Finite-State Technology in Natural Language Processing

Andreas Maletti*

Institut for Natural Language Processing, Universität Stuttgart
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
maletti@ims.uni-stuttgart.de

Summary

Finite-state technology is at the core of many standard approaches in natural language processing [1, 2]. However, the terminology and the notations differ significantly between theoretical computer science (TCS) [3] and natural language processing (NLP) [4]. In this lecture, inspired by [2, 4], we plan to illustrate the close ties between formal language theory as discussed in TCS and its use in mainstream applications of NLP. In addition, we will try to match the different terminologies in three example tasks. Overall, this lecture shall serve as an introduction to (i) these tasks and (ii) the use of finite-state technology in NLP and shall encourage closer collaboration between TCS and NLP.

We will start with the task of part-of-speech tagging [2, Chapter 5], in which given a natural language sentence the task is to derive the word category (the part-of-speech, e.g. noun, verb, adjective, etc.) for each occurring word in the sentence. The part-of-speech information is essential for several downstream applications like co-reference resolution [2, Chapter 21] (i.e., detecting which entities in a text refer to the same entities), automatic keyword detection [2, Chapter 22] (i.e., finding relevant terms for a document), and sentiment analysis [5] (i.e., the process of determining whether a text speaks favorably or negatively about a subject). Along the historical development of systems for this task [6] we will discuss the main performance breakthrough (in the mid 80s) that led to the systems that are currently state-of-the-art for this task. This breakthrough was achieved with the help of statistical finite-state systems commonly called *hidden Markov models* [2, Chapter 6], which roughly equate to probabilistic finite-state transducers [7]. We will outline the connection and also demonstrate how various well-known algorithms like the forward and backward algorithms relate to TCS concepts.

Second, we will discuss the task of parsing [2, Chapter 13], in which a sentence is given and its syntactic structure is to be determined. The syntactic structure is beneficial in several applications including syntax-based machine translation [8] or natural language understanding [2, Chapter 18]. In parsing, a major performance breakthrough was obtained in 2005 by adding finite-state information to probabilistic context-free grammars [9]. The currently state-of-the-art models

* Supported by the German Research Foundation (DFG) grant MA / 4959 / 1-1.

(for English) are *probabilistic context-free grammars with latent variables*, which are known as probabilistic finite-state tree automata [10] in TCS. We will review the standard process [11] (expectation maximization), which determines the hidden finite-state information in the hope that similar processes might be helpful also in the TCS community. In addition, we will recall a spectral learning approach [12], which builds on the minimization of nondeterministic field-weighted tree automata [13]. Similarly, advanced evaluation mechanisms like coarse-to-fine parsing [14] that have been developed in NLP should be considered in TCS.

Finally, we will cover an end-user application in NLP. The goal of machine translation [8] is the provision of high-quality and automatic translations of input sentences from one language into another language. The main formalisms used in NLP in this area are *probabilistic synchronous grammars* [15], which originate from the seminal syntax-based translation schemes of [16]. These grammars correspond to certain subclasses of probabilistic finite-state transducers [7] or probabilistic tree transducers [10]. So far, only local versions (grammars without latent variables) are used in state-of-the-art systems, so the effective inclusion of finite-state information remains an open problem in this task. However, the requirements of syntax-based machine translation already spurred a lot of research in TCS because the models traditionally studied had significant shortcomings [17]. In the other direction, advanced models like multi bottom-up tree transducers [18] have made reasonable impact in syntax-based machine translation [19].

References

1. Manning, C., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press (1999)
2. Jurafsky, D., Martin, J.H.: Speech and Language Processing. 2nd edn. Prentice Hall (2008)
3. Droste, M., Kuich, W., Vogler, H., eds.: Handbook of Weighted Automata. Springer (2009)
4. Knight, K., May, J.: Applications of weighted automata in natural language processing. [3] chapter 14 571–596
5. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval **2**(1–2) (2008) 1–135
6. Francis, W.N., Kučera, H., Mackie, A.W.: Frequency Analysis of English Usage: Lexicon and Grammar. Houghton Mifflin (1982)
7. Mohri, M.: Weighted automata algorithms. [3] chapter 6 213–254
8. Koehn, P.: Statistical Machine Translation. Cambridge University Press (2010)
9. Matsuzaki, T., Miyao, Y., Tsujii, J.: Probabilistic CFG with latent annotations. In: Proc. 43rd ACL, Association for Computational Linguistics (2005) 75–82
10. Fülöp, Z., Vogler, H.: Weighted tree automata and tree transducers. [3] chapter 9 313–403
11. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) **39**(1) (1977) 1–38

12. Cohen, S.B., Stratos, K., Collins, M., Foster, D.P., Ungar, L.: Spectral learning of latent-variable PCFGs. In: Proc. 50th ACL, Association for Computational Linguistics (2012) 223–231
13. Bozapalidis, S.: Effective construction of the syntactic algebra of a recognizable series on trees. Acta Informatica **28**(4) (1991) 351–363
14. Petrov, S.: Coarse-to-Fine Natural Language Processing. PhD thesis, University of California at Bekeley, Berkeley, CA, USA (2009)
15. Chiang, D.: An introduction to synchronous grammars. In: Proc. 44th ACL, Association for Computational Linguistics (2006) Part of a tutorial given with Kevin Knight.
16. Aho, A.V., Ullman, J.D.: The Theory of Parsing, Translation, and Compiling. Prentice Hall (1972)
17. Knight, K.: Capturing practical natural language transformations. Machine Translation **21**(2) (2007) 121–133
18. Arnold, A., Dauchet, M.: Morphismes et bimorphismes d'arbres. Theoretical Computer Science **20** (1982) 33–93
19. Braune, F., Seemann, N., Quernheim, D., Maletti, A.: Shallow local multi-bottom-up tree transducers in statistical machine translation. In: Proc. 51st ACL, Association for Computer Linguistics (2013) 811–821