Predicting Performance on MOOC Assessments using Multi-Regression Models

Zhiyun Ren, Huzefa Rangwala, Aditya Johri

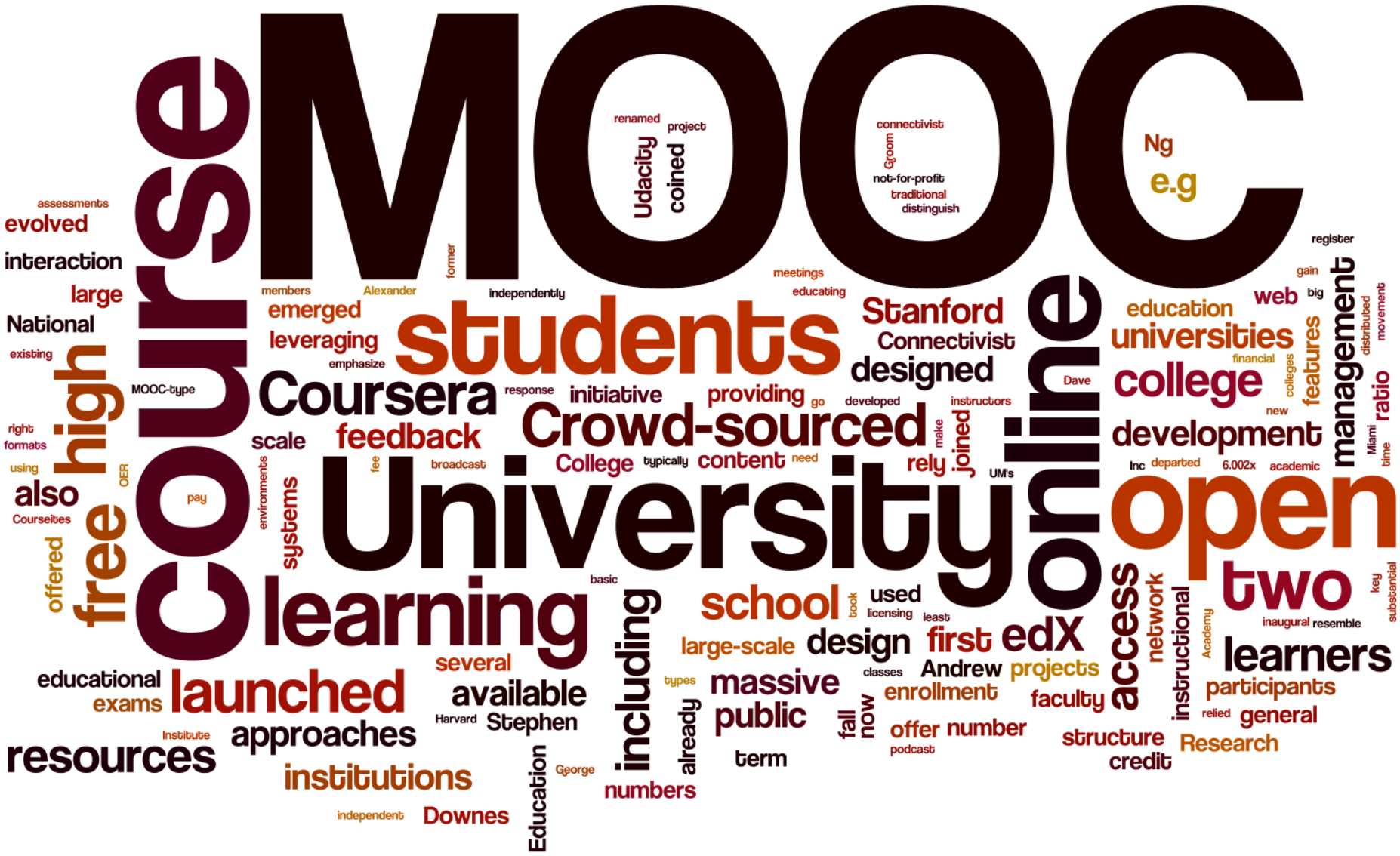
George Mason University
4400 University Drive, Fairfax, Virginia 22030

Outline

- ❑ Background
- ❑ Personal Linear Multi-Regression Models
- ❑ Feature selection
- ❑ Experiments and discussion
- ❑ Conclusion and future work

