

Applying activity classifications to UK charities

Chris Damm (Sheffield Hallam University – c.damm@shu.ac.uk)

David Kane (Independent data scientist - david@dkane.net)

VSSN conference 2021

Project aims

- Develop a new classification system for UK charities able to better capture the full range of charitable purposes in the UK context.
- Develop an automated process for applying categories, based on the regulatory information available on each charity (such as its name, description of activities, etc.)
- Make the classifications, methods and code freely available under a commons licence

Why classify?

- Great diversity within the Voluntary Sector (Kendal and Knap, 1995)
- Organisations environment may be more shaped by their activity-area than their sector (Kendal, 2003)
- Classification and conceptualisation two sides of the same coin - how VSOs differ as well as what unites them (Salamon and Anheier, 1992).

More specifically...

- Map 'supply and demand' geographically
- Help potential service users, members, collaborators, funder or donors find particular types of charities
- Sample from a sub-group for further research
- Include activity-area as a variable in quantitative modelling
- Help understand the 'shape' of the overall voluntary sector (would have been useful for the pandemic outbreak)

Why not to classify?

- Makes some groups visible and others invisible, inside 'other categories' or not covered at all (Bowker and Starr 1999)
- This can have implications for who gets resources (Barman, 2013)
- Can be quite a top-down process, imposing labels on others ('putting them in boxes')
- Being 'visible' might not be a good thing – first step towards being 'governed' (Nickel and Eikenberry, 2015; Carmel and Harlock, 2010).
- We argue that some classification schemes are 'better' than others, or at least more useful to people acting in good faith
- Bowker and Starr (1999, p.326) argue in favour of 'flexible classifications whose users are aware of their political and organisational dimensions and which explicitly retain traces of their construction' and that 'the only good classification is a living classification'

England and Wales classifications in the register of charities:

What does your charity do?

- General charitable purposes
- Education / training
- The advancement of health or saving of lives
- Disability
- The prevention or relief of poverty
- Overseas aid / famine relief
- Accommodation/ housing
- Religious activities
- Arts / culture / heritage / science
- Amateur sport
- Animals
- Environment / conservation / heritage
- Economic / community development /employment
- Armed Forces / emergency service efficiency
- Human rights / religious or racial harmony/equality or diversity
- Recreation
- Other charitable purposes
- People with disabilities
- People of a particular ethnic or racial origin
- Other charities / voluntary bodies
- Other defined groups
- The general public / mankind

How does your charity operate?

- Makes grants to individuals
- Makes grants to organisations
- Provides other finance
- Provides human resources

Who does your charity help?

- Children / young people
- Elderly / old people

Issues

- Charities self-select their categories and they are free to pick as many as they wish, which can allow quite peripheral inclusions, with no indication of which are most important.
- Some categories are very broad, including a 'general charitable purposes' category.
- More specific categories such as food banks, homelessness services or medical research are not included (though the CC are expanding their range of categories).
- There is overlap between the questions, which enables contradictions. Question one includes a 'disability' category, while question two includes 'people with disabilities'.

ICNPTSO

- The International Classification of Non-Profit and Third Sector Organisations (ICNP/TSO)
- Second iteration, developed as part of work with the UN to help account for non-profits as part of National Accounts.
- Designed to allow international comparison, so must work across national contexts

Issues

- Not tailored to the UK context
- Again, specific categories such as food banks, homelessness services or medical research are not included.
- Not collected routinely as part of charity registration or regulation
- Previous attempt to apply automatically (by David) a one-off and over a decade old

Methodology

Stage 1:

- Create a manually coded dataset (sample of UK charities)
 - Apply one ICNPTSO category per charity
 - Develop a bespoke system of UK charitable activity 'tags'
 - Apply as many 'UK-CATs' as applicable per charity

Stage 2:

- Create automated rules using code to assign the UK-CATs to all UK charities

Stage 3:

- Experiment with machine learning to automatically generate coding rules, to improve the process and automatically apply ICNPTSO categories (not the focus for this presentation)

Manual classification

- Manually classified 4,203 charities, from across the UK, based on what they write about their activities or objects in their application for charitable status.
- Classification done by 4 different people – weekly meetings to compare notes and agree principles and rules
- For each charity we added charitable activity / purpose ‘tags’
- We developed these tags as we went along, comparing as a group and adding to the list when needed.
- Also incorporated some from other schemes
- The list of tags became the UK-CAT (UK classification of charitable activities), a list of 254 tags arranged into a hierarchy of 24 groups.
- Mostly achieved early on, but not a ‘neat’ process as some backwards revision was inevitable
- Also applied one ICNPTSO category per charity

The UK-CAT

Stage 2: Automatic classification

- For each UK-CAT category, we developed a key word search rule to automatically apply the tag based on a charities' activities or objects.
- Used the manual dataset to provide word frequencies and to whittle down 'false negatives'
- At the same time, keeping a close eye on those being automatically classified to avoid 'false positives'

Table 1: Example keyword search rules

Classification category	Search terms ('regular expression')
Arts - Performing art - Choirs	<code>\b(choirs? choral chorus choristers singing singers?)\b</code>
Crime and Justice - Prevention and safety	<code>\b(crime (prevention reduction) public safety (prevention reduction) of crime)\b</code>
Education - scholarships	<code>\b(scholarships? bursar(y ies) grants for (student pupil)s?)\b</code>

Results – how well did it work?

