Predicting Social Modes of Knowledge Co-construction for Collaborative Learning

Yi-Chia Wang

Language Technologies Institute, School of Computer Science, Carnegie Mellon University

Introduction

Computer-supported collaborative learning (CSCL) mainly aims at engaging students in discourse activities which are related to knowledge acquisition. Analyzing process dimensions of knowledge construction would assist teachers to adjust instruction guidelines.

Motivation

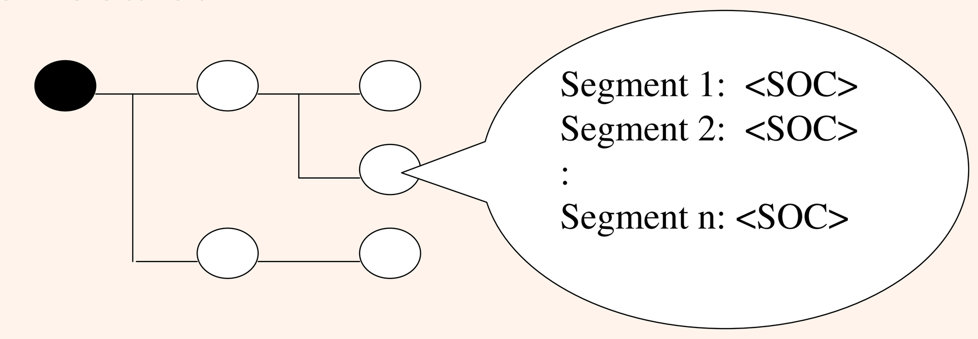
Social modes of argumentative knowledge construction (SOC) identify each student's contribution degree in one group

We propose approaches for detecting SOC by considering

- Information concealed between parent and current message in the same thread
- Introducing the extracted features into sequential learning algorithms

Problem Definition

The top black node represents the initial message of a *thread*, which is followed by replied messages. In our task, we assume that each message is divided into several segments. For each segment, it is assigned a SOC code. Our goal is to detect SOC for every message segment given the whole thread structure on a discussion board.



SOC Prediction Approaches

Feature Space Representation

Every segment was represented as a vector x:

$$\mathbf{x} = \langle a_1, \ldots, a_n, p_1, \ldots p_n, s, pmt, quo, len, fsm, deep \rangle$$

a $= \langle a_1, \ldots, a_n \rangle$, denotes a vector of terms (unigrams, bigrams, or part-of-speech tags) of current segment

p $= \langle p_1, \ldots, p_n \rangle$, represents the overlapping between current segment vector ac and parent message vector a_p

s computes the similarity (ex: cosine similarity) between ***a*** and ***p***

pmt, identifies whether the current message is a prompt

quo, denotes a cited message

len, is the length of a message

fsm, represents the states of a simple automaton

deep, identify how depth a message is now located in its belonging thread

Learning Algorithms for SOC Detection

We choose three different learning algorithms to demonstrate that the feature set is representative of capturing characteristics for predicting SOC codes and can be applied to improve the performance of various classifying algorithms.

SMO, a fast version of support vector machine learning algorithm.

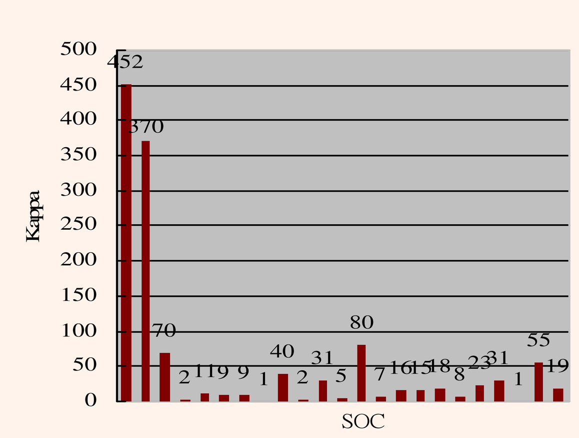
Naïve Bayes, a probabilistic model which takes Bayes' theorem as its theoretical basis.

Decision Tree, a method for approximating discrete value classifier.

