ACODEA: A Framework for the Development of Classification Schemes for Automatic Classifications of Online Discussions

Jin Mu, LMU München, Leopoldstrasse 13. 80802 München, GERMANY, jin.mu@psy.lmu.de
Karsten Stegmann, LMU München, Leopoldstrasse 13. 80802 München, GERMANY, karsten.stegmann@psy.lmu.de
Elijah Mayfield, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, elijah@cmu.edu
Carolyn Rosé, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, cprose@cs.cmu.edu
Frank Fischer, LMU München, Leopoldstrasse 13. 80802 München, GERMANY, frank.fischer@psy.lmu.de

Abstract: Research related to online discussions frequently faces the problem of analyzing huge corpora. Natural Language Processing technologies may allow automating the analysis. However, the state-of-the-art in machine learning and text mining approaches yields models that do not transfer well between corpora related to different topics. Also segmenting is a necessary step, but frequently, trained models are very sensitive to the particulars of the segmentation that was used when the model was trained. Therefore, in prior published research on text classification in a CSCL context, the data has been segmented by hand. We discuss work towards overcoming these challenges. We present a framework for developing coding schemes optimized for automatic segmentation and topic independent coding that builds on this segmentation. Our results show that our coding scheme can be fully automated by using a tool called SIDE. Finally, we discuss how fully automated analysis can enable context-sensitive support for collaborative learning.

Why should Online Discussions be Coded Automatically?

Online discussions have been widely used in the field of CSCL to foster collaborative knowledge construction. Learners work together to exchange ideas, negotiate meaning and formulate understanding (De Laat and Lally 2003). One important feature of online discussions is that this kind of communication produces a huge body of digital data as a byproduct of the interaction. Researchers are therefore confronted with the opportunity as well as the challenge of analyzing online discussions at multiple levels, such as quality of argumentation, or social modes of interaction (e.g., Weinberger & Fischer, 2006). A variety of multidimensional frameworks have been employed (cf. Clark et al., 2007). In this study, we focus specifically on analysis of what has previously been called micro-argumentation (Weinberger & Fischer, 2006), with the idea of expanding to other dimensions of analysis in future work.

Evaluation of discussion quality consumes a huge amount of resources in research projects related to online discussions. In order to address this problem, Rosé and colleagues (2008) reported a strand of experimental studies with about 250 online discussions (cf. Stegmann, Weinberger & Fischer, 2007; Weinberger et al., 2005; Weinberger, Stegmann & Fischer, 2010) where about 25% of all human resources in the research project were spent to analyze online discussions on multiple dimensions. Human coders had to be trained to annotate segments of these data using a multi-dimensional coding scheme that operationalized aspects of content as well as manner of argumentation and social modes of interaction. While uncovering findings related to how group knowledge construction works often make those efforts worth the time and energy they require, analyzing a huge body of online discussions by hand is an arduous task that slows down the progress of the research substantially. An automatic and thus faster classification of online discussions may affect the whole research process positively. One possible impact may be that an increasing number of researchers may be willing to analyze online discussions on multiple dimensions. Moreover, a part of the freed-up resources may be used to conduct follow-up studies or to try out pioneering approaches to data analysis.

Automatic classification may not only facilitate research on online discussions. Automatic classification also allows for adaptive collaborative learning support (ACLS; Kumar et al., 2007; Kumar & Rosé, in press; Walker, Rummel & Koedinger, 2009) to foster the quality of collaborative knowledge construction during online discussions (e.g. Gweon et al., 2006; Walker, Rummel & Koedinger, 2009). Online discussions could be analyzed in real-time and instructional support measures like hints or scaffolds could be adapted to the quality of certain aspects of the collaboration. For example, learners who are unable to provide warrants and grounds for their claims may get offered a scaffold to construct better arguments. Learners who fail to relate their contribution to other learning partners may be explicitly asked to provide feedback. Against this background, we report a use case on the application of a tool for Natural Language Processing (SIDE) in a multi-layer framework which has optimized for fully automatic segmenting and context independent coding. We begin by providing an overview of the state-of-the-art in the application of NLP technologies in CSCL research.

