COMP6590 Project Report - Active-Art

Trygve Sundby Ulsnes (tsu3)

`https://github.com/lopen/active-art`

Abstract

This is a report providing a system description and evaluation of Active-Art. A system that utilises Strava activity data to generate prompts used by GPT to write poems, which in turn are evaluated internally by the system. Active-Art is able to pose as a creative system, however it is limited in its knowledge-base and self-evaluation.

1 Introduction

Sharing is a natural part of what people do, we enjoy telling others what we have done or what we have experienced, and we do it through a multitude of mediums like speech, pictures, song or poems. The inspiration for this project comes from the activity app Strava, which lets users share their activities with others, posting pictures and numerical statistics to back it all up. When sharing how far or fast you have run for example, the natural way is to use numbers to back this up, "I ran 10km yesterday!" or "I did 4 intervals of 15km/h for 2 minutes this morning", but this can be limiting in sharing what that activity actually felt like. Active-Art aims to take all this numerical data and generate a poem that will represent not only the run but also the day you ran and the effort you exerted during it. This report will give a breakdown of the Active-Art system, other systems that inspired its creation and the results that Active-Art is able to produce. Finally the system will be evaluated through Colton's Creative Tripod and Ventura's three characteristics.

2 Background

There are two main papers that have contributed significantly to what Active-Art is today. The first is the 'Mere Generation' paper by Dan Ventura, and the second is 'Witscript 2: A System for Generating Improvised Jokes Without Wordplay' by Joe Toplyn. Both papers were discusses during our seminars and I am grateful that these papers were highlighted, as they brought about interesting discussions and gave rise to ideas for key features that are now the backbone of Active-Art.

2.1 Mere Generation

'Mere Generation' by Dan Ventura introduces an important discussion within the computational creativity community that seems to have been neglected to some extent. This is the discussion around what is merely generation and what is, or does it mean, to have something be *more* than mere generation. The significance of this paper cannot be understated, as someone new to the computational creativity scene, I benefited immensely from reading and discussing this paper. It forced me to think about where this line can be drawn, if it can be drawn at all, and allowed me to pose the question to what extent is Active-Art more than just mere generation.

The paper presents nine different algorithms, increasing in complexity and that seemingly stray further and further from mere generation. There are seven levels of generation that are introduced the more complex the algorithm gets, and around algorithm six and seven is when I believe the algorithms go from mere generation to something more, something creative.

While the paper gives these algorithms as examples, they are by no means a definitive answer to the question of what it means to be mere generation. The main takeaway Active-Art has utilised is the idea of a self-evaluation part in its main algorithm. This is what I believe helps set a creative system apart from something that is mere generation and will be highlighted further in section 3

2.2 Witscript 2

'Witscript 2: A System for Generating Improvised Jokes Without Wordplay' by Joe Toplyn is a system description of a creative system, Witscript 2. Witscript 2 is similar to its predecessor, Witscript, in that it uses the Basic Joke-Writing Algorithm created by Toplyn, however unlike the Witscript, Witscript 2 attempts to create jokes with wordplay.

Witscript 2 is powered by OpenAI's GPT-3, a large language model which as become increasingly popular due to its adaptability and capabilities. The system communicates back and forth with GPT-3 through the Basic Joke-Writing Algorithm, to select topics, 'topic handlers', creating punchlines and generating the angle between the topic and the punchline.

While the jokes generated perform similarly to the original Witscript system and outperforms GPT-3 on its own, Witscript 2 is outperformed by the human comedian. While it is not the perfect joke-writing system, it does given a look into how powerful a large language model can be and how a creative system could leverage this as a tool to generate creative artifacts.

3 Methodology and Design

As a system, Active-Art is written in Python and utilises the MetOffice Datapoint API, Strava API and OpenAI's large language model GPT-3 (model 3.5-turbo). Active-Art's design can be split into three separate stages that together produce seemingly creative artifacts:

1. Data Collection
2. Prompt Generation
3. Evaluation

A simple algorithm has been written, where the data collected from Strava gets turned into a written prompt that is fed to GPT. GPT's response is then evaluated by an evaluation function and if it passes the resulting artifact is saved to a separate text file, along with the primer and prompt used to generate the artifact. The rest of this section will cover the different steps in more depth. Before the prompt is submitted to GPT, a system message is sent to GPT to inform it of how it should interpret the coming prompt. The system message, or primer, used was: "You are a runner and poet, that enjoys writing poems about runs you have done. You also enjoy wordplay and clever rhyme schemes".

3.1 Data Collection

This is the first stage of the system, where the Strava API is queried and data from previously completed runs gets collected and saved in a CSV format. The system can be executed such that it collect es more than one run to create poems for, however the actual generation of the poems are separate and the data for each run is exclusively used for that single run. Active-Art is currently limited to the author's activities, due to account lookup and collection of other user's data without their express consent gives rise to ethical concerns regarding data privacy.

