Gewichtetes Multiple Sequence Alignment ist MAX-$\mathcal{SNP}$-hart

Bodo Siebert

Medizinische Universität zu Lübeck
Institut für Theoretische Informatik
siebert@tcs.mu-luebeck.de

Zusammenfassung

Multiple Sequence Alignment (MSA) ist eine wichtige Aufgabe der Bioinformatik. Die Berechnung guter Alignments spielt in der Biologie eine wichtige Rolle bei der Rekonstruktion der Evolution anhand phylogenetischer Bäume.

Eine Verallgemeinerung des Problems ist das sogenannte gewichtete MSA, bei dem jedem Paar von Sequenzen ein Gewicht zugeordnet wird. Damit kann bereits bekanntes Wissen über die Verwandtschaft der Sequenzen modelliert werden.

Wir zeigen, dass gewichtetes MSA MAX-$\mathcal{SNP}$-hart ist. Dieses Resultat gilt sogar für den eingeschränkten Fall, dass nur die Gewichte 0 und 1 zugelassen werden (binär gewichtetes MSA). Es wird eine explizite untere Schranke für die Approximierbarkeit dieses Problems angegeben.

Desweiteren wird gezeigt, dass jede untere Schranke für das allgemeine Problem bereits eine untere Schranke für das binär gewichtete MSA darstellt.