Applying NLP Technologies in CSCL

Natural Language Processing has long been used to automatically analyze textual data. The need for involving technology from NLP in the process of content analysis is growing in the presence of the Web and distance learning (Duwairi, 2006). For instance, the NLP methods of content analysis have been developed for the automatic grading of essays (Duwairi, 2006; Landauer, Laham, & Foltz, 2003;); and for the intelligent and cognitive tutoring (Rosé & VanLehn, 2005; Diziol, Walker, Rummel & Koedinger, 2010).

In the last few years, researchers have begun to investigate various text classification methods to help instructors and administrators to improve computer supported learning environments (Kumar & Rosé, in press; Romero & Ventura, 2006). Text classification, is an application of machine learning technology to a structured representation of text, which has been a major focus of research in the field of NLP during the past decade. Typically, text classification is the automatic extraction of interesting, frequently implicit, patterns within large data collections (Klosgen & Zytkow, 2002). Nowadays, text classification tools are normally designed mainly for power and flexibility instead of simplicity (Romero & Ventura, 2006), which can assess student's learning performance, examine learning behaviour, and provide feedback based on the assessment (Castro, Vellido, Nebot, & Minguillon, 2005). Consequently, most of the current text classification tools are too complex for educators to use, and thus their features go well beyond the scope of what an educator might require (Romero & Ventura, 2006).

Therefore, TagHelper (Rosé et al., 2008) and its successor SIDE (Mayfield & Rosé, 2010) were developed to automate the content analysis of collaborative online discussions. As a publically available tool, TagHelper has been downloaded thousands of times in over 70 countries. Recently, application of TagHelper for automated tutoring and adaptive collaboration scripts have been extensively researched (Kumar & Rosé, in press). In order to make TagHelper tools accessible to the widest possible user base, default behaviour has been set up in such a way that users are only required to provide examples of annotated data along with un-annotated data. TagHelper first extracts features like line length, unigrams, bigrams and part-of-speech bigrams from the annotated data. An interface for constructing rule-based features is also provided. In SIDE, more sophisticated support for feature construction is included, such as regular expressions, which are important in the area of information extraction and named entity recognition, which we make use of in the study reported in this paper. Recent work has also yielded approaches for automatic feature construction (Mayfield & Rosé, 2010b), which further enhances the ability to construct richer and more effective representations of text in preparation for machine learning. Tools such as TagHelper and SIDE then build models based on the annotated examples that it can then apply to the un-annotated examples. To get the best results, both tools allow users to switch easily between different machine learning algorithms provided by Weka (Witten & Frank, 2005). Namely, they are for example Naïve Bayes, SMO, and J48.

Despite the effectiveness of applying TagHelper to analyze text-based online discussions, at least two challenges associated with the current NLP approach are still needed to be addressed. First, the automatic approach has so far only been demonstrated on annotated examples from corpora that come from a single scenario, and the generated model is quite context sensitive and case dependent, and has not been demonstrated to transfer well to online discussions with different topics. Second, if no natural borders of segments are provided to the tool, the automatic segmentation that it has been able to get prior to the work reported here has not been satisfactory (Rosé et al., 2008). However, such an automatic segmenting is imperative for us to investigate the real-time adaptive fading, which is only possible when segmentation is also done automatically. Besides the shortcomings of applying TagHelper into the field of CSCL mentioned above, the accuracy and reliability is often not as optimal as we would like it to be, especially when the segmentation is done automatically (Rose et al., 2008). This motivates our investigation to explore whether the use of more advanced natural language processing technology can offer fully automatic and context independent automating techniques for content analysis.