3.2 Prompt Generation

Prompt generation is where Active-Art takes the collected data and creates a unique prompt which is further fed to GPT. This prompt is based on a template that gets tailored to the run which the prompt is being generated for.

Strava data	**Determines**
Round up((moving time / 60) / (distance / 1000))	Poem length in number of lines
Round up(cadence / 10)	Max number of syllables per line
Avg speed and max speed	Feel when read
Avg heart rate and max heart rate	Synonyms associated with HR zones
Weather	Tone of poem

Table 1. Shows what data determines what in the prompt generation.

3.2.1 Poem length

As show in figure 1, poem length is determined by calculating the average pace of the run in question. Pace is the time it takes to run one kilometer in minutes. A hard-limit of 10 lines has also been applied, such that the poems would not exceed this, due to GPT seeming to struggle with creating a coherent more than 10 line poem through testing. This hard-limit also makes the generation quicker.

3.2.2 Number of syllables

The number of syllables per line is determined by the average cadence of the run. Cadence, or stride rate, is the number of steps in one minutes when you run.

3.2.3 Descriptive synonyms

Descriptive words are selected to further describe the run. These words are selected by using max and average heart rate. When running there are 5 heart rate zones that inform us of how hard we are working during the activity. These zones are determined by our absolute max heart rate and are percentiles from 50-60 percent (zone 1) increasing by a magnitude of 10, up to 90-100 percent (zone 5) of our absolute max (Polar, 2016). Active-Art uses this information to select form a list of lists two words, one for whichever zone the average heart rate falls into and one for the zone the max heart rate of the run falls into. There are 10 words to choose from at random and the words used to find the 10 synonyms were very light, light, moderate, hard and maximum.

3.2.4 Poem feeling

Poem feeling is selected in a similar way to how the descriptive synonyms are selected. However, it uses the average and max speed from the run to select a negative, neutral or positive word, from a list of 10 words, increasing in positivity (negative -¿ neutral -¿ positive).

3.2.5 Tone of poem

The tone of the poem is based on weather and the algorithm takes the longitude and latitude of where the run ended and queries the MetOffice with this position to get a reading of current weather. As a result, the prompt asks GPT to generate the poem with tones of a cloudy day, for example.

3.3 Evaluation

For the evaluation function, there are two main parts: structural and thematic evaluation. The structural evaluation takes into account the limitations placed upon GPT in the prompt, so the limit on lines and syllables per line. The thematic evaluation uses Vader Sentiment analysis to analyse each line of the poem against its negative, neutral or positive feeling, and checks its collective score against the prompt. If either of these evaluations fail, the artifact will be regenerated by GPT.

4 Results

Active-Art as a system works and produces output as can be seen in figure 1. The primer used is the same as section 3 and the prompt generated for the poem in figure 1 was:

"Write a poem about a afternoon run you have just completed.
The poem must be 6 lines long, where the maximum number of syllables per line is 9.
The poem must express feelings of tough and at times feverish activity.
The poem must convey tones of a cloudy day and must convey a feeling of a pleasant time."

Parts of the system could definitely be implemented with more complexity, like for instance the evaluation function. As it is written right now, if the artifact fails the evaluation the entire artifact is scrapped and an entirely new one is generated. Further work here could be to make it so that only

Pounding pavement, muscles ache

Sweat drips down as I partake

I run with feverish glee

All while clouds form over me

The miles may be long and tough

But I feel alive, strong enough

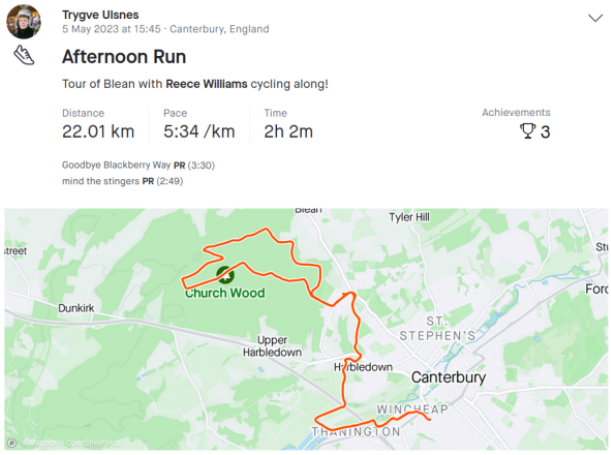

Figure 1: Artifact (poem) along with a screenshot of the activity

the offending part of the artifact gets re-generated. This would save time and could possibly produce better output.

