Using Opinion Mining Methods to Identify Students Experiencing Academic Difficulty in a Digital Education Environment.

By: Steve Harris
Supervisor: Dr. Vive Kumar
Dr. Kinshuk

Athabasca University
SCHOOL OF COMPUTING & INFORMATION SYSTEMS

INTRODUCTION

Why do we need to automatically detect academic difficulties?

- Learning in an online environment is becoming more common
- Class sizes are increasing
- Lack of face-to-face, or even direct contact with students
- Multiple Tutors per course
- Identify poor content that is causing student confusion
- Supply parameters to automated tutoring systems

RESEARCH QUESTION

Can an analysis of the text-based student interactions in a digital learning environment identify content deficiencies, or individual students experiencing academic difficulties?

Or, a more applied version:

Can we us data from the student’s use of Moodle’s Messaging and Discussion Forums to alert instructors to students who may be frustrated or confused? And can we identify those aspects or topics that are causing greater experience to the students overall?

RESEARCH APPLICATION

- **Text-based Interaction (Score or Probability)**
- Other data and metadata such as time spent on content, number of logins, etc.
- Assignment marks, final grades
- Student facial recognition

} Student Academic Difficulty Detection

There are a number of features that can be considered to detect student academic difficulty. This research focuses on text classification of the student's interaction with others, within the digital learning environment, which may ultimately be a single component in a great academic difficulty detection system.

ACADEMIC CONTRIBUTIONS

Detecting student confusion or frustration is typically more difficult than other text classification problems because language and words used between normal academic conversation and confused or frustrated students are often fairly similar.

Student 1: “Did you read this week’s material? The algorithms are interestingly complex!”

Student 2: “Did anyone read this week’s lecture – the algorithms are too complex”

Though, sometimes it’s easy:

Student 3: “HELP! I don’t understand this week’s course content!”

DETECTING CONFUSED OR FRUSTRATED STUDENTS

Similar to sentiment analysis, our approach to student confusion detection is based on Bing Lui's quintuple [4]:

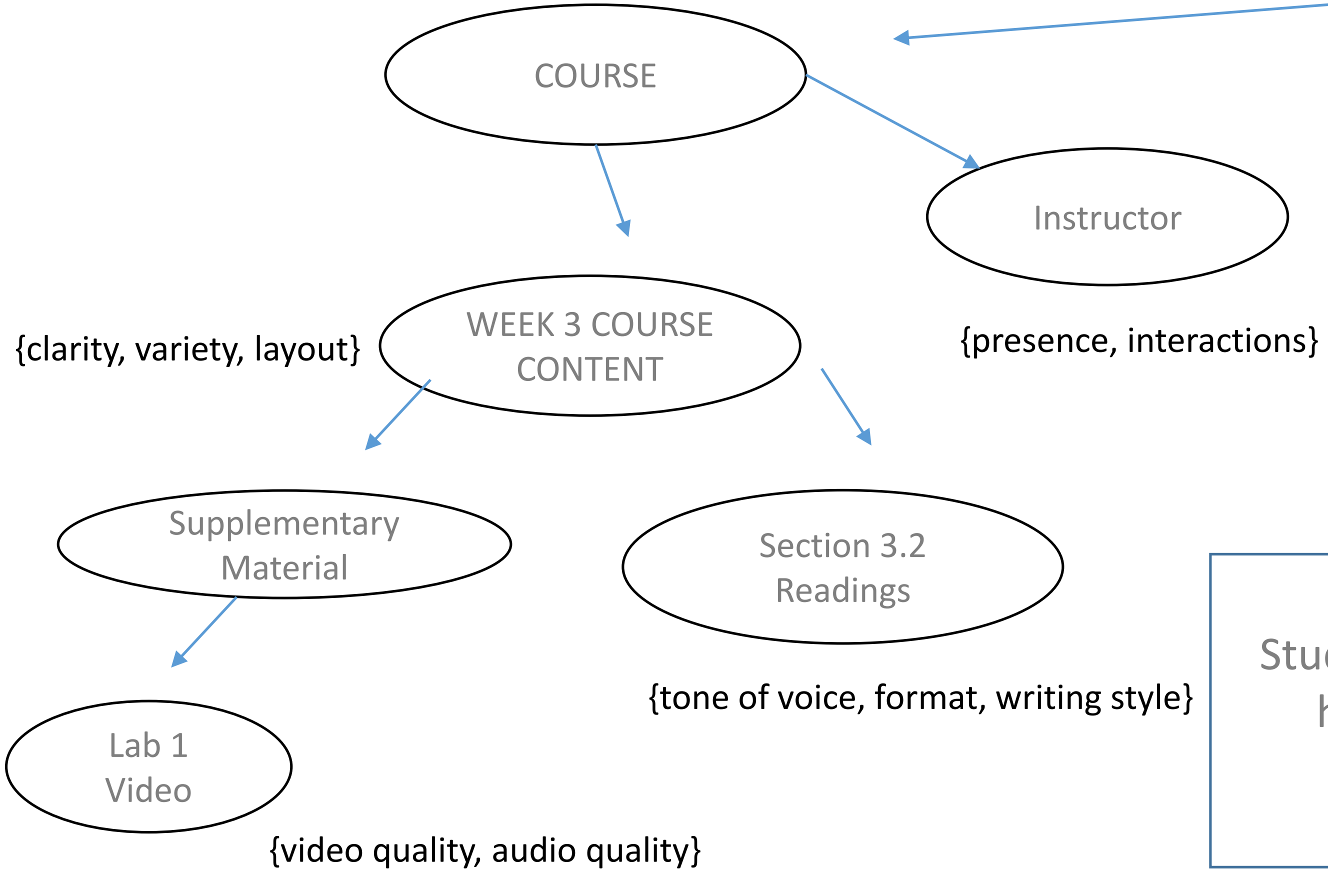
$$(e_j, a_{jk}, so_{ijkl}, h_i, t_i)$$

Where:

- **e** is the course and **a** is a component of the entity being described
- **so** is the orientation (sentiment) of the indicator, **h** is the holder of the emotion or state
- **t** is the time when the feeling or state was expressed

DETECTING CONFUSED OR FRUSTRATED STUDENTS

ENTITY & ATTRIBUTES



Student frustration or confusion can happen on any node, or about attributes of any nodes. [4]

attend jav
attend java lectur
chapter
i w
int
learn
nam
particip
post
pr
the exam
unit
wh
confus
creat
does not
error
exerc
man
pack
to find

DETECTING CONFUSED OR FRUSTRATED STUDENTS

Our data consists of 3,847 student messages and discussion forum posts taken from an Introductory Computer Programming Course delivered through the Moodle online learning environment

For training purposes, 400 messages are manually classified with NEUTRAL (200) or HELP (200). The rest are used for testing.

Student names have been replaced by a unique id for privacy purchases. And each post includes a timestamp.

BUILDING THE CLASSIFIER

1. Determine the tokenization and pre-processing method (stemming, multiple words, capitalization, emoticons, etc.). Test Principle Component Analysis.

2. Classification using different classifiers – Multivariate Naïve Bayes, MaxEnt, **Sequential Minimal Optimization (Support Vector Machines),** (J48) Decision Tree [5]

3. Feature Extraction (highly frequent phrases, especially after sentiment words). Plus custom feature identification (course/learning-related)

4. Develop custom domain-specific lexicon through semi-supervised learning

5. Use Cost-Sensitive learning in classifier to deal with class imbalance

PRELIMINARY FINDINGS - CLASSIFIER OPTIONS WITHOUT PCA

Classifier	Avg. Accuracy	Max Accuracy	Average F
Naïve Bayes	66.7%	77.4%	0.576
Naïve Bayes Multinomial	76.2%	83.3%	0.705
Sequential Minimal Optimization (SVM)	81.8%	84.8%	0.820
J48 (Decision Tree)	71.4%	81.3%	0.769

