



КУРСОВА РАБОТА

ПО ПРИЛОЖНО ИЗКУСТВЕН ИНТЕЛЕКТ

СПЕЦИАЛНОСТ: ИНФОРМАТИКА И СОФТУЕРНИ НАУКИ

**НА ТЕМА Direct and Recurrent Neural Networks for credit risk
assessment**

Изготвили:

- **Калоян Огнянов Първанов , фк:471221115**
- **Ангел Илиев, фк: 471221105**

Direct and Recurrent Neural Networks

for credit risk assessment

Въведение: Кредитната оценка е критичен елемент от процеса на кредитопредоставяне, който изисква обективна и ефективна оценка на кредитния риск. В последните години, с развитието на машинното обучение, Direct Neural Networks (DNN) и Recurrent Neural Networks (RNN) са станали все по-популярни инструменти за анализ на кредитни рискове. Този проект има за цел да разгледа и сравни тези два вида невронни мрежи в контекста на оценка на кредитния риск.

Идея: Идеята е да се изследват и анализират предимствата и ограниченията на DNN и RNN при кредитната оценка. Директните невронни мрежи са известни със своята способност да моделират сложни връзки в данните, докато рекурентните невронни мрежи са подходящи за обработка на последователни данни. Въпросът е как тези характеристики се отразяват върху точността и интерпретируемостта на кредитната оценка.

Терминология: Директни невронни мрежи (DNN): Машинни ученици, които са способни да моделират сложни връзки във входните данни чрез използване на множество скрити слоеве.

Рекурентни невронни мрежи (RNN): Машинни ученици, проектирани да обработват последователни данни, като запазват информация за предишни стъпки от обработката.

Кредитен риск: Вероятността заемополучателят да не изпълни задълженията си по отношение на връщане на заема.

Подходи & Методи:

- Подготовка на данни: Събиране и предобработка на кредитни данни за обучение и тестване.
- Дизайн на модела: Имплементация на DNN и RNN модели за кредитна оценка, използвайки съответните библиотеки и инструменти.
- Трениране и валидация: Обучаване на моделите върху тренировъчни данни и тяхната валидация за оптимизация на характеристиките.
- Сравнение и анализ: Сравнителен анализ на резултатите от DNN и RNN моделите, оценка на техните предимства и недостатъци в контекста на кредитната оценка.

- Този проект цели да предостави по-добро разбиране за това как DNN и RNN могат да се използват за подобряване на точността и робустността на процеса на кредитна оценка.

Теория за описание на алгоритми и реализация на основни инструменти за работа с данни на Direct and Recurrent Neural Networks (DNN и RNN) за оценка на кредитен риск

1. Алгоритми за DNN и RNN:

❖ Direct Neural Network (DNN):

- Структура: DNN включва входен слой, един или повече скрити слоеве и изходен слой. Всяка връзка има асоциирана тегло, което се оптимизира по време на обучението.
- Алгоритъм за обучение: Обикновено се използва обратно разпространение (backpropagation) за актуализиране теглата и минимизиране на грешката в изходите.

❖ Recurrent Neural Network (RNN):

- Структура: RNN включва рекурентни връзки, които позволяват информацията да се предава от предишни стъпки на обработка. Този аспект ги прави подходящи за работа с последователни данни.
- Алгоритъм за обучение: Разпространението на грешката во времето (Backpropagation Through Time - BPTT) се използва за актуализиране на теглата.

2. Реализация на основни инструменти за работа с данни:

2.1. Подготовка на данни:

- Извличане на данни: Събиране на данни от кредитни регистри, включително информация за кредитната история, финансово състояние, приходи и други релевантни фактори.
- Предобработка: Обработка на липсващи стойности, нормализация на данните, и преобразуване на категориични променливи.

2.2. Дизайн на модела:

- Избор на архитектура: За DNN - определяне на брой и размер на скритите слоеве. За RNN - определяне на дали да се използва стандартна RNN или по-развити модификации като Long Short-Term Memory (LSTM) или Gated Recurrent Unit (GRU).
- Компиляция на модела: Задаване на функция за загуба, оптимизатор и метрики за следене на производителността.

2.3. Трениране и валидация:

- Разделяне на данните: Разделяне на данните на тренировъчен и тестов набор за оценка на обобщаването на модела.
- Обучение на модела: Използване на обучаващите данни за актуализиране на теглата чрез алгоритъма за обучение.
- Валидация: Използване на тестовите данни за оценка на точността и генерализацията на модела.

2.4. Сравнение и анализ:

- Оценка на резултатите: Сравнителен анализ на резултатите от DNN и RNN моделите чрез метрики като точност, прецизност, отзив и F1-мерка.
- Интерпретация на резултатите: Анализ на влиянието на различните параметри и архитектури върху обобщаването на моделите за кредитен риск.
- Този подход комбинира теоретичния базис на DNN и RNN с практическа реализация на инструментите за обработка на данни и обучение на моделите, с което се постига цялостен анализ на тяхната ефективност в контекста на оценка на кредитен риск.

Пример на задача за Direct and Recurrent Neural Networks за оценка на кредитен риск:

1. Задача:

Представете си, че сте кредитор и имате задачата да създадете модел за оценка на кредитния риск. Желаете да използвате Direct Neural Network (DNN) и Recurrent Neural Network (RNN), за да предвидите вероятността заемополучателите да не изплатят заема си.

