Toward modeling a prototypical use of language learning strategies with decision tree-based methods

Atsushi MIZUMOTO
Osamu TAKEUCHI
(Kansai University, Japan)
Outline

• Background
  (Brief history of LLS research)
• Proposing the use of decision tree-based methods
• Introducing a practical web application
Brief History of LLS Research
Research History

- **1970’s “Good language learner’s studies”**
  → Naiman et al. (1978), Rubin (1975), Stern (1975)

- **1980’s–1990’s “Identification & classification”**

- **2000’s “Call for task-specific strategies”**

- **2010’s “Redefinition of LLS (inclusion of SRL)”**
Definition of LLS

- Oxford (1990, p. 8)  
  “learning strategies are specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations”

- O’Malley & Chamot (1990, p. 1)  
  “the special thoughts or behaviors that individuals use to help them comprehend, learn, or retain new information”
“Do learning strategies exit?”

Dörnyei (2005, p. 164) later solved in 2015…

• “The most fundamental one (issue) is this: What exactly is the difference between engaging in an ordinary learning activity and a strategic learning activity?”

• looking at a vocab list = (ordinary) learning? highlighting the words = strategy?
“What is a strategy?”

Cohen and Macaro (2007, p. 278)

“this question is still being asked after more than 30 years.”

→ Redefinition of LLS
Macaro (2006)

“goals and motivations are so integral to strategies that to change them changes the actual nature of the strategy itself” (p. 323).
Dörnyei (2005, p. 165)

- Distinguishing features of LLS: "if we define the strategic quality of learning with goal-oriented, intentionally evoked, and effortful behavior then we, in effect, equate 'strategic' with 'motivated,' because goal-oriented, intentionally evoked, and effortful are three key features of motivation."
Self-regulated Learning

Zimmerman (1989) “metacognitively, motivationally, and behaviorally active participants in their own learning process”

(a) Use of self-regulated learning strategies
(b) Self-efficacy
(c) Goal orientations
Griffiths (2008, 2013)

“activities consciously chosen by learners for the purpose of regulating their own language learning.”
“self-regulated L2 learning strategies are defined as deliberate, goal-directed attempts to manage and control efforts to learn the L2. These strategies are broad, teachable actions that learners choose from among alternatives and employ for L2 learning purposes” (p. 12).
Ehrman, Leaver, and Oxford (2003)

"A given learning strategy is neither good nor bad; it is essentially neutral until it is considered in context” (p. 315).

A strategy is useful when:
(a) the strategy relates well to the L2 task at hand,
(b) the strategy fits the particular student's learning style preferences to one degree or another
(c) the student employs the strategy effectively and links it with other relevant strategies
• Descriptions and taxonomies of LLS
• Relationship with other variables
• Strategy instruction = positive results
• LLS definition and measure, "elusive" and "fuzzy"
  → Redefinition of LLS by including SRL
Language Learning Strategies

Learners

Meta/Strategies
“Using the senses to remember”

Task

Context
Goals
Individual Differences

Tactics
“I write the new words many times.”
(qualitative > quantitative)
What teachers and learners actually want to know...

<table>
<thead>
<tr>
<th>Learners</th>
<th>Meta/Strategies</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“Using the senses to remember”</td>
<td>“I write the new words many times.”</td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td>(qualitative &gt; quantitative)</td>
</tr>
<tr>
<td>Goals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Differences</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
My students (teachers):

• “I gave SILL (Strategy Inventory for Language Learning) to my students to check their strategy use.”

• “I still teach them (tactics) as strategies because my students need to know them.”
Questionnaires still have a role in LLS research and practice.
Questionnaires still have a role in LLS research and practice.

- Strategy research
- Strategy instruction (initial repertoire)
- Awareness raising
- Introducing strategies (tactics)
Proposing the use of decision tree-based methods
Macaro (2006)

- Successful learning ≠ frequency of strategy use
- Successful learning = orchestration of strategies
- Orchestration = higher levels of metacognition
Strategy Chains and Clusters

Cohen & Weaver (2005)

- **Strategy chains** — in a predictable sequence
- **Strategy clusters** — simultaneously
Means do not tell us much.

<table>
<thead>
<tr>
<th>Questionnaire (Scale)</th>
<th>Subscales</th>
<th>No. of Items</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self-management</td>
<td>7</td>
<td>2.36</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Input-seeking</td>
<td>4</td>
<td>2.23</td>
<td>0.95</td>
</tr>
<tr>
<td>Vocabulary Learning</td>
<td>Imagery</td>
<td>5</td>
<td>2.80</td>
<td>0.87</td>
</tr>
<tr>
<td>Strategies</td>
<td>Writing Rehearsal</td>
<td>3</td>
<td>3.80</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Oral Rehearsal</td>
<td>3</td>
<td>2.69</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>Association</td>
<td>3</td>
<td>2.55</td>
<td>0.90</td>
</tr>
<tr>
<td>Motivation</td>
<td>Extrinsic Motivation</td>
<td>3</td>
<td>3.96</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Intrinsic Motivation</td>
<td>6</td>
<td>3.15</td>
<td>0.86</td>
</tr>
</tbody>
</table>
cluster analysis

Mizumoto & Takeuchi (2008)

Still, it’s means of each cluster (group).
Advantages of tree-based methods

• **Very flexible** (categorical, ordinary, continuous)
• Non-parametric method
• Visualizable
• Intuitive
Decision tree
Random forest
Example