GEORGE MASON UNIVERSITY

Background


MOOC
course
students
University
online
open
learning
Coursera
Crowd-sourced
high
free
launched
resources
approaches
institutions
feedback
Stanford
Connectivist
designed
universities
college
development
management
two
learners
access
school
design
first
edX
massive
public
including
available
Stephen
Downes
Education
numbers
renamed
project
Udacity
coined
connectivist
Groom
not-for-profit
traditional
distinguish
Ng
e.g
assessments
evolved
interaction
large
National
existing
MOOC-type
right
formats
using
also
Courseites
offered
OER
pay
emerged
leveraging
emphasize
members
Alexander
former
independently
meetings
educating
response
initiative
providing
go
developed
instructors
Dave
education
web
big
gain
register
features
colleges
financial
new
distributed
movement
ratio
Miami
time
scale
College
typically
content
need
rely
joined
make
UM's
Inc
departed
6.002x
academic
broadcast
fee
environments
systems
basic
used
licensing
took
least
large-scale
several
educational
exams
Institute
Harvard
George
independent
already
term
types
classes
Andrew
projects
enrollment
faculty
fall
now
offer
number
podcast
structure
credit
network
instructional
Academy
participants
relied
general
Research
inaugural
resemble
key
substantial

Background

coursera

Courses About Andre Gabriel Garza

Stanford ONLINE

Human-Computer Interaction

- Home
- Video Lectures
- Assignments
- Quizzes
- Discussion Forums
- About the Course
- Syllabus and Calendar
- Tracks, Grading, & Statements of Accomplishment
- Design Project Overview
- Design Briefs
- How to: Peer Assess

Video Lectures

Having trouble viewing lectures? Try changing your player in course preferences.

Week 1 -- Introduction

- Lecture 1.1: Human Computer Interaction (4:18)
- Lecture 1.2: The Power of Prototyping (13:49)
- Lecture 1.3: Evaluating Designs (12:15)
- Lecture 1.4: The Birth of HCI (8:48)

Week 2 -- Needfinding

- Lecture 2.1: Participant Observation (12:55)
- Lecture 2.2: Interviewing (11:37)
- Lecture 2.3: Additional Needfinding Strategies (11:54)

Week 3 -- Rapid Prototyping

- Lecture 3.1: Paper Prototypes and Mockups (12:47)

Overview

- **Information we have:** MOOC server log
- **Things we want to do:** Predict student's performance

GEORGE MASON UNIVERSITY

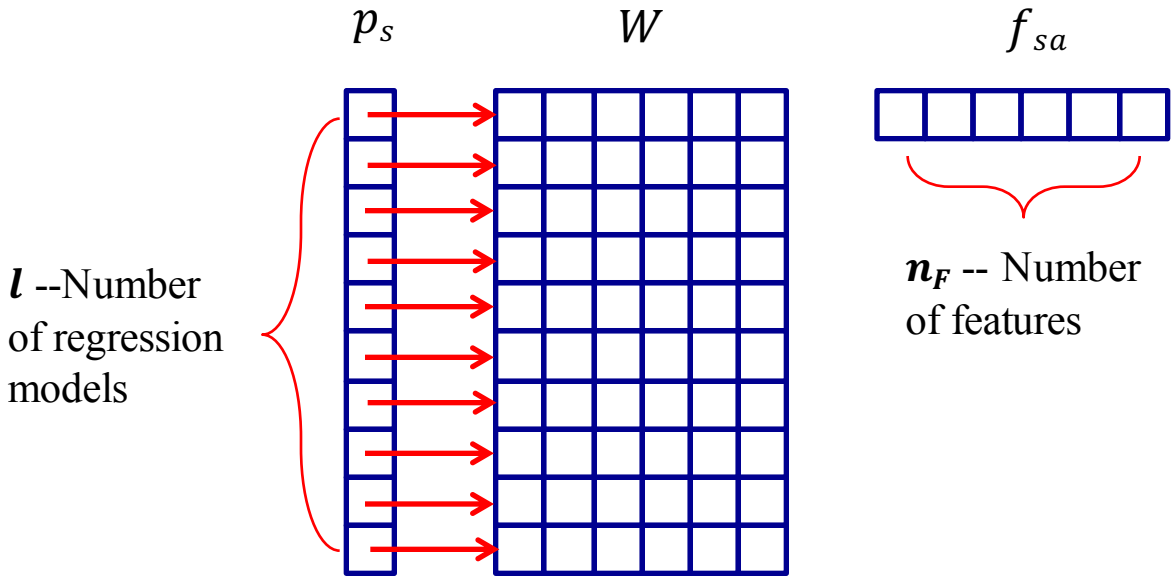
Challenge

- Various kinds of participants
- High attrition rate
- Flexible timetable
- **Baselines we have tried:** Linear regression model, meanscore

