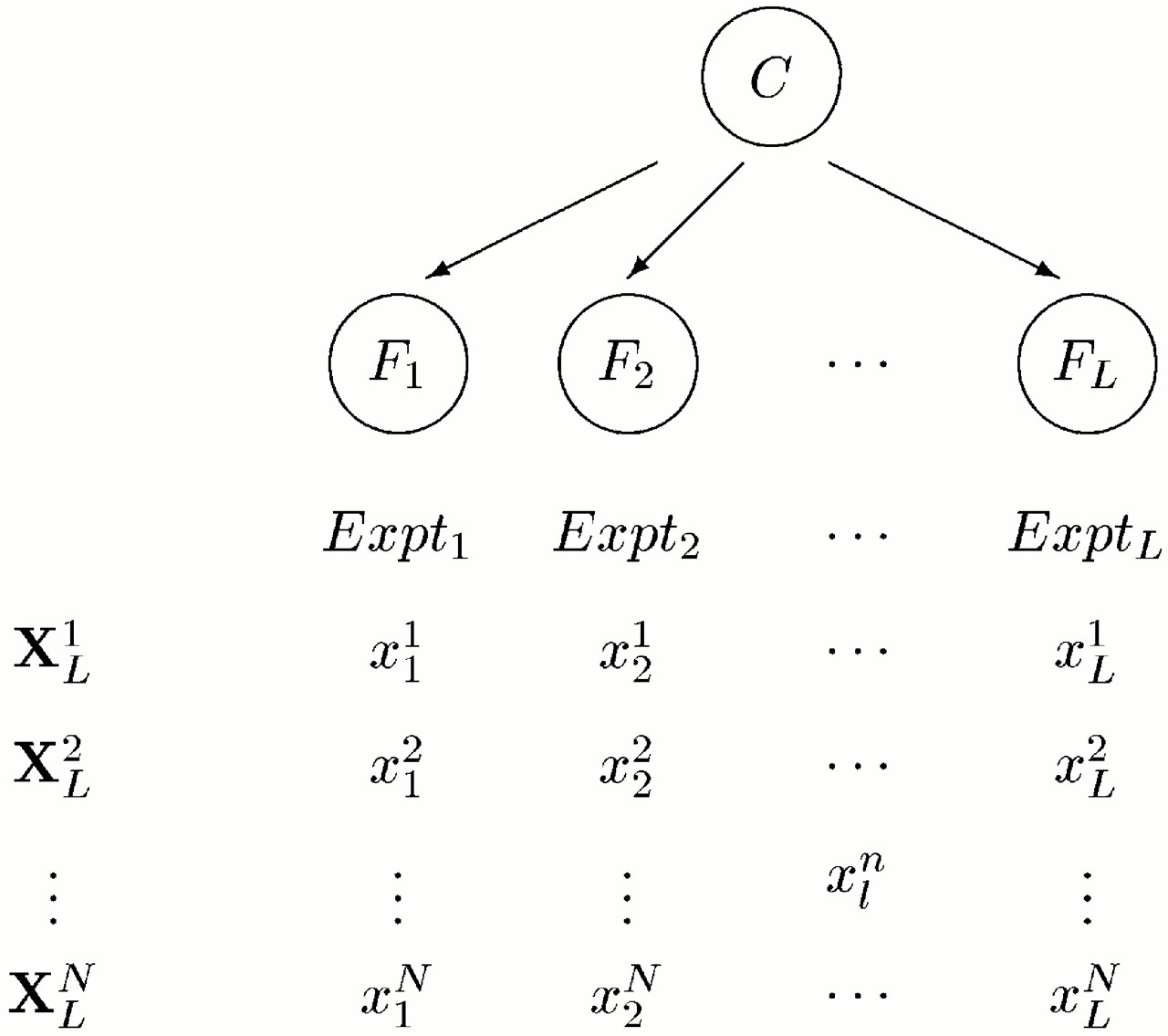
iReport Classification :

Naïve Bayes Linear Classification.



The first basic approach for any classification would be naïve bayes.

Naïve Bayes works better than any algorithm if it has huge and related training data for all classifiers uniformly.

Naïve Bayes works well with unigrams for classifications. i.e., it just based on the keywords provided in the report and simply ignore all grammatical words in report (is/was/the/for/with/ …).

But to get sense of out of text instead of relying on keywords, we need (deep learning)/(natural language processing) concept which is advanced area we all should focus on it.

The above picture shows the L-features and “N-word length report”.

L-feauters possible values can be from iReport mentioned below.

{ appolo health, lvprasad health, Narayana netralaya health, manipal health, …

Banglore crime, Hyderabad crime , delhi crime, Mumbai crime, ……..

