Adam Lopez:
Statistical machine translation.

Abstract

Statistical machine translation (SMT) treats the translation of natural language as a machine learning problem. By examining many samples of human-produced translation, SMT algorithms automatically learn how to translate. SMT has made tremendous strides in less than two decades, and new ideas are constantly introduced. This survey presents a tutorial overview of the state of the art. We describe the context of the current research and then move to a formal problem description and an overview of the main subproblems: translation modeling, parameter estimation, and decoding. Along the way, we present a taxonomy of some different approaches within these areas. We conclude with an overview of evaluation and a discussion of future directions.