A Mood-based Genre Classification of Television Content

Humberto Corona and Michael O'Mahony
{humberto.corona, michael.omahony}@insight-centre.org

Centre for Data Analytics

Insight

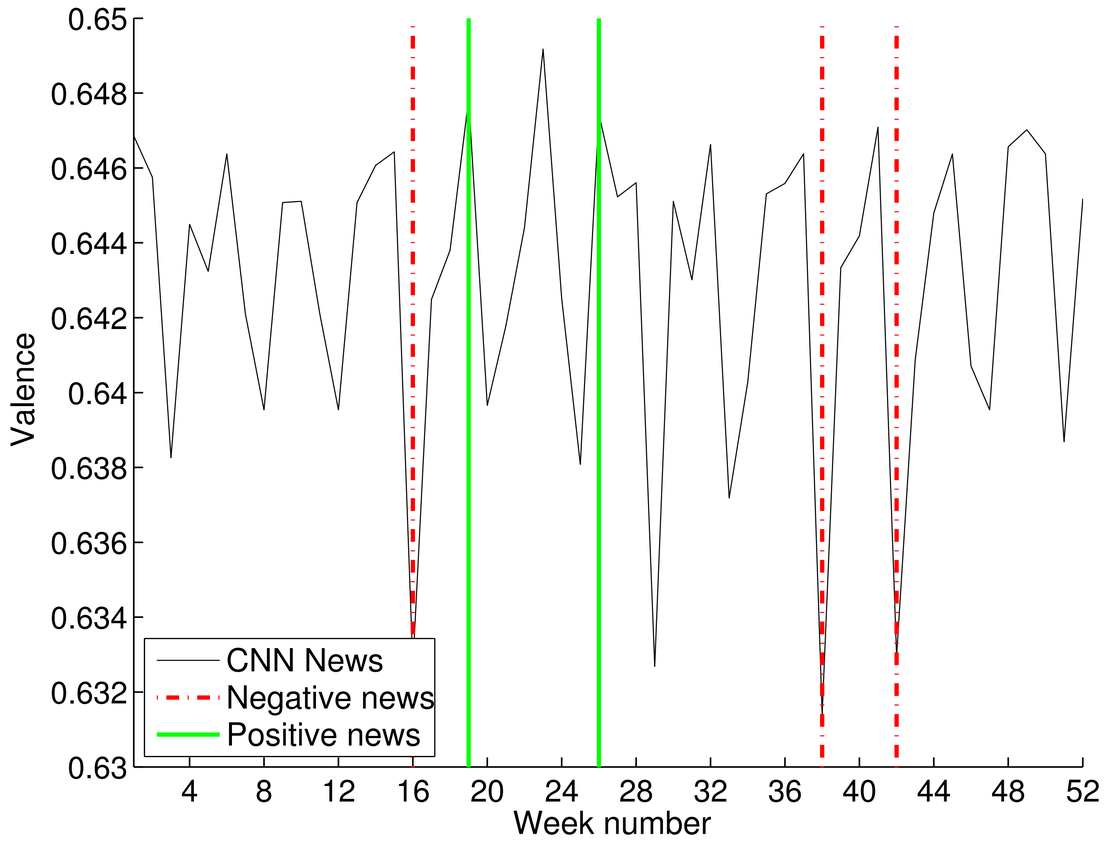
Television Classification

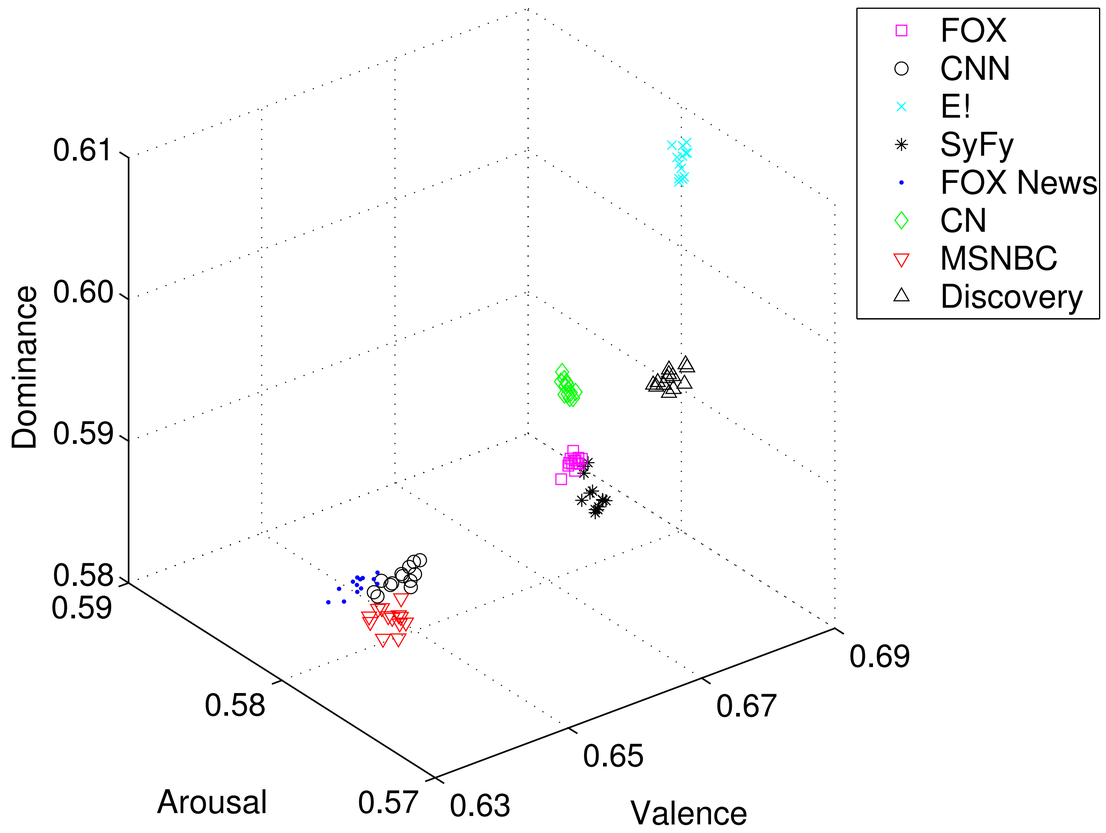
The classification of television content helps users organise and navigate through the large list of channels and programs now available. In this paper, we address the problem of television content classification by exploiting text information extracted from program transcriptions. We present an analysis which adapts a model for sentiment that has been widely and successfully applied in other fields such as music or blog posts. We use a real-world dataset obtained from the Boxfish API to compare the performance of classifiers trained on a number of different feature sets. Our experiments show that, over a large collection of television content, program genres can be represented in a three-dimensional space of valence, arousal and dominance, and that promising classification results can be achieved using features based on this representation. This finding supports the use of the proposed representation of television content as a feature space for similarity computation and recommendation generation.

The Data

All the television content information was obtained through the Boxfish API, which provides the electronic programming guide for various channels and the genre associated with each program. The *ANEW* dataset is a collection of 2,476 words annotated with emotional ratings in three dimensions — *valence*, *arousal* and *dominance*. The dataset was created using human assessment.

