A Study on the Interpretability of a Fuzzy System to Control an Inverted Pendulum

Bahadir Zeren∗, Muhammet Deveci∗, Simon Coupland†, Robert John∗, Ender Özcan∗

ASAP, School of Computer Science, University of Nottingham, Nottingham, UK∗
Institute of Artificial Intelligence, De Monfort University, Leicester, UK†

Email: {bahadir.zeren, muhammet.deveci, robert.john, ender.ozcan}@nottingham.ac.uk∗, simonc@dmu.ac.uk†

Abstract—Fuzzy systems mimic human reasoning and provide solutions to problems under uncertainty via ‘computing with words’. This particular strength of fuzzy systems is often discarded in some real world applications where the fuzzy sets are designed for control problems or created through training using historical data. This study explores the interpretability of fuzzy systems by generating ‘meaningful’ fuzzy sets using a dictionary constructed by humans and fuzzy transfer learning. The inverted pendulum control problem is used as a case study. The empirical results show that interpretability of a fuzzy system is achievable even for this problem at the expense of a ‘slightly’ reduced performance.

Index Terms—Interpretability, Fuzzy Transfer Learning

I. INTRODUCTION AND MOTIVATION

The motivation behind this research is that the fuzzy logic community often make claims that because the user/expert can interrogate both the fuzzy rules and fuzzy sets and are therefore in some way ‘meaningful’ or ‘interpretable’. For the fuzzy sets by interpretable we mean they have meaning. If a set ‘low’ is used, for example, we would expect it to appear towards the left hand side of the domain. If there are no fuzzy sets below low then one would expect the set to be either a shoulder set or (if triangular) the right hand side of the triangle. The supposed interpretability of fuzzy systems is one of the unique selling points over artificial neural networks that are essentially black box. However, in most real world applications, this is not true:

• Where the parameters of the system are trained using historical data (through ANFIS, Artificial Neural Networks for example), the final trained fuzzy sets are not usually sensible.

• In control applications the sets (whatever shape) have no meaning but are there to provide a mapping from inputs to outputs with the control engineer only interested in performance of the control system.

Zadeh introduced the notion that fuzzy logic can deliver Computing with Words [1] and many researchers have used this terminology in their work, however, we posit that they are not ‘doing’ Computing with Words, but jumping on the band wagon. This is initial work and we decided to see if we can investigate this through a control application and selected the inverted pendulum control problem.

The inverted pendulum problem on a cart is a classic control system problem and one of the most important problem in dynamics [2], [3]. It is an under actuated, high order, multi-variable, and unstable non-linear system, which is widely used as a benchmark problem in control theory. Therefore, it is an ideal model for testing various theories in the field of control engineering [4]–[7]. A flexible and intuitive way of expressing indefinite responses in system of many different stable controllers can be provided by fuzzy logic controller [7].

Some researcher on the inverted pendulum have been studied in recent years based on various methods including the state-space controller [8]–[10], the proportional integral derivative controller [11]–[13], the artificial neural network control theory [14]–[17], the sliding-mode control theory [15], [18]–[20], linear quadratic regular theory [13], [21], [22] and fuzzy logic controller [7], [23]–[26].

Fuzzy logic control that incorporates ambiguous human logic into computer programs [26] is based on fuzzy sets and fuzzy inference [27]. The fuzzy sets proposed by Zadeh [28] as a mathematical tool for useful modeling various types of uncertainty that include vague, ambiguous and imprecise data [29]. Fuzzy sets have been successfully utilised in many real-world applications, such as, controller design [30], [31], signal processing [32], decision making [33], medicine [34], risk analysis [35], electronic control system [36], [37], process planning [38], and so on.

The use of fuzzy systems for control problems is not new. The stabilisation of an inverted-pendulum is a complex problem requiring a nonlinear model [39]. The use of a fuzzy logic controller for a variant of the problem to stabilise an inverted pendulum on a cart is discussed in [4], [40], [41].

