Document Type Recognition Using Evidence Theory

Thierry Géraud\textsuperscript{1}, Geoffroy Fouquier\textsuperscript{1}, Quôc Peyrot\textsuperscript{1}, Nicolas Lucas\textsuperscript{2}, Franck Signorile\textsuperscript{2}

\textsuperscript{1} EPITA Research and Development Laboratory, France

www.lrde.epita.fr

\textsuperscript{2} SWT, France

www.swt.fr
**Who’s who?**

EPITA Research and Development Laboratory:
- software engineering,
- scientific computing in C++, meta-programming
- image processing, pattern recognition.

SWT:
- French company, editor of the “b-Wize” software product line
  "solutions to sort, index, read, retrieve and process contents from paper sources"
- winner of the European IST Prize 2003

http://www.ist-prize.org/
Outline

- introduction — context and intentions
- a running example
- first solutions:
  - Boolean logic approach
  - fuzzy approach
- evidence theory:
  - basics
  - modeling
  - comparative results
- conclusion and perspectives
Document type recognition:

- document types are known — a type database/knowledge base exists
- type = set of characteristics
- a characteristic can be featured by several document types
- evaluation “characteristic c / document d” \( \Rightarrow \) value \( \in [0, 1] \)
  - 1 means “d does feature c”
  - 0 means “d does not feature c”
  - 0.5 means “d more or less features c”

Example of characteristics:

- a flower-shaped logo is on top-left (W)
- document font is 12pt (F)
- there is a bar code somewhere (B)
- etc.
Within this context, we do not explain:

- how to build such a knowledge base
- how to select relevant features
- how to valuate couples such as “a characteristic / a document”.

We focus on how to handle information to proceed to document type recognition.

Keywords:

- information management
  - fusion
  - imprecision
- decision
  - uncertainty
  - conflict

Evidence theory is not new but is not well-known → let us be didactic...
Running example

<table>
<thead>
<tr>
<th>characteristics</th>
<th>document types</th>
<th>documents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>type 1 ($t_1$)</td>
<td>type 2 ($t_2$)</td>
</tr>
<tr>
<td>flower logo ($W$)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>12pt fonts ($F_{1^2}$)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>bar code ($B$)</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

This example is simple enough to be quickly solvable by a human.

Real applications are far more complicated:

- many characteristics,
- many document types,
- most of the characteristics are featured by several document types
### Boolean logic

<table>
<thead>
<tr>
<th></th>
<th>type 1 $(t_1)$</th>
<th>type 2 $(t_2)$</th>
<th>type 3 $(t_3)$</th>
<th>case 0 $(d_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>flower logo $(W)$</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>12pt fonts $(F_{12})$</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>bar code $(B)$</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
</tbody>
</table>

**Notation:**

1. $1_{t_i}(d) =$ “$d$ has type $t_i$”
2. $1_{t_i}(c_j) =$ “$c_j$ is a characteristic of $t_i$”
3. $1_{c_j}(d) =$ “$d$ features $c_j$”

$$1_{t_i}(d) = \bigwedge_j \left( 1_{t_i}(c_j) = 1_{c_j}(d) \right)$$

**Example:**

$1_{t_2}(d_0)$ is false since $d_0$ and $t_2$ does not perfectly match.
Main drawbacks:
- decisions are taken too early
- errors are propagated

No proper way to:
- handle imprecision
- measure ambiguity

Definitions:
- *Imprecision*: lack of precise knowledge (syn. inaccuracy).
- *Uncertainty*: incomplete knowledge.
- *Vagueness*: lack of clearness in contours or limits.
- *Fusion*: mixing several pieces of information.

Fuzzy approaches:
- well suited to model these notions
- decision is taken at the very end.
Fuzzy set theory

\( D \): set of documents
\( S_i \subset D \): fuzzy subset of \( D \)
\( d \in D \): a document

\( \mu_{S_i}(d) \in [0, 1] \): membership degree
\( \bigcup_i S_i = D \Rightarrow \sum_i \mu_{S_i}(d) = 1 \)
(normalization)

Fuzzy sets derived from characteristics:

\[
\begin{align*}
W &= W_{yes} \cup W_{no} \\
F &= F_{12} \cup \overline{F_{12}} \\
B &= B_{yes} \cup B_{no}
\end{align*}
\]
where \( \overline{F_{12}} = F_{<12} \cup F_{>12} \)

\[
\begin{align*}
scheme 1: \\
D &= W \times F \times B \Rightarrow t_1 = W_{yes} \times F_{12} \times B_{no} \\
or \\
scheme 2: \\
D &= W = F = B \Rightarrow t_1 = W_{yes} \cap F_{12} \cap B_{no}
\end{align*}
\]

denoting \( c_i^j \) the subset of (characteristic) \( c^j \) corresponding to \( t_i \):

when either \( t_i = \bigcap_j c_i^j \) or \( t_i = \bigcup_j c_i^j \), we have:

\( \mu_{t_i}(d) = \min_j \mu_{c_i^j}(d) \).
Fuzzy fusion

Generalization with a fuzzy fusion operator:

\[ \mu_{t_i}(d) = \bigoplus_j \mu_{c_{i}^{j}}(d) \]

\(\bigoplus\) can be conjunctive:

- "deciding to assign \(d\) to \(t_i\) means that we simultaneously well recognize every features \(c_{i}^{j}\) in document \(d\)"
- Conjunctive operators are T-norms and verify \(\bigoplus \leq \min\).

