 0*.*8*,*

where

*ω*1*,*1 = *µDP* (Flat)*, ω*2*,*1 = *µDW* (Working day)*,* and *ω*3*,*1 = *µDN*(MATIKO)

The function *fj* associated with crimej such that *j* ∈ {2*, . . . ,* 18} is deﬁned analogously, for all *j*.

The importance of each segment follows from an expert’s opinion, while the interaction of segments will be calculated. The following fuzzy measure *ν* : P({P*,* W*,* N}) → [0*,* 1] is considered:

*ν*({P}) = 0*.*9 *ν*({P*,* W}) = @2(*ν*({P})*, ν*({W})) = 0*.*9

*ν*({W}) = 0*.*6 *ν*({P*,* N}) = @2(*ν*({P})*, ν*({N})) = 0*.*9

*ν*({N}) = 0*.*8 *ν*({W*,* N}) = @2(*ν*({W})*, ν*({N})) = 0*.*8

This example is simple and needs only two levels of aggregation. The ﬁrst level is done by using the t-conorm *S*M : [0*,* 1]2 → [0*,* 1] deﬁned as *S*M(*x, y*) = max{*x, y*}, for all *x, y* ∈ [0*,* 1], as the aggregation operator @2 for calculating measure of two element sets. The second level is based on @1 = *S*L, where

*S*L-

: [0*,* 1]2 → [0*,* 1] is the

L ukasiewicz t-conorm deﬁned as *S*L- (*x, y*) = min{*x* +

*y,* 1}, for all *x, y* ∈ [0*,* 1] and it provides *ν*({P*,* W*,* N}) = *ν*(*A*1) = *ν*(*X*) = 1.

The values of Choquet integral for functions *fj*, being *j* ∈ {1*, . . . ,* 18}, with respect to the fuzzy measure *ν* given in Step 3 are collected in Table 3.

Therefore, the most threatening crimes are 2 and 4.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *j* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| (*C*) | 0.97 | 0.99 | 0.98 | 0.99 | 0.79 | 0.98 | 0.87 | 0.79 | 0.79 | 0.68 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *j* | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| (*C*) | 0.74 | 0.79 | 0.96 | 0.96 | 0.95 | 0.95 | 0.94 | 0.86 |

Table 3: Values of Choquet Integral

The presented research considers integral aggregation operators, namely the Cho- quet integral, as a tool for ranking forensics dataset.

The approach is based on fuzzy measures that are being used for modeling

importance and interaction of all relevant segments of observed dataset.

This procedure is adaptable to personal opinions based on experience of in- vestigators.

The focus of the further research is on construction of the most appropri- ate fuzzy measure that will reflect amplification of importance of the relevant segments when they are observed as a group. Further on, this method will be used on more complex real life problems.

Also, the expert veriﬁcation of membership functions will be done.

References

1. Benvenuti, P., Mesiar, R., Vivina, D.: Monotone set functions-based integrals. In: Pap, E. (ed) Handbook of Measure Theory, pp. 1329-1379, Elsevier, Ams- terdam (2002)
2. Choquet, G.: Theory of capacities. Annales de l’Institut Fourier 5, 131–295 (1953)
3. Grabisch, M., Marichal, J., Mesiar, R., Pap, E.: Aggregations Functions. Cam- bridge University Press (2009)
4. Takaˇci, A.,

Sˇtajner-Papuga, I., Drakuli´c, D., Mari´c, M.: An Extension of

Maximal Covering Location Problem based on the Choquet Integral. Acta Polytechnica Hungarica 13(4), 205–220 (2016)

1. Sugeno, M.: Theory of fuzzy integrals and its application. Doctoral thesis. Tokyo Institute of Technology (1974)
2. Wang, Z., Klir, G.J.: Generalized Measure Theory. Springer (2000)
3. Zadeh, L. A.: Fuzzy sets. Information and Control 8, 338–353 (1965)