GEORGE MASON UNIVERSITY

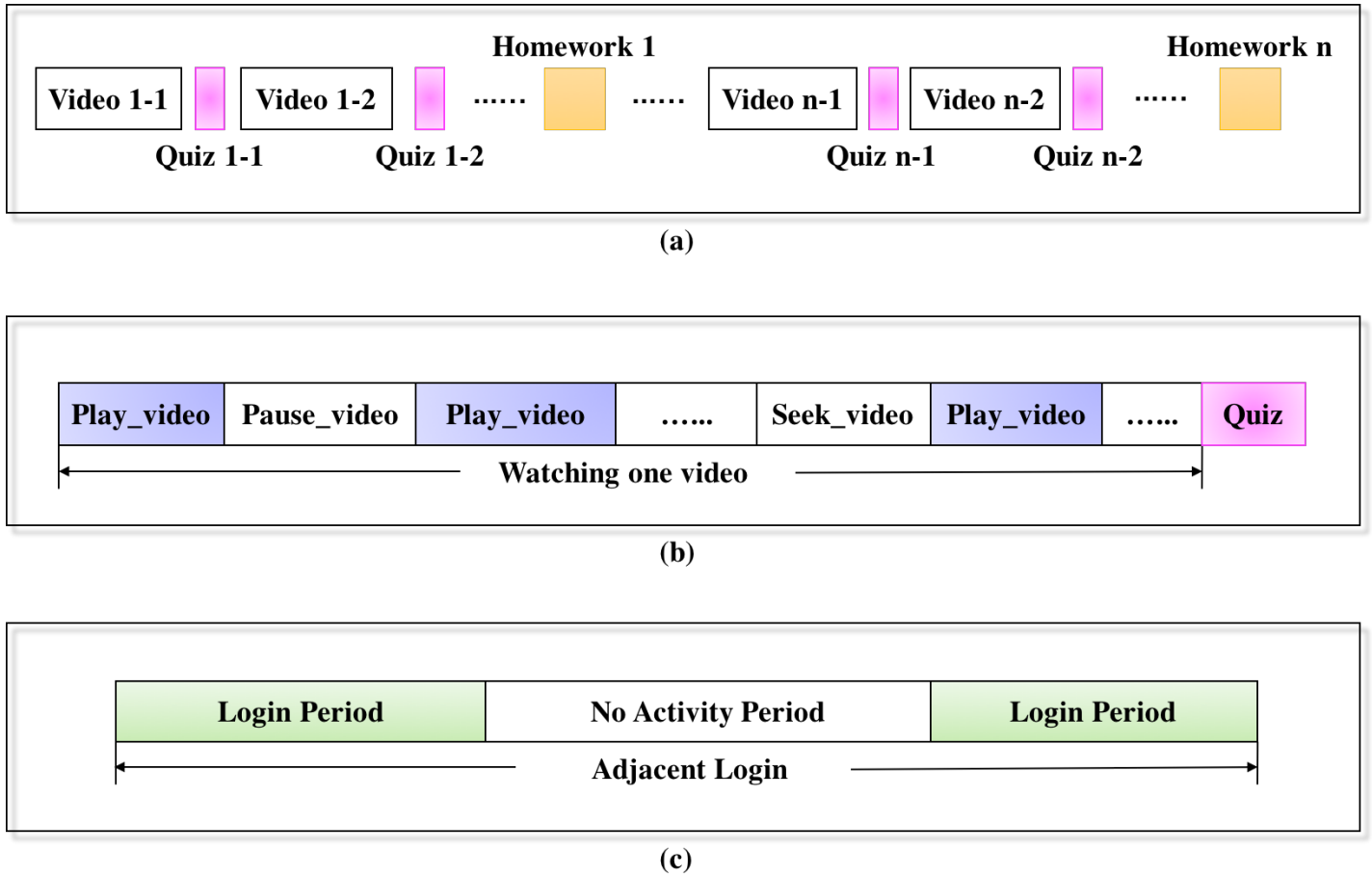
Personal Linear Multi-Regression Models

$$\widehat{g_{s,a}} = b_s + b_c + p_s^t \mathcal{W} f_{sa} = b_s + b_c + \sum_{d=1}^{l} (p_{s,d} \sum_{k=1}^{n_F} f_{sa,k} \omega_{d,k})$$



$$\underset{(W,P,B)}{minimize} \frac{1}{2N} \sum_{i=1}^{N} (g_{s,a} - \widehat{g_{s,a}})^2 + \lambda(\|P\|_F^2 + \|W\|_F^2) + \gamma(\|P\| + \|W\|)$$

