HEXTac: the Creation of a Manual Extractive Run

Pierre-Etienne Genest
Guy Lapalme
Mehdi Yousfi-Monod

RALI, Université de Montréal

November 16 2009
Introduction

What is a manual extractive run?
- Extractive summaries written by humans
- Pure sentence extraction from source documents
- No editing, no re-writing, no compression

Why create a manual extractive run?
- Determine how well/poorly extracts perform compared to abstracts
- Verify if human extracts beat the best current systems
- Provide a model similar to current systems
- Applications in domains which require quotable summaries
Experimental Context of HEXTAC

- 5 human extractors
- 88 extractive summaries (18 each on average)
- 3.88 extracted sentences per summary on average
- 20 minutes spent to complete a summary on average
- 30 man-hours of work in total
Interactive Human Extraction Interface

- Make manipulating sentences during extraction easy and fast
- Guarantee pure sentence extraction and respect of 100-word limit
- Automatize data handling
  - Loading clusters, saving completed summaries
- Keep records of metadata
  - User name, date, time spent, ID of selected sentences
Interactive Human Extraction Interface

User name: Pierre-Etienne

Topic: Indian Pakistan conflict. Describe efforts made toward peace in the India-Pakistan conflict over Kashmir.

D0901A-A

XIN_ENG_20041113.0001, 2004/11/13
Roundup: Kashmiris waiting for festival and peace to come

Markets are overcrowded, traffic jam is heavy and the shops are jostling with shoppers in the capital city of Srinagar in the Indian-administered Kashmir as the holy Moslem festival of Eid approaches here.

Kashmiris are known as incorrigible festive shoppers and because of that reputation, unscrupulous shopkeepers have been mining money by over-charging the locals for everything from a chop of mutton to the chickens and hosiery items that the locals must buy to protect themselves from the biting cold of the winter.

But, around this Eid festival, there is more to the happy public mood than just the urge to over spend during the festival.

Indian Prime Minister Manmohan Singh, who arrives here on a two-day visit next week, announced on Thursday that orders had been given for the reduction of troops inside Kashmir.

This measure has obviously been taken to take care of the building tensions between the Indian army and local people who often come into unpleasant contact during encounters, crackdown operations and search and cordon exercises that have become so routine in Kashmir ever since the present armed struggle against the Indian rule started here 18 years back.

It must also be mentioned that the Indian army on Friday suspended and took into custody an officer who had been accused of having raped a woman and her daughter in a north Kashmir village last week.

It is, therefore, clear that the Prime Minister’s decision has not been influenced by any dramatic improvement in the ground situation.
Guidelines for Human Extractors

1. Begin with part A
2. Always read the topic and all 10 articles
3. Extract sentences that answer the topic and summarize the cluster
4. Favor sentences that can be understood on their own
5. Respect the limit of 100 words
6. Maximize the information content
7. Re-order the sentences of the extract to improve readability
8. Complete part B the same way immediately after part A
9. Avoid repetition of information that appears in cluster A
Feedback from HEXTAC participants

- **Thankful** for the interface
  - It saves time and helps with the task
- **Frustration** at the inability to make modifications
  - Solving referential clarity problems (time, person)
  - Removing a few words
- Difficulty to choose how to answer **list-like topic requests**
- Difficulty with the tradeoff between **content and linguistic quality**
- **Boredom**, repetitiveness of the task
Manual extraction performs better than any automatic system

Pure extraction performs very significantly worse than abstraction

Greatly superior linguistic quality, even with pure extraction

Shows room for improvement in automatic sentence extraction models
Inter-Extractor Agreement

- Based on 12 additional, redundant extractive summaries
- Very low inter-extractor agreement
- Roughly 15% sentence-agreement between human extractors
- Widely varying scores between extractors, though using a small sample

<table>
<thead>
<tr>
<th></th>
<th>Pyramid Score</th>
<th>Linguistic Quality</th>
<th>Overall Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE1</td>
<td>0.278</td>
<td>8.222</td>
<td>7.556</td>
</tr>
<tr>
<td>HE2</td>
<td>0.297</td>
<td>7.611</td>
<td>5.333</td>
</tr>
<tr>
<td>HE3</td>
<td>0.340</td>
<td>7.000</td>
<td>5.917</td>
</tr>
<tr>
<td>HE4</td>
<td>0.378</td>
<td>7.583</td>
<td>7.125</td>
</tr>
<tr>
<td>HE5</td>
<td>0.392</td>
<td>6.063</td>
<td>4.125</td>
</tr>
</tbody>
</table>
**HexTAC as a ROUGE model**

- **HexTAC-ROUGE**
  - ROUGE score of a system using HexTAC as the model

- Similarity with other metrics

<table>
<thead>
<tr>
<th>Correlation coefficients</th>
<th>Part A</th>
<th>Part B</th>
</tr>
</thead>
<tbody>
<tr>
<td>HexTAC-ROUGE–ROUGE</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>HexTAC-ROUGE–Overall Responsiveness</td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td>ROUGE–Overall Responsiveness</td>
<td>0.97</td>
<td>0.94</td>
</tr>
</tbody>
</table>

- A less costly alternative to other evaluation metrics?
  - Made from only one manual ”run”
  - Extracts easier to make than abstracts
Conclusion

HexTac

- Successful, reusable methodology to manual extraction
- Requires 30 man-hours for 88 extracts using an interactive interface
- Approximation of an upper-bound on purely extractive summarization
  - Better extracts/extractors definitely exist

Food for thought

- A tool for supervised training of sentence selection?
- Manual sentence ranking / sentence evaluation, the next step?
- Necessity of a linguistically richer approach than sentence selection to achieve significant improvements?
Questions? Comments?