In this paper we explore several enhancements to this technology in order to overcome these challenges. For example, one promising direction to consider is the integration of information extraction techniques for improving content analysis. Previous work on applying NLP in the field of CSCL, generally accepted raw text as input for segmenting and coding, and the features used for classification were very low-level and simplistic. The new approaches used in the current study have strengths in information extraction, which allow the construction of a more sophisticated representation of the text to build the classification models on. Such technology includes named entity recognition, which is an active area of research in the field of language technologies.

Part-of-Speech tagging (PoS) is the process of assigning a syntactic class marker to each word in a text (Brill, 1992; Mora & Sánchez Peiró, 2007), therefore, a PoS tagger can be considered as a translator between two languages: the original language that has to be tagged and a "machine friendly" language formed by the corresponding syntactic tags, such as noun or verb. As Poel et. al.(2007) proposed, PoS tagging is often only one step in a text processing application. The tagged text could be used for deeper analysis. Instead of using PoS as the default generalized features, it makes sense to apply modified and specialized PoS categories and thereby to facilitate automatic segmentation if the unit of analysis is syntactically meaningful.

The goal of *Named Entity Recognition* (NER) is to classify all members of certain categories of "proper names" appearing in the raw text, into one of seven categories: person, organization, location, date, time, percentage, and monetary amount (MUC6, 1995). Core aspects of NER are entity and mentions. Mentions are specific instances of entities. For example, mentions of the entity class "location" are New Brunswick, Rhodes, and Hong Kong. Therefore NER provides not only additional features based on extracted entities for each word, but also a more context-independent way to train automatic classifiers. The mentions of New Brunswick, Rhodes, Hong Kong are cities in, for example, the discussions about the past and upcoming three CSCL conferences, while Bloomington, Utrecht, and Chicago would have the same semantic function within discussions about the past three ICLS conferences. As an initial step of pre-processing in information extraction applications, an automatic classifier that had been trained with predefined entities (e.g. "location") instead of specific mentions (e.g. Hong Kong) might have more flexibility for modelling contextual information, potentially improving classification performance. More recently, there have been tasks developed to deal with different practical problems (IREX and CoNLL-2002), in which every word in a document must be classified into one of an extensive set of predefined categories, rather than only identifying names of people and locations.

With the support of current approaches in information extraction, the input to SIDE is assumed to be enhanced in a fully automatic way to be less context dependent. In the following section, we will present the multi-layer framework for the development of classification schemes for automatic segmentation and coding.

The Automatic Classification of Online Discussions with Extracted Attributes (ACODEA) Framework

Typical text classification for online discussions in CSCL is made to be applied by humans. They rely strongly on implicit knowledge held by human coders (e.g., understanding sentences with misspelled words or wrong grammar) to reach an acceptable level of reliability. Text classification that should be applied automatically has to account for the more limited features that are usually used to train automatic classifiers. Our following framework may support the development of such classification schemes.

Before delving into the specific processes of how the machine learning tool operates, we further clarify the concepts that are to be learned by SIDE. Witten and Frank (2005) details how data can be associated with classes or concepts, which should be reproducible by SIDE, intelligible to human analysts, and operational to be applied to actual examples. The starting point for understanding online discussion analysis is to define the coding schemas. In choosing the coding schemas, the researcher needs to determine what sized segments (which range from single word, sentence, paragraph, to the entire message) match with the desired and target activities to be coded (Strijbos et al., 2006). Thus the first target concept to learn is to classify, at each word, whether a segment boundary occurs. Similar to an earlier segmentation approach (Rosé et al., 2008), the concept of segmentation is implemented as a "sliding window" consisting of a specific number of words. In this way, any segmentation is possible since the boundary between any neighboring pair of words is a possible site for a segment boundary. The second concept considered here is to sort each unit of analysis (segment) to one or more categories (dimensions of analysis). For instance, a specific sentence, utterance or message is classified according to quality of argumentation or social mode of interaction (cf. Weinberger & Fischer, 2006).

