

فرادرس

فرا تراژیک کلاس درس
www.faradars.org

آموزش یادگیری عمیق با پایتون TensorFlow و Keras

فصل هفتم: NLP و پردازش متن

مدرس:

سعید محقق

برنامه‌نویس، متخصص هوش مصنوعی و دانشجوی دکتری مهندسی پزشکی

faradars.org/fvdl104

فهرست مطالب

روش‌های پیش‌پردازش متن در پایتون و Keras

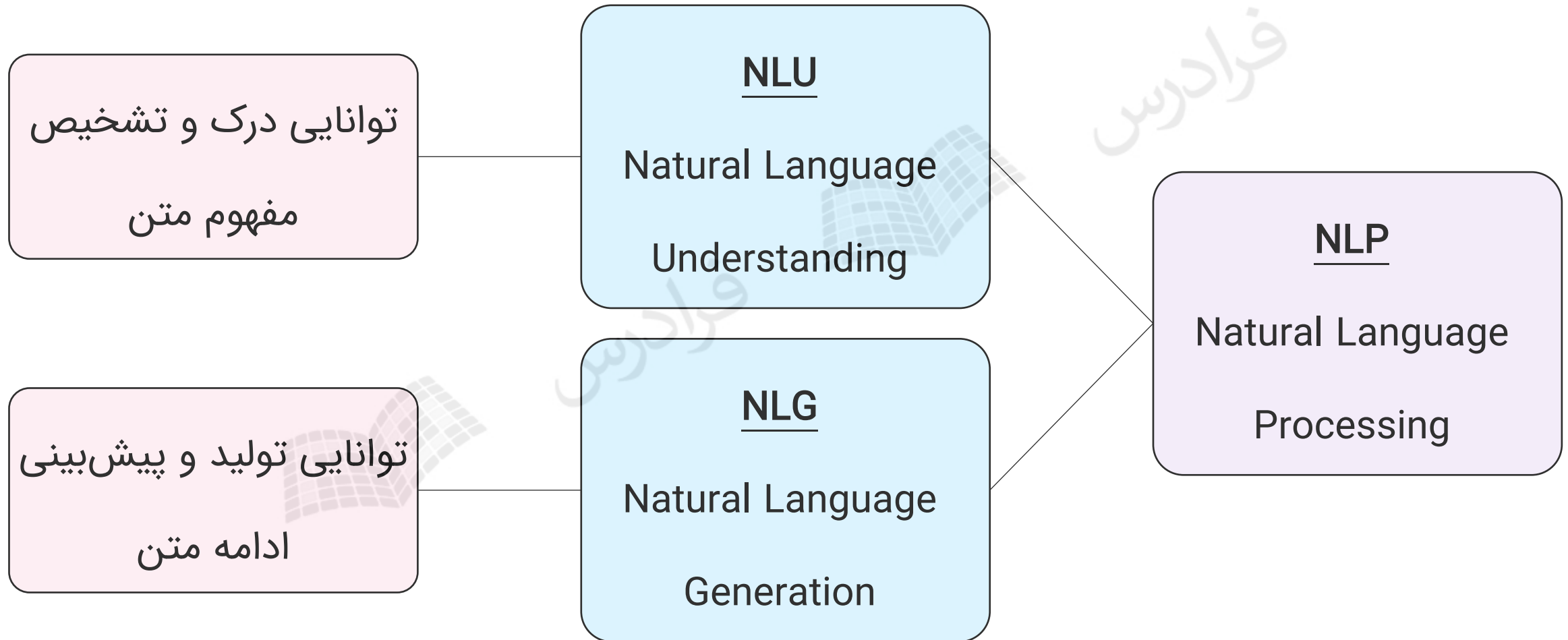
معرفی لایه Embedding برای استخراج ویژگی معنادار از متن