Data structure

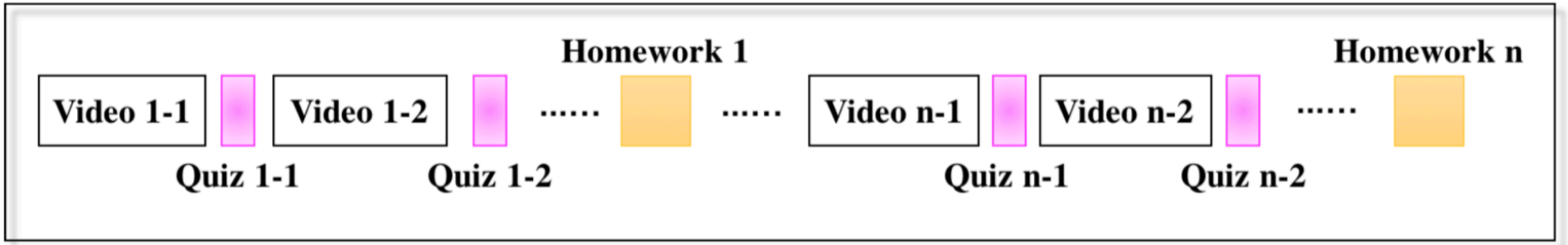


(a) Homework and quiz

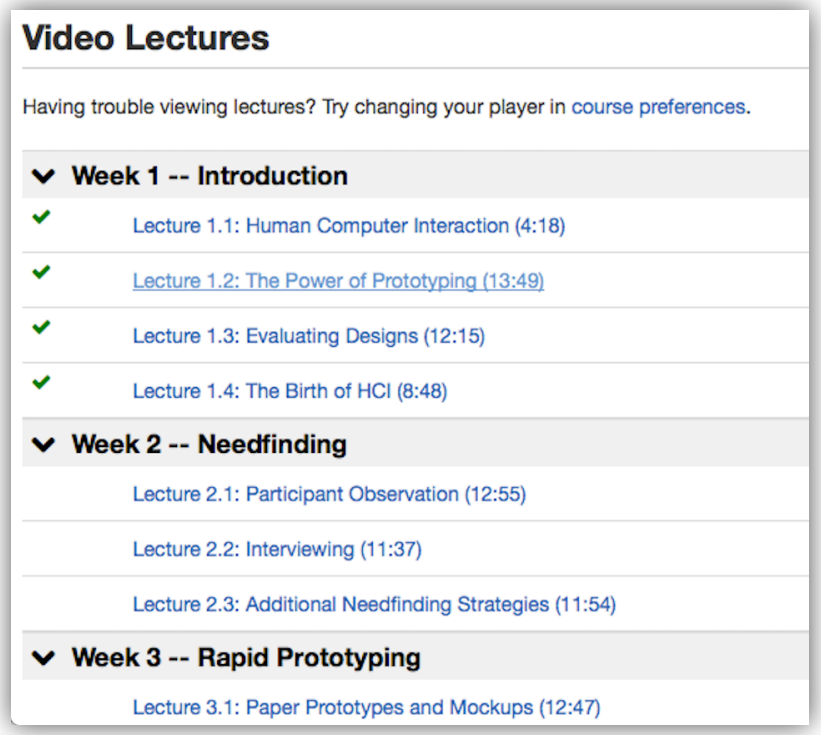
(b) Video

(c) Study session

Feature selection



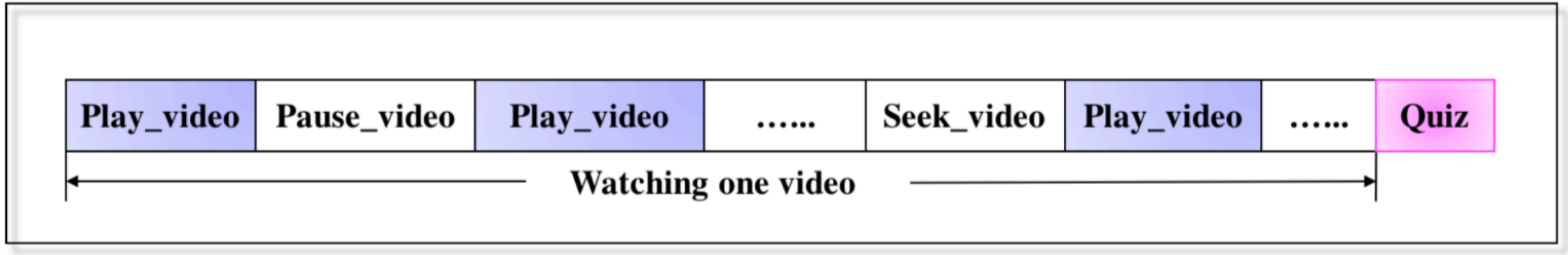
(a)



- ❑ quiz related features
- ❑ time related features
- ❑ interval-based features
- ❑ homework related features

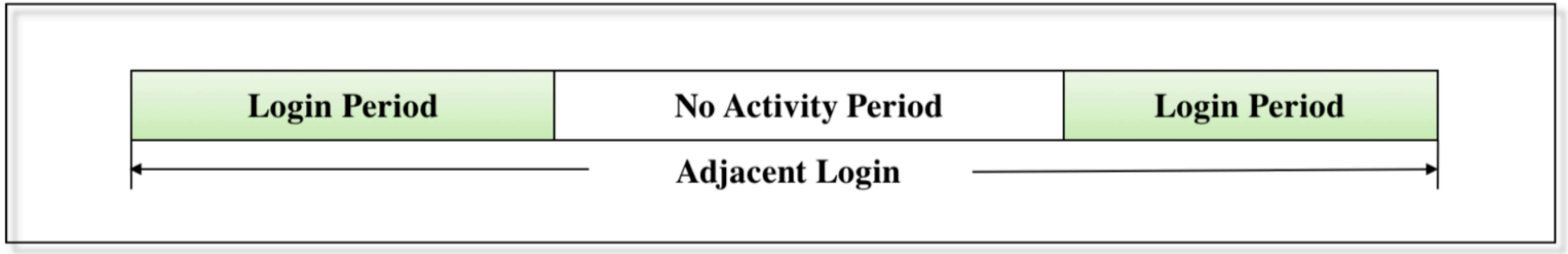
GEORGE MASON UNIVERSITY

Feature selection



(b)