Each individual instance (word in the text to be segmented and then coded) provides an input to machine learning, which is characterized by a fixed and predefined set of features or attributes. Text classification often requires data transforming into appropriate forms (Han & Kamber, 2006). Attribute construction (or feature construction), where new attributes are constructed and added from the given set of attributes, can help provide richer, more effective features for representing the text prior to text classification, consequently, ease the training of automatic classifiers as well.

Here we present our architecture for extracting features from the text in order to construct a representation suitable for applying machine learning to, either for the segmenting layer or the coding layer. The basic rules are to apply part-of-speech tagging and named entity recognition to extract features that are abstract enough to make interesting patterns apparent to machine learning algorithms and yield models that generalize well. On both the syntactic and semantic levels, rather than use predefined categories, we design customized sets of labels which extract information about the specific tasks or target activities we wish to classify. These labels align with behaviors which participants are expected to do during the discourse. In addition, the entire architecture is structured to cascade from one layer to the next, incorporating information from the previous layers to improve the current classifier's performance. Extracting attributes on the syntactic level benefits from the use of off-the-shelf grammatical part of speech taggers, while the layer related to semantic representation benefits from the inclusion of named entities and techniques from information extraction. The outputs of these layers is used as the attributes for the final classification layers of segmenting and coding. In this paper we want to explore the positive impact on performance of adding in abstract syntactic and semantic features above baseline feature spaces consisting of word-level representations such as unigrams and bigrams.

The automatic framework to analyze the content of online discussion can be further illustrate in the detailed flow processes below. At first, each single word in the raw data for training must be pre-processed by human coders to extract the syntactic and semantic features. These annotated examples, which reach acceptable reliability, can then be used to train classifiers for all defined ontologies. On the first layer, the part-of-speech tagger and named entity recognition are both applied independently. Secondly, human coders have to classify the borders between the segments. These human coded examples are used to train the automatic segmentation. However, the input to the segmentation classifier would be the preprocessed concepts from the syntactic attributes, instead of the raw text. After successful training of automatic classifiers for segmentation (i.e. high reliability regarding the identification of borders between segments), an automatic classifier segments all the data. This is required to provide equal training material for humans and automatic classifiers. Once the data is segmented, human coders have to classify all segments in the training data regarding the dimensions defined for the coding layer. These classifications are used to train classifiers for all defined dimensions. Again, the concepts from the semantic attributes are used to train the automatic classifiers. Finally, evidence is required that the automatic classifier (classifying and training with concepts from the predefined attributes) is reliable compared with human coders (classifying and training with raw data). This is needed because the whole procedure has one major risk: Several classifications are made in a row, and thus errors on the different layers are cascaded. Therefore, the final automatic classification must ultimately be checked against pure human coding.

Research Questions

In the following, we will present a use case of the multi-layer framework. Our research questions are:

RQ1: Can the automatic classification of attributes be trained reliably compared with human coding?
RQ2: Can the automatic segmentation be implemented reliably compared with human segmenting?
RQ3: Can the framework be used to train context independent segmentation with sufficient reliability?
RQ4: Can the automatic coding be implemented reliably compared with human coding?
RQ5: Can the framework be used to train context independent classification with sufficient reliability?

Methods

Participants and Learning Task

Eighty-four (84) students of Educational Science at the University of Munich participated in this study. Students were randomly assigned to groups of three. Each group was randomly assigned to one of three experimental conditions. The task of the groups was to join a collaborative, argumentative online discussion and solve five case-based problems with the help of an educational theory. The computer-based learning environment used in this experiment is a modified version of the one employed by Stegmann, Weinberger and Fischer (2007).

The subject of the learning environment is Weiner's attribution theory (1985) and its application in education. The students read the text of attribution theory and the text of introducing argumentation individually. In the collaborative learning phase, three problem cases from practical contexts were used as a basis for online discussions. The case "Math" describes the attributions of a student with respect to his poor performance in mathematics. In the case "Class reunion" a math tutor talks about how he tries to help female students deal with success and failure in assignments. The case "Asia" describes differences in school performance between Asian and American/European students that were explained by the attribution theory. Another two cases were used in the pre and post test, which mainly concern the factors which affect a student's "Major choice" at the university and student's explanation for failure in the exam of "Text analysis".

