

ラフ集合を用いた土石流発生・非発生 Rule 抽出に関する研究

山口大学大学院	学生会員	○杉原成満	山口大学工学部	正会員	竹本大昭
(株)エイトコンサルタント	正会員	佐藤文晴	中電技術コンサルタント(株)	正会員	荒木義則
甲南大学理工学部	非会員	中山弘隆	京都大学大学院	正会員	水山高久
			山口大学工学部	正会員	古川浩平

1. 本研究の背景と目的

豪雨による土石流やがけ崩れ等に代表される土砂災害は未だに後を絶たず、甚大な被害を与えることも少なくない。未だ明確とされていない土石流発生のメカニズムが解明できれば、事前に危険な溪流と安全な溪流が特定できる為、溪流調査数の絞込みや、対策工施工サイトの優先順位の設定に役立つものと考えられる。

そこで、本研究では1982年7月に日本観測史上最も激しい集中豪雨により長崎市周辺で激甚な災害をもたらした土石流災害（以下 S57.7 長崎災害）を対象とし、ラフ集合を用いたデータマイニング手法を土石流発生・非発生データに適用することで土石流災害に関する発生・非発生 Rule の抽出を行った。

2. ラフ集合

ラフ集合論の基本概念は種別と近似である。本研究ではこの2つの概念を用いた土石流発生・非発生データの縮約により、データベースから簡潔かつ有効な知識（Rule）を得ることを目的とした。

図-1 にラフ集合の概念図を示す。土石流の発生溪流を●、非発生溪流を○とすると、発生溪流のみの集合は図-1 に示される部分集合となる。全体集合を2つの要因を用いて9個の領域に区分しても発生溪流の部分集合は完全に分離されない。このように完全に分離されない集合をラフ集合という。これらの要因によって集合が区分された時、同じ領域内のデータがすべて同種であるデータ領域を整合データ（図-1 の影の部分）とし、混在するデータ領域を矛盾データと定義する。要因の組合せを評価する指標として次式に定義する整合度を用いた。

$$\text{整合度} = \frac{\text{整合データ数}}{\text{全データ数}}$$

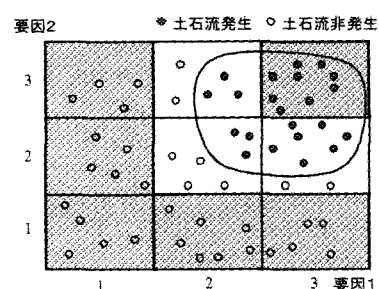


図-1 ラフ集合の概念

3. Rule の拡張

ラフ集合により抽出した Rule は、Rule 領域に含まれる溪流が少なく、有効な Rule であると言い難い。そこで Rule の適応領域の拡張を行った。図-2 (a) に示す色のついた領域が発生 Rule である場合、より危険と考えられる部分集合を含む領域を発生 Rule とすることで、Rule に含まれる溪流数が増加し、有効な Rule となる。（図-2 (b) 参照）非発生 Rule では、より安全と考えられる領域を拡張領域とする。

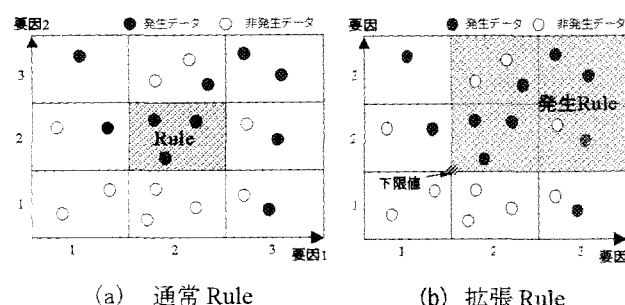


図-2 拡張の概念図

4. 離散データの拡張

溪流方位のような離散データは、もともと物理的な順位を持たない為、拡張の対象とならない。しかし、図-3 (a) のように各カテゴリの発生率（発生溪流/全溪流）を算出することで、危険もしくは安全なカテゴリランクを判断することができる。したがって、離散データは発生率の順に並び替えることで、拡張が可能となる。図-3 (b) のように溪

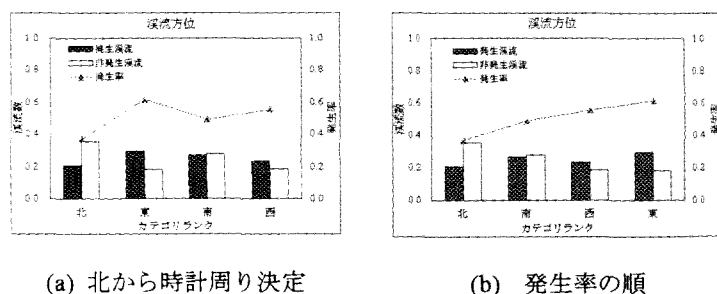


図-3 溪流方位のカテゴリ分布

流方位を発生率の順に並び替えると東向きの溪流が最も危険となる。

5. 要因選定

Rule はデータの持つ規則を縮約した知識である為、規則を持たないデータから有効な Rule を得ることは困難である。土石流発生・非発生データの持つ規則を考えると、「カテゴリランクの増加に伴う発生溪流数の増加」は発生と非発生を分離する上で最も有効な規則であり、拡張に伴う精度低下の抑制に有効であると推測される。

