Shared Task on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid Machine Translation (ML4HMT-2011)

(http://www.dfki.de/ml4hmt/)

Barcelona (Spain) · Saturday, November 19th, 2011

The "Shared Task on Optimising the Division of Labour in Hybrid MT" is an effort to trigger systematic investigation on improving state-of-the-art Hybrid MT, using advanced machine-learning (ML) methodologies. The main focus of the shared task is trying to answer the following question: *Can Hybrid/System Combination MT techniques benefit from extra information (linguistically motivated, decoding and runtime) from the different systems involved?*

Participants of the challenge are requested to build hybrid translations by combining the output of several MT systems of different types. Four participating combination systems, each following a different solution strategy, have been submitted to the shared task. We have computed automated metric scores and conducted an extensive manual evaluation campaign to assess the quality of the hybrid translations. Interestingly, the system winning nearly all the automatic scores only reached a third place in the manual evaluation. Vice versa, the winning system according to manual rankings ranked last place in the automatic metric scores based evaluation. This clearly indicates that more systematic investigation of hybrid system combination approaches, both on a system level and on the evaluation of such systems, needs to be undertaken.

We will work on an updated version of the corpus for the next edition of this shared task, and we will further focus on the integration of advanced machine learning techniques as these are expected to support better exploitation of our corpus' data properties.

We are looking forward to an interesting workshop and want to thank the participants for their efforts during the ML4HMT-2011 Shared Task.

Acknowledgments

This work has been funded under the Seventh Framework Programme for Research and Technological Development of the European Commission through the T4ME contract (grant agreement no.:249119).

We thank the organisers of LIHMT 2011 for their support.

Organisation committee

Toni Badia (Pompeu Fabra University, Spain)

Christian Federmann (German Research Center for Artificial Intelligence, Germany)

Josef van Genabith (Dublin City University, Ireland)

Maite Melero (Barcelona Media Innovation Center, Spain)

Eleftherios Avramadis (German Research Center for Artificial Intelligence, Germany)

Pavel Pecina (Dublin City University, Ireland)

Marta R. Costa-jussà (Barcelona Media Innovation Center, Spain)

Venue

Barcelona Media

Av. Diagonal, 177, 9th floor

Barcelona, Spain

Programme

09:15 Welcome

09:30 Toni Badia (BM) "Introduction to the ML4HMT Shared Task Workshop"

09:40 Patrick Lambert (LIUM) "The MANY System @ML4HMT-2011"

10:30 Tsuyoshi Okita (DCU) "DCU System Combination @ML4HMT-2011"

11:00 Eleftherios Avramidis (DFKI) "DFKI System Combination with sentence ranking @ML4HMT-2011"

11:30 *Coffee break*

12:00 Christian Federmann (DFKI) "DFKI System Combination using Syntactic Information @ML4HMT-2011"

12:30 Christian Federmann (DFKI) "Comparison of overall results @ML4HMT-2011"

12:40 Alon Lavie (CMU) "MEMT: Alignment-based MT System Combination with Linguistic and Statistical Features"

13:10 Discussion Panel chair: Patrick Lambert (LIUM), Alon Lavie (CMU), Cristina España-Bonet (UPC) and Christian Federmann (DFKI). Topics include:

(i) Two Hybrid paradigms: Multi- vs Single-system
(ii) In the Multi-system approach: can Hybrid/System Combination MT techniques benefit from extra information (linguistically motivated, decoding and runtime) from the different systems involved?
(iii) Evaluation in the Multi-system approach: do we evaluate the output in isolation or do we use evaluation information from the different systems involved?

14:00 *Lunch*

The ML4HMT workshop is supported by META-NET

About META-NET

META-NET is a Network of Excellence dedicated to fostering the technological foundations of a multilingual European information society. Language Technologies will:

- enable communication and cooperation across languages,
- secure users of any language equal access to information and knowledge,
- build upon and advance functionalities of networked information technology.

A concerted, substantial, continent-wide effort in language technology research and engineering is needed for realising applications that enable automatic translation, multilingual information and knowledge management and content production across all European languages. This effort will also enhance the development of intuitive language-based interfaces to technology ranging from household electronics, machinery and vehicles to computers and robots.

To this end META-NET is building the Multilingual Europe Technology Alliance (META). Bringing together researchers, commercial technology providers, private and corporate language technology users, language professionals and other information society stakeholders. META will prepare the necessary ambitious joint effort towards furthering language technologies as a means towards realising the vision of a Europe united as one single digital market and information space.

http://www.meta-net.eu/