The right AIs for the rights of citizens

Francesca Lagioia, Marco Lippi, Paolo Torroni
Bologna, Alma AI Seminar, June 17th, 2020
PART III

Methodologies
Machine learning methodologies

• The enactment of consumer-empowering AI requires to address a number of relevant (and foundational) tasks in AI…

• Natural language understanding

• Knowledge representation and reasoning

• Computational linguistics across different languages
Unfair clause detection in Terms of Service

• From a machine learning point of view, unfair clause detection can be seen just as a sentence classification task…

• **Classic algorithms** can be used: SVM with bag-of-words, deep networks (CNN or LSTM), structured approaches (SVM-HMM)

• The best performing system is an ensemble…
Experimental results for Terms of Service

[Lippi et al., AI & Law, 2019]

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Method</th>
<th>$P$</th>
<th>$R$</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>SVM – Single Model</td>
<td>0.729</td>
<td>0.830</td>
<td>0.769</td>
</tr>
<tr>
<td>C2</td>
<td>SVM – Combined Model</td>
<td>0.806</td>
<td>0.779</td>
<td>0.784</td>
</tr>
<tr>
<td>C3</td>
<td>Tree Kernels</td>
<td>0.777</td>
<td>0.718</td>
<td>0.739</td>
</tr>
<tr>
<td>C4</td>
<td>Convolutional Neural Networks</td>
<td>0.729</td>
<td>0.739</td>
<td>0.722</td>
</tr>
<tr>
<td>C5</td>
<td>Long Short-Term Memory Networks</td>
<td>0.696</td>
<td>0.723</td>
<td>0.698</td>
</tr>
<tr>
<td>C6</td>
<td>SVM-HMM – Single Model</td>
<td>0.759</td>
<td>0.778</td>
<td>0.758</td>
</tr>
<tr>
<td>C7</td>
<td>SVM-HMM – Combined Model</td>
<td>0.848</td>
<td>0.720</td>
<td>0.772</td>
</tr>
<tr>
<td>C8</td>
<td>Ensemble (C1+C2+C3+C6+C7)</td>
<td>0.828</td>
<td>0.798</td>
<td>0.806</td>
</tr>
<tr>
<td></td>
<td>Random Baseline</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>Always Positive Baseline</td>
<td>0.123</td>
<td>1.000</td>
<td>0.217</td>
</tr>
</tbody>
</table>

$P = \text{Precision}$, $R = \text{Recall}$, $F_1 = \frac{2PR}{P+R}$
Moving to privacy policies…

What is different?

• Ensure **compliance** with articles 13-14 of GDPR
• Detect **vague** language
• Detect **problematic** clauses for data processing
Can we improve CLAUDETTE with background knowledge?

• So far, we used purely data-driven approaches…

• Legal experts are able to recognize potentially unfair clauses thanks to their background knowledge of the domain

• Typically, when detecting a potentially unfair clause, a legal expert also knows the reason for which the clause is unfair…

• These explanations should be useful also for CLAUDETTE!
Can we improve CLAUDETTE with background knowledge?

- Symbolic vs. sub-symbolic approaches…
- Most recent research directions
  - Neural-symbolic learning
  - Statistical relational learning
  - Connectionist models for reasoning
Memory-Augmented Neural Networks

• Process input and store the information in some memory
• Understand pieces of knowledge relevant to a given query
• Retrieve concepts from memory
• Combine memory and query to make a prediction
Memory-Augmented Neural Networks

KB

Similarity

Content Extraction

Reasoning Module

Yes

Keep reading?

No

opt_out

«the consumer must first opt out for the arbitration not to be obligatory»

domicile

«the arbitration is to take place in country different than the consumer’s domicile»

«By accepting these Terms of Service, you agree to be bound by this arbitration clause and class action waiver»
Memory-Augmented Neural Networks in CLAUDETTTE

Potentially unfair clause #4
EXCEPT FOR CERTAIN TYPES OF DISPUTES MENTIONED IN THE ARBITRATION CLAUSE, YOU AND HEADSPACE AGREE THAT DISPUTES RELATING TO THESE TERMS OR YOUR USE OF THE PRODUCTS WILL BE RESOLVED BY MANDATORY BINDING ARBITRATION, AND YOU WAIVE ANY RIGHT TO PARTICIPATE IN A CLASS-ACTION LAWSUIT OR CLASS-WIDE ARBITRATION.
Unfairness categories: Arbitration
Hide/show rationales

Potentially unfair clause #5
1.4 CHANGES TO TERMS Headspace reserves the right to change or update these Terms, or any other of our policies or practices, at any time, and will notify users by posting such changed or updated Terms on this page.
Unfairness categories: Unilateral Change
Hide/show rationales
The clause is potentially unfair for Unilateral Change since the provider has the right for unilateral change of the contract, services, goods, features for any reason at its full discretion, at any time (score = 0.834)

Potentially unfair clause #6
Your continued use of the Products constitutes your agreement to abide by the Terms as changed.
Unfairness categories: Contract by Using
Hide/show rationales

http://claudette.eui.eu/demo
The challenge of multilingualism

• Both Terms of Service and Privacy Policies are available in (almost) each of the languages of the European Union

• Can we develop a CLAUDETTE for each language?

• Do we need to start from scratch for any language?
The challenge of multilingualism

• Can we use English tagging for other languages?
• Can we automatically compare a document in two languages?
• Do we need competence for each and every language?
• Is it easier (or more difficult) to detect potential unfair clauses in some of the languages, maybe due to lexicon or style?
Two main possible approaches
1/2 - Use just English version of Claudette
Two main possible approaches
2/2 - Retrain a German version of Claudette

- ENGLISH CORPUS (GOLDEN LABELS)
- GERMAN CORPUS (NOISY LABELS)

INPUT PRIVACY POLICY (GERMAN VERSION)

CLAUDETT SYSTEM (German version)
Two main possible approaches
Pros and cons

• System #1: **only one** (English) version

• System #2: **multiple systems** to maintain

• System #1: **only one** golden corpus

• System #2: need (advantage?) to **retrain** a corpus

• System #1: processing **noisy** documents

• System #2: using **noisy** labels
Looking to the future

• Which **kind of AIs** do consumers need?

• **Explainability** vs. performance accuracy…

• Capability to digest and exploit **background knowledge**

• Impact on civil society and **cooperation** among consumers

• **Multilingualism** can be an **incentive to cooperate**…