

Automatic Identification of Patent Claim Types: Enhancing Efficiency in Patent Analysis

Rima Dessi¹, Hidir Aras¹, Mark Prince² and René Hackl-Sommer¹

¹ FIZ Karlsruhe, Karlsruhe, Germany / firstname.lastname@fiz-karlsruhe.de

² CAS - Chemical Abstracts Service, Columbus, Ohio, USA

Motivation

Patent Claim Type Recognition

Traditional Approaches

Data Labeling

Machine Learning Models In An Artificial Intelligence Infrastructure

Abstract
Improving machine learning models in an artificial intelligence infrastructure includes: storing, within one or more storage systems ...

Claims
What is claimed is:
1. A method of storage system query processing, the method comprising: receiving, at a storage system from one or more graphical processing units (GPUs) via a storage system application program interface (API) provided by the storage system directly to the one or more GPUs and configured...
2. The method of claim 1, wherein the query conforms to an application programming interface implemented by the storage system.
3. The method of claim 2, wherein the application programming interface supports specifying a function and one or more function parameters, and wherein responsive to determining that the function has been received, the storage system applies the function to one or more of the multiple data objects.
4. The method of claim 3, wherein the function is a data transformation that is applied to the one or more of the multiple data objects, and wherein the one or more of the multiple data objects are selected in accordance with the one or more function parameters.
5. The method of claim 3, wherein the one or more function parameters including an indication to the storage system to cache results of applying the function to the one or more of the multiple data objects.
6. The method of claim 5, further comprising: in response to receiving the function and the specified indication to cache results, caching the results of applying the function to the one or more of the multiple data objects, wherein the results include one or more of a duplicate of the one or more of the multiple data objects or metadata describing or referencing the one or more of the multiple data objects; receiving, via the application programming interface, another invocation of the function and the one or more function parameters; and in response to determining that the one or more of the multiple data objects have not changed, transmitting the cached results of applying the function to the one or more of the multiple data objects.
7. The method of claim 1, wherein the query is a database query, and the query specifies one or more parameters including one or more of: data location, data size, data ownership, data access history, data reduction ratio, access permission, data type, data content, or one or more metadata characteristics. ...

Identify Type Information

Method

System

Process

Apparatus

Labeled Data

Train a Classifier

Make a Prediction

Challenges:

- Require labeled data
- Designed for traditional text (e.g., news) classification

Labeling Training Data:

Regex baseline has been designed and which relies on start and end markers. In-between such markers are the targets.

E.g.,

68. A method according to claim 67, with "68. A" being a start marker, "according to" an end marker, and "method" the target.

Labeling Test Data:

- 500 samples were randomly selected from the internal and external data sources
- Test data then labeled by overall 4 SMEs following an iterative process in order to achieve the required level of agreement, utilizing a majority vote system.

Patent Claim Type Recognition (PCTR)

