CS 6320.002: Natural Language Processing
Fall 2019

Project Milestone 1 – 45 Points
Due 8:30am 04 Nov. 2019

Deliverables: A PDF writeup. You can submit just one per team. Please put the names of ALL members of your team on your writeup.

0 Planning Ahead

After Milestone 1, you have the following milestones:

- Milestone 2, due 1 week after Milestone 1. Finish finding or collecting your training, tuning, development, and test data.
- Milestone 3, due 1.5 weeks after Milestone 2. Implement a baseline from an existing paper and evaluate it on your test data.
- Milestone 4, due 2.5 weeks after Milestone 3. Implement your improvements over the baseline and evaluate. Complete your final project paper.
- Presentation, also 2.5 weeks after Milestone 3.

Each milestone is worth 8% of your final grade; the proposal and presentation are worth 4% each.

You don't have to (and shouldn't!) wait for the milestone due dates to work on each stage of your project! Keep working steadily so that you aren't rushing at the end of the semester. There are no more homeworks in this class, so set aside the time you would have spent each week doing homework to work on your project instead.

1 Instructions

Do the following for **each** of the 3 (or more) papers you are using for this milestone. You can use the same papers you put in the project proposal, or you can use other papers you found since then.

1.1 Citation – 1 point per paper

Give the full citation for the paper. For example,

> Orasan and Chiorean (2008). Evaluation of a Cross-lingual Romanian-English Multi-document Summariser. In *Proceedings of the 6th International Conference on Language Resources and Evaluation.*

1.2 Task – 2 points per paper

Write 1-2 sentences describing the task addressed by the paper. Explain how it relates to your project and to other work in the area. Continuing to use cross-lingual summarization

for low-resource languages as an example,

> Orasan and Chiorean (2008) experimented with cross-lingual summarization, using a relatively low-resource language, Romanian, as the source language, and English as the target language. In contrast, most work on cross-lingual summarization uses high-resource languages for both the source and target languages – most commonly English and Chinese.

1.3 Data – 2 points per paper

Write 1-2 sentences describing the dataset used in the paper. For example,

> Orasan and Chiorean collected a dataset of Romanian news articles, published between 2001 and 2005. The articles covered five topics relating to major events in Romania that were widely reported in Romanian news media, but were largely unknown outside the country. They collected 50 documents per topic, where each document consisted of the first 10,000 characters of a news article (truncated for length).

1.4 Approach – 5 points per paper

Write 1-2 paragraphs describing the approach used in the paper. What learning paradigm was used? What learning algorithm? What features or neural architecture? Etc. For example,

> Orasan and Chiorean first performed extractive multi-document summarization for each of the five topics using the 50 Romanian documents for that topic. They used the Maximal Marginal Relevance method, which maximizes selected sentences' similarity to the topic while minimizing redundancy among selected sentences. The MMR algorithm iteratively scores candidate sentences based on similarity and redundancy, then selects the highest-scoring sentence, recalculates the scores for the remaining candidates, and so on until a predifined length limit is reached.
>
> They then translated the extracted summaries into English using a freely-available online translation system, eTranslator. They chose this system because it was the only free Romanian-to-English translator at the time. They used the system off-the-shelf, with no modifications or additional training.

1.5 Evaluation – 5 points per paper

Write 1-2 paragraphs describing the evaluation setup and results. What was the evaluation metric? Was it an intrinsic or extrinsic evaluation? What baselines did they compare against? For example,

> Orasan and Chiorean used a task-based evaluation in which human judges were asked to read the generated summaries and answer a series of questions. The questions were designed so that they could be answered using information

from the full document; a good summary should also allow all of the questions to be answered. The judges performed the evaluation twice, once with the extracted Romanian summaries, and once with the translated English summaries, in order to capture the effect of translation on summary quality. The human judges were also asked to rate the coherence of the summaries on a scale from 1 to 5.

Orasan and Chiorean compared with two baseline systems. The first used an information retrieval system to rank the documents for each topic in order of relevance, then extracted the first sentence from each document until the length limit was reached. The second was a set of human-written summaries. They found that, as expected, human-written Romanian summaries were both the most coherent and allowed the judges to answer the most questions correctly; the first-sentence baseline performed the worst, with the MMR system in the middle. On the translated English summaries, the percentage of questions answered correctly dropped from 57% on the Romanian summaries to just 43%, and feedback from the human judges indicated that translation errors made it difficult to figure out what the summaries were about.