پروژه ۶: تشخیص کنایه‌آمیز بودن یا نبودن متن

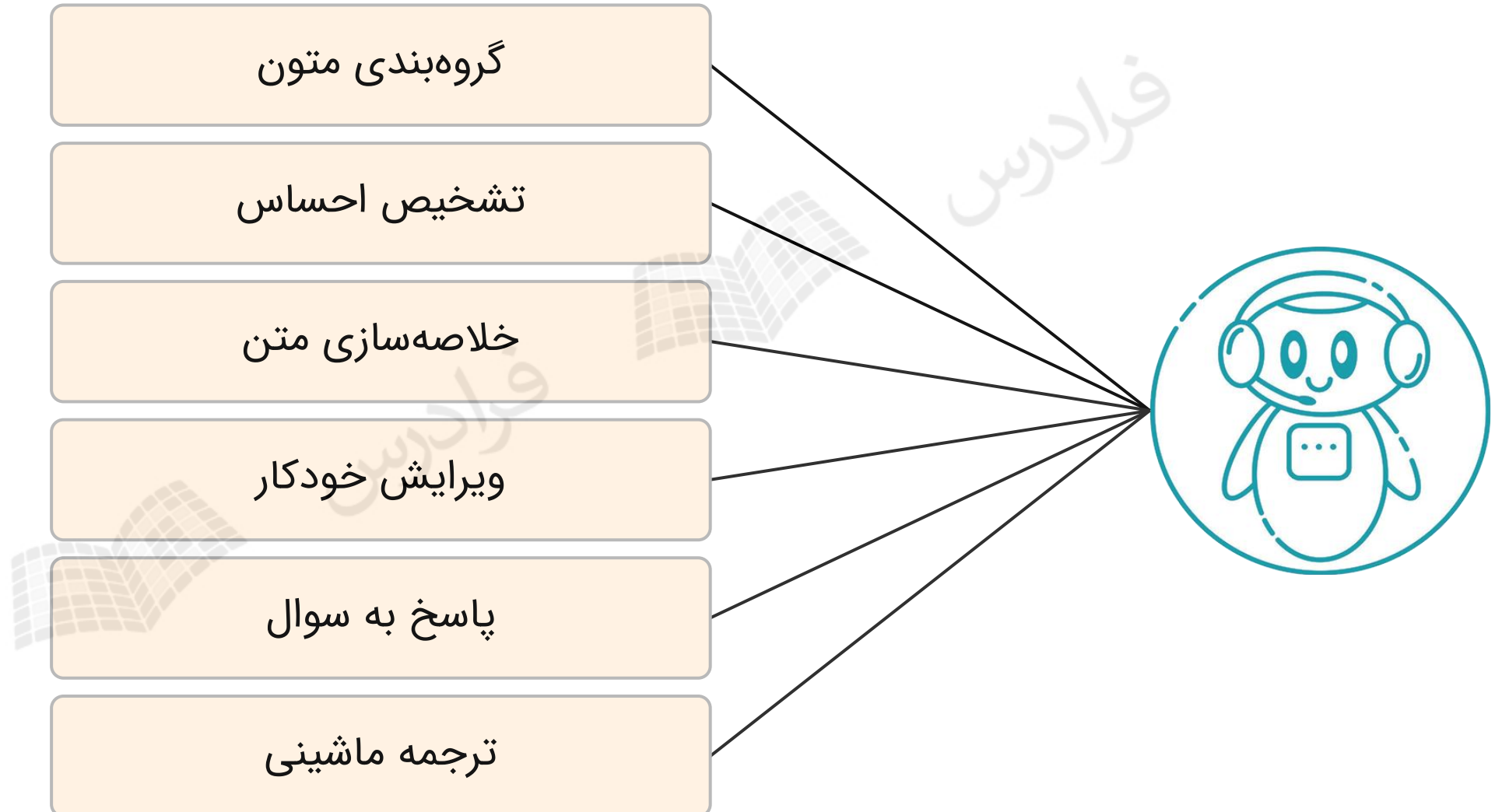
استفاده از لایه Embedding از پیش آموزش دیده برای بهبود نتایج پروژه ۶

استفاده از مدل RNN برای بهبود نتایج پروژه ۶

مدل‌های زبانی در NLP



مدل‌های زبانی در NLP



چالش‌های داده‌های متنی

نبود قوانین و قواعد ثابت و مشخص

داده‌های غیر عددی و بدون ساختار

مدل یادگیری عمیق فقط با داده
عددی آموزش می‌بیند.

وابستگی داده‌ها به یکدیگر

تشخیص مفهوم متن با توجه به
متون قبل یا بعد از آن

مراحل آنالیز متن در مدل‌های یادگیری عمیق

پیش‌پردازش متن

- استانداردسازی، جداسازی، حذف کلمات اضافی و...

تبدیل نوع داده

- تبدیل متن به بردارهای عددی

طراحی لایه‌ها و مدل‌های متناسب

- لایه‌ها و مدل‌های یادگیری عمیق مناسب برای آنالیز متن

پیش‌پردازش متن در Keras

استفاده از پایتون و NLTK برای پیش‌پردازش متن
(جداسازی جملات، حذف لغات اضافی (stopwords) و...)

Tokenization: تبدیل جملات به کلمات
Indexing: تبدیل کلمات به اعداد
Padding: هم اندازه کردن بردارها

روش ۱: استفاده از ماژول‌های `Tokenizer` و `pad_sequences`

روش ۲: استفاده از لایه `TextVectorization`

پروژه ۶: تشخیص کنایه آمیز بودن یا نبودن متن

آنالیز داده متنی

- روش: مدل‌های MLP و RNN و مقایسه نتایج

چالش‌ها

- پیش‌پردازش داده متنی
- طراحی مدل مناسب و استفاده از دانش پیشین

تکنیک‌ها

- استفاده از لایه Embedding از پیش آموزش دیده
- نمایش فضای یادگیری شده توسط لایه Embedding



این اسلایدها بر مبنای نکات مطرح شده در فرادرس «آموزش یادگیری عمیق با پایتون – TensorFlow و Keras» تهیه شده است.

برای کسب اطلاعات بیشتر در مورد این آموزش به لینک زیر مراجعه نمایید.

faradars.org/fvdl104