The function for selecting the tone of the poem could also use some work, as it only gets the weather forecast of the current day when the system is run. A better implementation would be to use the timestamp of when the run was started or completed, to grab historical weather observation data to inform the prompt generation, which could give a much better representation of the run. This was not implemented, as limitations to the MetOffice API made it difficult to implement beyond what it currently is.

5 Evaluation

How you have tested/evaluated your creative software; results of testing; a discussion of the evaluation in terms of the end goal of producing creative software (and any relevant issues arising). You should use at least one of the creativity evaluation frameworks discussed in the lectures.

I will be using Colton's Creative Tripod and Ventura's three characteristics to better evaluate if Active-Art is a creative system or not. While I am not a poet, I am a runner and the artifacts produced are based on runs I have completed. This has allowed me to see if the artifact produced actually convey feelings or tones of the activity I did, or if it is a complete miss. While this is a good point to bring up, it also means that no third-party has had the chance to read the poem and that it might not be up to a certain poetic standard. These are all points to take into consideration when evaluating a system such as Active-Art.

Colton's Creative Tripod is comprised of three parts: skill, appreciation and imagination. (Colton, 2008) Through its use of GPT and its evaluation function, Active-Art is able to display a limited skill-set and appreciation within the domain it is working in (poetry / language). While GPT is able to create rhyme schemes to a certain extent, it is by no means able to create anything as complex or vivid as a seasoned poet or the doer of the activity. This can be attributed to the limited knowledge that is passed to GPT via the prompt and could most likely be expanded with more time and effort.

Through its use of the evaluation function, Active-Art is able to enforce rules such as line count, syllable count and even sentiment of the poem, which shows the slight appreciation of the system being able to assess the artifacts it produces. However, while the system is able to evaluate its output, this is only at a very basic level. The system is unable to modify its output and re-generates a new

output if the previous output fails in any regard. While the evaluation is rudimentary in its nature, it still does evaluate the output.

In terms of imagination, Active-Art is limited as it follows a template for its prompt and a limited knowledge-base to populate this template with. The imaginative part of the system lies in its data-driven approach and how the metadata of the run and weather data informs the prompt. While the knowledge-base could be expanded and the prompt could be made even more descriptive, this could prove to make it too convoluted for GPT to generate a poem that has any meaning.

Ventura's three characteristics - novelty, value and intentionality - are used as surrogate indicators for the existence of (or lack of) some form of creativity (Ventura, 2016). Novelty is perhaps Active-Art's weakest point, as mentioned earlier the system possesses a limited knowledge-base and restrictive template, and while the system is able to create poems the overall poetic structure and rhyme schemes can be lacking.

Active-Art provides a unique value to the user of the system, as it creates a unique rendition of their run in form of a poem. It gives the user another way to represent their activity other than numerical data. Active-Art is intentional in its creation. The system sets out to create a poem that represents a run and the result is just that. While it is obviously up for interpretation if the poem is good or not, it cannot be denied that the system does not use data collected about a run to create a poem.

6 Conclusion

While Active-Art does produce artifacts that could be called poems, there are obvious further steps that could be taken to improve this. For example specifying poem types in the prompt so that GPT is more informed about the structure of the poem that it is trying to write. The evaluation of the artifacts can also be improved upon, it would be interesting to have GPT evaluate its own output. Fixing the weather collection and adding more data points like time of day and elevation could help tailor the prompt and poem it generates to be an even more convincing representation of a given run.

Further work on the system would be to improve the functions that are already there. The structure of a creative system is in place, however I think the depth and complexity required to write what could be called a creative system is not. With that said, I find that Active-Art is able to pose as a creative system and that the poems produced are able to represent the run completed to a certain extent.

7 References

Colton, Simon. (2008). Creativity Versus the Perception of Creativity in Computational Systems. AAAI Spring Symposium - Technical Report. 14-20.

Colton & Wiggins. (2013). Computational Creativity: The Final Frontier?. accessed 20th March 2023, https://computationalcreativity.net/iccc2014/wp-content/uploads/2013/09/ComputationalCreativity.pdf

Polar (2020) "Heart Rate Zones: The Basics," Polar Journal. Available at: https://www.polar.com/blog/running-heart-rate-zones-basics/.

Toplyn, Joe. (2022). Witscript 2: A System for Generating Improvised Jokes Without Wordplay. ICCC 2022, no. 92. accessed 8th February 2023, https://computationalcreativity.net/iccc22/papers/ICCC-2022_paper_92.pdf

Ventura, Dan. (2016). Mere Generation: Essential Barometer or Dated Concept?. ICCC 2016. accessed 8th February 2023, https://www.computationalcreativity.net/iccc2016/wp-content/uploads/2016/01/Mere-Generation.pdf