

# 機械学習を用いた打音による容器内水量の推定

○大蔵侑生（電気通信大学） 董晨宇（電気通信大学） 滝澤優（電気通信大学）  
工藤俊亮（電気通信大学） 末廣尚士（電気通信大学）

## 1. 緒言

これまで料理ロボットに関する研究は多く行われてきたが、そのほとんどがカメラから得た視覚情報を用いており、音情報を取り扱った研究は視覚情報を扱った研究に比べて少ない。しかし、例えば餃子のように焼く音が調理のポイントとなる料理も存在し、音情報も料理において重要である。また音は状態推定に有用な情報であり、缶詰の打検では打音から良品及び不良品の分類を行っている。他にも、メロンの果肉硬度の推定 [1] や樹木の内部欠損の推定 [2] を打音から行う研究も行われてきた。これは打音から内部状態の差異を識別することが可能であることを示しており、様々な応用が考えられる。ほかにも音情報を用いた容器内の状態推定として、液体の注ぎタスクにおける注ぎ先容器内での液体の高さの推定の研究 [3] や液体の重量、溢水の有無の推定を行う研究 [4] が行われてきた。

本研究では、容器内の液体の量を打音から推定する手法を提案する。経験的に、容器の側面を叩いたときの打音は、液体の量や叩く位置によって変化することがわかっているため、そこから液体量を推定できると期待できる。

以上を踏まえ、本研究では液体が入った容器の側面を叩いたときの打音から液体量を推定することを目的とする。この目的のため、容器の側面を叩いたときの打音から水面に対する叩き位置を分類する SVM を作成する。分類する叩き位置のクラスは図 1 に示すように上部、下部、水面の 3 クラスに設定し、この SVM でちょうど水面と判定される位置を求め、その位置を推定水量とする。また、SVM を用いた推定手法を複数考え、どのように推定すればより良い結果が得られるかを検証する。

## 2. 手法

容器の側面を叩いたときにその叩き位置を判定する SVM を応用して、次に示す手順で容器内の水量を推定することを考えた。

1. ペットボトルの高さを  $n$  等分し、叩き位置を複数

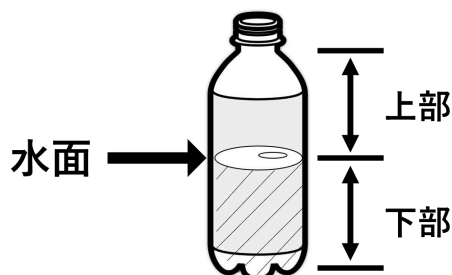


図1 容器の叩き位置の3クラス分類

設ける。図2に、 $n = 10$  としたときの図を示す。 $n = 10$  であれば叩き位置は1～9の9段階で、推定される水量は0～10の11段階になる。

2. 現在の叩き位置  $h$  を真ん中（図2では5の位置）として、その叩き位置  $h$  の打音を SVM で分類する。
3. SVM で水面と判定された場合、その叩き位置  $h$  を水量として推定を終了する。
4. SVM で上部と判定された場合は下へ、下部と判定された場合は上へ二分法で叩き位置  $h$  をずらす。下方向にずらすときは  $\lfloor (h + \text{最低叩き位置}) / 2 \rfloor$ 、上方向にずらすときは  $\lceil (h + \text{最高叩き位置}) / 2 \rceil$  を次の叩き位置  $h$  とする。なお、最低叩き位置で下部と判定された時は推定水量を最低水量である0とし、最高叩き位置で上部と判定された時は推定水量を最高水量（ $n = 10$  の場合 10）として推定を終了する。
5. 分類回数が4回目に達した場合、水面判定が生じずにループとなっている（例えば叩き位置  $h$  の打音では下部と判定され、 $h+1$  の打音では上部と判定されている）ため、分類回数3回目のときの叩き位置と現在の叩き位置  $h$  の平均値を推定水量として推定を終了する。
6. 新しい叩き位置  $h$  の打音を SVM で分類し、手順3に戻る。