- Applying the key-word search rules to the dataset of 201,990 active charities resulted in over 805,837 matches across all 254 UK-CAT tags.
- General ‘eyeball’ suggests that the tags do a reasonably good job, but are less efficient / parsimonious than human researchers.
- We ran comparison using 50 manually classified charities and found a ‘false negative rate’ of 24 per cent (though this is probably quite a harsh measure for a number of reasons).
- Many more tags provided by the automatic sampling than the human sampling, but the proportion of these that are ‘wrong’ (true false positives) is small (~4 per cent).

5 examples and their automated 'tags'

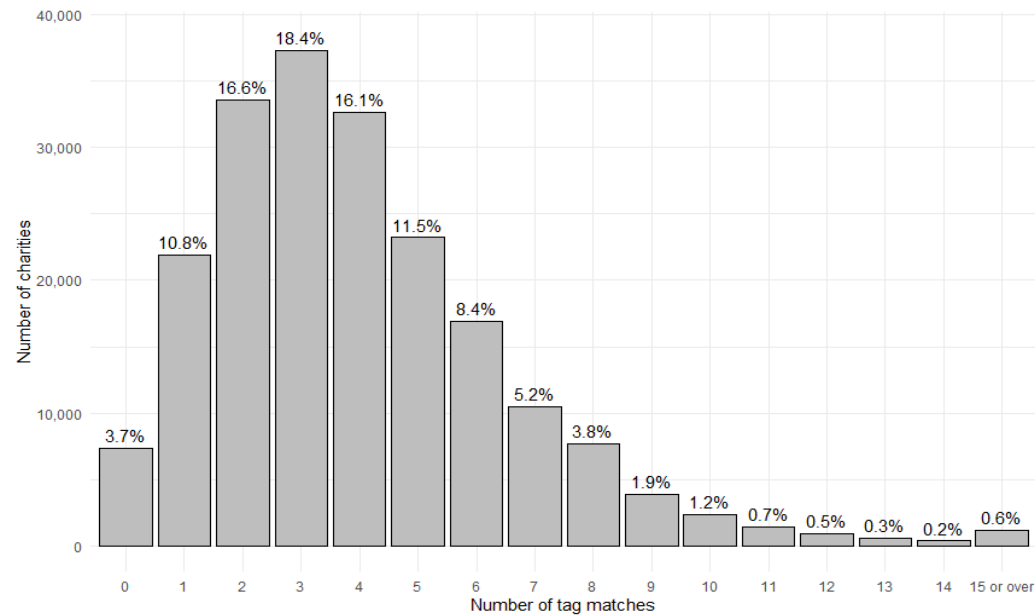
Table 2: Example charities and their UK-CAT tags

Charity name	Activities	UK-CAT tags
Corporation of The High School of Dundee	"The advancement of education."	Education, Schools, Secondary education
Northampton Scottish Association Fund	"Providing charitable donation to local charities on an annual basis"	Associations, Charity and VCS support
Craven police charity fund	"Supporting local causes in raising money"	Fundraising, Emergency services
The windfall centre limited	"The Windfall Centre is a not for profit organisation of professionals [sic] with expertise in the field of children's and young people's health, welfare and development. We provide therapeutic support to children and young [sic] people through the medium of play and creative activities."	Children, Young people, Health
1st Culter Rainbow Unit	Promoting the instruction of girls of all classes in the principles of discipline, loyalty and good citizenship.	Girls, Women, Citizenship

