APEHR: Automated Prognosis in Electronic Health Records using multi-head self-attention

Alexander YC Florez, <alexanderchf1@gmail.com>
(speaker)
Lucas Scabora, Danilo M Eler, Jose F Rodrigues-Jr
Presentation outline

1. Introduction and motivation
2. Problem definition
3. Proposed methodology
4. Experiments and results
5. Conclusion
Introduction and motivation

- Researchers have begun to focus on Clinical Decision Support Systems;
- The number of applications of DL in healthcare has increased in the last years.
Introduction and motivation

- Transformer was proposed in the NLP area, it avoids the use of recurrence and convolution;
- The use of Transformer has the potential for a better performance, since it is based on a different principle
Introduction and motivation

- Transformers have been applied successfully in different areas such as NLP, speech recognition, among others. It has initiated new architectures, each with specific benefits for certain domains.
Presentation outline

1. Introduction, motivation, and problem definition
2. Problem definition
3. Proposed methodology
4. Experiments and results
5. Conclusion
Problem definition

- Given a sequence of patient admissions to a hospital, we want to predict the most probable clinical events to which she/he is subject to in the future.
Problem definition

● Problem statement

Tuberculoma of meninges
Problem definition

- Hierarchical structure of medical coding systems
Presentation outline

1. Introduction, motivation, and problem definition
2. Problem definition
3. Proposed methodology
4. Experiments and results
5. Conclusion
Proposed Methodology

- We propose a new architecture that employs only the **Decoder component**, since is sufficient for a one-direction classification task;
- In contrast to recurrent approaches, our method processes the entire clinical history at once.
Proposed Methodology

- Overview
Proposed Methodology

- Overview
Proposed Methodology

- Overview
Proposed Methodology
● Overview - Positional Encoding
Proposed Methodology

- Overview
Proposed Methodology

- Overview
Proposed Methodology

- Overview
Proposed Methodology

● Overview
Presentation outline

1. Introduction and motivation
2. Problem definition
3. Proposed methodology
4. Experiments and results
5. Conclusion
Experiments and results

- **Datasets**

<table>
<thead>
<tr>
<th></th>
<th><strong>Mimic-III</strong></th>
<th><strong>InCor</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
<td>Boston, Massachusetts</td>
<td>Sao Paulo, Brazil</td>
</tr>
<tr>
<td><strong>Encoding</strong></td>
<td>ICD-9/CCS</td>
<td>ICD-10</td>
</tr>
<tr>
<td><strong># admissions</strong></td>
<td>58,976</td>
<td>820,819</td>
</tr>
<tr>
<td><strong># patients</strong></td>
<td>48,520</td>
<td>89,048</td>
</tr>
</tbody>
</table>
Experiments and results

- We perform fine-tuning of the main hyperparameters;
- We also test our method against over state-of-the-art architectures (DoctorAI, LIG-Doctor) on Mimic-III public dataset.
Experiments and results

- Fine-tuning of parameters $A$ and $H_d$

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$A$</th>
<th>$H_d$</th>
<th>R@10</th>
<th>R@20</th>
<th>R@30</th>
<th>P@1</th>
<th>P@2</th>
<th>P@3</th>
<th>AUC-ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIMIC-III-CCS</td>
<td>10</td>
<td>30</td>
<td>0.53</td>
<td>0.70</td>
<td>0.79</td>
<td>0.81</td>
<td>0.78</td>
<td>0.76</td>
<td>0.94</td>
</tr>
<tr>
<td>MIMIC-III-ICD9</td>
<td>8</td>
<td>30</td>
<td>0.45</td>
<td>0.62</td>
<td>0.70</td>
<td>0.78</td>
<td>0.76</td>
<td>0.73</td>
<td>0.95</td>
</tr>
<tr>
<td>InCor</td>
<td>8</td>
<td>30</td>
<td>0.71</td>
<td>0.75</td>
<td>0.77</td>
<td>0.65</td>
<td>0.40</td>
<td>0.25</td>
<td>0.97</td>
</tr>
<tr>
<td>MIMIC-III-CCS</td>
<td>10</td>
<td>30</td>
<td>0.53</td>
<td>0.71</td>
<td>0.79</td>
<td>0.81</td>
<td>0.78</td>
<td>0.75</td>
<td>0.94</td>
</tr>
<tr>
<td>MIMIC-III-ICD9</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>0.61</td>
<td>0.70</td>
<td>0.79</td>
<td>0.76</td>
<td>0.72</td>
<td>0.95</td>
</tr>
<tr>
<td>InCor</td>
<td>10</td>
<td>30</td>
<td>0.70</td>
<td>0.74</td>
<td>0.77</td>
<td>0.64</td>
<td>0.39</td>
<td>0.28</td>
<td>0.97</td>
</tr>
<tr>
<td>MIMIC-III-CCS</td>
<td>10</td>
<td>30</td>
<td>0.53</td>
<td>0.71</td>
<td>0.79</td>
<td>0.81</td>
<td>0.78</td>
<td>0.76</td>
<td>0.94</td>
</tr>
<tr>
<td>MIMIC-III-ICD9</td>
<td>8</td>
<td>30</td>
<td>0.47</td>
<td>0.63</td>
<td>0.71</td>
<td>0.76</td>
<td>0.74</td>
<td>0.72</td>
<td>0.96</td>
</tr>
<tr>
<td>InCor</td>
<td>8</td>
<td>15</td>
<td>0.70</td>
<td>0.74</td>
<td>0.76</td>
<td>0.64</td>
<td>0.39</td>
<td>0.28</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Experiments and results

- Comparison to related work

⇒ Our results were pronouncedly superior to those of former 4 works in the literature
Presentation outline

1. Introduction and motivation
2. Problem definition
3. Proposed methodology
4. Experiments and results
5. Conclusion
Conclusion

• We demonstrate Transformer's adaptability to the diagnostic problem using only diagnostic codes;

• The experiments over InCor dataset verify that our model is feasible in different medical settings;

• Our method surpasses other works by as much as 4% regarding metric Recall@k;

• Future work: extended to more techniques, such as Transformer XL.
Thank you for your attention!

Alexander YC Florez, <alexanderchf1@gmail.com>  
(speaker)  
Lucas Scabora, Danilo M Eler, Jose F Rodrigues-Jr