Data Source and Procedure

We collected 140 conversation transcripts, each of which contained the full interaction from one group, and was targeted to a single scenario. Altogether, there are 74764 words in the corpus. Two human coders analyzed almost one tenth of the raw data (distributed over five cases), which have been further used for training the customized algorithms for the classification on the extracing attributes, segmenting and coding layers by SIDE (Mayfield & Rosé, 2010). SIDE is the successor of TagHelper (Rosé et al., 2008) and includes an annotation interface allowing for automatic and semi-automatic coding more easily. To train such classifiers with SIDE we had to provide examples of annotated data. SIDE extracted multiple features from the raw data, like line length, unigrams, bigrams, Part-of-speech bigrams, etc. Machine learning algorithms used these features to learn how to classify new data. As output, SIDE builds a model based on the human annotated data. This model can then be easily applied to classify un-annotated data, and then the assigned codes can be further reviewed on the annotation interface by modifying the codes the human coders disagree with. Furthermore, SIDE employs a consistent evaluation methodology referred to as 10-fold cross-validation, where the data for training the models can be randomly distributed into 10 piles. Nine piles are combined to train a model. One pile is used to test the model. Self-reported reliability is calculated by averaging across the ten iterations.

Statistical Tests

The reliability of the coding was measured using Cohen's Kappa value and percent accuracy. Both of them have been regarded as accepted standards for measuring coding reliability. The criterion for success is reaching a level of agreement with a gold standard as measured by Cohen's Kappa that is .7 or higher (Strijbos et al., 2006).Here it is worthwhile to further clarify that the present study was undertaken to evaluate different types of Kappa in the distinguishable phases, including (1) inter-rater agreement between human coders to evidence the reliability of training examples; (2) internal generated Kappa by the 10-fold cross-validation to certify the reliability of the SIDE training models, and (3) the conclusive Kappa between human coders and SIDE.

Application of the Framework

(i) Our coding layer was defined with respect to the approach of argumentative knowledge construction. Learners construct arguments in interaction with their learning partners in order to acquire knowledge about argumentation as well as knowledge of the content under consideration (Andriessen, Baker, & Suthers, 2003). Therefore on this layer we are mainly concerned with the following categories, based on the micro-argumentation dimension of the multidimenional framework generated by Weinberger and Fischer (2006): (a) *Claim* is a statement that advances the position learners take to analyze case with attribution theory. (b) *Ground* is the evidence from case to support claim. (c) *Warrant* is the logical connection between the grounds and claims that present the theoretical reason why a claim is valid. Consequently, (d) the *Inadequate claim* should be differentiated in the coding, which concerns other related educational theory to explain case. (e) *Evaluation* is the agreement with learning partner or not. There are more dimensions to indicate the (f) *Empty Message* and (g) *Scripts*, both of them are the computer generated to structure the argumentative discourse, and finally (i) *Others*, which cannot be sorted by any other dimensions.

(ii) The unit of analysis was defined as a sentence or part of a compound sentence that can be regarded as 'syntactically meaningful in structure, regardless of the meaning of the coding categories' (cf. Strijbos et al., 2006, p. 37). For instance, according to these rules of segmentation, punctuation and the special words like 'and' are boundaries to segment compound sentences if the parts before and after the bounder are 'syntactically meaningful' segments. This size of segment has been proved to be reliable (Strijbos et al., 2006), and suitable for the coding dimension conducted in the current study.