Feature Analysis, A Year of Television Broadcasting





Program Genre Classification

Feature Group	Feature
ANEW features (valence, arousal, dominance)	Minimum value
	Maximum value
	Mean value
	Standard deviation
	Median value
Stylistic features	Num. words
	Num. unique words
	Num. unique ANEW words
	Max. word frequency

Table : Meta-feature representation: ANEW and stylistic feature groups.

Classification Results

Genre	VSM			Meta-features		
	TP	FP	AUC	TP	FP	AUC
Animated (120)	0.725	0.004	0.942	0.792	0.161	0.885
Documentary (65)	0.554	0.025	0.870	0.447	0.144	0.780
Horror (24)	0.208	0.013	0.649	0.583	0.107	0.865
Newscast (41)	0.976	0.053	0.972	0.707	0.056	0.905
Reality (93)	0.882	0.260	0.845	0.333	0.064	0.729
Weighted Average	0.729	0.084	0.885	0.583	0.115	0.824

Table : Classification results: Naïve Bayes classifier — 5-fold crossvalidation.

Conclusions

Television content, at least at a high (i.e, channel) level, can be discriminated by the proposed three-dimensional space of affect. Moreover, there is evidence to suggest that the meta-features based approach have the potential to contribute to enhanced classification performance, particularly if used in combination with other feature types. Additional information (including source code) can be found at `https://github.com/hcorona/recsystv-2014`.

This work was supported by Science Foundation Ireland through the CLARITY Centre for Sensor Web Technologies under grant number 07/CE/I1147 and through the Insight Centre for Data Analytics under grant SFI/12/RC/2289

DCU · NUI Galway OÉ Gaillimh · UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh · UCD DUBLIN · sfi Science Foundation Ireland