Identification of Learning Goals in Forum-based Communities

Julian Krenge
Information Management
FIR at RWTH Aachen University
Aachen, Germany
Julian.Krenge@fir.rwth-aachen.de

Zinayida Petrushyna, Milos Kravcik, Ralf Klamma
Chair of Computer Science 5
RWTH Aachen University
Aachen, Germany
{Petrushyna; Kravcik; Klamma}@dbis.rwth-aachen.de

***Abstract*—When Internet users search for information, surf on websites or discuss with others, their actions are driven by certain goals. Extraction of users' goals can enable higher effectiveness and accuracy of web services. Supporting users in based on their goals can be highly beneficial, especially supporting of learners in the preparation for an exam as a learning process, Different phases of learning are identified when users learn collaboratively. We scrutinize how goals are constructed and achieved within a community, examining not only social activities based on patterns of behavior, but also emotions and intents users express in their posts. As a result we elicit users' goals. We achieved good accuracy in defining emotions of users and recognizing their intents and social patterns in our case. Here we discuss how the obtained results contribute to mining of learning community goals.**

Goal mining; Sentiment Analysis; Intent Analysis

I. INTRODUCTION

Nowadays learners capitalize on the benefits of web-based collaboration. Here we deal with an e-learning environment, using a web-based forum. Investigated learners are non-native English speakers, who want to sharpen their profile by taking an established English test such as TOEFL, GRE or GMAT. In order to efficiently deal with the vast supply of learning material on the web, learners organize in groups, where they exchange experience and knowledge.

Objective of the paper is to create a thorough concept of the specific learning goals within an e-learning system for enabling systems automatically support learners by supplying them with additional material, tools, advice on next actions or even suggestions for learning partners. We create foundation for a method for matching learning advice and implement the method for supporting learners' needs.

II. DEFINITION OF A LEARNING GOAL

Goals have been a field of interest for a long time [1]. The definition of the term "goal" is different from many points of view coming from psychology to computer science. We choose Strohmaier et al.'s definition and consider a goal as "condition or state of affairs in the world that the agent would like to achieve or avoid, how the goal is to be achieved or avoided is typically not specified, allowing alternatives to be considered" [2].

III. RELATED WORK

Goal mining is a relatively new domain and its approaches are focusing on narrow target fields. Some of the projects and methods of goal mining dealing with the identification of user goals are applying Natural Language Processing (NLP) techniques. One of them is Sentiment Analysis (SA). SA aims to analyze the sentiment encapsulated in user generated content that can include goals. Users make goals explicit with specific emotions. SENTAL (http://www.ukp.tu-darmstadt.de/?id=2664) by TU Darmstadt specifically focuses on SA in the context of e-learning.

Intent Analysis is an approach to the identification of user goals based on the analysis of user-written texts and text fragments. Specific phrases or parts of speech are supposed to hold intentional artifacts, therefore traces of the original user intent.

In order to establish and understand the interrelation of goals, users' texts are queried directly. Goal concepts are partially included in knowledge repositories such as Princeton's WordNet or MIT's ConceptNet. WordNet focuses on the semantic and syntactic relations of words, allowing for inferences of goal relation. ConceptNet explicitly lists goal-related connections between common-sense knowledge artifacts, e.g.: *MotivatedByGoal* or *UsedFor* [3].

The other aspect of analysis of behavior in our context is supported by the Social Network Analysis. Combined with Actor Network Theory, SNA is able to identify concrete patterns of behavior of human agents within a certain environment. The PALADIN GraphService, developed at the Chair of Computer Science 5 at RWTH Aachen University, allows for such a user classification in forums. Relevant behavior models are the questioner, conversationalist and answering person. Their identification is based on occurrence of posts and threads, not on the content of posts [4].

IV. GOAL MINING IN STAGES OF LEARNING

Established models of learning processes such as Self-Regulated Learning (SRL) and the Eight Learning Events Model (8LEM) define three coarse stages of learning: plan, learn and reflect.

Plan: learners seek support in the goal formulation by more experienced members of a community. When shifting to learning material selection, the help-seeking of learners

978-0-7695-4346-8/11 $26.00 © 2011 IEEE
DOI 10.1109/ICALT.2011.95

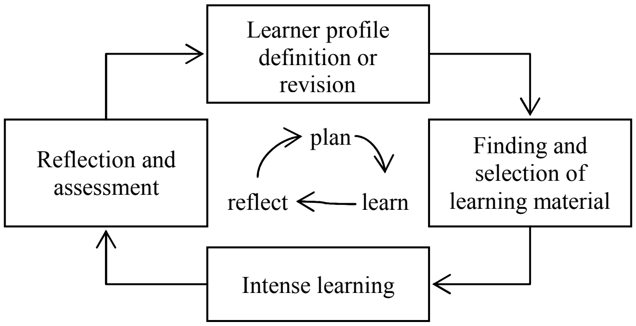

Figure 1: Learning phases [6]

increases significantly, as relevant sources and peers are requested [5, 6]. The sentiment is initially neutral while activity starts to increase. When asking for help, a learner also leaves intentional traces giving a first idea on his learning strategies and specific learning goals to be identified by SA. Due to the lack of expertise the learner generally does not engage in discussions.