- Video related features



(c)

- Session features

Experimental setup

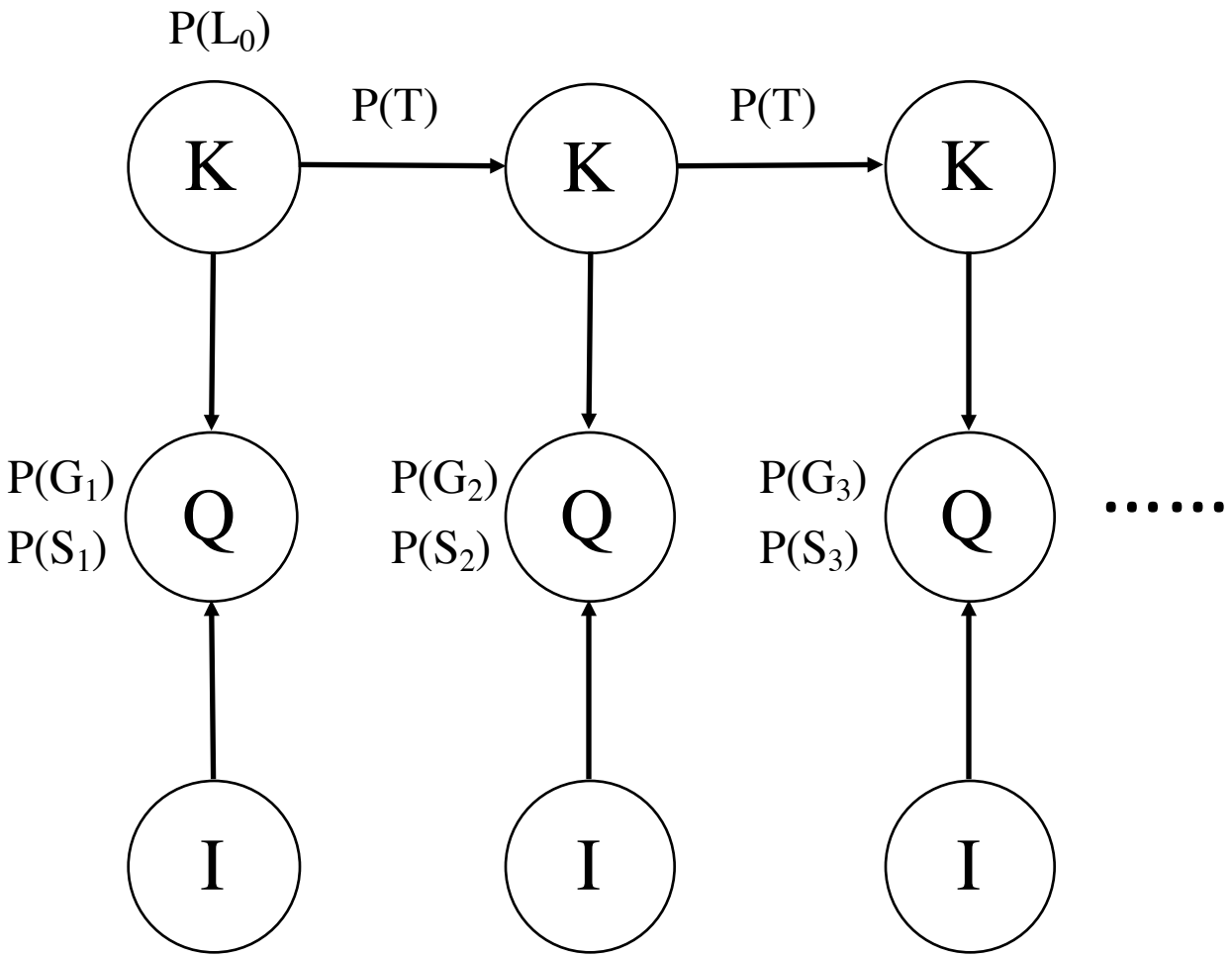
- Different motivations part the data into two groups.
- Different models are applied for different data types.