したがって本研究では、整合度の要求水準（95%）を満足する要因の組合せをラフ集合により抽出し、その中で、カテゴリの増加と共に発生溪流数及び、発生率が増加する要因（図-4 参照）を多く含む組合せを選定した。最適な要因の組合せの例は表-2 に示すとおりである（整合度 95.1%）。

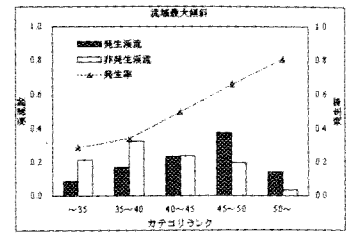


図-4 要因の例
流域最大傾斜

6. 土石流発生・非発生 Rule の抽出

前章で選定した要因を用いた拡張 Rule の抽出を行うと、223 個の Rule（発生 90、非発生 133）が抽出できた。Rule の有効性を示す為、次式で定義される評価指標を算出し、拡張前の Rule との比較を行った。

$$\text{確信度} = \frac{\text{Ruleに適合する整合データの溪流数}}{\text{Ruleに適合する溪流数}} \quad (\text{Ruleの精度})$$

$$\text{サポート} = \frac{\text{Ruleに適合する溪流数}}{\text{全溪流数}} \quad (\text{Ruleの汎用性})$$

表-1 要求水準を満足する Rule 数、及び最良 Rule

	要求水準		拡張無し		拡張有り	
	確信度	サポート	Rule数	最良Rule 確信度 サポート	Rule数	最良Rule 確信度 サポート
発生Rule	0.7	0.02	0	-	6	0.719 0.056
非発生Rule	0.9	0.05	2	1	45	0.934 0.319

表-1 より、確信度とサポートの要求水準を満足する Rule は、拡張前の 2 個に対し、拡張後では 51 個と増加した。また、拡張前では有効な発生 Rule が存在しなかったが、拡張後では確信度 70%、サポート 6%の Rule を抽出できた。また、非発生 Rule に関しては、サポートが約 5 倍に増加した。したがって、Rule 領域の拡張はサポートの向上と発生 Rule の抽出に有効であると考えられる。

また、更に多くの溪流のサポートを目的とし、表-1 の要求水準を満たす非発生 Rule、発生 Rule のサポート上位 5 つの Rule を対象とした組合せ土石流 Rule（表-2）を設定したところ、88.5%の確信度で全溪流の 59.8%（339 溪流）を特定することができた（図-5 参照）。非発生 Rule で特定可能な溪流は、時間雨量 57 mm/h 以上の激しい豪雨でも土石流の発生確率が極めて低く、土石流に対して強い地形であると考えられる。また、発生 Rule に関しては、流域最大傾斜の重要度が高いなど、定性的な見解と一致しており、0 次谷の数、溪流の地質などの要因が抽出され長崎で発生した土石流の特徴をよく捉えた Rule といえる。また、実効雨量が Rule に関与しない原因（全て*）は、S57.7 長崎災害が集中豪雨による災害であった為と推察できる。

表-2 組合せ土石流 Rule

	地形・地質要因					降雨要因		判定	適合データ	整合データ	矛盾データ	残りの溪流	確信度	サポート
	流域長	流域方位	0次谷の数	流域最大傾斜	地質小分類	時間雨量	実効雨量							
												567		
Rule69	2以下	*	*	3以下	*	*	*	非発生	181	169	12	386	0.934	0.319
Rule56	*	2以下	1	*	*	2以下	*	非発生	217	202	15	350	0.931	0.383
Rule98	3以下	*	*	1	*	4以下	*	非発生	232	216	16	335	0.931	0.409
Rule118	2以下	*	*	4以下	3以下	*	*	非発生	240	224	16	327	0.933	0.423
Rule116	*	2以下	*	*	1	1	*	非発生	245	229	16	322	0.935	0.432
Rule196	4以上	3以上	3以上	3以上	*	*	*	発生	277	252	25	290	0.910	0.489
Rule223	3以上	2以上	*	3以上	6以上	*	*	発生	322	285	37	245	0.885	0.568
Rule177	*	*	*	4以上	7以上	4以上	*	発生	329	291	38	238	0.884	0.580
Rule150	*	4以上	*	5以上	*	*	*	発生	334	296	38	233	0.886	0.589
Rule163	*	4以上	3以上	*	*	2以上	*	発生	339	300	39	228	0.885	0.598

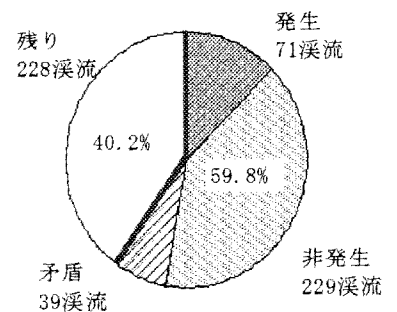


図-5 特定できた溪流の分布

7. 結論

- ラフ集合で抽出した Rule 領域を拡張すること及び、拡張 Rule の組合せはサポートの向上につながる。
- カテゴリの増加と共に発生溪流数及び、発生率が増加する要因の選定は Rule の拡張に有効である。また物理的な順位を持たない離散データは、発生率の順に並び替えることで、拡張が可能となる。
- 設定した土石流 Rule で全 567 溪流の約 6 割の 339 溪流を特定可能であり、調査溪流数の絞込みや、対策施工サイトの優先順位などの防災対策に役立つと考えられる。