\(\bigoplus\) can be a compromise:

- "deciding to assign \(d\) to \(t_i\) means that we globally well recognize all features \(c_{i}^{j}\) in the document \(d\)"
- Compromise operators are means and verify \(\min < \bigoplus < \max\) (between T-norms and T-conorms).
Fuzzy decision

Decision function:

\[ \omega(d) = \arg \max_i \mu_{t_i}(d) \]

2\textsuperscript{nd} best decision:

\[ \omega_2(d) = \arg \max_{i \neq \omega(d)} \mu_{t_i}(d) \]

No decision is taken when:

- confidence is too low, i.e. \( \mu_{t_{\omega}(d)} < h_1 \)
- ambiguity is noticed, i.e.
  \[ \mu_{t_{\omega}(d)} - \mu_{t_{\omega_2}(d)} < h_2 \]

or

\[ \frac{\mu_{t_{\omega}(d)}}{\mu_{t_{\omega_2}(d)}} < h_3. \]
Fuzzy fusion results

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$\mu_{\text{yes}}(d_1)$</th>
<th>$\mu_{\text{yes}}(d_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>0.1 → no</td>
<td>0.2 → no</td>
</tr>
<tr>
<td>$F_{12}$</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>0.8 → yes</td>
<td>0.7 → yes</td>
</tr>
<tr>
<td>$B$</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>0.7 → yes</td>
<td>0.5 → ?</td>
</tr>
</tbody>
</table>

intuitive results → $t_3$ → $t_3$

---

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_1$</td>
<td>min</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>0.22</td>
<td>0.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_2$</td>
<td>min</td>
<td>0.20</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>0.28</td>
<td>0.32</td>
</tr>
</tbody>
</table>

where $\mu$ is the normalized arithmetical mean.
When $\oplus$ is conjunctive, false estimations of feature presence can lead to false results;

$\oplus$ should be a compromise but then a lot of false ambiguities appear...
Main problem:

- different types can have several characteristics in common;
- until now, each document type is handled separately;
- actually we valuate singletons...

A simple illustration:

- set of people = \{ Greg, Jack, Tom \}
- statement = “I can’t remember who’s the biggest fool but I’m positive that it’s either Greg or Tom.”

  - Fuzzy modeling = $0.5 / \text{Greg} + 0.5 / \text{Tom} + 0 / \text{Jack}$
  - Drawback = 0.5 for Greg means “half a fool”
  - Proper translation = $1 / (\text{Greg or Tom}) + 0 / \text{Jack}$. 
Evidence theory

Hypothesis set: $\Theta = \{ t_1, \ldots, t_n \}$.

Mass function:

\[
\begin{aligned}
&\forall A \subset \Theta, \; m(A) \in [0, 1] \\
&\sum_{A \subset \Theta} m(A) = 1 \quad A \subset \Theta \text{ is a focal element if } m(A) \neq 0. \\
&m(\emptyset) = 0.
\end{aligned}
\]

Several functions $A \subset \Theta \rightarrow [0, 1]$ are defined.

Belief function (amount of evidence which implies $A$):

\[
bel(A) = \sum_{B \subset A} m(B).
\]

Uncertainty about $A$:

Interval $[bel(A), pls(A)]$

Ignorance: $ign(A) = pls(A) - bel(A)$.

Plausibility function (amount of evidence that does not refute $A$):

\[
pls(A) = 1 - bel(\overline{A}) = \sum_{B \cap A \neq \emptyset} m(B).
\]

Doubt about $A$ (amount of evidence that does refute $A$):

\[
dou(A) = bel(\overline{A}).
\]
Measure of conflict between \( s \) sources \( (m_i, i = 1..s) \):

\[
K = \sum_{\cap_{i=1}^s B_i = \emptyset} \left( \prod_{i=1}^s m_i(B_i) \right).
\]

Mass combination (Dempster’s rule):

\[
\left( \bigoplus_{i=1}^s m_i \right)(A) = \frac{1}{1 - K} \sum_{\cap_{i=1}^s B_i = A} \left( \prod_{i=1}^s m_i(B_i) \right).
\]

Property:

\[
m = \bigoplus_{i=1}^s m_i \text{ is a mass.}
\]

Finally, we compute from \( m \):

\[
\forall i, \text{ } bel(\{t_i\}) \text{ and } pls(\{t_i\}).
\]
Decision rules

1. maximum of belief:

\[ \omega_{bel}(d) = \arg \max_i bel(\{t_i\})(d) \]