Experimental protocol

- PreviousHW-based prediction

- PreviousOneHW-based prediction

Experimental baseline: KT-IDEM

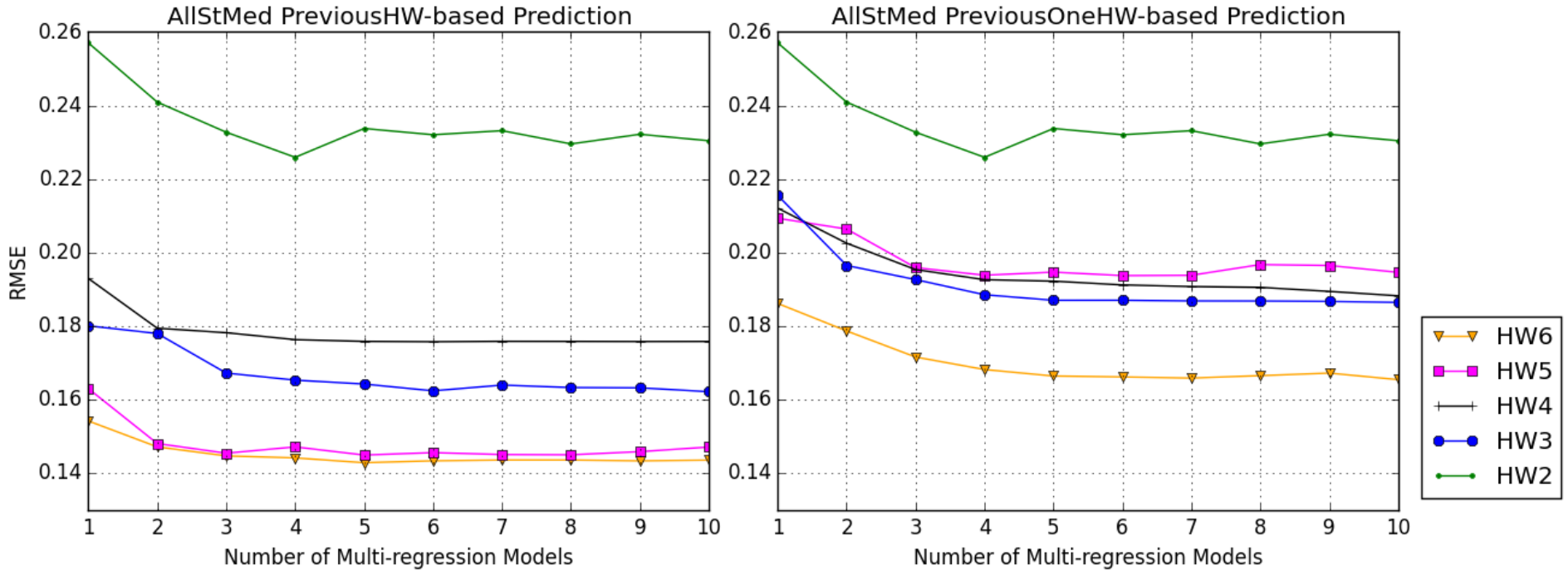


Model parameters

$P(L_0)$ = Initial Knowledge
$P(T)$ = Probability of learning
$P(G_{1...n})$ = Probability of guess per question
$P(S_{1...n})$ = Probability of slip per question

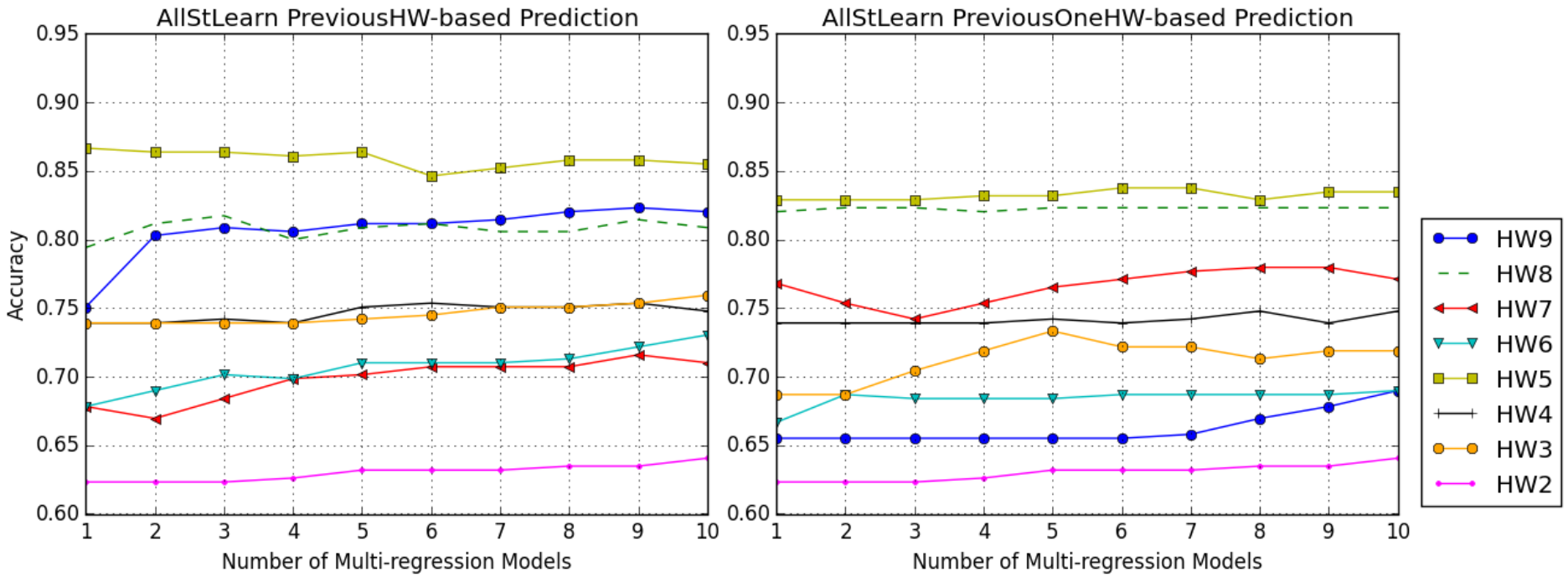
n denotes the number of all questions.