Learn: the learner is actively involved in the community when he is discussing studied knowledge. His contributions can also include intentional traces; therefore the level of help-seeking stays elevated [6], but the support that is asked for changes. Whereas in *planning* short questions are stated, in *learning* a more active discussion is desired in order to reflect on the learning success so far. We assume a number of sentiments are raised to show more personal commitment to the domain, also activity is peaking.

Reflect: The learner reflects on his learning success and results [6]. Therefore his sentiment gives indication about his state; it is assumed that he is emotionally attached to the result of his work. Furthermore, learners can act as supporters for other learners which are at earlier stages of learning the same topic.

Figure 1 illustrates the sequence of the learning phases. As it is shown, the learning process is a cycle. After finishing on one topic, learners are assumed to continue on a different topic and repeating the circle.

V. Data Set and Experimental Set-Up

The data set used in this thesis was acquired by crawling a popular online e-learning bulletin board. The board is one of the largest e-learning communities on English test preparation: Urch Forums. In the most relevant sub-forums, we collected over 67k threads with over 428k posts, created by nearly 21k members.

To investigate learning processes of local communities, we establish the concept of a clique of users: those are formed by the users of one thread, which are – in graph terms – a clique. The clique furthermore fulfills following characteristics: it has at least 4 users and occurs at least ten times. This occurrence is defined as follows: if at least three quarter of users of a clique post in at least 10 threads, the clique is relevant to consider. This refines our data set to 12,881 cliques with an average size of five users and an average occurrence of 14 threads.

A. Activity

Activity is the most basic characteristic: starting threads, writing blog posts, giving comments in a discussion etc. Here, all posts and created threads are counted as activities. We plot activities over a timeline, quantized to weeks and identify the phases with high activity. Time is a crucial factor for our analysis and therefore all following aspects are set into a temporal dimension as well.

B. Sentiment

Sentiments expresses the emotions encapsulated in learners' contributions, therefore they reflects his attitude towards the e-learning community and the personal commitment to his own goals. Increasing stress and involvement in later phases of the learning process can result in emotional attachment to the topics and therefore more frequent use of sentiment expressions. As we want to cover merely the existence of the expression of strong sentiments within a post, we limit ourselves to polarity analysis. We deploy a machine learning algorithm, more precisely a Naïve Bayes Classifier, to identify the polarity of a post, ranging from highly emotional to not emotional at all, neglecting the direction of the emotion. The training set is based on Language Inquiry and Word Count (http://www.liwc.net/).

C. Help-Seeking(SNA patterns and Intents)

The aspect of help-seeking is a complex concept. Help-seeking is the set of activities when "a learner identifies and calls upon outside resources" [5], which can refer to peers in a community or external sources such as books or websites. Two aspects are used to identify the level of help-seeking: 1) the behavior within a community, e.g. starting threads as an indication of seeking help or opinions from the community, which is identified by the PALADIN GraphService [4]; 2) the detection of expressions of intent and specific questions and seeking help from the other learners. Our Intent Analysis bases on the syntactic patterns of language that can be observed when humans express goals. These are the VB_1toVB_2 and the *WRBtoVB* patterns, where *VB* is a verb in base form such as "go, try, plan" and *WRB* refers to an Wh-adverb like "how, where", as proposed by Tatu [7]. She states that more than 70% of intentions are expressed with one of the first two of the patterns.

VI. Results

We apply the defined criteria for user characteristics on the data set and assess the significance of each characteristic on the identification of learning goals. Evaluation is pursued by direct user feedback, i.e. questionnaires, asking for validation of the learning goal fragments identified for each of them. We received feedback from 18 users of the data set. We briefly go into aspect of method correctness and investigate deeper significance showing whether our mechanisms were suitable to cover learners' goals.

A. Correctness

For SA, precision is 0.67 and recall is 1.0. For the SNA pattern analysis, we assume precision and recall of 1.

Possible threat of this module is rather the construction of patterns and roles, since the expressiveness and validity of the patterns has to be proven individually. For IA, we incorporated a PoS tagger based on a machine learning algorithm. As its classification is rather simple and the training set as well as a classifier is proven in a number of contexts, we do not explicitly check precision and recall of it. The evaluation will further focus on significance.

