

Regresor Profundo: Materia Cruzada Académica

Sistema de Predicción de Rendimiento para la Universidad

Estudiantes de nivel

B. Raveendran Pillai, GauthamJ

Resumen— La predicción del rendimiento académico de los estudiantes ha sido un importante tema de investigación en el campo de la Educación. El objetivo principal de una institución de educación superior es proporcionar una educación de calidad a los estudiantes. Una forma de lograr un mayor nivel de calidad de la educación es prediciendo el rendimiento académico de los estudiantes y tomando acciones correctivas tempranas para mejorar el mismo. Este artículo presenta un sistema que utiliza técnicas de aprendizaje automático para clasificar y predecir el rendimiento académico de los estudiantes en el momento justo antes de que ocurra la deserción. El sistema primero acepta los parámetros de rendimiento de los cursos de nivel básico que el estudiante ya había aprobado, ya que estos parámetros también influyen en el estudio posterior. Para predecir el desempeño del programa actual, el sistema acepta continuamente los parámetros de desempeño académico después de cada proceso de evaluación académica. El sistema emplea técnicas de aprendizaje automático para estudiar el rendimiento académico de los alumnos tras cada proceso de evaluación. El sistema también aprende las reglas básicas que sigue la Universidad para evaluar a los estudiantes. Basado en el desempeño actual de los estudiantes, el sistema clasifica a los estudiantes en diferentes niveles e identifica a los estudiantes de alto riesgo. Una predicción más temprana puede ayudar a los estudiantes a adoptar medidas adecuadas por adelantado para mejorar el rendimiento. Ce. Los sistemas también pueden identificar el factor que afecta el rendimiento de los mismos estudiantes, lo que les ayuda a tomar medidas correctivas con anticipación.

Palabras clave: aprendizaje automático, rendimiento académico, profundo aprendizaje, métodos de clasificación-regresión, keras, tensorflow

I. INTRODUCCIÓN

Muchos estudios se han realizado en el campo del e-learning para analizar el rendimiento académico de los estudiantes y la predicción temprana de la deserción en los cursos académicos. Esto ayudaría a los estudiantes a tomar las medidas correctivas adecuadas antes de que se produzca la deserción. En paper[1] los autores S. Kotsiants. C Pierrakeas y P Pintelas investigaron el rendimiento académico del programa de educación a distancia para predecir estudiantes desertores. Se utilizan técnicas de ganancia de máquina como árboles de decisión, redes neuronales y algoritmo Naïve Bayes. Después del proceso de entrenamiento y prueba, se compara el rendimiento. Finalmente, se concluyó que el algoritmo Naive Bayes resultó ser muy eficiente en la predicción de abandonos, ya que predice con un 83 por ciento de precisión. En el proceso de predicción propuesto por Lykourantzou, Giannoukos, Nikolopoulos y Loumosinpaper[2], el aprendizaje automático

técnicas; Se han utilizado máquinas de vectores de soporte y redes neuronales de avance. Para propósitos de entrenamiento y prueba, se usaron datos tanto en el tiempo como invariantes en el tiempo. En la clasificación de los alumnos basada en el rendimiento se adopta la técnica de minería de datos. Se puede realizar una evaluación continua del rendimiento de los alumnos en cada asignatura. Esto ayuda al profesor a prestar más atención al estudiante. afeitado bajo rendimiento. Un estudio utilizando Machine learning[3] sobre el rendimiento académico de Moseley y Mead para predecir la deserción de estudiantes de enfermería. En este caso se utilizó un árbol de decisión en el que se utilizó la detección automática de interacciones Chi-cuadrado. El conjunto de datos utilizado en las fases de prueba y validación es invariable en el tiempo (p. ej., edad, sexo) y variable en el tiempo (p. ej., calificaciones, asistencia). El sistema identificó a los estudiantes en riesgo y la eficiencia de predicción es del 70 por ciento. El inconveniente de este método es que la eficiencia de la predicción está solo en un nivel medio y el método no determina el rendimiento académico, sino que el sistema solo detecta a los estudiantes que abandonan.

