Supervisor's Review of Diploma Thesis

title of the thesis:

Combining Text-based and Vision-Based Semantics

author:

Giang Binh Tran

---

1. *Task specification*

The submitted thesis focuses on automatic semantic similarity measurement with the support of unsupervised statistical vector space models. High quality similarity measurement becomes more and more important in many applied fields, which has been a good motivation for this work.

The goal specific for this thesis was to use and integrate both vision-based and text-based semantics to create a multimodal semantic space from large amounts of images and texts, in order to improve measurement of semantic similarity. Student's task was to use both text and visual corpora, to extract bags of "visual words" from images that share some topic characteristics or themes, and then to build a multimodal semantic model. To evaluate the emerging multimodal semantic space, the model should applied to tasks such as measuring word similarity or concept clustering. The results should be compared with the traditional approach without using visual data.

2. Thesis structure

The submitted thesis is well and clearly structured. Its structure, containing 6 chapters, is quite standard. In Introduction (Chap. 1) the author explains the scope and objectives of his work. In the second chapter he reviews related literature. Then Chapter 3 provides some background of image processing and advanced approaches in describing visual features extracted from images. The core of student's work is presented in next two chapters. In Chapter 4 he describes the high-level architecture of his proposed framework for building a multimodal semantic model. Then in Chapter 5 he concentrates on evaluation tasks. Finally, Chapter 6 concludes the results and analysis, and gives some discussion about future work.

The list of bibliography has 81 items. Delivered programs and data developed by the author are attached to the thesis on CD. Unfortunately contents of the CD has not been documented in the text at all.

To consider the quality of the submitted work as a whole, crucial part is Chapter 5, which describes experiments, parameter tuning, and evaluation of the proposed multimodal

semantic model. Obviously the author is well aware of the fact that without careful evaluation he could not make any serious statement about the proposed model. Therefore he performs an extensive series of experiments and evaluates his semantic model using a number of tests and reference datasets. He compares the proposed model with a text-based model, namely Distributional Memory, which is considered to be one of the state-of-the-art models for semantic tasks. Moreover, the results obtained from "standard" experiments (using widely used reference datasets) are also compared with a number of other recently published models.

3. Author's contributions

Contributions are summarized in chapter 1.2 and in Conclusion, namely:

- proposed framework for creating multimodal learning system;
- proposed strategy for representing concept meaning based on suitable visual features extracted from images;
- proposed algorithms for combining vision-based and text-based semantic models;
- empirical findings about the potential of information extracted from images: visual features can capture semantic relations among words and can enhance state-of-the-art text-based semantic models; there is some evidence that vision-based features are complementary to text-based ones ("the image-based semantic model tends to capture semantic relations among concrete concepts while the text-based semantic model tends to capture semantic relations among more abstract concepts" (ch. 6.1)).

In my view, the main contribution of the submitted thesis is the fact that the author has proposed and developed "the first framework to integrate state-of-the-art text-based semantic models and vision-based semantic models" in order to create a multimodal semantic model; "The framework is designed as an open prototype for further studies/analysis in both computer vision and computational linguistics." (ch. 6.1).

4. Formal aspects

The submitted thesis is written in English. Unfortunately, the language is not perfect, minor grammatical mistakes are not quite rare as well as grammatically incomplete sentences. Also the list of bibliography is a bit neglected, containing a number of inconsistencies. However, all text is well comprehensible.

The submitted thesis fulfills all general formal requirements.

As partial results of his work on the thesis the author has already published 2 papers at scientific conferences/workshops (co-authored by his supervisors and other co-workers).

5. Conclusion

I highly appreciate that the author has successfully contributed to quite a new research area, namely multimodal corpora exploitation, which has recently emerged as a new and promising source of possible progress (not only) in the field of lexical semantics or general computational linguistics.

Although the imperfections mentioned above are certainly not negligible, I consider them quite minor.

I recommend the submitted thesis for the defense and I suggest accepting this work as a diploma thesis.

Prague, September 2011

RNDr. Martin Holub, Ph.D.

Institute of Formal and Applied Linguistics
Charles University in Prague