Combining human and machine approaches to classification

UNSILO

Michael Upshall
STM Digital Publishing, Dec 5 2017
UNSILO

Finding relevant knowledge from text

- Founded 2012
- Based Denmark
- Machine Learning, Text Analytics and NLP
Classification

Document classification

The task is to assign a document to one or more classes or categories.
Limitations of taxonomies?

1. They change
2. They grow
3. They impose a world view
4. New ones come along
5. They are endless
Imposing a world view

London Street Lighting  use: Street Lighting England London

Thermoregulation  use: Body temperature regulation

Permafrost  use: Frozen ground

“the easiest way to find and use LCSH is to start with a ‘keyword’ search and then look at the Subject Headings to locate other related material.”
New taxonomies come along
Taxonomies are endless
Endless Labour

Aye, and I saw Sisyphus in violent torment, seeking to raise a monstrous stone with both his hands

Homer, Odyssey, Book 11
Serum consists of water (93% of serum volume) and nonaqueous components, mainly lipids and proteins (7% of serum volume). Sodium is restricted to serum water. In states of hyperproteinemia or hyperlipidemia, there is an increased mass of the nonaqueous components of serum and a concomitant decrease in the proportion of serum composed of water. Thus, pseudohyponatremia results because the flame photometry method measures sodium concentration in whole plasma. A sodium-selective electrode gives the true, physiologically pertinent sodium concentration because it measures sodium activity in serum water. Whereas the serum sample is diluted in indirect potentiometry, the sample is not diluted in direct potentiometry. Because only direct reading gives an accurate concentration, we suspect that indirect potentiometry which many hospital laboratories are now using may mislead us to confusion in interpreting the serum sodium data. However, it seems that indirect potentiometry very rarely gives us discernibly low serum sodium levels in cases with hyperproteinemia and hyperlipidemia. As long as small margins of errors are kept in mind of clinicians when serum sodium is measured from the patients with hyperproteinemia or hyperlipidemia, the present methods for measuring sodium concentration in serum by indirect sodium-selective electrode potentiometry could be maintained in the clinical practice.
Granularity

The image shows a bar chart comparing the number of terms in different taxonomies. The x-axis represents the taxonomies, and the y-axis shows the number of terms. The chart includes the following taxonomies:

- Typical corporate taxonomy
- BISAC Book Industry Classification
- INSPEC
- MeSH
- Unsupervised machine learning

The number of terms in each taxonomy is as follows:

- Typical corporate taxonomy: 0
- BISAC Book Industry Classification: 0
- INSPEC: 0
- MeSH: 0
- Unsupervised machine learning: 1,250,000

The chart indicates that unsupervised machine learning has a significantly higher number of terms compared to the other taxonomies.
## Classification methods compared

<table>
<thead>
<tr>
<th></th>
<th>Manual taxonomy</th>
<th>Rules</th>
<th>Unsupervised machine learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual labour to maintain</td>
<td>high</td>
<td>fair</td>
<td>zero</td>
</tr>
<tr>
<td>Requires cataloguing expertise</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Requires IT expertise</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Add new term after completion</td>
<td>slow</td>
<td>quick</td>
<td>quick</td>
</tr>
<tr>
<td>Speed of updating</td>
<td>slow</td>
<td>fair</td>
<td>quick</td>
</tr>
<tr>
<td>Can the system learn?</td>
<td>no</td>
<td>no</td>
<td>potentially</td>
</tr>
<tr>
<td>Human intervention?</td>
<td>yes</td>
<td>cumbersome</td>
<td>no</td>
</tr>
</tbody>
</table>
Combining human and machine capabilities
Human-machine interaction
Diabetes mellitus and tuberculosis: programmatic management issues

The Diabetes Interactive Atlas is a recently released Web-based collection of maps that allows users to view geographic patterns and examine trends in diabetes and its risk factors over time across the United States and...

A Rare Case of Congenital Diabetes Insipidus

The Diabetes Interactive Atlas is a recently released Web-based collection of maps that allows users to view geographic patterns and examine trends in diabetes and its risk factors over time across the United States and...
**Gold Documents**

**Comparison of skin microvascular reactivity with hemostatic markers of endothelial dysfunction and damage in type 2 diabetes**

Pollock, Jonathan P.; Patel, Hardikumar M.; Randolph, Brittney J., et al. in Physiological Reports (2014)

Effective medications are a cornerstone of prevention and disease treatment, yet only about half of patients take their medications as prescribed, resulting in a common and costly public health challenge for the US health care system...
Developmental dyslexia affects almost 10% of school-aged children and represents a significant public health problem. Its etiology is unknown. The consistent presence of phonological difficulties combined with an inability to manipulate language sounds and the grapheme–phoneme conversion is widely acknowledged. Numerous scientific studies have also documented the presence of eye movement anomalies and deficits of perception of low contrast, low spatial frequency, and high frequency temporal visual information in dyslexics. Anomalies of visual attention with short visual attention spans have also been demonstrated in a large number of cases. Spatial orientation is also affected in dyslexics who manifest a preference for spatial attention to the right. This asymmetry may be so pronounced that it leads to a veritable neglect of space on the left side. The evaluation of treatments proposed to dyslexics whether speech or oriented towards the visual anomalies remains fragmentary. The advent of new explanatory theories, notably cerebellar, magnocellular, or proprioceptive, is an incentive for ophthalmologists to enter the world of multimodal cognition given the importance of the eye’s visual input.
Improving diabetes medication adherence: successful, scalable interventions

Effective medications are a cornerstone of prevention and disease treatment, yet only about half of patients take their medications as prescribed, resulting in a common and costly public health challenge for the US health care system...

Most likely belongs to one of these classes
- Diabetes
- Diseases
- Diets

- Infusional antibody
- Elispot assay
- Peritoneal relapse
- Infusional antibody
- Elispot assay
- Peritoneal relapse
- Infusional antibody
- Elispot assay
- Peritoneal relapse
- Infusional antibody
- Elispot assay
- Peritoneal relapse
UNSILO

Finding relevant knowledge from text