

## ナীবベイズ分類による気象害種別の判定手法の開発

## Identifying the types of forest damage by meteorological factors using the Naive Bayes classifier

鈴木 覚<sup>\*1</sup>・吉武 孝<sup>\*1</sup>・高橋正義<sup>\*1</sup>・伊藤香里<sup>\*2</sup>・平浪浩二<sup>\*2</sup>・糸川結花<sup>\*2</sup>・佐藤宏行<sup>\*2</sup>・宮澤國明<sup>\*2</sup>  
Satoru SUZUKI<sup>\*1</sup>, Takashi YOSHITAKE<sup>\*2</sup>, Masayoshi TAKAHASHI<sup>\*2</sup>, Kaori ITO<sup>\*2</sup>, Koji HIRANAMI<sup>\*2</sup>, Yuka ITOKAWA<sup>\*2</sup>,  
Hiroyuki SATO<sup>\*2</sup>, Kuniaki MIYAZAWA<sup>\*2</sup>

\* 1 森林総合研究所

Forestry and forest products research institute, Tsukuba 305-8687

\* 2 森林保険センター

Forestry insurance center, Kawasaki 212-0013

**要旨:** 森林気象害は被災から被害発見までに時間的なずれが大きく、かつ被害発見時には被害をもたらした気象条件を知る術がない。こうした状況のため、被害原因が特定しにくい場合が生じる。そこで、ナীবベイズ分類で被害状況から被害原因を特定する手法を開発した。既存の文献および森林保険センターの損害調査報告書類から抽出した被害形態、土壌、地形、林況等の被害状況を表すデータセットをつくり、それを訓練データとした。風害、雪害、寒風害、干害は正答率が8割を超えたが、凍害、霜害、潮害は正答率が低かった。凍害と霜害は寒風害に誤分類されたものが多かったことから、その原因を地形条件で検討した。凍害や霜害でありながら寒風害に特徴的な地形因子をもったデータセットが誤分類されていたことから、凍害・霜害と寒風害の区別が可能な新たな選択肢を追加する必要があると考えられた。潮害についてはデータセット数が少ないことが誤分類の原因と考えられた。

**キーワード:** ナীবベイズ・風害・雪害・凍害・寒風害

**Abstract:** Usually, there is a large time lag between the occurrence and the discovery of forest damage by meteorological factors, and there is no method to know the weather conditions at the time and place of the occurrence of damage. We developed a technique to identify damage types from such situations using the naive Bayes classifier. Data sets that contain damage situations, soil, terrain, and vegetation information were prepared from previous articles and damage reports collected by the Forestry Insurance Center. Training data for the naive Bayes classifier were obtained from the data sets. As a result of cross-validation, more than 80% of data sets for wind, snow, cold wind, and drought damages were correctly classified. On the contrary, lower percentage of data sets for salt (sea water) damages were correctly classified. More than 20% of data sets for freeze and frost damages were misclassified in cold wind damage. We investigated the reason for the misclassification of freeze and frost damages. The data set for freeze and frost damages with topographic factors typically seen in cold wind damage was misclassified. It is considered that we need to prepare additional choices of words to separate damages correctly.

**Key-word:** Naive Bayes, Wind damage, Snow damage, Freeze damage, Cold wind damage

### はじめに

森林気象害は被災から被害の発見までに時間的なずれが大きという特徴がある。そのため、被害原因の特定が困難な場合が見られる。被害はあったが何が原因かわからないという状況は、研究においては森林気象害研究のボトルネックとなってきたものであり、森林管理においては管理技術の向上を阻むものである。また、森林保険においては保険請求手続きに困難さをもたらすものである。さらに、今後の再生林の増加に伴って幼齡林が増

加すると、過去に頻発した災害が再び増加すると予想されるが、世代交代が進んで森林管理技術の伝承が不十分な現場では、被害原因の特定が一層困難になる事態も危惧される。

一方、(国研)森林研究・整備機構森林保険センターでは災害が発生して保険請求がなされたとき、保険センターの担当者が現地調査を行い、被害の発生状況を確認する。このときに使われる紙媒体によるデータ収集をデジタル化する事によって、手続きの簡素化、データとして

のさらなる活用が可能になると考えられる。

本研究では、現場における被害状況把握の支援、保険センターの調査業務における現行様式の代替および調査の支援を目的に、被害状況を記録するとともに被害状況から被害種別が特定できるシステムを開発した。