また、この推定方法を実装するためには次の3通りの SVM の実装方法が考えられる。それらをすべて実装し、液体量推定に適した方法はどれか検証する。

### 叩き位置基準 SVM

それぞれの叩き位置（ $n-1$  パターン）の場合についてそれぞれの水量（ $n+1$  パターン）まで水を入れて叩いたときの打音を学習させ、 $n-1$  個の SVM を作成する。

### 水量基準 SVM

最低水量と最高水量を除くそれぞれの水量（ $n-1$  パターン）まで水を入れた場合についてそれぞれの叩き位置（ $n-1$  パターン）で叩いたときの打音を学習させ、 $n+1$  個の SVM を作成する。

なお、水量基準 SVM を使用して学習時と同一の条件で推定するためには水量を事前に分かっている必要があるが、当然推定前には未知であり、同一条件の SVM を用いて推定することはできない。そのため前述の叩き位置を基準とした手順で水量の推定を行うことになる。

### 全打音学習 SVM

それぞれの叩き位置（ $n-1$  パターン）の場合についてそれぞれの水量（ $n+1$  パターン）まで水を入

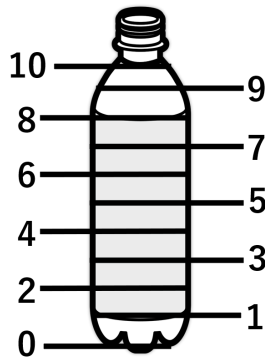


図2 高さを10等分したペットボトル

れて叩いたときの打音を全てひとまとめに学習させ、1個のSVMを作成する。

### 3. 実験

本研究では、ペットボトルの側面を叩いた打音を次の条件で録音、学習してその精度を検証した。

**ペットボトル** 500mlの炭酸飲料のペットボトルに水道水を入れ、蓋を開けた状態で叩く。図2のようにペットボトルの底から曲面部分と飲み口直下の垂直部分の境界線までを10等分し、水量を0から10、叩き位置を1から9のそれぞれに設定する。これらの組み合わせの全99パターンについて20音ずつ、計1980音を録音し、15音ずつを学習用に、5音ずつを評価用に用いる。

**叩き動作** 打検棒を使用し、水平にペットボトルを叩く。ペットボトルは台に置き、叩く面と逆側を軽く手で抑え固定する。なお1, 2の位置を叩くときは反動で床を叩いてしまうことがあるためペットボトルを持ち上げて録音する。

**録音** マイクから15cm程度の場所で叩き作業を行いUSBコンデンサマイクで録音する。設定は、モノラル、サンプリングレート44.1kHz、ビット深度16bitとする。またノイズ処理等は行わない。

**FFT** 録音した音声は $2^{16} = 65536$ 点でFFT処理する（予め録音音声を65536点（ $\approx 1.486$ 秒）で切り出しておく）。窓関数は用いず、全ての点を最大値で割り正規化する。

**SVM** 実装に当たり、Python3の機械学習ライブラリであるscikit-learnを使用した。FFT点数の半分である32768次元のスペクトルをSVMに学習させる。学習させるスペクトルのラベル付けは、図2で水量<叩き位置のときの打音では「上部」、水量=叩き位置のときの打音では「水面」、水量>叩き位置のときの打音では「下部」となる。カーネル関数はRBFガウシアンとし、複数の手法を同条件で比較するためにパラメータは変更しない。

