TECHNISCHE UNIVERSITÄT MÜNCHEN
LEHRSTUHL FÜR STEUERUNGS- UND REGELUNGSTECHNIK
ORDINARIUS: UNIV.-PROF. DR.-ING./UNIV. TOKIO MARTIN BUSS

April 24, 2015

M A S T E R ' S T H E S I S
for
Simon Kaltenbacher
Student ID 10025865, Degree Computer Science (LMU)

An Empirical Approach to Corpora Dimensionality Reduction

Problem description:

It is well known and best practice to *make use as much meaningful training data as possible* in any statistical pattern learning application. However, big mentionable drawbacks in the use of large scale data are cost of retrieval of ground truth labels and processing times, which make many applications in symbolic artificial intelligence not possible in real time. Many large symbolic data applications in the robotics domain can benefit from dimensionality reduction, reducing the information necessary to represent hidden states of motion [1], symbolic OAC streams [2], or natural language sources [3]. The student, given a preliminary and empirical algorithm, will research how to perform text dimensionality reduction, i.e. reducing the size of a given text-based corpus while maintaining the majority of its symbolic learning potential.

Tasks:

- Literature research on symbolic data compression
- Understand the preliminary algorithm, and devise a more efficient and effective selection criteria
- Evaluate the algorithm on large sets of well-known available text-based corpora
- Formalize an information metric which characterizes relative informativeness of text-based corpora
- Transpose and verify the effectiveness of such information characterization in the proto-symbolic domain of hidden states of motion [1] [optional]

Bibliography:

[1] Tetsunari Inamura, Iwaki Toshima, Hiroaki Tanie, and Yoshihiko Nakamura. Embodied symbol emergence based on mimesis theory. *The International Journal of Robotics Research*, 23(4-5):363–377, 2004.

[2] Nicholas H. Kirk, Karinne Ramírez-Amaro, Emmanuel Dean-León, Matteo Saveriano, and Gordon Cheng. Predicting modular robotic plans via probabilistic semantic reasoning on context and sequence. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany*. IEEE RAS, 2015 [submitted].

[3] Wataru Takano and Yoshihiko Nakamura. Integrating whole body motion primitives and natural language for humanoid robots. In *8th IEEE-RAS International Conference on Humanoid Robots*, pages 708–713. IEEE, 2008.

Supervisor:	Nicholas H. Kirk, M.Sc.
Start:	15.04.2015
Intermediate Report:	15.07.2015
Delivery:	15.10.2015

(D. Lee)
Carl-von-Linde Fellow