Dataset Statistics

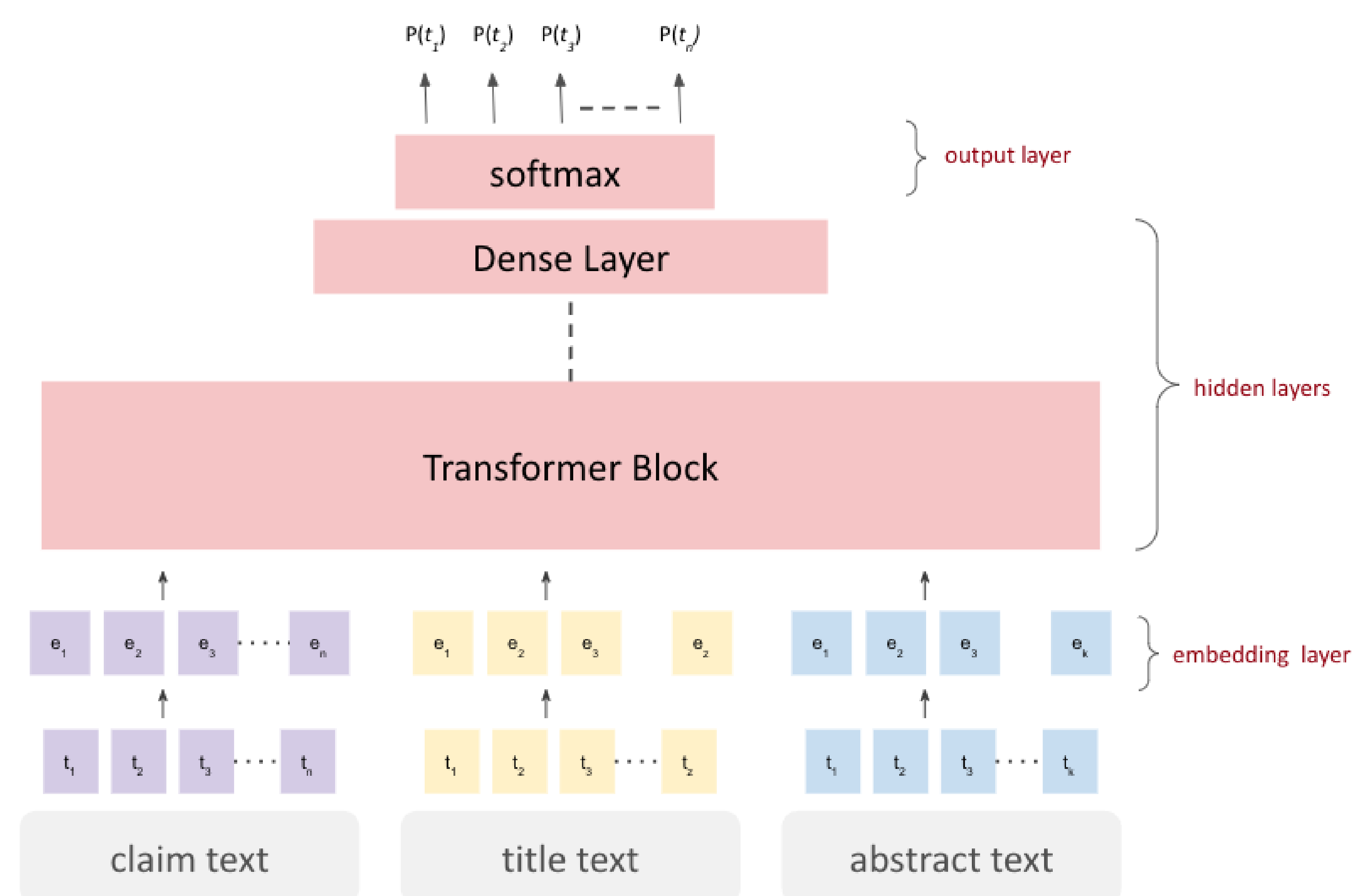
Claim Type	#Sample
System	322,848
Process	84,847
Method	988,687
Composition	141,273
Apparatus	540,025
Use	14,705

Table 1: Statistics of labeled Training Data

Claim Type	#Sample
System	98
Process	37
Method	83
Composition	152
Apparatus	89
Use	31

Table 2: Statistics of labeled Test Data

Architecture



Experimental Results

Model Type	Features	Accuracy
PCTR_V1	claim	0,561
PCTR_V2	claim + title	0,576
PCTR_V3	claim + title + abstract	0.602

Model Type	Features	Accuracy incl. feedback
PCTR_V1	claim	0.808
PCTR_V2	claim + title	0.816
PCTR_V3	claim + title + abstract	0.824

Lessons Learned:

- *Data from different Patent Authorities:* As claim type definitions have been discovered to vary by jurisdiction, it is needed to customize the developed models per patent office (or collections of patent offices) accordingly.
- *The importance of feature selection:* The experiments indicate that including more features, specifically the title, and abstract, can significantly improve the accuracy.
- *The role of SMEs:* The results also demonstrate that incorporating the feedback and insights from SMEs can greatly enhance the accuracy of the PCTR models.