

今会いに行ける声優: ブログに登場する位置情報単語を用いた女性声優の出現位置予測

@y_benjo

Abstract

声優オタクの関心として、声優は普段どこにいるのか、どこに行けば「出待ち」や「入り待ち」などのアナーキーな手段を用いずとも、「偶然」声優と出会うことができるのか、というものがある。今回我々は、声優のブログを入力として、「声優がブログ中において言及した都内の駅名」を「その駅に出現した」とし、声優の出現位置予測問題に取り組む。99名の女性声優のブログデータを用いた実験の結果、22%の精度で最後に出現した駅を当てる事に成功した。

Keywords: 声優統計, 出現位置予測, トピックモデル

1. はじめに

声優オタクの関心として、声優は普段どこにいるのか、どこに行けば「入り待ち」や「出待ち」¹などのアナーキーな手段を用いずとも、「偶然」声優と出会うことができるのか、声優があの時ラジオで言及していた飲食店はどこなのかを知りたい、というものがある。

しかし通常、声優は自分の位置情報を明示的に発言することはなく、訪問した店舗の写真やランチの写真などをブログやツイッターなどに発信するのみに留まっている。もし、それらの写真がどの場所で撮影されたのかが判明すれば、声優の出現位置がピンポイントで判定可能である。この問題は Location Estimation [1] と呼ばれるタスクであるが、機械学習で取り組むのは非常に困難であり、人手での予測結果とあまり変わらない(そして低い)精度しか実現できてないのが現状である [2]。この問題に対する解決案としては、今年著しい成果を上げている「声優の携帯電話」²に代表されるような、クラウドソーシングによる人手でのアノテーションが考えられるが、それでもなお、携帯電話と異なり飲食店、特に店内の画像は曖昧性が高く、前述の研究 [2] で示された精度よりも低いものしか実現できないと思われる。

もう一つのアプローチとして、記述された文章から位置情報に相当する固有表現を抽出するという方

法が考えられるが、こちらについては予備実験を行ったが、やはり曖昧性が高く、やはり最終的には人手でのアノテーションを必要とすることが判明したため、次回取り組むタスクとする。

今回は声優の出現位置予測の第一歩として、もっとも単純な問題として、声優のブログ記事から都内の駅名に一致する単語を抽出し、それらを声優の出現位置とした上で、声優の出現位置予測に取り組む。これにより、

- そもそもこのタスクがどの程度難しいのか
- すなわち声優の出現位置に規則性はあるようなのか

という観点からの見通しを立てることを目標とする。

2. 手法

今回我々は、

1. 声優には、それぞれ「出現範囲」があり、近い位置に出現しやすい。例えば、中央線沿線が好きな声優は中野、吉祥寺、西荻窪に出現しやすい、大江戸線のヘビーユーザーであれば新宿から麻布十番、六本木に出現しやすいと考えられる。
2. 声優は各自の興味に従って訪問する位置を選ぶ。オタク趣味が強い声優ならば、池袋や秋葉原に出現しやすく、酒を飲むのが好きな声優ならば新橋や上野に出現しやすいと考えられる。

Email address: ybenjo.repose@gmail.com (@y_benjo)

URL: <http://d.hatena.ne.jp/repose/> (@y_benjo)

¹ イベント会場などに声優が入る/出るタイミングを見計らい出入口で待機する行為。通常禁止されている。

² <http://seiyunokeitaidenwa.ssig33.com/>