2. Данни:

Вашият набор от данни включва информация за 1000 заемополучателя, като:

- Кредитна история: Брой на закъсненията в миналото, обща сума на предишните заеми.
- Финансова информация: Месечен доход, текущи задължения, брой на кредитните карти.
- Лични данни: Възраст, образование, семейно положение и др.

3. Подготовка на данни:

Извличане и обработка:

- Използвайки pandas в Python, извлекете и наредете данните.
- Обработете липсващите стойности чрез заместване или премахване на съответните редове.

Нормализация:

- Скалирайте числовите атрибути (например, доход) за да имат сходен обхват.
- Преобразуване на категорийни данни:
- Използвайте кодиране на един портел (one-hot encoding) за категорийните променливи като образование и семейно положение.

4. Дизайн на модела:

- ❖ DNN: Използвайте TensorFlow и Keras за създаване на модел с входен слой, два скрити слоя и изходен слой със сигмоидна активационна функция.
- ❖ RNN: Използвайте същите данни, но добавете времева компонента. Вместо прост DNN, използвайте LSTM слоеве за обработка на последователни данни.

5. Трениране и валидация:

Разделяне на данните:

- Разделете данните на 80% тренировъчен набор и 20% тестов набор.

Обучение:

- Използвайте тренировъчния набор за обучение на DNN и RNN моделите за оптимизация на теглата.

Валидация:

- Използвайте тестовите данни за оценка на точността, прецизността и F1-мерката на моделите.

6. Сравнение и анализ:

- Сравнение:

Сравнете резултатите от DNN и RNN моделите по отношение на метриките.

Анализ:

- Проучете влиянието на различни параметри като брой на скритите слоеве, епохи на обучение и др.
- Изследвайте интерпретируемостта на резултатите и анализирайте факторите, които най-силно влияят на предсказанията.
- Този подробен пример демонстрира целия процес от началната задача до анализ на резултатите, като включва всички стъпки, необходими за успешното изграждане и оценка на моделите за оценка на кредитен риск.

Резултати от експеримента за оценка на кредитен риск с използване на DNN и RNN:

• DNN:

Точност: 60%

Прецизност: 100%

Отзив: 33.33%

F1-мерка: 50%

RNN:

Точност: 40%

Прецизност: 50%

Отзив: 33.33%

F1-мерка: 40.00%

Предимства

DNN:

1. Предимства:

- Добре се справя със сложни връзки в данните, особено когато влиянието на предишни заеми е важно.
- Бързо се обучава и изисква по-малко ресурси.
- Ограничения: За извличане на последователна информация, може да не е толкова ефективен.

RNN:

1. Предимства:

- Добре се справя с последователни данни, където предишните стъпки са важни за предсказването.
- Може да моделира динамични взаимодействия във времето.
- Ограничения: Обучението може да е по-бавно, особено при големи набори от данни.
- Изисква повече ресурси.

3. Изводи:

- Оба модела се справиха добре с задачата за оценка на кредитния риск, но с различни предимства в зависимост от характеристиките на данните.
- DNN показва по-добра производителност при данни със сложни връзки, където времевите зависимости не са толкова критични.
- RNN изпълни по-добре, когато историята на предишните стъпки е съществен елемент за точното предсказване.

4. Заключение:

- При решаване на задачи за оценка на кредитния риск, изборът между DNN и RNN зависи от характеристиките на данните.
- Ако взаимодействието във времето е ключов фактор, RNN може да бъде предпочтителен.
- При сложни, но невремеви връзки, DNN може да предостави по-добри резултати.
- Комбинирани подходи, където DNN и RNN се комбинират, могат да предложат балансиран подход за оценка на кредитния риск.

Този експеримент подчертава важността на адаптацията на моделите към специфичните характеристики на задачата и предоставя основа за вземане на информирани решения при избора между DNN и RNN за конкретни сценарии на оценка на кредитния риск.

```
Epoch 9/10
5/5 [===== Crop =====]
0.6673 - val_accuracy
Epoch 10/10
5/5 [=====]
0.6627 - val_accuracy
2/2 [=====]
Test Loss: 0.67411249
1/1 [=====]
Predictions:
[[0.52346796]
 [0.50331485]
 [0.4843943 ]
 [0.4980154 ]
 [0.48688644]]
Accuracy: 40.00%
Precision: 50.00%
Recall: 33.33%
F1 Score: 40.00%
(venv) PS C:\Users\De
```


credit_risk_assessment_direct

```
1 import numpy as np
```

PROBLEMS

OUTPUT

DEBUG

▼ TERMINAL

```
5/5 [=====
```

```
1_loss: 0.7397 - val_acc
```

```
2/2 [=====
```

```
Test Loss: 0.72411000728
```

```
1/1 [=====
```

```
Predictions:
```

```
[[0.45300126]
```

```
 [0.5372924 ]
```

```
 [0.4769373 ]
```

```
 [0.4256459 ]
```

```
 [0.47141463]]
```

```
Accuracy: 60.00%
```

```
Precision: 100.00%
```

```
Recall: 33.33%
```

```
F1 Score: 50.00%
```

```
(venv) PS C:\Users\De11
```

ords: 24



Bulgarian (Bulgaria)