II. REGRESO PROFUNDO O MODELO

2.1 Definición del problema

En el sistema educativo algunos estudiantes incluso con un buen la formación académica también puede debilitarse debido a muchas razones; dureza en los temas de estudio, mala enseñanza y método de presentación adoptado en la situación educativa, condiciones ambientales, falta de interés, malas condiciones de salud, influencia de los compañeros, etc. Pero si el

La institución cuenta con un sistema para monitorear consistentemente el desempeño académico de los estudiantes y adopta medidas correctivas adecuadas a tiempo para que los estudiantes con bajo desempeño puedan mejorar. Pero en un sistema manual, es muy difícil Vigilar constantemente las actividades académicas de cada estudiante, especialmente cuando el número de estudiantes es grande. Además con el sistema manual, es prácticamente muy difícil predecir con precisión la deserción de los estudiantes en base a los datos disponibles; los parámetros de valoración del desempeño adquiridos hasta la etapa actual. Los algoritmos de aprendizaje automático, incluido el aprendizaje profundo, se pueden emplear para clasificar a los estudiantes en función del rendimiento académico y predecir las deserciones con mayor precisión.

Predicción cruzada de materias Además, es difícil para un ser humano rastrear la relación entre las materias del primer semestre y las materias del segundo, tercero o cualquier otro semestre, ya que las materias no son las mismas pero

Manuscrito revisado recibido el 22 de septiembre de 2019.

Prof. (Dr.) B. Raveendran Pillai Director, PRS Facultad de Ingeniería, Trivandrum, India. ravipoykali@gmail.com

GauthamJ, 2do año de Ingeniería Mecánica, Instituto Nacional de Tecnología, Calicut, India. gauthamastro@gmail.com.

diferentes materias de la misma rama. Allí, al dificultar el análisis de un estudiante por su desempeño en el próximo semestre en función de su desempeño en el anterior
semestres. Hay varios modelos como GritNet[4] para los cuales la entrada son datos variables en el tiempo de la misma materia para la que se predice la salida. Nuestro modelo toma datos de calificaciones de algunas materias en el semestre anterior y luego predice para completamente diferentes materias. Esto se hace aprendiendo el nivel de conocimiento del estudiante de diferentes fuentes y este es uno de los puntos que destacan de nuestro modelo.

	subject0	subject1	subject2	subject3	subject4	subject5	subject6	subject7	subject8	subject9	...
0	0.8467	0.815	0.83	0.835	0.835	0	0	1	0.6	0.6	...
1	0.7741	0.815	0.66	0.705	0.665	0	0	1	0.7	0.7	...
2	0.82	0.77	0.745	0.745	0.845	0	0	1	0.85	0.7	...
3	0.8066	0.9	0.615	0.805	0.835	0	0	1	0.7	0.6	...
4	0.733	0.735	0.7	0.735	0.585	0	0	1	0.7	0.5	...

Figura 2: Representación de datos de entrada para red neuronal

2.3 Arquitectura

Objetivo El objetivo principal de este estudio es desarrollar un modelo de aprendizaje profundo para predecir tempranamente con mayor precisión la deserción de los estudiantes del 5to semestre de una carrera de ingeniería en la Universidad. Los datos de entrada son los parámetros de desempeño de los estudiantes de cada asignatura en cada semestre hasta el cuarto semestre.

Modelo El modelo Deep Regressor trata de predecir el punto de calificación para todas las materias en el semestre de salida que es una salida continuamente valorada que va de 0 a 10. Los puntos de calificación correspondientes se convierten en calificaciones según los criterios de calificación de la universidad mediante los cuales podemos predecir si un estudiante abandonará o no.

2.2 Modelo de línea de base

Para acceder a cuánto valor aporta nuestro modelo, estamos comparando nuestro modelo con los algoritmos de regresión basados en árboles de decisión más efectivos disponibles.