2. maximum of plausibility:

\[ \omega_{pls}(d) = \arg \max_i pls(\{t_i\})(d) \]

3. absolute decision rule = maximum of belief without overlapping of belief intervals:

\[ \omega_{abs}(d) = \omega_{bel}(d) \text{ if } \forall i \neq \omega_{bel}(d), \text{ } pls(\{t_i\})(d) < bel(\{t_{\omega_{bel}(d)}\})(d) \]

4. compromise = maximum of \((bel + pls)/2\):

\[ \omega_{cpm}(d) = \arg \max_i \frac{bel + pls}{2}(\{t_i\})(d) \]
### Evidence modeling

With global uncertainty:

<table>
<thead>
<tr>
<th></th>
<th>type 1 ((t_1))</th>
<th>type 2 ((t_2))</th>
<th>type 3 ((t_3))</th>
<th>focal elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>flower logo ((W))</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>(m_W({t_1}))</td>
</tr>
<tr>
<td>12pt fonts ((F_{12}))</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>(m_{F_{12}}({t_1, t_3}))</td>
</tr>
<tr>
<td>bar code ((B))</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>(m_B({t_2, t_3}))</td>
</tr>
</tbody>
</table>

\((*)\) this means: “according to \(W\), when it is not \(t_1\), it is either \(t_1\), \(t_2\), or \(t_3\)”;
we then have: \(m_W(\Theta) = 1 - m_W(\{t_1\})\).

Fusion step:

\[
m_u = m_W \oplus m_{F_{12}} \oplus m_B.
\]

Without global uncertainty:

e.g., \(m_{\neg W}(\{t_1\}) = m_W(\{t_1\})\) and \(m_{\neg W}(\Theta - \{t_1\}) = m_W(\Theta)\)
means: “according to \(W\), when it is not \(t_1\), it is either \(t_2\) or \(t_3\)”.

---

Fifth IAPR International Workshop on Graphics Recognition, Barcelona, Spain, 2003. – p.18/22
Three different approaches ⇒ results having three different flavors.

\[ d_1 \]

<table>
<thead>
<tr>
<th></th>
<th>( {t_1} )</th>
<th>( {t_2} )</th>
<th>( {t_3} )</th>
<th>( {t_1, t_3} )</th>
<th>( {t_2, t_3} )</th>
<th>( {t_1, t_2, t_3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_u )</td>
<td>0.03</td>
<td>0.00</td>
<td>0.54</td>
<td>0.23</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>( m_{\phi} )</td>
<td>0.04</td>
<td>0.19</td>
<td>0.77</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \mu )</td>
<td>0.22</td>
<td>0.33</td>
<td>0.44</td>
<td>undef</td>
<td>undef</td>
<td>undef</td>
</tr>
</tbody>
</table>

\[ d_2 \]

<table>
<thead>
<tr>
<th></th>
<th>( {t_1} )</th>
<th>( {t_2} )</th>
<th>( {t_3} )</th>
<th>( {t_1, t_3} )</th>
<th>( {t_2, t_3} )</th>
<th>( {t_1, t_2, t_3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_u )</td>
<td>0.11</td>
<td>0.00</td>
<td>0.31</td>
<td>0.31</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>( m_{\phi} )</td>
<td>0.15</td>
<td>0.26</td>
<td>0.60</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \mu )</td>
<td>0.28</td>
<td>0.32</td>
<td>0.40</td>
<td>undef</td>
<td>undef</td>
<td>undef</td>
</tr>
</tbody>
</table>
Comparison “fuzzy / evidence” (decision = compromise)

\[
\begin{array}{|l|c|c|c|}
\hline
 & \{t_1\} & \{t_2\} & \{t_3\} \\
\hline
bel & 0.03 & 0.00 & 0.54 \\
pls & 0.32 & 0.19 & 0.97 \\
evidence & 0.18 & 0.10 & 0.75 \\
fuzzy & 0.22 & 0.33 & 0.44 \\
\hline
\end{array}
\]

\[
\begin{array}{|l|c|c|c|}
\hline
 & \{t_1\} & \{t_2\} & \{t_3\} \\
\hline
bel & 0.11 & 0.00 & 0.31 \\
pls & 0.56 & 0.27 & 0.89 \\
evidence & 0.33 & 0.13 & 0.60 \\
fuzzy & 0.28 & 0.32 & 0.40 \\
\hline
\end{array}
\]
Conclusion

Evidence theory:

- is well-suited to handle both imprecision and uncertainty in document type recognition;
- allows to describe document types by (fuzzy) characteristics.

Effective application:

- several thousand documents to be processed;
- about one hundred different document types;
- quasi-perfect recognition results.
Implementation

Materials:

- we provide free software libraries under the GNU PUBLIC LICENSE (GPL)
- downloadable from www.lrde.epita.fr

- Mathematical Theory of Evidence project eVidenZ
- Image Processing and Pattern Recognition project Olena

Thanks for your attention; any questions?