Kotak insurance, lic insurance, icici insurance , ……….

Telangana govt , Andhra pradesh govt, karnataka govt, india govt, USA govt, china govt, mississipi govt,……….

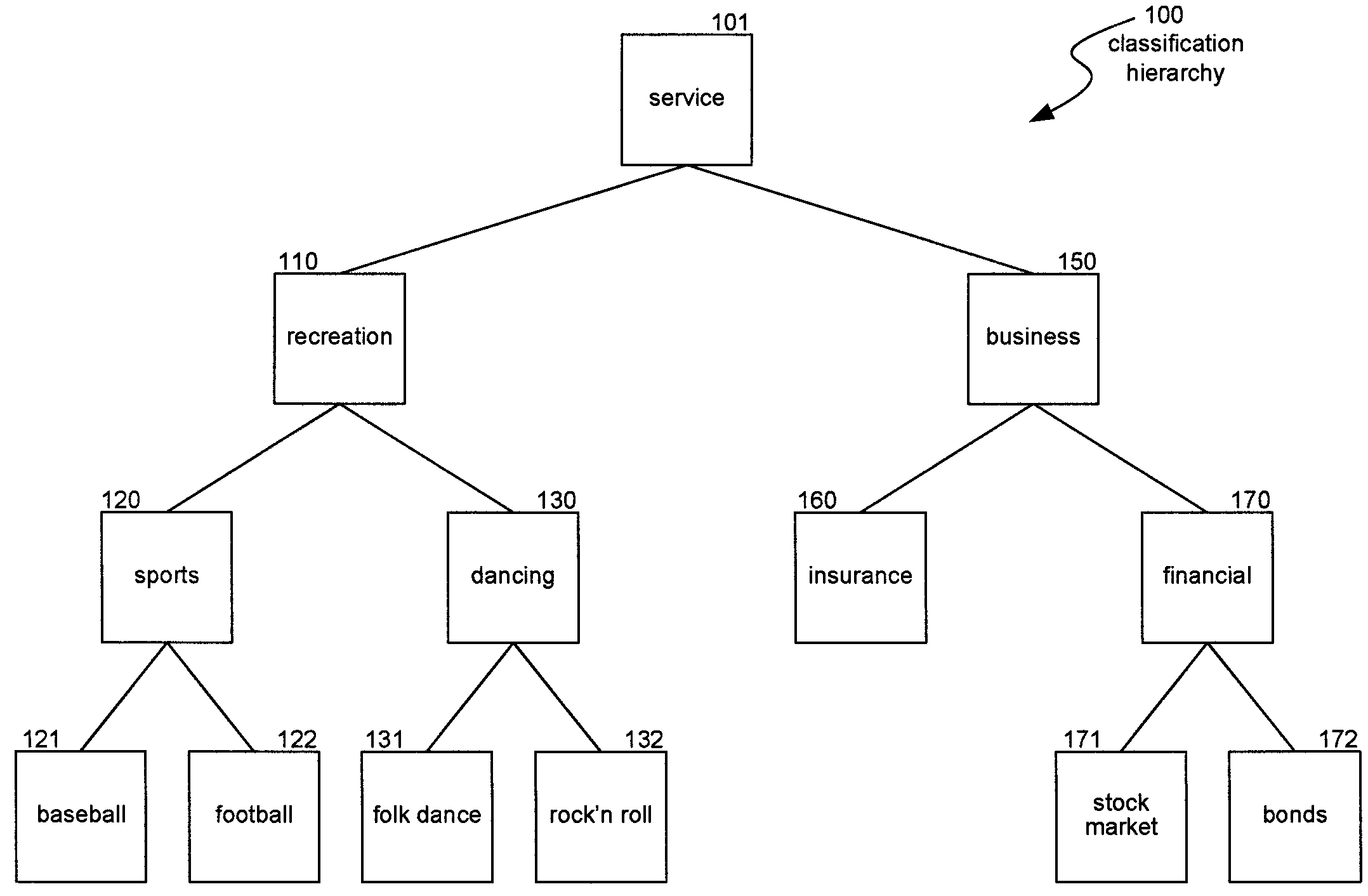
}

Given n-word sentence from report , we will have to check for all classifiers and have to assign best suitable classifier based on minimum threshold. The reports which are not supporting minimum threshold may have to be classified after adding new classifiers (we will have to come up with some solution.)

Cons : it is very difficult to support new classifiers dynamically and train them accordingly.

iReport Proposed Classification :

Hierarchical Classification.



Hierarchical classification is the best suitable structure for our requirement (at least I think.)

There are very few publications and resources available online regarding this particular scenario.

The main challenging part with this solution is training the data for all classifiers.

Please go through http://www.google.com/patents/US7809723.