(iiia) Regarding the syntactic attributes: The set of syntactic annotation consists of 12 different tags, which denote the grammatical features of a word class. Each word in the computerized data can be pre-processed into multiple and syntactic categories. An example of such a tag is: Term, Verb, Property, Conjunction, Comma/Stop Symbol, and so on. These tags are a reduced version of the full tag set, making it more suitable for machine learning. Some stop words like Pronoun are clustered into the class of Others. For example, "The math-failure of the son is external stable, because the entire family is not good at math." can be replaced into "Others / Terms / Others / Others / Terms / Verbs / Property / Property /CommaSymbol /Others / Others / Property / Terms / Verbs / Others / Property / Others / Terms /StopSymbol /".

(iiib) Regarding the semantic attributes, each single word in the text can fall into one of the multiple categories, either (a) *Case*, key words from problem space, (b) *Theory*, key words from the concerned conceptual space (actually, attribution theory in the present study), or (c) *Extraneous theory*, from the related educational theory. In addition, there are words that are important in reflecting the (d) *evaluation* either positive or Negative among partners (which refers to key indicator of Counterargument), (e) *Empty Message*, and even (f) *Others activities*, can be extracted in this phase. Therefore the mentioned above example can be represented on this layer as Others (The) / Case (math-failure) / Others (of) / Others (the) /Case (son) / Others (is) /Theory (external) /Theory (stable) /Others (,) /Others (because) /Others (the) / Others (entire) /Case (family)/Others (is) /Others (not) /Others (good) /Others (at) /Case (math) / Others (,). All of the categories are choosen because they might support the coding on the classification layer. For instance, according to our learning task a claim would typically contain both case and theory information, while a ground mainly includes case information, and a warrant only includes elaborations on the attribution theory.

Results

Two coders created the training material for SIDE. The inter-rater agreement between two human coders was Cohen´s Kappa =.93 on the syntactic-attributes layer and Cohen´s Kappa = .97 on the semantic-attributes layer. We achieved a high value of Cohen's Kappa = .96 for the segmentation layer and Cohen's Kappa = .71 for the coding layer. These results indicate acceptable human baseline performances for SIDE to be trained to analyze the un-annotated data regarding the extracting attributes, segmentting and coding layers.

RQ1: Training of the Layer of Extracting Attributes

SIDE achieved an average Cohen's Kappa = .94 (accuracy = 91.7%) with the training material on the syntactic layer, and an average of Cohen's Kappa = .93 (accuracy = 96.0%) on the semantic layer. A human coder and SIDE achieved an agreement of Cohen's Kappa = .89 (accuracy = 92.0%) on the syntactic layer, and Cohen's

Kappa = .94 (accuracy = 97.4%) on the semantic layer. As shown in Table 1, the reliability of SIDE to analyze text on the syntactic and semantic layers is satisfactory across all 5 cases. Because the precision on the layer of extracting attributes greatly influences the performance of the steps further in the chain of linguistic treatments, the accuracy of the PoS tagger and named entity recognition is very important.

Table 1: Reliability of Automatic attributes extraction within 5 Cases (SIDE vs. Human).

Case	Syntactic-attributes Layer		Semantic-attributes Layer	
	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy
Major choice	0.90	92.2%	0.95	98.1%
Math	0.85	88.1%	0.92	96.9%
Class reunion	0.91	92.9%	0.90	96.0%
Asia	0.93	94.4%	0.96	98.3%
Text analysis	0.88	90.3%	0.94	97.7%

RQ2 & RQ3: Training of the Segmentation Layer

Table 2: Comparison without and with the layer of extracting attributes to automate the content analysis (SIDE vs. Human).

	Without extracting attributes		With extracting attributes	
Segmentation	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy
Internal SIDE Kappa	0.84	96.7%	0.98	99.6%
Kappa between SIDE and Human	0.86	97.0%	0.97	99.3%
Major choice	0.80	96.7%	0.95	99.1%
Math	0.86	96.6%	0.96	98.9%
Class reunion	0.87	97.0%	0.97	99.3%
Asia	0.90	97.7%	0.99	99.7%
Text-analysis	0.83	96.9%	0.98	99.6%
Classification	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy
Internal SIDE Kappa	0.70	75.6%	0.77	81.3%
Kappa between SIDE and Human	0.61	67.8%	0.81	84.5%
Major choice	0.63	71.2%	0.77	82.9%
Math	0.67	72.3%	0.78	82.6%
Class reunion	0.47	58.5%	0.76	81.0%
Asia	0.53	63.1%	0.85	87.5%
Text-analysis	0.68	75.0%	0.87	89.2%