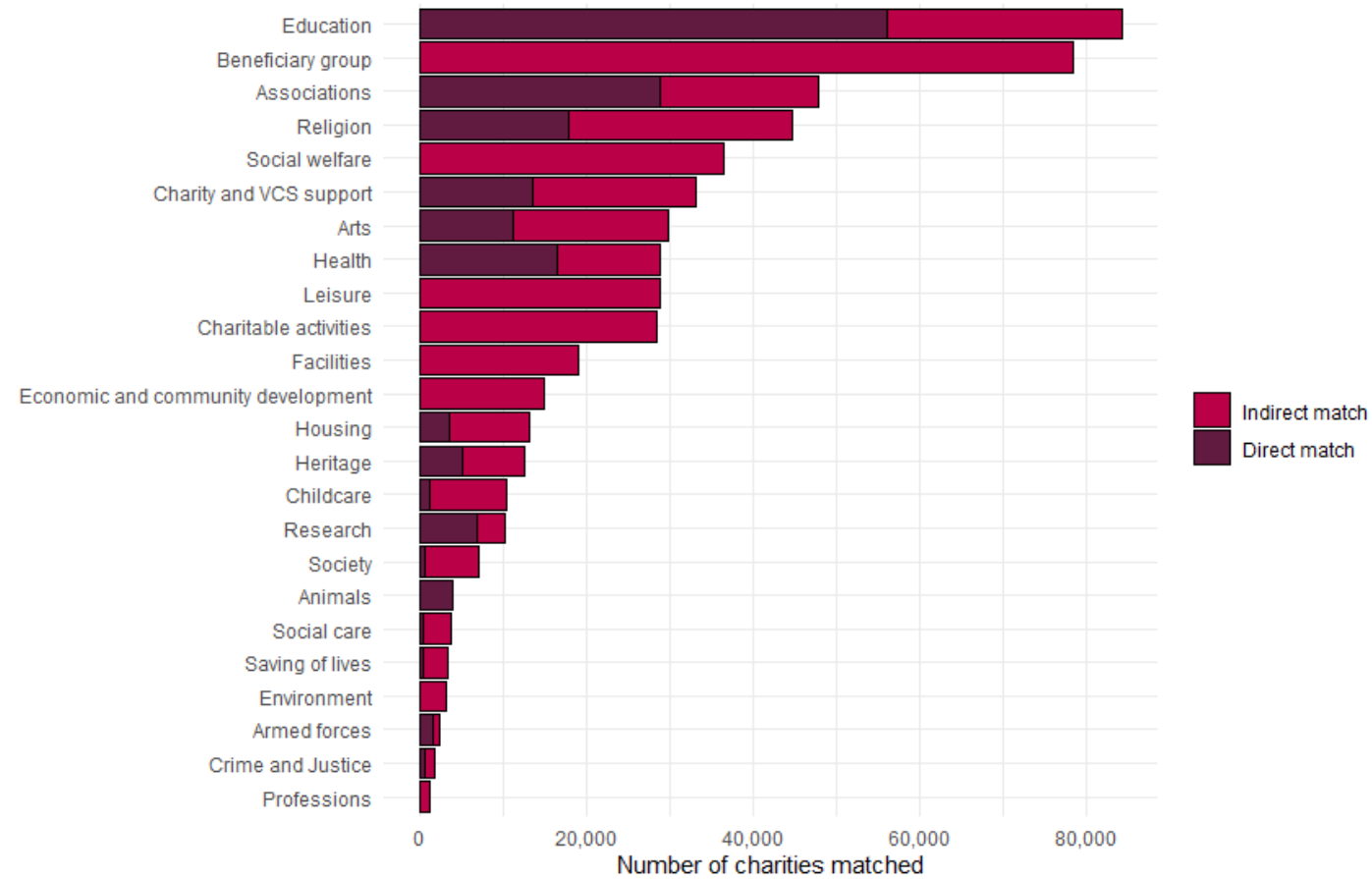
Tags per charity

- The mean and median number of tag matches per charity is four.
- 3.7 per cent have no tags. Issues include Welsh language, spelling errors, or insufficient detail.
- Possibly can be rectified by using information from the charity register to fill in some blanks.
- Some charities have loads of tags, especially when multi-purpose charities have very long activity descriptions

Figure 1: Number of tags per charity



Charities per tag



Most common individual tags (for English and Welsh charities)

Table 3: 20 most commonly applied tags

Tag	Matches	Percent of all charities	Tag	Matches	Percent of all charities
Education	84,283	41.7%	Health	28,953	14.3%
Beneficiary group	78,367	38.8%	Leisure	28,944	14.3%
Associations	47,857	23.7%	Charitable activities	28,498	14.1%
Religion	44,763	22.2%	Children	26,778	13.3%
Social welfare	36,580	18.1%	Church or place of worship	22,671	11.2%
Charity and VCS support	33,226	16.4%	Individual poverty	22,563	11.2%
Christianity	33,220	16.4%	Young people	21,725	10.8%
Schools	31,367	15.5%	Training	20,472	10.1%
Arts	29,867	14.8%	Facilities	19,019	9.4%
Religious activities	29,862	14.8%	Grant making	18,884	9.3%

Not as common as expected?

- Animals (2.0%)
- Social care (1.8%)
- Environment (1.6%)
- Crime and justice (0.9%)
- Abuse (0.8%)
- Social enterprise (0.2%)
- Racial justice (0.1%)

Next steps

- Finalise analysis and reporting
- Make some amendments to the UK-CAT
- Proceed with machine learning for UK-CAT and ICNPTSO
- Share for others to use and improve

Conclusion

- Keyword matching seems to have worked reasonably well overall
- But far from a perfect process, lots of room for feedback and improvement
- Hopefully helped to provide a bit more insight about what charities do and why
- Lots of caveats, but perhaps some disconnect between how we see charities and the reality