**1. ナイーブベイズ分類** ベイズの定理によれば、被害状況を表すデータセット ( $DS$ ) が、ある災害種別 ( $Dist$ ) に分類される確率 ( $P(Dist|DS)$ ) は(1)式で表される。

$$P(Dist|DS) = \frac{P(Dist)P(DS|Dist)}{P(DS)} \quad \dots (1)$$

ここで、 $P(Dist)$  は災害種別が発生する確率、 $P(DS|Dist)$  はある災害種別のときに  $DS$  が生起する確率である。 $P(DS)$  はデータセットが生起する確率であるが、どの災害を対象としても同じなので無視し、(2)の関係で示される。

$$P(Dist|DS) \propto P(Dist)P(DS|Dist) \quad \dots (2)$$

さらに、ナイーブベイズでは  $P(DS|Dist)$  の計算において、 $DS$  を構成するある単語を  $word_i$  としたとき、(3)式で示される。

$$P(DS|Dist) = \prod_i P(word_i|Dist) \quad \dots (3)$$

また、(2)式右辺で得られる値はもはや確率値ではないが、(4)式で確率値とした。

$$P(Dist|DS) = \frac{P(Dist)P(DS|Dist)}{\sum_i P(Dist_i)P(DS|Dist_i)} \quad \dots (4)$$

被害種別ごとに確率値が計算されるが、事後確率が最大となる被害種別をシステムの出力結果とした。

**2. 訓練データ** 本研究で分類対象とした被害種別は森林保険センターにおける対象災害である風害、雪害、凍害、霜害、寒風害、干害、潮害であった。既存の論文や報告書を主体に情報を抽出して訓練データとしたが、特に干害は既存の論文や報告書がわずしか見つからなかったため、森林保険センターにおいて保険業務で使用している「損害調査報告書類」を主体とした。損害調査報告書類は保険契約者が保険請求する際に被害の状況や被害種別の判定根拠を示す添付資料であり、現場写真、被害概況の記述、標準地法による被害率算定データ、被

害発生日とその時の気象データ等が記載されている。ただし、これらデータのうち被害発生日とその時の気象データは訓練データに使用しなかった。その理由は、冬季にアクセスできないなどの理由で被害発生日と発見日が大きくずれることが珍しくなく、過去の気象データから被害発生日が推定される場合があるからである。また、アメダスデータが気象データとして使われるのが通常であるが、被害発生日とアメダス観測所とは地形や標高が異なるのが一般的であり、たとえ発生日が正しかったとしても気象データが被害発生日の状況を表しているとは限らない。このように、客観的に「確実であること」を目安にデータを取捨選択して訓練データとした。

収集したデータ項目は被害形態、土壌、地形、林況、その他に関する合計 30 カテゴリ、238 の選択肢からなっていた(表-1)。立木密度と樹高のデータがある場合は、収量比数と幹材積を推定した。収量比数と幹材積は森林総合研究所「収量比数  $R_y$  計算プログラム」(<https://www.ffpri.affrc.go.jp/database/yieldindex/index.html> 2016年10月3日参照)を使用した。収集したデータセットは凍害と干害が 120 データセット以上あったが、それ以外は 60 データセット以下で、潮害は文献が少なく 10 データセットにとどまった(表-2)。合計で 428 の被害事例を訓練データとした。各データセットを構成するデータ数は論文や報告書をメインにした風害が多く、損害調査報告書類をメインとした干害は少ない傾向にあった(表-2)。

**3. 精度評価** 判定精度を検証するため、Leave-one-out 法によるクロスバリデーションを行った。すなわち、データセットのある 1 データセットを判定対象データとし、残りの 427 データを訓練データとして正しい判定が行われた比率を計算した。判定対象データを順次入れ替える事によって 428 回の判定を行い、正答率で精度を検証した。

## 結果と考察

**1. 正答率** 正答率の最も高かったのは風害であり、風害、雪害、寒風害、干害は 8 割以上の正答率であった。凍害と霜害の正答率は他種別への誤答率よりも高かったが、それぞれ 55%、41%にとどまっていた。誤答先は凍害の 20%が寒風害、16%が干害であった。霜害は 34%が寒風害に誤分類された。正答率が最低だったのは潮害であり、10%の正答率にとどまり、40%は凍害に誤分類された。

**2. 誤分類の原因検討** 凍害と霜害の正答率が低く、寒風害に誤分類される傾向がみられた原因を検討した。

表 - 1 収集したデータのカテゴリとカテゴリを構成するデータの種類や階級数

Table1 Categories of collected data and constituting contents and number of classes for the categories

