NovelPerspective

Anonymous ACL submission

Abstract

We present a proof of concept for a tool to allow consumers to subset ebooks, based on the chapters main character. Many novels have multiple main characters, and vary with each chapter which character the story is focused on. A well known example is George R. R. Martin's "Game of Thrones" novel, and others from that series. The NovelPerspective tool detects which character the chapter is about, and allowed the user to generate a new ebook with only those chapters. The detection of main character can be done by many means. We present two simple baselines, and several machine learning based methods.

1 Introduction

Many books have multiple main characters, often each character is written from the perspective of a different main character. Different sections are written from the perspective of different characters. Generally, these books are written in limited third-person point of view (POV); that is to say the reader has little or or more knowledge of the situation described than the main character does. Having a large cast of character, in particular POV characters, is a hallmark of the epic fantasy genre.

We propose a method here to detect the main/POV character for each section of the book. Detecting the main character is not a difficult task, as the strong baseline result shows. However to our knowledge there does not exist any current software to do this. We attribute this lack to it being impractical to physically implement until recent times. The surge in popularity of ebooks has opened a new niche for consumer discourse processing. Tools such as the one present here, give the reader new freedoms in controlling how they consume their media.

We focus here on novels written in the limited third-person point of view. In these stories, the main character is the point of view (POV) characters. Some examples include: Across its 15 books, Robert Jordan's "Wheel of Time" series which has 146 POV characters[1]). Only about one fifth of the total word count was from the POV of the "main character". George R.R. Martin's "A Song of Ice and Fire", have over 30 POV characters in the books published so far [2]. Other well-known books meeting this requirement include: Robert Jordan's "Wheel of Time" series, all the novels from Brandon Sanderson's "Cosmere" universe, Brent Week's "Nightangel" and "Lightbringer" series, Steven Erikson's "Malazan Book of the Fallen" series, and thousands of others. This is also of interest for works written in omniscient third person point of view, such as J. R. R. Tolkien's "Lord of the Rings", which also may feature a focus on different main characters however the correct split is much less clear in these cases.

The utility of dividing a book in this way varies with the book in question. Some books will cease to make sense when the core storyline crosses over different character. Other novels, particularly the large epic fantasy stories we are primarily considering, have many parallel story lines focused on the different characters that only rarely intersect. While we are unable to find formal study on this, many readers speak of "skipping the chapters about the boring characters", or "Only reading the real main character's sections". Particularly on a re-read, or after already having consumed the media in some other form such as watching a movie adaptation, or reading a summary.

[1] http://wot.wikia.com/w3iki/Statistical_analysis

[2] http://awoiaf.westeros.org/index.php/POV_character

2 Character Detection Systems

2.1 Baseline systems: First and Most Common

3 Experimental Setup

3.1 Datasets

We make uses of two series of books selected from our own personal collections. The first four books of George R. R. Martin's "A Song of Ice and Fire" series (hereafter referred to as ASOIAF); and the two books of Leigh Bardugo's "Six of Crows" duology (hereafter referred to as SOC).

The requirements of the books to use in the training and evaluation of the NovelPerspective system is that they provide ground truth for the chapter's main characters. These books do so in the chapter names – each matching to a character name.

3.2 Evaluation Details

In the evaluation, the books chapters are pre-separated into body text and chapter name (character name). The detection systems are given the body text and asked to predict the character names. To mimic the human users ability to select multiple aliases of a character, before final classification the scores of character's nicknames are merged. For example merging `Ned` into `Eddard`.

3.3 Implementation

The text is preprocessed using NLTK (**?**) to added features. It is XGBoost tree ensemble's are used for the machine learning methods. The full implementation is available at https://github.com/oxinabox/NovelPerspective/

4 Results and Discussion

5 Conclusion