XGBoostRegressor[6] Este es un método de aprendizaje de conjunto y es una implementación popular de aumento de gradiente. Una de las características que hace que XGBoost esté fuera es la opción de penalizar modelos complejos a través de la regularización L1 y L2. Esto previene la modificación.

$$L_1 Loss = \sum_{i=1}^n (Y_i - \sum_{j=1}^m X_{ij} B_j)^2 + \lambda \sum_{j=1}^m B_j^2$$

$$L_2 Loss = \sum_{i=1}^n (Y_i - \sum_{j=1}^m X_{ij} B_j)^2 + \lambda \sum_{j=1}^m |B_j|$$

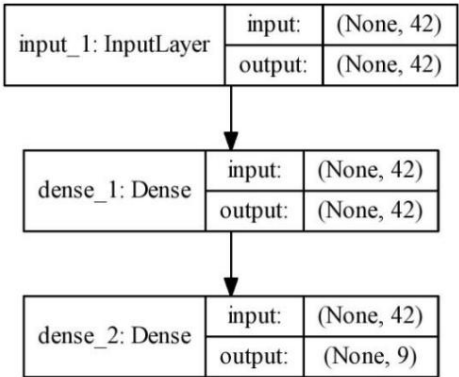


Figura 1: Arquitectura para la predicción académica de materias cruzadas del estudiante como se describe en la sección 2.3

2.3.1 Representación de entrada

Primero se limpian los datos de cualquier ausentismo y todas las calificaciones se convierten en puntos de calificación según las normas de la universidad. Luego, esto se divide por 10 para volver a escalarlo al rango [0,1]. Este tipo de representación ayuda a que la red neuronal converja rápidamente. La forma de los datos de entrada es (cantidad de estudiantes, cantidad de materias). Cada uno representó a un estudiante en particular y las columnas de sus calificaciones.

2.3.2 Arquitectura del modelo

La parte central de aprendizaje ocurre en la capa oculta (dada en la fig. 1). Esta capa aprende la importancia de las materias individuales para las materias de salida y cómo se relacionan con el grado, esto incorpora las marcas de entrada en una representación mucho más densa.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 42)	0
dense_1 (Dense)	(None, 42)	1806
dense_2 (Dense)	(None, 9)	387
Total params: 2,193		
Trainable params: 2,193		
Non-trainable params: 0		

Figura 3: Parámetros entrenables en diferentes capas

que luego es utilizado por la capa de salida que contiene 9 neuronas, ya que tenemos 9 materias diferentes para el semestre que estamos tratando de predecir. El optimizador que usamos es Adam con una tasa de aprendizaje = 0.001 y la función de activación es tanh. Esta combinación demostró ser la mejor de las múltiples combinaciones que probamos: ingrelu, sigmoid y tanh.

optimizador, pero elegimos a Adam porque obtuvimos mejores resultados en el conjunto de prueba. Se demostró que el error cuadrático medio es la mejor función de pérdida para este caso de uso.

$$L(Y_i, \hat{Y}_i) = (1/n) * \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Aquí, Yi es el valor verdadero, Y^l son las predicciones del modelo y siendo n el número de muestras utilizadas para la evaluación. La arquitectura completa se ilustra en la figura 1



tercero DATOS Y FORMACIÓN

3.1 Datos de la Universidad Técnica de Kerala (KTU)

Recopilamos nuestros datos de la Facultad de Ingeniería de PRS afiliada a la Universidad Técnica de Kerala. Tomamos puntajes de los 5 semestres anteriores y el desempeño del estudiante en más dos, que fue preprocesado y limpiado de cualquier anomalía antes de enviarlo a la red. Nos aseguramos de que los datos no contuvieran ausencias o valores nulos y que un estudiante en particular estuviera presente durante todos los semestres en el mismo orden en que apareció en los semestres anteriores y todos los semestres se concatenaron en una matriz numérica, que se introdujo en la red para la formación. De manera similar, las materias del último semestre se organizaron en orden y se redujeron y alimentaron como etiquetas. Todo el conjunto de datos se dividió en 90 por ciento y 10

Todo el conjunto de datos se dividió en un 90 por ciento y un 10 por ciento para los conjuntos de entrenamiento y prueba respectivamente. Esto aseguró que tanto los datos de entrenamiento como los datos de prueba provinieran del mismo distribución.

3.2 Formación

Tanto la línea de base como nuestro modelo se entrenaron con los mismos datos de entrenamiento y se validaron en el 50 por ciento de los datos de prueba. Basilea en emodel se entrenó con el parámetro predeterminado sas descrito en su artículo de investigación y nuestro modelo se entrenó para 3000 épocas con un tamaño de lote de 64 y los datos se mezclaron después de cada época. Se guardó el modelo de ponderación suave que mostró el error cuadrático medio más bajo en la prueba de validación.