なお、テスト打音は各水量、叩き位置の組み合わせについて5音ずつ存在するが、1から5のインデックス

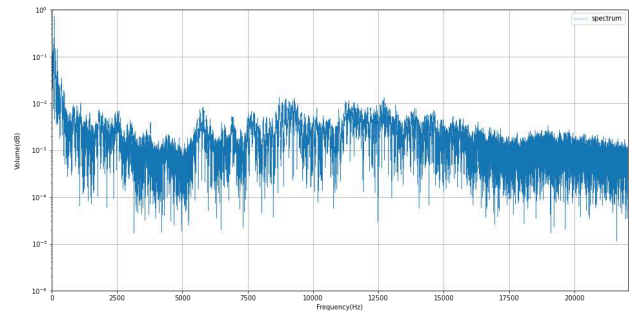


図3 水量0, 叩き位置9のときの打音スペクトル

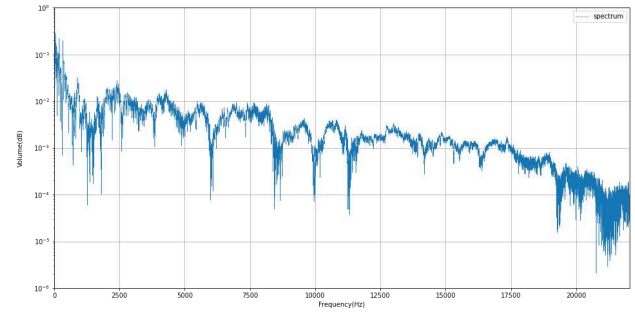


図4 水量5, 叩き位置5のときの打音スペクトル

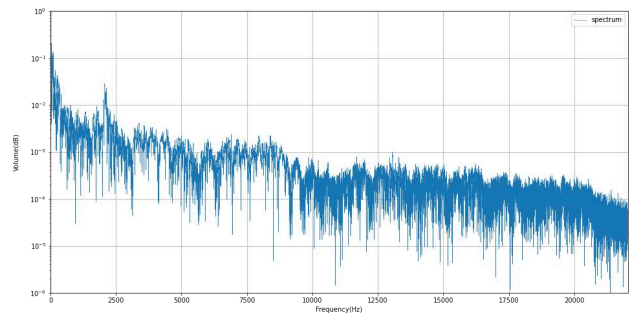


図5 水量10, 叩き位置1のときの打音スペクトル

スを付け、最初に叩き位置5、インデックス*i*の打音を用いたときはその推定では他の叩き位置についても全てインデックス*i*の打音を使用する。

録音データのスペクトルの典型的な例として、図3に水量0, 叩く位置9の場合（上部の打音）、図4に水量5, 叩く位置5の場合（水面の打音）、図5に水量10, 叩く位置1の場合（下部の打音）のスペクトルを示す。比較すると水面、下部のスペクトルは上部のスペクトルに比べて右下がり、水面のスペクトルは上部、下部に比べて分かりやすい谷が現れる、などの特徴が見受けられる。また同じクラスの打音でも、水量と叩き位置の組み合わせ次第で特徴の違いを確認できた。

### 4. 結果

まず、3通りのSVMの精度を示す。表1に叩き位置基準SVMの学習条件（叩き位置*h*で学習したSVMで叩き位置*h*の打音を分類した結果）での評価用打音5音中の分類の正解数を示す。同様に、表2に水量基準SVMの学習条件（水量*a*で学習したSVMでの水量*a*

表1 叩き位置基準 SVM の同一条件での正解数

a	h								
	1	2	3	4	5	6	7	8	9
0	5	5	5	5	5	5	5	5	4
1	5	5	5	5	5	5	5	5	2
2	5	5	5	5	5	5	5	5	5
3	5	5	5	5	5	5	5	5	5
4	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
6	5	5	4	4	5	5	5	5	5
7	5	5	5	5	5	5	5	4	0
8	5	5	5	5	5	5	5	5	4
9	5	5	5	5	5	5	5	5	4
10	5	4	5	5	5	5	5	5	5

表2 水量基準 SVM の同一条件での正解数

h	a										
	0	1	2	3	4	5	6	7	8	9	10
9	2	5	5	5	5	5	5	5	5	0	0
8	5	5	5	5	5	5	5	5	2	4	0
7	5	5	5	5	5	5	4	5	5	5	1
6	5	5	5	5	5	0	3	2	5	5	4
5	5	5	5	5	5	5	4	2	5	5	1
4	5	5	5	1	1	5	5	5	5	5	0
3	5	5	5	5	5	5	5	5	5	5	0
2	5	5	5	5	5	5	5	5	5	5	4
1	5	5	4	5	5	5	5	5	5	5	3

の打音の分類結果)での分類に正解数を,表3に全打音学習 SVM での打音分類の正解数を示す.なお,水量基準 SVM では水量が0または10のときは該当する SVM を構成できない(打音が全て上部または全て下部となりクラス分類 SVM が作成できない)ため,それぞれ近い水量の水量1基準の SVM,水量9基準の SVM による分類結果を示している.

