HeadlineCause: A Dataset of News Headlines for Detecting Causalities
Ilya Gusev and Alexey Tikhonov

Abstract
Detecting implicit causal relations in texts is a task that requires both common sense and world knowledge. Existing datasets are focused either on commonsense causal reasoning or explicit causal relations. In this work, we present HeadlineCause, a dataset for detecting implicit causal relations between pairs of news headlines. The dataset includes over 5000 headline pairs from English news and over 9000 headline pairs from Russian news labeled through crowdsourcing. The pairs vary from totally unrelated or belonging to the same general topic to the ones including causation and refutation relations. We also present a set of models and experiments that demonstrates the dataset validity, including a multilingual XLM-RoBERTa based model for causality detection and a GPT-2 based model for possible effects prediction.

Examples
A: Exclusive: NextVR acquired by Apple (Updated)  
B: Apple Buys Virtual Reality Company NextVR  
Label: same
A: Oklahoma spent $2 million on malaria drug touted by Trump  
B: Gov. Kevin Stitt defends $2 million purchase of malaria drug  
Label: first causes second
A: Report: Microsoft acquiring Microvision  
B: Microsoft denies MicroVision acquisition  
Label: second refutes first
A: Meizu 17 teaser poster confirms 64MP Sony IMX686 quad-camera setup  
B: Meizu 17 Pro has Super wireless mCharge support  
Label: other relation

Definitions
Same headlines. A and B are about the same things, but can differ in minor details.
Causality. A causes B if B is impossible without A. If A did not happen, then B must not be happening too.
Refutation. B refutes A if B makes A irrelevant.

Pairs sampling
Heuristic filters:
- A presence of a hyperlink between two documents
- An affiliation of documents to the same website
- A cosine distance between LaBSE embeddings with a threshold
- A presence of different locations in headlines

Aggregation
- Majority vote: 7 votes or more
- Two modes: full and simple
- Removed pairs with cause and effect reversed in time

Classification models

<table>
<thead>
<tr>
<th>Test type and description</th>
<th>English</th>
<th>FR</th>
<th>Russian</th>
<th>English</th>
<th>FR</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV: Adding types</td>
<td>3.5%</td>
<td>2.9%</td>
<td>4.9%</td>
<td>3.9%</td>
<td>2.6%</td>
<td>3.9%</td>
</tr>
<tr>
<td>INV: Swapping order of not causal pairs</td>
<td>2.8%</td>
<td>2.0%</td>
<td>3.1%</td>
<td>2.8%</td>
<td>2.0%</td>
<td>3.1%</td>
</tr>
<tr>
<td>DIR: Swapping order of causal pairs</td>
<td>22.0%</td>
<td>12.2%</td>
<td>22.9%</td>
<td>13.5%</td>
<td>22.2%</td>
<td>14.1%</td>
</tr>
<tr>
<td>MFT: Refutations with different locations</td>
<td>9.5%</td>
<td>2.9%</td>
<td>10.1%</td>
<td>10.5%</td>
<td>3.2%</td>
<td>10.9%</td>
</tr>
</tbody>
</table>

GPT-2 experiment

The Michigan capital clash with armed protesters
- The Michigan capital clash with armed protesters
- Govt defends lockdown in Michigan state capital
- State capital protesters demand lockdown of state capital, condemn lockdown.

Table: GPT generation examples

Results
We present a publicly available dataset for detecting implicit causal relations between pairs of news headlines.
Classification and generation models were trained on this dataset.
We believe that it can be successfully used to train causal relation detection models with the subsequent composition of causation graphs

Problems
- Unclear method of candidates sampling
- Majority vote annotation aggregation instead of Dawid-Skene method
- Poor annotation of refutations
- Disparity between Russian and English datasets