

Text Mining with SAS

Dominique Haughton

Maria Skaletsky

Agenda

- Data Set
- What is Text Mining?
- Why use Text Mining Techniques/Software?
- Introduction to SAS Text Miner
 - Unfair grading perceptions example
 - Emc example

Data Set

- Used to help faculty gain insight into the student perspective on fairness or unfairness in grading
- 416 undergraduate students enrolled in an expository writing courseware given a questionnaire
- Questionnaire:
 - Two open-ended questions
 - Think about a previous class—in any subject—in which you felt the teacher graded students' work in an unfair way.
 - Discuss how your own work was evaluated and why you felt the grading was unfair
 - Provide examples of fair grading
 - Age, gender , expository writing course and semester

Text Mining

- *Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools.*
- *Text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns*

-The Text Mining Handbook

Text Mining

- Aims to identify
 - Themes
 - Relevance
 - Interestingness
 - Relationships
 - categorization
 - Summary information
 - Illustrative Quotes
- Some text mining tasks
 - Text categorization
 - Text clustering
 - Concept/entity extraction
 - Entity relation

Why use Text Mining Techniques/ Software?

- Saves Time in Analyzing Qualitative Data
- Helps Researcher Deal with piles of Qualitative Data

SAS Text Miner

- Developed by SAS
- Is a tool within SAS enterprise Miner
- \$10,551 for academics and \$154,200 for commercial organization.
- The bulk of the SAS text Miner cost are those of adding SAS Enterprise Miner and SAS text Miner to a Typical SAS base.
- <http://www.sas.com/technologies/analytics/data-mining/textminer/>

Data Preparation

SAS TextMiner

- Interface uses nodes
- Clusters documents
 - Builds a matrix
 - terms as rows
 - documents as columns
 - Uses Singular Value Decomposition to reduce size
 - By default creates groups of equivalent words

Reporting of Results

SAS TextMiner

- Shows frequencies of key words (based on own analysis)
- Clustering
 - Two Types
 - Expectation Maximization (EM)
 - Hierarchical
 - Clusters based on Documents and then lists more frequent terms that appear
 - Users have to select
 - SVD resolution
 - Number of descriptive terms
 - Maximum number of clusters
 - Graphing/Cluster Analysis
 - Initial Results
 - Dendrogram
 - Tree View

Additional Features

SAS TextMiner

- Search for similar terms, documents or clusters
- Concept Link

Evaluation: SAS Text Miner

The Good

- Ability to cluster documents and not just terms inside documents
- Can run frequencies on words
- Searching for words in context is possible