カテゴリ	選択肢の内容
被害形態	被害形態 (根返り等 6 種類), 枯れの状況 (全体, 部分等), 被害痕跡 (凍傷痕有無等), 発生月, 発見月, 腐朽有無, 被圧木・偏奇木
土壌	土性 (壤土等 6 種類), 土壌分類 (BA 等 7 種類), マサ土
地形	斜面方位 (8 方位), 斜面上の位置 (上中下), 谷, 尾根等 (6 種類), 斜面傾斜 (4 階級), 標高 (7 階級)
林況	樹高 (7 階級), Ry (4 階級), 胸高直径 (3 階級), 幹材積合計 (5 階級), 樹種 (11 種類), 人天別, 齢級 (10 階級), 形状比 (4 階級), 針広混交, 単層・複層別
その他	都道府県, 間伐時期 (6 種類), 道路・作業道隣接, 地位級 (5 階級), 汀線距離 (2 種類)

表 - 2 被害種別ごとのデータセット数およびデータセットあたりの平均データ数

Table2 Number of data set for damage types and average number of data per data set

	風害	雪害	凍害	霜害	寒風害	干害	潮害
データセット数	54	28	128	29	55	125	10
データセット中の平均データ数	19.8	16.1	15.2	13.7	14.2	11.8	9.3

表 - 3 分類先ごとのデータセット数および割合 (上段はデータセット数, 下段の括弧書きは割合)

Table 3 Number and ratio of classified data set for damage types (upper: number, lower: ratio)

	分類結果						
	風害	雪害	凍害	霜害	寒風害	干害	潮害
風害	52 (0.96)	0 (0.00)	1 (0.02)	1 (0.02)	0 (0.00)	0 (0.00)	0 (0.00)
雪害	4 (0.14)	23 (0.82)	0 (0.00)	0 (0.00)	1 (0.04)	0 (0.00)	0 (0.00)
凍害	0 (0.00)	2 (0.02)	71 (0.55)	9 (0.07)	25 (0.20)	21 (0.16)	0 (0.00)
霜害	0 (0.00)	0 (0.00)	4 (0.14)	12 (0.41)	10 (0.34)	3 (0.10)	0 (0.00)
寒風害	0 (0.00)	0 (0.00)	3 (0.05)	0 (0.00)	47 (0.85)	5 (0.09)	0 (0.00)
干害	2 (0.02)	0 (0.00)	8 (0.06)	0 (0.00)	14 (0.11)	100 (0.80)	1 (0.01)
潮害	0 (0.00)	0 (0.00)	4 (0.40)	2 (0.20)	0 (0.00)	3 (0.30)	1 (0.10)

注: 左上から右下への対角線上のデータが正答であり, それ以外は誤分類されたデータセットである。

Note: The data on the diagonal from the upper left to the lower right is the correct answer, and the others are misclassified data sets.

ここで, 被害種別の違いが表れやすいと考えられる斜面方位, 地形の記述, 斜面傾斜と比較した。斜面方位は 8 方位に分けられており, 寒風害は 55 データセット中の 53 データセット (以下 53/55 と表記する), 凍害は 107/128, 霜害は 27/29 に斜面方位に関するデータが含まれていた。寒風害は北向き斜面, 霜害は南東向きが卓

越し, 凍害ははっきりした傾向は見られなかった (図 - 1)。地形の記述は, 寒風害は 32/55, 凍害は 92/128, 霜害は 22/29 にみられた。寒風害が尾根筋, 凍害は凹地・窪地, 緩傾斜, 平坦部, 霜害は凹地・窪地や平坦部に多かった (図 - 2)。また, 被害地の斜面傾斜は寒風害が 38/55, 凍害が 81/128, 霜害が 23/29 に記述が見られた。

寒風害は「中」が、凍害と霜害は「緩」が最も多かった。以上の結果は、寒風害が北向き尾根部、凍害と霜害が凹地や緩傾斜、平坦部に多いという一般的傾向とよく一致した(1, 2, 3)。したがって、平均的傾向として凍害・霜害と寒風害には一般的傾向に矛盾しない違いがデータセット間にあった可能性が高いと考えられた。