- 711 Japanese university EFL learners
- Explanatory variables
  Self-management, Input-seeking, Imagery, Writing, Oral, Association, Goal-setting, Volitional control, Satisfaction
- Self-efficacy as an criterion (outcome) variable.
Random forest

```r
randomForest(formula = SE ~ ., data = dat)

Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 3

Mean of squared residuals: 0.8298364
% Var explained: 30.37

```

![Graph showing error decrease with increasing number of trees](graph.png)
Random forest

![Graph showing IncNodePurity vs. various attributes including Satisfaction, Volition, Self.manage, Input, Association, Goal, Oral, Imagery, and Write.](image-url)
Multiple regression analysis

| Coefficients: | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|---------|
| (Intercept)   | 0.40383  | 0.13634    | 2.962   | 0.003159|
| Self.manage   | 0.11300  | 0.04366    | 2.588   | 0.009840|
| Input         | 0.17739  | 0.02993    | 5.927   | 4.83e-09 |
| Imagery       | -0.06185 | 0.03738    | -1.655  | 0.098380|
| Write         | -0.07244 | 0.02448    | -2.959  | 0.003190|
| Association   | 0.09261  | 0.03152    | 2.938   | 0.003407|
| Volition      | 0.13737  | 0.04078    | 3.368   | 0.000798|
| Satisfaction  | 0.35576  | 0.04609    | 7.718   | 4.06e-14|
Yet other advantages of tree-based methods

• prototypical use of LLS (in a given context)

• IF-THEN rule — more diagnostic feedback for teachers and learners

• Applicable to any quantitative LLS measures (e.g., a practical classroom tool such as Language Strategy Use Inventory: Cohen, Oxford, & Chi, 2005)
Tree-based methods?

Easier said than done?
Introducing a practical web application
Cronbach's Coefficient Alpha

Option:
- The input data includes variable names (header).

Note: Input values must be separated by tabs. Copy and paste from Excel/Numbers.

Basic statistics of the scale (test)

<table>
<thead>
<tr>
<th>Item</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>3.45</td>
<td>0.77</td>
<td>3.5</td>
<td>3.48</td>
<td>0.37</td>
<td>1.75</td>
<td>4.75</td>
<td>3</td>
<td>-0.22</td>
<td>-0.09</td>
<td>0.2</td>
</tr>
</tbody>
</table>
R as a Lingua Franca: Advantages of Using R for Quantitative Research in Applied Linguistics

Atsushi Mizumoto and Luke Plonsky

Abstract

In this article, we suggest that using R, a statistical software environment, is advantageous for quantitative researchers in applied linguistics. We first provide a brief overview of the reasons why R is popular among researchers in other fields and why we recommend its use for analyses in applied linguistics. In order to illustrate these benefits, we report recent works and developments in quantitative data analysis seeking to move the field toward more appropriate practices, many of which take advantage of the flexibility and functionality of R. Finally, in order to facilitate the use of R, we also introduce an R-based web application developed by the first author.

© Oxford University Press 2015
Tools and Resources

The following links provide access to very user-friendly programs for conducting many of the analyses described in this chapter. The first, the langtest.jp developed by Atsushi Mizumoto, is an R- and web-based app (http://langtest.jp/); the second, ESCI (http://www.latrobe.edu.au/psy/research/cognitive-and-developmental-psychology/esci), is a set of freely downloadable Excel macros designed by Geoffrey Cumming to help researchers consider and report results with an emphasis on effect sizes and CIs.

Second Edition

Helpful Websites

langtest.jp

Atsushi Mizumoto, Professor of Applied Linguistics at Kansai University, has created a very helpful website for using R without having to actually download R or R Commander. The Web App page gives ways to calculate descriptive statistics, statistical tests, appropriate graphics, tests of normality, effect sizes, and confidence intervals for correlation, t-tests, regression, chi-square, non-parametric tests, one-way and two-way ANOVA, cluster analysis, exploratory factor analysis, principal component analysis, correspondence analysis, structural equation modeling, classical test theory item analysis, and IRT models, among others. The user enters their own data and results appear. R code is also available if you want to take the code and modify it yourself. I would have made use of this resource in the book had I been aware of it earlier, but as it is, I heartily recommend it for taking much of the work out of doing statistics in R.

http://langtest.jp:3838/mes/

This specific website, within Mizumoto’s “langtest” calculates effect sizes as well as any other online effect size calculator I know and uses summary stats of sample size, mean and standard deviation. In addition, it returns a 95% CI of the mean of the differences, a t-test result, and effect sizes.
Web App

*Cronbach's Coefficient Alpha* (mirror)

*Basic Statistics Calculator* (mirror)

*Learning by Doing Stats (t-test Tutorial)* (mirror)

*Comparing Two Independent Samples* (mirror)

*Comparing Paired Samples* (mirror)

*Effect Size Calculator 1 (Means, SDs, Ns)* (mirror)

*ANOVA* (mirror)

*Non-parametric Tests* (mirror)

*Correlation* (mirror)

*Regression Analysis* (mirror)

*Chi-square Test* (mirror)

“Copy, paste and you’re done.”
Summary

Changing definition of LLS

Unchanged questionnaire use

Decision tree-based methods as an alternative to reflect individual differences in a given context