Internal Cohen's Kappa = .98 (accuracy = 99.6%) was achieved by SIDE when it attempted to automatically segment the text, when the raw text has been pre-processed to extract syntactic attributes. A human coder and SIDE achieved an agreement of Cohen's Kappa = .97 (accuracy = 99.3%). In addition, the inter-rater reliability within the five different cases is displayed in the Table 2. The algorithm for segmenting generated on the base of the layer of extracting attributes achieved sufficiently high Cohen's Kappa across the tested cases.

RQ4 & RQ5: Training of the Classification Layer

SIDE achieved an internal Cohen's Kappa = .77 (accuracy = 81.3%) using the extracted semantic attributes across all cases during training. The reliability across all cases comparing SIDE with a human coder (based on raw text) was sufficiently high (Cohen's Kappa = .81; accuracy = 84.5%). Table 2 shows the results within five different cases. Sufficient Cohen's Kappa values were achieved for all of the cases.

In summary, a domain-independent framework consisting of multiple layers is presented, which has been used successfully for the design, implementation and evaluation of a methodology for automatic classification of a large German text corpora. The reliability of automatic segmentation and coding has been reported across 5 cases with sufficient Cohen´s Kappa and accuracy. Compared with previous work, in which the raw data was directly used for the segmentation and classification, SIDE have been demonstrated to yield improved performance with the pre-processed layers to extract external attributes.

Discussion

The performance of SIDE to automate the content analysis has been demonstrated to be enhanced, by applying the multi-framework of the Automatic Classification of Online Discussions with Extracted Attributes (ACODEA). However, the language processing framework embedded in the argumentative knowledge construction and specific thematic domains is not usually designed to analyze other learning activities, such as problem solving or thought-provoking questioning. Depending on the domain as well as the type of target learning process, different sets of categories for the layers of extracting attributes and coding may be used for maximum performance. It remains to be seen whether such an automatic approach aimed at case independence will be further transferred to be able to analyze other collaborative activities, such as social interactivity and so on.

So far, the encouraging results indicate that it is possible to reach our ultimate goal of realizing adaptive collaboration scripts. In the proposed framework, the raw text was annotated using the techniques of part-of-speech tagging and named entity recognition before segmenting and classifying on the desired dimentions. The benefit is twofold. On one hand, the most challenging methodological problems have been successfully resolved, and the analysis of online discussion in the concerned domain has been evidenced to be fully automatic and context independent. On the other hand, one disadvantage of this information extraction is a vast reduction in the stored information, potentially losing some valuable features from the raw data that we had ignored on the layer of extracting attributes.

In addition, one interesting issue should be nevertheless still investigated to advance SIDE. It would be useful to be able to assess how "correct" the argumentation is, rather than only how complete the argumentation is, as we have done so far. From an epistemic perspective, an appropriate argument is more than a simple pile-up of information from problem and conceptual space, which includes a structurally appropriate connective between specific case and concerned theory. One possibility is that in the pre-processing step, the keywords from case and theory, which are correctly connected corresponding to an expert model, can be weighed automatically. This way, scaffolds provided by an adaptive collaboration script assisted by the automated and customized approach of qualitative content analysis can be much more powerful in its facilitation role, supporting valuable learning processes.

References

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to learn: confronting cognitions in computer-supported collaborative learning environments* (Vol. 1, pp. 1-25). Dordrecht: Kluwer Academic Publishers.

Brill, E. (1992). A Simple Rule-Based Part-of-speech Tagger. *Proceedings of the Third Conference on Applied Natural Language Processing*. ANLP.