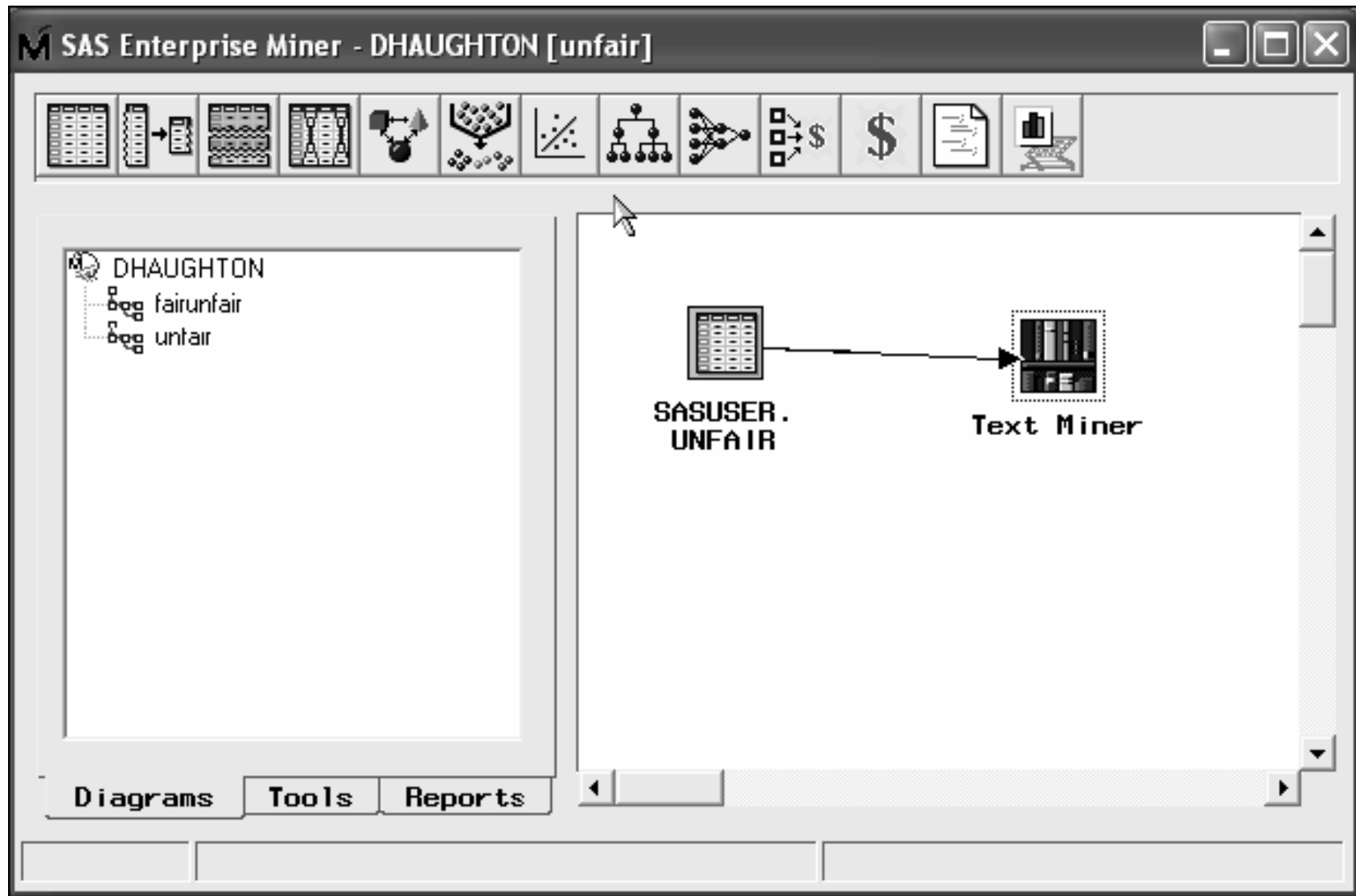
The Bad

- Ability to cluster documents has two loopholes
 - Not always meaningful
 - the researcher still needs to recheck identified theme
- Frequencies of words not that helpful

Further Exploration

- Base analysis units on something else other than terms
 - “Event” vs “key word”
 - InFact by Insightful Corporation

SAS Interface



Listing of Equivalent Terms

The screenshot shows a software interface with a dialog box titled "View and Edit Equivalent Terms" and a table of terms below it. A red circle highlights the dialog box.

View and Edit Equivalent Terms

Select terms to treat separately:

Equivalent terms to grade

- graded
- grading
- grades

OK Cancel

2,246 Terms ☒ Display dropped terms ☒ Display kept terms Filter Find Similar

Term	Freq	# Documents	Keep	Weight	Role
in	638	309	Y	0.075	Prep
+ teacher	470	283	Y	0.085	Noun
+ not	511	277	Y	0.093	Part
+ have	545	272	Y	0.099	Verb
+ class	413	246	Y	0.114	Noun
+ grade	390	227	Y	0.128	Noun
on	345	218	Y	0.128	Prep
+ grade	282	208	Y	0.131	Verb
+ do	254	187	Y	0.151	Aux
unfair	205	170	Y	0.159	Adj
+ would	258	144	Y	0.202	Aux

SAS Text Miner: EM Clustering

The screenshot displays the SAS Text Miner Results window, which is divided into several sections for analyzing document data.