Rank	Firstword	Keyverb	Count
1	Want	Know	1264
2	Need	Know	1027
3	Planning	Take	885
4	Like	Know	854
5	Going	Take	788

Table 1: Top 5 combinations of *firstword* and *keyverb*

B. Significance

How many intent expressions are covered by our patterns? To determine this, we would need a list of all possible patterns of expressions of intent and their probability distribution in average human language. In the following, we call the first word of an expression of intent *firstword* and the central verb, i.e. VB_2, *keyverb*. As *noun* we define the first noun mentioned after an expression of intent.

Rank	Keyverb	Noun	Count
1	Take	Test	837
2	Get	Score	478
3	Spend	Time	366
4	Solve	Problem	333
5	Take	Exam	253

Table 2: Top 5 combinations of *keyverb* and *noun*

We now focus on the evaluation of the patterns of intent expression. What pattern is most likely to expresses an intention and what *keyverbs* and *nouns* are the most significant when determining the learners' intentions? In the whole data set, 132,328 expressions of intent were detected.

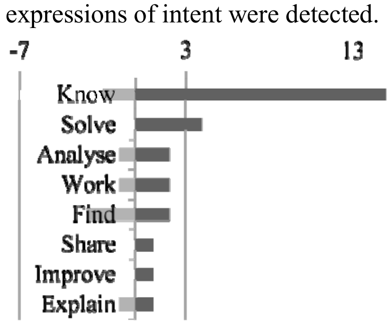

Figure 2: Significance of keyverbs

First and most obvious is the *keyverb* "know", which accounts for the majority of occurrences and clearly indicates that a learner needs to improve his knowledge. Expressions such as "how to solve" can be counted to the same group of expressions as well. Second significant *keyverb* is "take", mostly related to taking a test or improving its results or aspects of understanding.

We asked the learners which combinations actually reflected their goals. The learners themselves marked 14 *keyverbs* at least once as expressions of their goals. An overview is given in Figure 2. The listing of these *keyverbs* supports our assumptions of the most important terms. Focus of the community is clearly on improving skills in English language

VII. Conclusion and Outlook

We investigated the method to extract goals from learning communities. Using sentiment analysis, intent analysis and social network analysis we defined and described the characteristics of phases of learning goals. In order to refine automatic understanding of learning goals, our approach can be extended and build upon. The system's improvements are apparent: the correlation of sentiment, expression of cognitive mechanisms and patterns of behavior need to be found and the actual learning phase need to be further investigated. As yet, the assumptions are valid in a relevant fraction of the local communities, but not universally applicable. By knowing the exact start of learning through the user's explicit input, tracking of the phases might be more reliable. Creating goal ontologies might even enable system to refine a goal such as "learn English" to "improve English grammar".

Acknowledgment

The research work described in this paper is partially funded through the ROLE Integrated Project (grant agreement no 231396) of the European Union.

References

[1] R.C. Schank and R.P. Abelson, Scripts, plans, goals, and understanding: an inquiry into human knowledge structures. Psychology Press, 1977.

[2] M. Strohmaier, P. Prettenhofer, and M. Lux, "Different Degrees of Explicitness in Intentional Artifacts: Studying User Goals," *Proceedings of the CSKGOI*, 2008.

[3] M. Strohmaier, M. Kröll, and C. Körner, "Intentional query suggestion: making user goals more explicit during search," Proc. Workshop on Web Search Click Data, New York, New York, USA: ACM, 2009, pp. 68–74.

[4] R. Klamma, M. Spaniol, and D. Denev, "PALADIN: A Pattern Based Approach to Knowledge Discovery in Digital Social Networks," Proc. I-KNOW '06, 6th International Conference on Knowledge Management, Springer, 2006, pp. 457 - 464.

[5] N. Dabbagh and A. Kitsantas, "Supporting Self-Regulation in Student-Centered Web-Based Learning Environments," International Journal on E-Learning, vol. 3, 2004, pp. 40 - 47.

[6] A. Nussbaumer, K. Fruhmann, D. Albert, P. Ferdinand, A. Kiefel, U. Kirschenmann, C. Ullrich, S. Wills, A. Mikroyannidis, B. von Der Heiden, M. Werkle, and A. Naeve, "Deliverable 6.1: Common psycho-pedagogical framework," Responsive Open Learning Environments, 2009, pp. 1-72.

[7] M. Tatu, "Automatic discovery of intentions in text and its application to question answering," ACL '05: Proceedings of the ACL Student Research Workshop, Morristown, NJ, USA: Association for Computational Linguistics, 2005, pp. 31-36.