Comparative Performance



- Prediction results with varying number of regression models for student group with continuous grade value

Comparative Performance



- Prediction results with varying number of regression models for student group with binary grade value

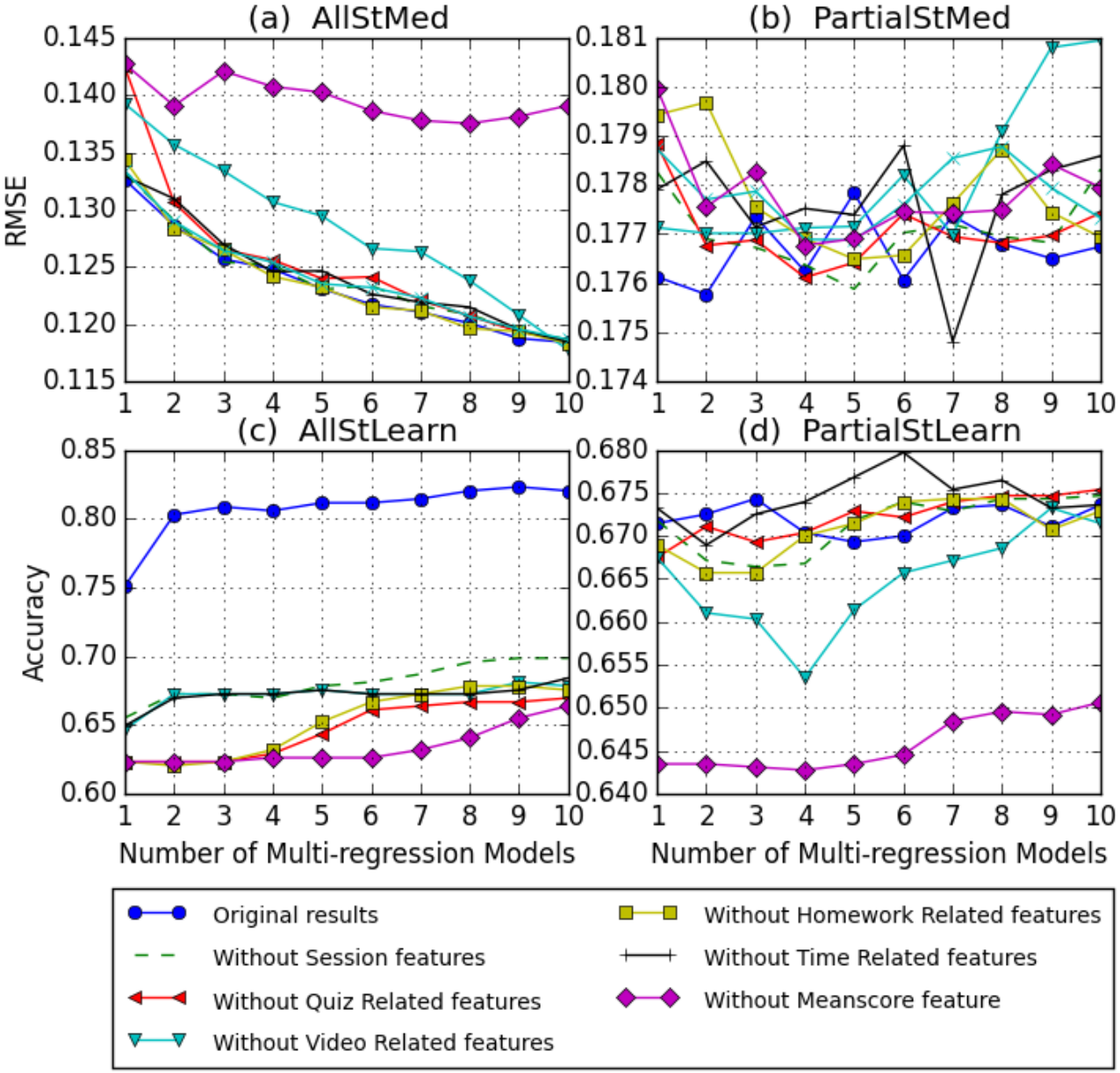
Comparative Performance

HW#	Accuracy (↑)			F_1 (↑)		
	PLMR	Baseline		PLMR	Baseline	
		Meanscore	KT-IDEM		Meanscore	KT-IDEM
2	0.641	**0.646**	0.623	0.775	**0.777**	0.768
3	**0.760**	0.580	0.681	**0.821**	0.805	0.810
4	**0.754**	0.710	0.739	0.838	0.706	**0.850**
5	**0.867**	0.809	0.829	**0.920**	0.880	0.906
6	**0.730**	0.678	0.667	**0.808**	0.776	0.800
7	0.716	0.675	**0.730**	**0.887**	0.878	0.844