416 Documents

Unfair

	Student#	Age	Gender	Pres
In my senior year English class, every essay I wrote received roughly a "B" for a grade. I began to see a pattern th	1	19	1	1
In a previous English class the teacher did not deduct points from papers that were handed in late and also allowe	2	18	2	1
My Sociology class, senior year in High school was a class in which the teacher graded unfairly. This teacher grad	3	18	1	1
In my statistics class last year my teacher never graded our homework the only thing he would base our grades o	4	18	2	1
In my senior year, I took a DECA class, which is business and marketing concepts. My teacher was new becaus	5	18	2	1
In my senior year English class, I felt my teacher graded our essays & tests unfairly. The instructor never informed	6	18	1	1
In one of my high school classes my history teacher would always grade papers/projects unfairly. He would not c	7	18	1	1
In Sociology, we were given an assignment sheet at the beginning of the course and it listed due dates for variou	8	18	1	1
In my history class, 30% of the final grade is our midterm, 30% is the final exam, and 30% is 1 project. The other 10	9	18	2	1
In my math class senior year my teacher would take off random points from questions for no apparent reason. She	10	18	2	1
An experience that I have had with unfair grading is when a professor of mine basically graded me on all of my pa	11	18	1	1
I remember a previous writing class in which the teacher did not evaluate improvement and solely on the skill in w	12	18	1	1
In high school I had a teacher whose policy was to only give a few A's. She thought that it would make everyone	13	18	1	1

2,248 Terms

☐ Display dropped terms
☒ Display kept terms

Term	Freq	# Documents	Keep	Links	Weight	R
in	638	309	Y	0	0.075	P
+ teacher	470	283	Y	2	0.085	N
+ not	511	277	Y	0	0.093	P
+ have	545	272	Y	0	0.099	V
+ class	413	246	Y	1	0.114	N
+ grade	390	227	Y	0	0.128	N
on	345	218	Y	0	0.128	P
+ grade	282	208	Y	0	0.131	V
+ do	254	187	Y	0	0.151	A
unfair	205	170	Y	0	0.159	A
+ would	258	144	Y	0	0.202	A
+ student	245	143	Y	3	0.205	N
+ feel	190	134	Y	1	0.205	V
+ paper	276	130	Y	17	0.221	N

Clusters

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	+ paper, + class, + teacher, + do, + have	78	19%	*****
2	+ grade, unfair, + paper, on, + teacher	51	12%	*****
3	+ test, + do, + not, on, + class	78	19%	*****
4	response, no, + woman, + class, + have	10	2%	*****
5	+ give, + grade, unfair, + class, in	83	20%	*****
6	+ grade, + paper, + do, + not, + teacher	55	13%	*****
7	+ group, + work, + do, + grade, + class	37	9%	*****
8	wrong, + answer, + test, + would, + not	24	6%	*****

Output - (Untitled) Log - (Untitled) Program Editor - (U... Text Miner Results

SAS Text Miner: EM clustering

20 descriptive terms

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	off, + point, + writing, + essay, + homework, english class, + make, + time, + assignment, + answer, + paper, + feel, + good, + write, + give, should, no, + class, + teacher, unfair	78	19%	0.1066333394
2	other students, english teacher, english, own, + style, english, + opinion, + work, + writing, + receive, + write, grading, other, with, + grade, unfair, + paper, + base, high school, + like	51	12%	0.103921704
3	+ test, quizzes, + test, + question, + fail, + thing, can, well, + do, + not, but, + hard, on, when, all, fair, should, + good, + like, + give	78	19%	0.1065503221
4	response, no, + woman, + class, + have	10	2%	0.0512804746
5	+ assignment, only, + year, + make, same, grading, + know, could, more, + give, + grade, other, as, + bad, + receive, grading, + essay, never, unfair, with	83	20%	0.1035279399
6	always, unfairly, fair, + look, never, + grade, + like, 's, into, + work, like, + paper, + feel, + do, + not, + student, + bad, + would, more, + year	55	13%	0.1023508415
7	+ group, + project, + professor, + work, different, people, out, + evaluate, + do, + opinion, when, + time, unfairly, + base, should, other, as, + work, + receive, same	37	9%	0.1047414592
8	wrong, partial, partial credit, choice, multiple choice, wrong answers, + answer, credit, multiple, math, + test, bentley, + problem, but, + would, + thing, high school, + give, no, + feel	24	6%	0.0970475959