叩き位置基準 SVM では99パターン中90パターンで正解率100%,97パターンが80%以上と高い分類精度が得られた.これは面の傾きなど細かい条件が異なる叩き位置ごとに学習することで細かい特徴を捉えて正確な分類ができたと考えられる.

水量基準 SVM では99パターン中77パターンで正解率100%,85パターンで正解率60%以上となっておりある程度の精度で分類を行えることが分かった.また精度が低くなるのは叩き位置と水量が近い場合や水量が10のときであることも分かる.

全打音学習 SVM では99パターン中85パターン,60%以上であるのは93パターンであり,精度は高めであるが叩き位置が高くなったときに正解率が低い傾向

表3 全打音学習 SVM での分類正解数

a	h								
	1	2	3	4	5	6	7	8	9
10	5	5	5	5	5	5	5	5	5
9	5	5	5	5	5	5	5	4	1
8	5	5	5	5	5	5	5	0	0
7	5	5	5	5	5	5	5	5	1
6	5	5	5	2	3	5	4	5	5
5	5	5	5	5	5	3	4	5	5
4	5	5	5	1	5	5	5	5	4
3	5	5	5	5	5	5	5	5	5
2	5	5	5	5	5	5	5	5	5
1	5	4	5	5	5	5	5	5	4
0	5	5	5	5	5	5	5	5	5

表4 叩き位置基準 SVM による水量推定結果

i	a										
	0	1	2	3	4	5	6	7	8	9	10
1	0	1	2	3	4	5	6	7	8	9	10
2	0	1	2	3	4	5	6	7	8	10	10
3	0	1	2	3	4	5	6	7	8	9	10
4	0	1	2	3	4	5	6	7	8	9	10
5	0	1	2	3	4	5	6	7	8	9	10

があることが分かる.この原因としては図2から分かる通り叩き位置8は円筒部と曲面位置の境界,9は曲面位置であり,スペクトルが不安定になりやすいためパラメータを変更しない SVM では特徴を吸収しきれないといったことが考えられる.

次に,3通りの SVM を用いた推定結果を示す.表4に叩き位置基準 SVM を用いて水量を推定した結果,表5に水量基準の SVM を用いて水量を推定した結果,表6に全打音学習 SVM を用いて水量を推定した結果を実際の水量と比較した表を示す.なお,iは評価用打音のインデックスを示している.

叩き位置基準 SVM の推定では55パターン中54パターン(=98%)で完全に正解し,残りの1パターンも実際の水量から $\pm 1$ で収まる非常に精度の高い結果となった.

水量基準 SVM の推定では誤差が $\pm 1$ 以内で収まっているのは55パターン中21パターン(=38%)のみ,完全に正解したのはわずか14パターン(=25%)というとても実用的とは言えない結果となった.

全打音学習 SVM の推定では完全に正解したのは55パターン中38パターン(=69%), $\pm 1$ で収まっているのは52パターン(=95%)となった.

これら3種類の SVM を用いた推定について,その推定水量の精度と平均絶対誤差(MAE)をまとめたものを表7に示す.

表5 水量基準 SVM による水量推定結果

$i$	$a$										
	0	1	2	3	4	5	6	7	8	9	10
1	9	8	8	3	8	5	8	7	5	5.5	0
2	9	1	10	5	3	5	8	7	5	5.5	0
3	9	1	8	3	3	5	8	7	9	6.5	0
4	9	1	8	5	3	5	8	0	7.5	6.5	0
5	9	9	8	8	3	5	3	0	8	0	0