Applying the feature set to sequential learning algorithm

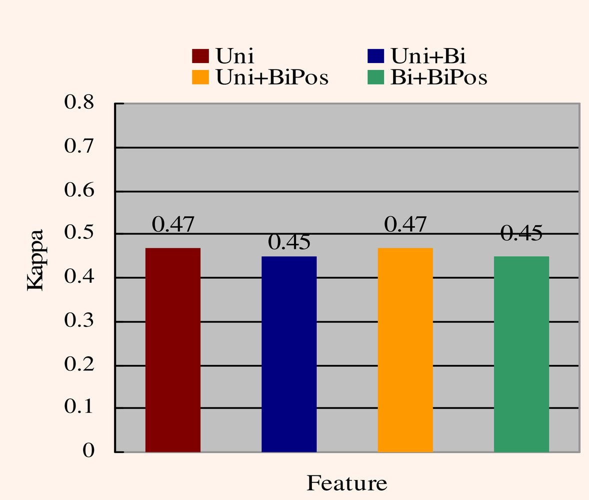
It seems reasonable that a following message segment would be related to its previous segment. Based on this assumption, we arrange an experiment by applying a sequential alogrith`m, Colli.

Experiments and Evaluation Results



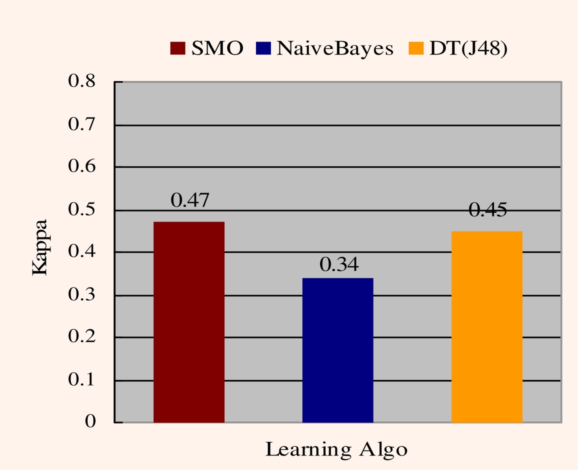
Data Source

Utilizing a German discourse as my data set. The corpus was collected in an experimental online learning environment where students collaborated via web discussion boards.



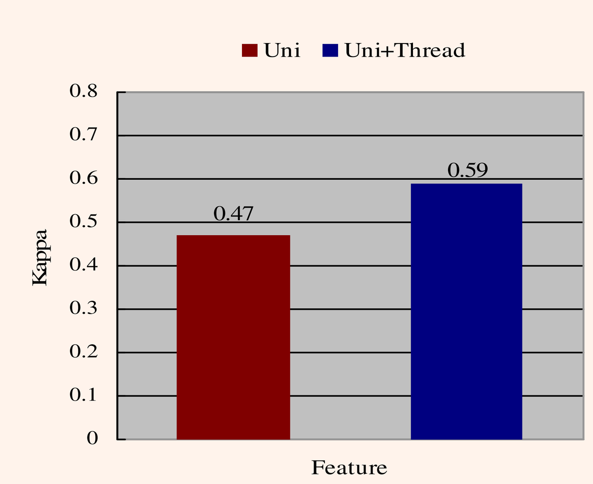
The Baseline

Considering four different syntactic characteristics as basic feature set without taking any structure information available from discussion boards into concern yet



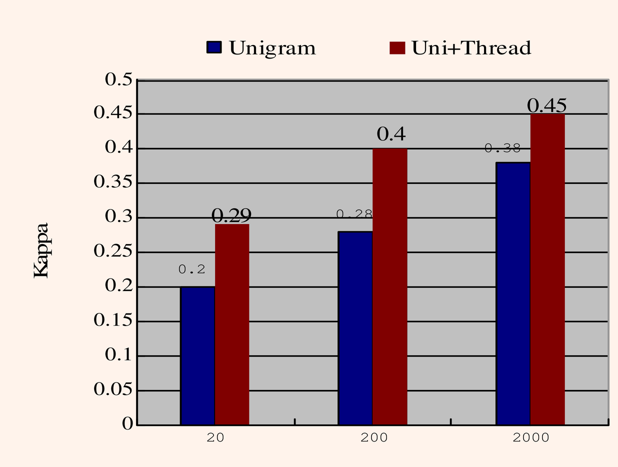
Three Algo

Compared results for the three algorithms. By selecting top 100 unigrams as features, we can conclude that SMO has the highest Kappa value (0.47) among there three learning algorithms.



Meta Feature

We can see Kappa value increases from 0.47 to 0.59. This tells us that adding the meta features does help to improve the performance on predicting SOC codes.



Sequential Algo

We can observe the improvement benefits from thread structure. But, the highest Kappa value (0.45) is still lower than 0.59. This result could be caused by inefficient feature selection method.

Note: **Cohen's Kappa Statistic** is a commonly used metric in assessing inter-rater reliability on the defined analyses units among target categories. In our case of detecting SOC codes, the golden standard of Kappa is 0.81.

2006 Fall Machine Learning Final Project Poster, Nov. 30