Hierarchical Clustering

416 Documents

Unfair

In my senior year English class, every essay I wrote received roughly a "B" for a grade. I began to see a pattern th
In a previous English class the teacher did not deduct points from papers that were handed in late and also allowe
My Sociology class, senior year in High school was a class in which the teacher graded unfairly. This teacher grad
In my statistics class last year my teacher never graded our homework the only thing he would base our grades on
In my senior year, I took a DECA class, which is business and marketing concepts. My teacher was new because
In my senior year English class, I felt my teacher graded our essays & tests unfairly. The instructor never informed o
In one of my high school classes my history teacher would always grade papers/projects unfairly. He would not car
In Sociology, we were given an assignment sheet at the beginning of the course and it listed due dates for various
In my history class, 30% of the final grade is our midterm, 30% is the final exam, and 30% is 1 project. The other 10%
In my math class senior year my teacher would take off random points from questions for no apparent reason. She
An experience that I have had with unfair grading is when a professor of mine basically graded me on all of my pap
I remember a previous writing class in which the teacher did not evaluate improvement and solely on the skill in whi
In high school I had a teacher whose policy was to only give a few A's. She thought that it would make everyone
Last year I was enrolled in a computer class in which I felt the teacher's grading was extremely unfair. I remember d
In a previous math class in high school. My teacher would grade tests by looking at the final answer from each pro
During my sophomore year in English a teacher almost failed me because he believed my work on an essay was un
In my high school math class my teacher graded some of us unfairly. I worked very hard in high school especially i
I had an English class in high school where the teacher graded work unfairly. She only liked one style of writing an
I had one teacher who never seemed to take the time to read an entire paper. He would make comments on the fi
I believe that my English teacher, senior year of high school, graded everyone's papers unfairly. She would read m
A previous class in which I felt the teacher graded students' work in an unfair way was in American History AP class

