Partial Word Representation

F. Blanchet-Sadri

Fields Institute Workshop

This material is based upon work supported by the National Science Foundation under Grant No. DMS–1060775.
Partial words and compatibility

- A partial word is a sequence that may have undefined positions, called holes and denoted by $\diamond$’s, that match any letter of the alphabet $A$ over which it is defined (a full word is a partial word without holes); we also say that $\diamond$ is compatible with each $a \in A$.

  \[a \diamond b \diamond a a b\] is a partial word with two holes over \{a, b\}

- Two partial words $w$ and $w'$ of equal length are compatible, denoted by $w \uparrow w'$, if $w[i] = w'[i]$ whenever $w[i], w'[i] \in A$.

  \[
  \begin{align*}
    a & \diamond b \diamond a \\
    \uparrow & \\
    \diamond & \diamond b a a a
  \end{align*}
  \quad
  \begin{align*}
    a & \diamond b \diamond a \\
    \uparrow & \\
    \diamond & \diamond a a a a
  \end{align*}
  \]
Factor and subword

- A partial word $u$ is a **factor** of the partial word $w$ if $u$ is a block of consecutive symbols of $w$.

  $\diamond a\diamond$ is a factor of $aa\diamond a\diamond b$

- A full word $u$ is a **subword** of the partial word $w$ if it is compatible with a factor of $w$.

  $aaa$, $aab$, $baa$, $bab$ are the subwords of $aa\diamond a\diamond b$
  corresponding to the factor $\diamond a\diamond$
Some computational problems

- We define $\text{REP}$, or the problem of deciding whether a set $S$ of words of length $n$ is representable, i.e., whether $S = \text{sub}_w(n)$ for some integer $n$ and partial word $w$.

- If $h$ is a non-negative integer, we also define $h\text{-REP}$, or the problem of deciding whether $S$ is $h$-representable, i.e., whether $S = \text{sub}_w(n)$ for some integer $n$ and partial word $w$ with exactly $h$ holes.
Rauzy graph of $S = \{aaa, aab, aba, baa, bab\}$

$S$ is 0-representable by $w = aaababaa$
Why partial words? (Compression of representations)

\[ S = \{ aaa, aab, aba, baa, bab \} \]

is representable by

\[ aaababaa \]
Why partial words? (Compression of representations)

\[ S = \{ aaa, aab, aba, baa, bab \} \]

is representable by

\[ \diamond aabab \]
Why partial words? (Compression of representations)

\[ S = \{ \text{aaa, aab, aba, baa, bab} \} \]

is representable by

\[ a \Diamond a \Diamond \]
Why partial words? (Representation of non-0-representable sets)

- **Set $S$ of 30 words of length six:**

<table>
<thead>
<tr>
<th></th>
<th>aaaaaa</th>
<th>6</th>
<th>aabbaa</th>
<th>11</th>
<th>abbbba</th>
<th>16</th>
<th>baabbb</th>
<th>21</th>
<th>bbabab</th>
<th>26</th>
<th>bbbabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>aaaaab</td>
<td>7</td>
<td>aabbbba</td>
<td>12</td>
<td>abbbbab</td>
<td>17</td>
<td>bababbb</td>
<td>22</td>
<td>bbabbb</td>
<td>27</td>
<td>bbbaba</td>
</tr>
<tr>
<td>3</td>
<td>aaabbb</td>
<td>8</td>
<td>aabbbb</td>
<td>13</td>
<td>abbbbb</td>
<td>18</td>
<td>bababa</td>
<td>23</td>
<td>bbbbaa</td>
<td>28</td>
<td>bbbbab</td>
</tr>
<tr>
<td>4</td>
<td>aaabba</td>
<td>9</td>
<td>ababbb</td>
<td>14</td>
<td>abbbbb</td>
<td>19</td>
<td>babbba</td>
<td>24</td>
<td>bbaaab</td>
<td>29</td>
<td>bbbbaa</td>
</tr>
<tr>
<td>5</td>
<td>aaabbb</td>
<td>10</td>
<td>abbaab</td>
<td>15</td>
<td>ababba</td>
<td>20</td>
<td>bbaabb</td>
<td>25</td>
<td>bbabba</td>
<td>30</td>
<td>bbbbbb</td>
</tr>
</tbody>
</table>

- **Rauzy graph $(V, E)$ of $S$, where $E = S$ and $V = \text{sub}_S(5)$ consists of 20 words of length five:**

<table>
<thead>
<tr>
<th></th>
<th>aaaaa</th>
<th>5</th>
<th>aabbb</th>
<th>9</th>
<th>abbb</th>
<th>13</th>
<th>bbaaa</th>
<th>17</th>
<th>bbbba</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>aaaaab</td>
<td>6</td>
<td>ababb</td>
<td>10</td>
<td>baabb</td>
<td>14</td>
<td>bbaab</td>
<td>18</td>
<td>bbbab</td>
</tr>
<tr>
<td>3</td>
<td>aaabb</td>
<td>7</td>
<td>abbaa</td>
<td>11</td>
<td>babab</td>
<td>15</td>
<td>baba</td>
<td>19</td>
<td>bbbba</td>
</tr>
<tr>
<td>4</td>
<td>aabba</td>
<td>8</td>
<td>abbaa</td>
<td>12</td>
<td>babbb</td>
<td>16</td>
<td>bbaab</td>
<td>20</td>
<td>bbbba</td>
</tr>
</tbody>
</table>
4  aaabba
5  aaabbb
   aaabb
4  aaabba
5  aaabbb
   aaabb
Membership of $h$-REP in $\mathcal{P}$

Theorem

$h$-REP is in $\mathcal{P}$ for any fixed non-negative integer $h$.

Membership of REP in $\mathcal{P}$

**Theorem**

REP is in $\mathcal{P}$.

---

F. Blanchet-Sadri and S. Munteanu, Deciding representability of sets of words of equal length in polynomial time. Submitted
Open problems

- Characterize the sets of words that are representable.

- Characterize minimal representing partial words (can they be constructed efficiently?)