Proseminar Maschinelles Lernen und Experimentelles Design, UdS, SS11 Hausaufgaben für den 7.7.2011

Caroline Sporleder

July 1, 2011

Aufgabe 1: Literaturverzeichnis Student Wendelin Wissbegierig hat für seine Hausarbeit zum Thema "Genetische Algorithmen" das folgende Literaturverzeichnis abgegeben. Leider hat Wendelin in seinem Proseminar nicht richtig aufgepaßt, daher enthählt das Literaturverzeichnis mehrere formale Fehler. Findet diese! (Achtung: Es geht nur um **formale** Fehler, d.h. ihr braucht z.B. nicht die Korrektheit der Seitenzahlen etc. zu prüfen.)

Literaturverzeichnis

Marina Litvak, Mark Last und Menahem Friedman (2010): *A new Approach to Improving Multilingual Summarization using a Genetic Algorithm*. In: Proc. of ACL-2010, S. 927-936.

Echizen-ya, H., K. Araki und Y. Momouchi: Machine Translation Method Using Inductive Learning with Genetic Algorithms. In: COLING 1996, S. 1020-1023

Raquel Hervás & Pablo Gervás (2005): An Evolutionary Approach to Referring Expression Generation and Aggregation, S. 168-173.

Steven Bird (1994): Automated Tone Transcription, S. 1-12, In: *Proceedings of the First Meeting of the ACL Special Interest Group in Computational Phonology*.

Milen Kouylekov und Matteo Negri (2010): An Open-Source Package for Recognizing Textual Entailment. In: "Proc. of the ACL 2010 System Demonstrations", S. 42-47.

Jinyun Ke, Mieko Ogura und William S.-Y. Wang (2003): Optimizing Models of Sound Systems Using Genetic Algorithms. In: Computational Linguistics 29:1, S. 1-18.

ai-junkie: Genetic Algorithms in Plain English. http://www.ai-junkie.com/ga/intro/gat1.html

Weka. http://www.cs.waikato.ac.nz/ml/weka/

Melanie Mitchell (1998): Introduction to Genetic Algorithms.

Leila Kallel, Bart Naudts und Alex Rogers (Hg.) (2001): Theoretical Aspects of Evolutionary Computing. Berlin: Springer Verlag.

J. E. Rowe (2001): The Dynamical Systems Model of the Simple Genetic Algorithm. In: Theoretical Aspects of Evolutionary Computing.

J. K. Hart (2004): Automatic control program creation using concurrent Evolutionary Computing. PhD Thesis.

Constantin Orăsan (2003): An evolutionary approach for improving the quality of automatic summaries. In: Proc. of the ACL 2003 Workshop on Multilingual Summarization and Question Answering, S. 37-45.

Aufgabe 2: Korrektes Zitieren Im Folgenden findet ihr einige Zitate aus Wendelins Hausarbeit (direkte Zitate, indirekte Zitate, Umschreibungen), jeweils mit der dazugehörigen Passage im Orginal. Leider hat Wendelin auch bei den korrekten Zitiertechniken nicht aufgepaßt. Was ist korrekt, was falsch?

Orginal 1: Table 1 shows that for 3% summaries, the context information has little influence on the number of the discourse ruptures present in a summary. ... The situation changes when longer summaries are considered. As can be observed in Table 1, the continuity principle reduces the number of DR; ...
Wendelin 1: Orăsan (2003) found that context information can reduce the number of discourse ruptures for longer summaries (e.g., 5% of the original text), though not for shorter ones.

Orginal 2: Karamanis and Manurung (2002) used the *continuity principle* in text generation to choose the most coherent text from several produced by their generation system. ... We take a similar approach in our attempt to produce coherent summaries, trying to minimise the number of violations of the principle they contain.
Wendelin 2: Orăsan (2003) take an approach that is similar to Karamanis and Manurung (2002) in their attempt to produce coherent summaries, trying to minimise the number of violations of the principle they contain.

Orginal 3: After investigating our corpus we can definitely say that the *continuity principle* is present in human produced abstracts, and therefore by trying to enforce it in automatic summaries, we might produce better summaries.
Wendelin 3: Orăsan (2003) carried out a corpus study and found that one "can definitely say that the continuity principle is present in human produced abstracts".

Orginal 4: In our corpus almost 75% of the pairs of consecutive utterances (614 out of 835) satisfy the principle.
Wendelin 4: Orăsan (2003, p. 38) found that in their corpus "almost 75% of the pairs of consecutive utterances satisfy the principle".