表6 全打音学習 SVM による水量推定結果

$i$	$a$										
	0	1	2	3	4	5	6	7	8	9	10
1	0	1	2	3	3.5	5	6	7	10	10	10
2	0	1	2	3	3.5	5	5	7	9	10	10
3	0	1	2	3	3.5	5	7	7	9	10	10
4	0	1	2	3	3.5	5	6	7	10	9	10
5	0	2	2	3	4	5	5	7	7.5	7.5	10

最も精度が低いのは水量基準 SVM による推定手法であり、推定水量が実際の水量の  $\pm 1$  以内に収まったものが半分にも達せず、また絶対値誤差の平均が 3.86 と非常に大きくとても実用的とは言えない。誤差の大きさについては、表5からも分かる通り少ない水量 (0 ~ 2) のときに非常に多く判定されてしまったり、逆に水量 9, 10 で空 (0) と判定されてしまったりしていることが原因である。しかし、水量基準の SVM で考え得る他の手法をいくつか試したもののこれ以上精度を上げることができなかった。また叩き位置を更に細かく求めたり、500ml サイズより大きいペットボトルを用いたりすることで等分が 10 以上になると、その分だけ作成する SVM の数が増えてしまうデメリットも存在している。以上より水量推定のために水量基準で学習した SVM を用いることのメリットは無いと言える。

最も精度が高いのは叩き位置基準 SVM による推定手法であり、ほぼ全てで正しく推定できた。しかし水量基準 SVM と同様に、等分を増やしただけ作成する SVM も増やす必要がある。一方で、全打音学習 SVM では等分を変えても作成する SVM は 1 つのみで済み、精度についても完全正解こそ 2/3 程度であるもののほぼ全てで推定水量が実際の水量の  $\pm 1$  以内で収まっており、表6から分かるように  $\pm 2$  を超える誤差は発生せず平均絶対値誤差は少ない。そのため高い精度を求めるのであれば高さ基準 SVM による水量推定が、1 つの分類器で十分な精度で推定したいのであれば全学習 SVM による水量推定が向いていると考えられる。

## 5. 結言

本研究では、容器内の液体の量を打音から推定する手法を提案した。水が入ったペットボトルの側面を叩いたときの打音から水面に対してどの位置を叩いたのかを判別する SVM を作成し、SVM を用いて水量を推

表7 水量推定手法別の精度

SVM	$\pm 1$ 以内	完全正解	MAE
叩き位置基準	55(100%)	54(98%)	0.02
水量基準	21(38%)	14(25%)	3.86
全打音学習	52(95%)	38(69%)	0.31

定する手法を 3 種類実装、比較してどの手法が優れているかを検証した。その結果、高い精度を求めるのであれば叩き位置を基準に学習した SVM、精度を大きく失わずに分類器の数を増やしたくなければ全打音学習 SVM が優れているという知見が得られた。

将来課題としては、本研究は側面の凹凸がほぼ一定な 500ml の炭酸ペットボトルにおける推定結果を示したのみで、明確な角がある四角いペットボトルなどでは同様の推定が行えるかは分からないという問題点がある。同様に、容積の違うペットボトル、またペットボトル以外の容器について推定結果がどのようなになるかも検討を要する課題である。

## 参 考 文 献

- [1] 林節男, 杉山純一, 乙部和紀, 菊池佑二, 臼井支朗: “打音によるメロンの熟度の非破壊評価”, 日本食品工業学会誌, vol. 39, no. 6, pp. 465-470, 1992.
- [2] 白石洋一, 岡田佑樹, 吉澤光三: “ウェーブレット解析を用いた、打音による樹幹内部の欠陥推定方法”, 日本森林学会誌, vol. 4, no. 90, pp. 223-231, 2008.
- [3] Hongzhuo Liang, Shuang Li, Xiaojian Ma, Norman Hendrich, Timo Gerkmann, Fuchun Sun, Jianwei Zhang: Making Sense of Audio Vibration for Liquid Height Estimation in Robotic Pouring IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 5333-5339
- [4] Justin Wilson, Auston Sterling, Ming C. Lin: Analyzing Liquid Pouring Sequences via Audio-Visual Neural Networks, IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), 2019, pp. 7702-7709