

研究テーマ：Rを用いたテレビドラマの視聴率予測

名列番号:006 門端卓朗

1. 研究内容

統計処理ソフトウェア R を用いて、回帰分析によるテレビドラマの視聴率予測を行い、ドラマの視聴率とキャストやスタッフ等の特徴との関連性について考察する。

2. 研究に使用したデータ

本研究ではwebサイト「Audience Rating TV」(<http://artv.info/>)より、2003年1月から2009年9月までに放送されたドラマのデータを抽出し行列にして用いた。行名をドラマのタイトル、列名を平均視聴率、キャスト名、時間帯などの特徴とした358×2741行列で、人物のいる、いないを”1”, ”0”で表したものと、出現度に応じて重み付けしたものの2通りで予測を行った。

3. 実験方法

今回の実験では、「キャスト」「プロデューサー」「脚本」「演出」「曜日と時間帯」の5つ特徴でそれぞれ平均視聴率の回帰予測を行う。予測には関数lm(線形回帰)とksvm(サポートベクター回帰)を使用する。ksvmについてはrbfdot(ガウシアンカーネル)とvanilladot(線形カーネル)の2つのカーネルについて予測する。

また予測した結果と実際の視聴率との誤差を求め、その平均を比較することでどの特徴との関連性が高いかを考察する。

4. 実験結果・考察

図1, 2は予測結果と実際の視聴率との誤差の絶対値の平均を示したものである。最も誤差の絶対値の平均が小さかったのは「曜日と時間帯」のksvm予測(rbfdotカーネル)であり、視聴率への影響がより大きい特徴であると考えられる。逆に最も誤差が大きかったのは「キャスト」のlm予測で、他の特徴に比べて要素数(人数)が多かったため予測精度が低くなってしまったと思われる。用いたデータによる違いは特に見られず、原因としては要

素のほとんどが0であり重み付けの影響が小さかったことが考えられる。またどの特徴においても、lmでは予測値の振れ幅が比較的大きく、ksvmでは触れ幅が小さくなるような傾向が実験中に見られた。

5. まとめ

実際の視聴率を決める要因には様々なものがあり、現実での影響を考えないデータでの予測は難しいと思われたが、特定の時間帯によってある程度の傾向が見られたというのは意外な結果であり、視聴率は決してキャストだけでは決まらないということが分かった。今回の実験では5つの特徴に絞って予測を行ったが、このほかの特徴にも視聴率を決める要因はありそうだと実験を終えて感じた。

	lm	ksvm(rbfdot)	ksvm(vanilladot)
プロデューサー	4.26	3.21	3.65
脚本	3.45	3.11	3.30
演出	3.64	2.96	3.64
キャスト	40.81	3.04	-
曜日と時間帯	2.97	2.85	2.94

図1 重みなしデータでの予測の誤差平均

	lm	ksvm(rbfdot)	ksvm(vanilladot)
プロデューサー	4.66	3.22	4.16
脚本	3.51	3.16	3.48
演出	3.74	2.87	3.57
キャスト	26.88	3.07	2.98
曜日と時間帯	-	-	-

図2 重み付けしたデータでの予測の誤差平均