Based on subset of 400 COMP268 student posts

CLASSIFIER OPTIONS WITH AND WITHOUT PCA

Table 3 Classification accuracy

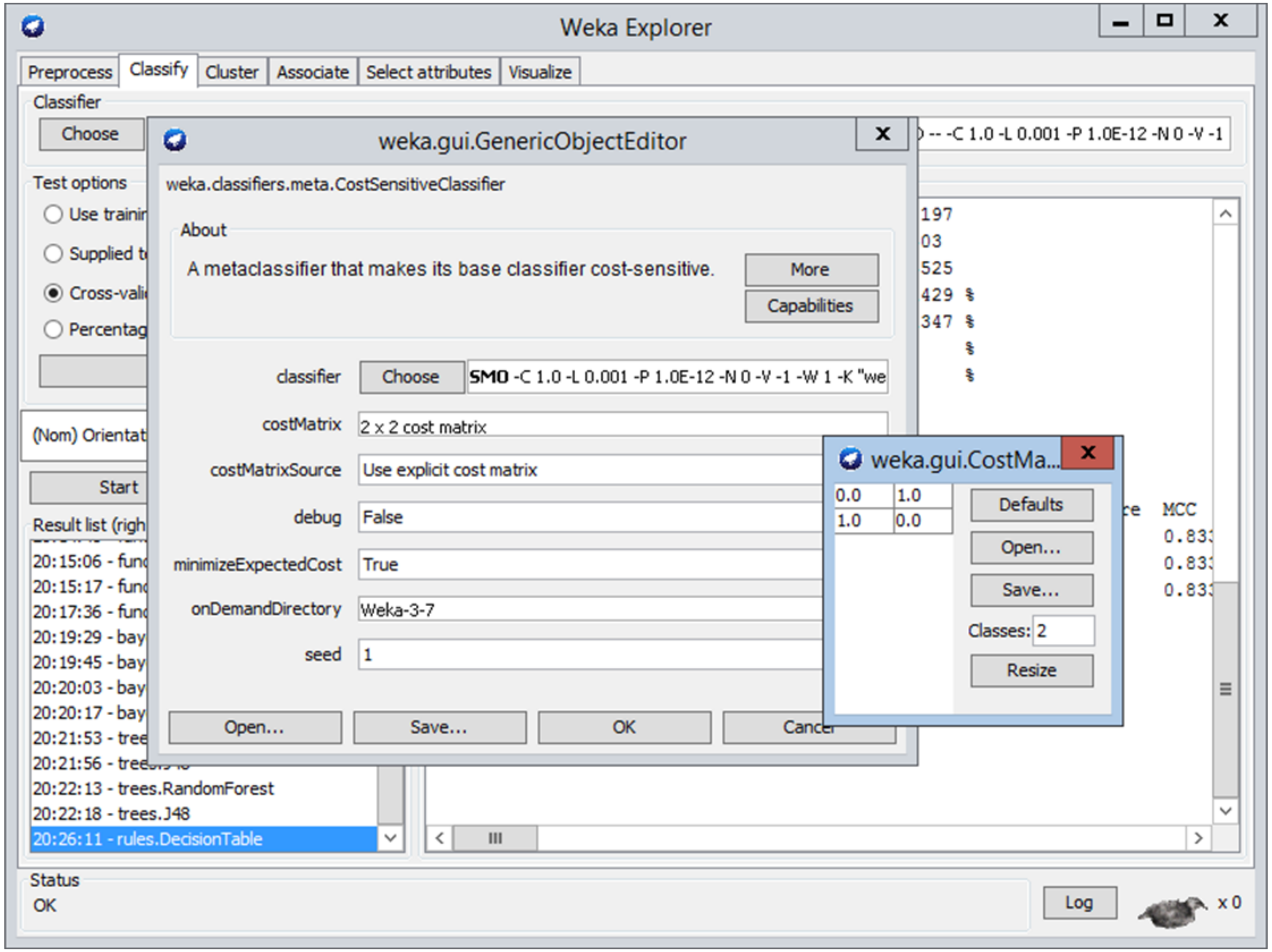
	Accuracy (%)		
Method	Camera	Mobile phone	Music player
Single SVM	86.99	85.34	88.73
Hybrid SVM—Majority voting	87.55	86.05	90.51
Hybrid SVM—LSE-based weighting	87.82	86.98	90.87
Hybrid SVM—Hierarchical SVM	88.01	87.57	91.08

Taken from Vinodhini, G.; Chandrasekaran, R.M. *Sentiment Mining Using SVM Hybrid Classification Model [6]*