IV. RENDIMIENTO DE PREDICCIÓN

una. Medida de evaluación

Para demostrar el poder real de este modelo, nos enfocamos en la cercanía de la predicción de puntos de calificación al punto de calificación real que obtuvo el estudiante. El conjunto de datos resultó estar desequilibrado, considerando la clasificación de grado de los estudiantes. En nuestro caso, la precisión no es una métrica precisa ya que algunas materias están dominadas por un grado en particular.

Utilizamos el error cuadrático medio (RMSE) en la salida escalada y no escalada de los modelos para evaluarlo. El RMSE escalado mostró cuánto puede variar el punto de calificación real del valor predicho. Baje el RMSE, mejor es el modelo.

4.2 Resultados

Tabla 1: Comparaciones de resultados

Modelo	Rango de error de RMSE escalado	
Base	1.5185	2
Regresor profundo	1.0657	1

Como hemos demostrado en la tabla anterior, el modelo de referencia tenía un error de predicción en el rango de 2 puntos de calificación, lo que conduce a predicciones inexactas, pero nuestro modelo simple pero poderoso pudo reducir el rango de error a 1. Esta reducción es el rango de error en una unidad tienen grandes ventajas cuando

convirtiendo de nuevo a grados donde nuestro modelo pudo predecir con una compensación de error máxima de solo 1 grado.

CONCLUSIÓN V

En este artículo, hemos aplicado con éxito el aprendizaje profundo al desafiante problema de predicción del rendimiento de los estudiantes de materias transversales que hasta ahora no ha sido explotado. También demostramos la mayor capacidad del modelo Deep Regresor sobre los algoritmos de aumento de gradiente como XG-Boost. Además, en lugar de utilizar directamente los métodos de clasificación existentes, hemos realizado la clasificación mediante un modelo de regresión para predecir las calificaciones de los estudiantes. Este método demostró ser mucho más eficiente y preciso.

Cuatro propiedades novedosas de nuestro modelo son (1) que no necesita ningún paso de preprocesamiento complejo o la división de r por 10 (2) puede operar para obtener predicciones incluso cuando el conjunto de datos es pequeño y desequilibrado. (3) no captura tendencias debido al desequilibrio como los modelos de referencia (4) La capacidad de predicción cruzada con un rango de error mínimo que no se muestra en cualquier otro algoritmo.

Como trabajo futuro, la incorporación de más datos de diferentes universidades reducirá aún más el rango de error escalado de 1 a valores mucho más bajos y mejorará el impresionante rendimiento del regresor profundo.

REFERENCIAS

- 1) S. Kotsiants, C. Pierrakeas y P. Pintelas, Prevent ingstudentdropoutindistancelearningsystemusing Machine Learning Techniques, AI Techniques in Webbased Educational system en la séptima Conferencia Internacional sobre Sistemas Inteligentes de Información e Ingeniería basados en el aprendizaje, septiembre de 2003.
- 2) Lykourantzou, Giannoukos, Nikolopoulos y Loumosin, Predicción de abandono en cursos de aprendizaje electrónico mediante la combinación de técnicas de aprendizaje automático, Informática y educación, vol. 53 número 3, páginas 950-965, noviembre de 2009.
- 3) LGMoseley y DM Mead, Predecir quién abandonará los cursos de enfermería: un ejercicio de aprendizaje automático Nurse Education Today, vol. 28, Número 4, páginas 469-475 Mayo 2008
- 4) Byung-HakKim, Ethan Vizitei y Varun Ganapathi, GritNet: Predicción del rendimiento estudiantil con aprendizaje profundo, agosto de 2018.
- 5) F. Miand DY Yeung, Modelos temporales para predecir la deserción estudiantil en cursos masivos abiertos en línea, en las actas de la 15.ª Conferencia Internacional IEEE sobre el taller de simulación de datos (ICDMW 2015), páginas 256-253, 2015.
- 6) Tianqi Chen y Carlos Guestrin, XGBoost: un árbol escalable Sistema de Impulso, 2016.
- 7) DiederikP. Kingma, Jimmy Lei Ba, ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION, publicado como documento de conferencia en ICLR, 2015.
- 8) Timothy Dozat, Incorporating Nesterov Momentum into Adam.