Student#

Age

Gender

Present Writing Class

Semester at Bentley

Fair

2,248 Terms

☐ Display dropped terms
☒ Display kept terms

Term

Freq

Documents

Keep

Links

Weight

Role

Attribute

in

638

309 Y

0

0.075

Prep

Alpha

+ teacher

470

283 Y

2

0.085

Noun

Alpha

+ not

511

277 Y

0

0.093

Part

Alpha

+ have

545

272 Y

0

0.099

Verb

Alpha

+ class

413

246 Y

1

0.114

Noun

Alpha

+ grade

390

227 Y

0

0.128

Noun

Alpha

on

345

218 Y

0

0.128

Prep

Alpha

+ grade

282

208 Y

0

0.131

Verb

Alpha

+ do

254

187 Y

0

0.151

Aux

Alpha

unfair

205

170 Y

0

0.159

Adj

Alpha

+ would

258

144 Y

0

0.202

Aux

Alpha

+ student

245

143 Y

3

0.205

Noun

Alpha

+ feel

190

134 Y

1

0.205

Verb

Alpha

+ paper

276

130 Y

17

0.221

Noun

Alpha

+ give

153

121 Y

3

0.217

Verb

Alpha

when

136

111 Y

0

0.230

Conj

Alpha

+ do

143

107 Y

9

0.239

Verb

Alpha

but

120

102 Y

0

0.244

Conj

Alpha

with

116

100 Y

0

0.244

Prep

Alpha

+ work

131

99 Y

9

0.253

Noun

Alpha

+ good

112

97 Y

0

0.249

Adj

Alpha

high school

90

88 Y

14

0.259

Noun

Unknown

Clusters

TOP

+ not, + grade, + grade, on, + do
406

+ paper, + teacher, + grade, + grade, + do
297

+ point, + feel, on, + teacher, + not
40

+ have, + grade, + paper, + do, + grade
257

+ test, + have, + class, in, on
109

+ group, + work, + do, + feel, + grade
13

+ test, on, + not, + teacher, + class
96

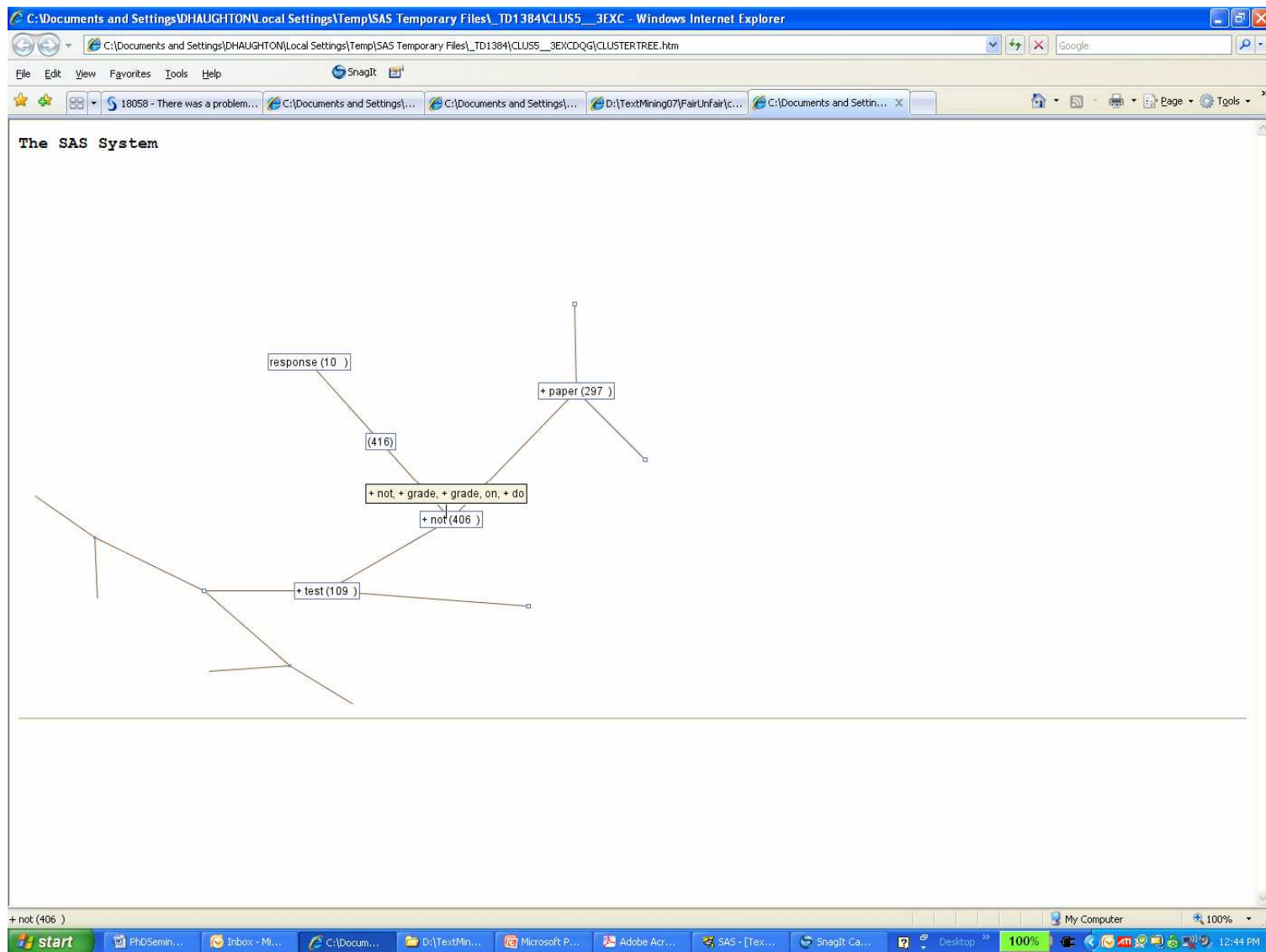
+ do, + teacher, + grade, in, + not
34

+ assignment, + grade, + problem, + do, + grade
22

partial, partial credit, credit, wrong, + would
12

+ test, on, + grade, unfair, + class

Hierarchical Clustering: Tree View



Concept Link Tree

