

INTISARI

ALGORITME GENETIKA UNTUK MENSINTESIS RESPON PADA CHATBOT AIML BERBASIS GENERASI

Fonda Krista Nurhanditya
11/316781/PA/13908

Chatbot adalah program yang dapat mensimulasikan percakapan manusia dalam bahasa alami. Berdasarkan cara suatu respon dihasilkan, ada 2 jenis *chatbot*, *retrieval-based* dan *generation-based chatbots*. Sistem *retrieval-based* hanya semata-mata memperoleh respon dari basis pengetahuan yang ada, sehingga hanya dapat menghasilkan respon yang umum atau telah ditentukan sebelumnya. Sebaliknya, sistem *generation-based* memodelkan bagaimana respon dapat dihasilkan untuk mempelajari fitur alami dari basis pengetahuan yang diberikan, sehingga dapat mensintesis teks-teks baru melalui konteks yang belum teramati sebelumnya.

Dalam penelitian ini disajikan suatu *generation-based chatbot* kombinasi dari komponen *retrieval-based* dan komponen *generation-based*. Komponen *retrieval-based* dikembangkan dengan arsitektur AIML. Sedangkan, komponen *generation-based* adalah rutin algoritme genetika yang secara generatif dapat mentransformasi teks yang diberikan. *Chatbot* menghasilkan beberapa respon sesuai spesifikasi AIML sebagai populasi awal dari rutin algoritme genetika. Respon-respon tersebut kemudian dievolusi selama beberapa generasi dengan hanya menggunakan operator mutasi untuk membangkitkan respon-respon baru yang berbeda dari nilai awalnya. Sistem mengevaluasi setiap respon tersebut berdasarkan nilai entropy dan memilih yang tertinggi sebagai respon akhir.

Sistem diuji untuk setiap peningkatan parameter probabilitas mutasi dan generasi maksimum terhadap nilai *entropy* dan *word error rate (WER)*. Nilai *entropy* meningkat sebanding terhadap nilai-nilai parameter dengan nilai rata-rata maksimum tercapai 10.81 untuk probabilitas mutasi 5% dan generasi maksimum 20, sebaliknya standar deviasi yang menurun menandakan entropy akan memuncak pada suatu nilai maksimum. Sementara itu, *WER* meningkat dengan nilai rata-rata maksimum 67.5% untuk probabilitas mutasi 5% dan 20 generasi maksimum, beberapa pengujian bahkan mencapai *WER* 100%. Hasil pengujian menunjukkan bahwa sistem dapat menghasilkan respon-respon yang lebih *random* dan lebih beragam dengan setiap mutasi dan generasi yang dilakukan.

Kata Kunci: *Algoritme genetika, chatbot, retrieval-based chatbot, generation-based chatbot, AIML, entropy, word error rate*

ABSTRACT

GENETIC ALGORITHM FOR RESPONSE SYNTHESIZATION ON GENERATION-BASED AIML CHATBOT

Fonda Krista Nurhanditya
11/316781/PA/13908

Chatbot is a program that could simulate human conversation in natural language. Based on how responses are delivered, there are 2 kinds of *chatbots*, *retrieval-based* and *generation-based chatbots*. *Retrieval-based* system solely derive its response from an existing knowledge base, as such it would only produce predefined or generic responses. Whereas *generation-based* system models how the response could be generated to learn the natural feature of a given knowledge base, thus is able to synthesize new texts from unseen contexts.

This paper presents a *generation-based chatbot* system that is a combination of both *retrieval-based* and *generation-based* components. The *retrieval-based* component is developed using AIML architecture. On the other hand, the *generation-based* component is a genetic algorithm routine that generatively transforms given texts. The *chatbot* generates some responses using AIML specification based on a given input and treats them as initial population for the genetic algorithm routine. The responses then evolved several times through generations using only mutation operator to generate new responses that varied from its initial value. The system then evaluate each responses by its entropy and choose the highest one as the final response.

The system was tested on increasing mutation probability and maximum generation parameters to evaluate the responses generated based on the value of *entropy* and *word error rate (WER)*. The *entropy* shows steady increase, proportionate to the parameters, with the maximum average attained being 10.81 for mutation probability of 5% and maximum generation of 20, while it's standard deviation declines, signifying that the entropy would be capped at a maximum value. Meanwhile, *WER* tends to get higher as the parameters went up with the maximum average being 67.5% for 5% mutation probability and 20 maximum generation, moreover, some of the test sample even reach word error rate of 100% which particularly create an entirely new sentence. These result show that the system is able to generate more *random* and more diverse responses with every mutation and every generation occurred.

Keywords: *Genetic algorithm, chatbot, retrieval-based chatbot, generation-based chatbot, AIML, entropy, word error rate*