Castro, F., Vellido, A., Nebot, A., Minguillon, J., 2005. Detecting atypical student behaviour on an e-learning system. *In Simposio Nacional de Tecnologas de la Informacin y las Comunicaciones en la Educacion*, Granada, Spain (pp. 153–160).

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*(3), 343-374.

De Laat, M. F., & Lally, V. (2003). Complexity, theory and praxis: Researching collaborative learning and tutoring processes in a networked learning community. *Instructional Science*, *31*, 7–39.

Diziol, D.; Walker, E.; Rummel, N.; & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, *22*(1), 89-102.

Duwairi, R. M. (2006). A framework for the computerized assessment of university student essays. Computers in Human Behavior, *22*, 381–388.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. *In Proceedings of EdMedia*. .

Gweon, G., Rose, C. P., Zaiss, Z., & Carey, R. (2006). Providing Support for Adaptive Scripting in an On-Line Collaborative Learning Environment, *Proceedings of CHI 06: ACM conference on human factors in computer systems*. New York: ACM Press.

Han, J. & Kamber, M. (2006) *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.

Klosgen, W., & Zytkow, J. (2002). *Handbook of data mining and knowledge discovery.* New York: Oxford University Press.

Kumar, R. & Rosé, C. P. (in press). Architecture for building Conversational Agents that support Collaborative Learning, *IEEE Transactions on Learning Technologies special issue on Intelligent and Innovative Support Systems for Computer Supported Collaborative Learning*.

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of Artificial Intelligence in Education*.

Landauer, T. K., Laham, D., & Foltz, P. (2003). Automatic essay assessment. *Assessment in Education: Principles Plociy and Practice*, *10*(3).

Mayfield, E. & Rosé, C. P. (2010). An Interactive Tool for Supporting Error Analysis for Text Mining. *Proceedings of the North American Chapter of the Association for Computational Linguistics*.

Mayfield, E. & Rosé, C. P. (2010b). Using Feature Construction to Avoid Large Feature Spaces in Text Classification, *Proceedings of the Genetic and Evolutionary Computation Conference*.

Mora, G. & Sánchez Peiró, J. A. (2007). Part-of-Speech Tagging Based on Machine Translation Techniques. *Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis, Part I* (pp. 257 -264). Girona, Spain.

MUC6. Morgan Kaufmann Publishers, Inc. Proceedings of the Sixth Message Understanding Conference (MUC-6). Columbia, Maryland. 1995.

Poel, M., Stegeman, L. & op den Akker, H. J. A. (2007). A support vector machine approach to dutch partof-speech tagging. In M.R. Berthold, J. Shawe-Taylor, and N. Lavrac(Eds), *Proceedings of the 7th International Symposium on Intelligent Data Analysis* (pp. 274–283.) Springer Verlag, Berlin.

Romero, C., & Ventura, S. (2006). *Data mining in e-learning*, Southampton, UK: Wit Press.

Rosé, C. P., & VanLehn, K. (2005). An Evaluation of a Hybrid Language Understanding Approach for Robust Selection of Tutoring Goals, *International Journal of AI in Education*,*15*(4).

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Fischer, F., Weinberger, A., & Stegmann, K. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning 3*(3), 237-271

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(4), 421-447.

Strijbos, J-W, Martens, R. L., Prins, F. J., Jochems, W.M.G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*, 29–48

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge Universty Press.

Walker, E., Rummel, N., & Koedinger, K. R. (2009). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction*, *19*(5), 387-431.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computersupported collaborative learning. *Computers & Education*, *46*(1), 71-95.

Weinberger, A., Reiserer, M., Ertl, B., Fischer, F., & Mandl, H. (2005). Facilitating computer-supported collaborative learning with cooperation scripts. In R. Bromme, F. W. Hesse & H. Spada (Eds.), *Barriers and Biases in netwerk-based knowledge communication in groups*. Dordrecht: Kluwer.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*, 506–515.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548-573.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier: San Francisco