

tf.contrib.layers.sparse_column_with_vocabulary_file

```
sparse_column_with_vocabulary_file(  
    column_name,  
    vocabulary_file,  
    num_oov_buckets=0,  
    vocab_size=None,  
    default_value=-1,  
    combiner='sum',  
    dtype=tf.string  
)
```

Defined in [tensorflow/contrib/layers/python/layers/feature_column.py](#).

See the guide: [Layers \(contrib\) > Feature columns](#)

Creates a `_SparseColumn` with vocabulary file configuration.

Use this when your sparse features are in string or integer format, and you have a vocab file that maps each value to an integer ID. `output_id = LookupIdFromVocab(input_feature_string)`

Args:

- `column_name`: A string defining sparse column name.
- `vocabulary_file`: The vocabulary filename.
- `num_oov_buckets`: The number of out-of-vocabulary buckets. If zero all out of vocabulary features will be ignored.
- `vocab_size`: Number of the elements in the vocabulary.
- `default_value`: The value to use for out-of-vocabulary feature values. Defaults to -1.
- `combiner`: A string specifying how to reduce if the sparse column is multivalent. Currently "mean", "sqrtn" and "sum" are supported, with "sum" the default. "sqrtn" often achieves good accuracy, in particular with bag-of-words columns.
 - "sum": do not normalize features in the column
 - "mean": do l1 normalization on features in the column
 - "sqrtn": do l2 normalization on features in the column For more information: [tf.embedding_lookup_sparse](#).
- `dtype`: The type of features. Only string and integer types are supported.

Returns:

A `_SparseColumn` with vocabulary file configuration.

Raises:

- `ValueError`: `vocab_size` is not defined.
- `ValueError`: `dtype` is neither string nor integer.

Stay Connected

- Blog
- GitHub
- Twitter

Support

- Issue Tracker
- Release Notes
- Stack Overflow

English

Terms | Privacy