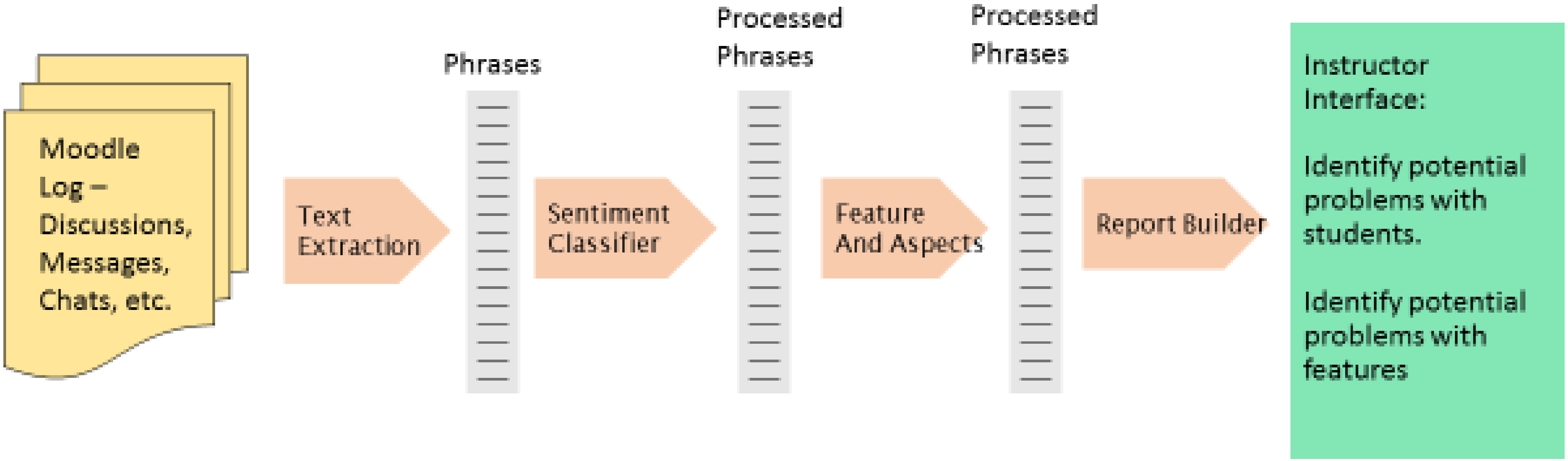
CLASSIFIER OPTIONS WITH AND WITHOUT PCA

Classifier	Avg. Accuracy	Max Accuracy	Average F
Sequential Minimal Optimization (SVM) With PCA			
Sequential Minimal Optimization (SVM) without PCA	81.8%	84.8%	0.820

THE WEKA TOOLSET



SUMMARY FLOW



Based on [5], [6]

REFERENCES

[1] Vaassen, F., & Daelemans, W. (2011, June). Automatic emotion classification for interpersonal communication. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 104-110). Association for Computational Linguistics.

[2] Feidakis, M.; Daradoumis, T.; Caballe, S., "Endowing e-Learning Systems with Emotion Awareness," *Intelligent Networking and Collaborative Systems (INCoS), 2011 Third International Conference on* , vol., no., pp.68,75, Nov. 30 2011-Dec. 2 2011

[3] Pedro Calais Guerra, Wagner Meira, Jr., and Claire Cardie. 2014. Sentiment analysis on evolving social streams: how self-report imbalances can help. In *Proceedings of the 7th ACM international conference on Web search and data mining* (WSDM '14). ACM, New York, NY, USA, 443-452.

[4] Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012

[5] Jurafsky, Dan. Sentiment Analysis – What is Sentiment Analysis. PowerPoint Presentation. Accessed March 10, 2014: *www.stanford.edu/class/cs124/lec/sentiment.pptx*

[6] Vinodhini, G.; Chandrasekaran, R.M. *Sentiment Mining Using SVM Hybrid Classification Model*. Computational Intelligence, Cyber Security and Computational Models, p155-162. Springer, Sept 2014.

[7 Priyanthan, Gokulakrishnan, Ragavan, and Prasath. *Opinion Mining and Sentiment Analysis on a Twitter Data Stream*, Accessed Feb 16, 2015: http://www.icter.org/conference/sites/default/files/icter/B03--95-P.Priyanthan.pdf