次に、誤分類されたデータセットを検討した。凍害のデータセットで寒風害に誤分類されたものは25データセットあったが(表3), そのうち8データセットは寒風害である確率が9割以上として誤分類された。寒風害の特徴を表す平均的条件は、上述したように斜面方位が北西・北向き、尾根部、中傾斜であった。これら3条件の8データセットへの出現数は、3条件のうち1条件を満たしたのが2データセット、2条件を満たしたのが5データセット、3条件を満たしたものが1データセットであった。同じく、霜害でありながら90%以上の確率で寒風害であると誤分類されたものは4データセットあり、寒風害に特徴的な3条件のうち2条件を満たしたものが2データセット、3条件とも満たしたものが2データセットであった。このように、高い確率で誤分類されたデータのほとんどは2条件以上を満たしていたことから、凍害や霜害であるが、寒風害に似た被害状況であったものが寒風害に誤分類されたと考えられた。

(1)式で明らかのように、 $P(Dist)$ と事後確率は比例関係にある。潮害はデータセット数が最も少なかったため(表-2), 正答率が小さかったと考えられた。

### おわりに

本手法では一部の被害種別を除いて十分な判定精度が得られた。誤分類が多かった凍害、霜害は、寒風害の被害発生状況に類似したデータセットが誤分類されていた。したがって、誤分類を減らすにはデータ数を増やすのは必ずしも効果的でなく、凍害・霜害と寒風害の間に違いがあって、互いの区別が可能な選択肢を追加する必要があると考えられた。潮害はデータセット数が少なかったことが正答率の低さの原因と考えられ、被害事例を収集し訓練データに追加する必要があると考えられた。地形条件における寒風害、凍害、霜害の被害傾向が明瞭に見られたように(図-1, 図-2, 図-3), 訓練データは被害データベースとしても利用できることがわかる。すなわち、本手法を保険請求手続きに組み入れれば、被害データの収集システムになりえることを示している。保険請求は年に数百件~数千件あるので、被害状況の収集システムとして、気象害研究や保険業務に重要なインフラになる可能性がある。

### 引用文献

- (1) 堀内孝雄・宮内 宏(1963)寒風防除基礎調査. 茨城県林試昭和 38 年度業報: 45-56 .
- (2) 酒井 昭(1963)寒風害のおこるしくみとその防除法. 北方林業 15: 381-221 .
- (3) 笹沼たつ・橋本武雄(1970)阿武隈山系におけるスギの凍害と地形. 日林誌 2: 283-288 .

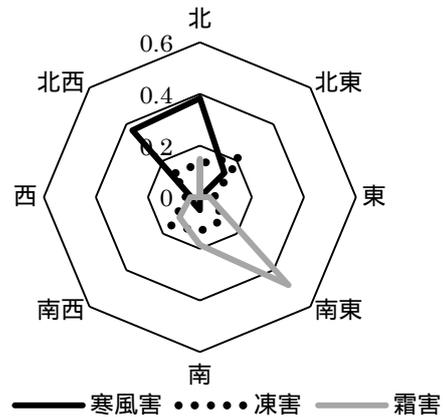


図-1 寒風害、凍害、霜害における被害が発生した斜面方位  
Fig.1 Azimuth of slope that cold wind, frozen, frost damage were occurred

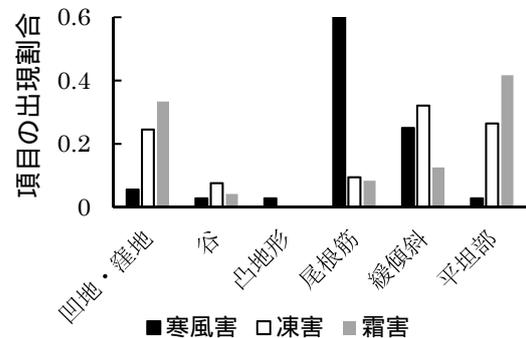


図-2 寒風害、凍害、霜害における地形の記述の出現割合  
Fig.2 Geological factor for the place that cold wind, frozen, frost damage were occurred

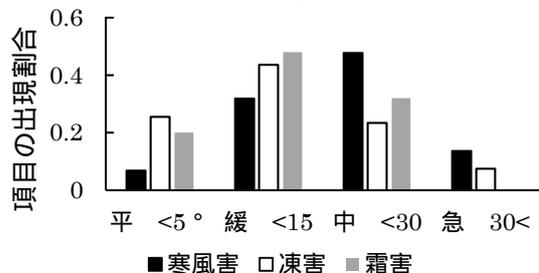


図-3 寒風害、凍害、霜害における被害が発生した斜面の傾斜量の出現割合  
Fig.3 Magnitude of slope that cold wind, frozen, frost damage were occurred.