8	**0.817**	0.762	0.817	**0.903**	0.849	0.886
9	**0.823**	0.794	0.777	**0.864**	0.856	0.853
Avg	**0.764**	0.707	0.759	**0.852**	0.816	0.848

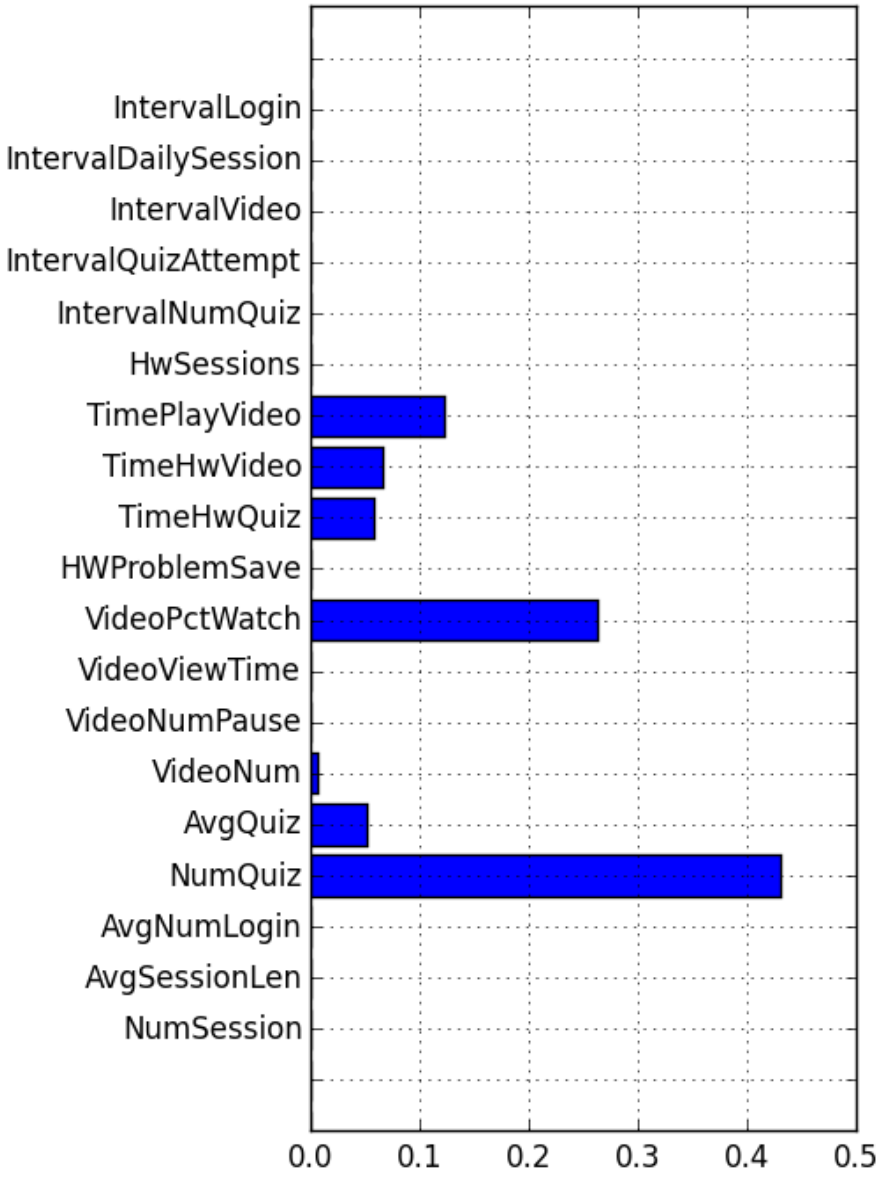
- The comparison of the accuracy and F1 scores with baseline approaches.

GEORGE MASON UNIVERSITY

Feature Importance



Feature Importance



Conclusion and future work

- **Predict algorithm:** personalized multiple linear regression model.
- **Experimental results:** improved performance compared to baseline methods.
- **Other contribution:** analysis of feature importance.
- **Future work:** to set up an early warning system to help improve student's performance

GEORGE MASON UNIVERSITY

Thank you!