There has been a growing number of studies on the control of inverted pendulum by using fuzzy logic control. Ma et al. [42] addressed the analysis and design of the fuzzy controller and the fuzzy observer on the basis of the Takagi-Sugeno (T-S) fuzzy model. The main contribution of the study is to improve the separation: fuzzy controller and fuzzy observer can be designed independently. Wang et al. [43] attempted a design methodology for stabilisation of non-linear system for fuzzy control utilising Takagi-Sugeno fuzzy model. Magana and Holzapfel [44] proposed an experimental setup of a fuzzy logic controller for an inverted pendulum problem using vision feedback. Ochoa et al. [45] investigated the effect of performance tuning the absolute error value of the fuzzy system to control the position of a pendulum. They used seven membership
functions for the position and three membership functions for the speed. El-Bardini and Al-Nagar [46] studied a type-2 fuzzy logic controller for inverted pendulum and then compared it to a type-1 controller. The results showed that the performance of the type-2 controller was significantly improved upon the other one. Jang [47] examined temporal back propagation based self-learning fuzzy controllers. Lee and Takagi [48] built an automatic fuzzy system design method that used a genetic algorithm for the classical inverted pendulum problem. Su et al. [49] presented event-triggered fuzzy control design for the inverted pendulum system. They showed the advantages and effectiveness of the proposed design using three simulation experiments.

In this study, an ‘interpretable’ fuzzy controller for the inverted pendulum problem is proposed. The next section describes the inverted pendulum problem. Sections III and IV present how meaning is assigned to the fuzzy sets based on the concept of ‘computing with dictionary of words’. The results are provided in Section V. Finally, Section VI provides the concluding remarks.

II. THE INVERTED PENDULUM PROBLEM

The inverted pendulum is a typical nonlinear control problem that is of importance in a range of areas, such as robotics, rockets and industrial processes [49]. The schematic diagram of inverted pendulum system is shown in Fig. 1, which consists of a cart, straight line, driving unit and pendulum [4], [7], [41], [50]. The cart can move freely to the right or left. It is assumed that there is no friction in the system [41]. It is clear that the pendulum in the upright position has an unstable balance leading to a non-linear control problem for its stabilisation. The notation for the pendulum system is given in Table I.

![Fig. 1: The schematic diagram of inverted pendulum system on a cart.](image)

The closed-loop fuzzy control system for stabilising the inverted pendulum on a cart is also shown in Fig. 1.

The non-linear dynamic equations of the inverted pendulum system are based on the physical laws [49]. Eqs. 1 and 2 are the differential equations used to simulate system response to controllers for the inverted pendulum.

\[
\ddot{x} = \frac{p \times g \times \cos \Theta \times \sin \Theta + p \times l \times \dot{\Theta}^2 \times \sin \Theta + F}{M + p \times (1 + \sin \Theta^2)} 
\]

\[
\ddot{\Theta} = \frac{(p \times l \times \dot{\Theta}^2 \times \sin \Theta - F) \times \cos \Theta - g \times (M + 2p) \times \sin \Theta}{l \times (M + p \times (1 + \sin \Theta^2))} 
\]

III. THE DICTIONARY

In this paper we are interested in the interpretability of a typical engineering focused fuzzy system. To do this, we shall investigate whether fuzzy sets in this application can be replaced with fuzzy sets obtained from surveys in a general context. In doing this, we are bringing together two previously separate strands of research computing with words dictionaries [51]–[53] and fuzzy transfer learning [54]. In this section we consider the construction of a dictionary of words from interval data which can be used to replace the more synthetic sets in the inverted pendulum application.

In previous work [51], [52] a large scale web based survey was undertaken to obtain interval data about where people feel a set of word should placed on a somewhat arbitrary scale of [0, 10]. In this paper we shall use this data to construct a type-1 fuzzy set with a Gaussian membership function for each word in the survey data. The steps used to obtain the parameters for the membership functions were:

1) Clean the data using the preprocessing step given in [51].
2) Construct a histogram (we used a bin count of 201) for each word based on the cleaned interval data.
3) Calculate the mean (\(\mu\)) and standard deviation (\(\sigma\)) of this histogram.

These steps are depicted in Figure 3.

This process produced a dictionary of 32 words, each modelled by a fuzzy set with a Gaussian membership function. Table II lists these 32 words with the parameters and Figure 4 depicts these fuzzy sets.
Fig. 3: The Construction of a Fuzzy Set for a Word from Interval Data. (a) The Cleaned Intervals for the Word Medium. (b) The Histogram and Fitted Gaussian Membership Function for the Word Medium.

TABLE II: Parameters of the Fuzzy Set which make up the Dictionary.

<table>
<thead>
<tr>
<th>Word</th>
<th>c</th>
<th>σ</th>
<th>Word</th>
<th>c</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A bit</td>
<td>1.00</td>
<td>3.23</td>
<td>Modest amount</td>
<td>4.18</td>
<td>1.82</td>
</tr>
<tr>
<td>A lot</td>
<td>8.30</td>
<td>1.96</td>
<td>None to very little</td>
<td>0.08</td>
<td>2.19</td>
</tr>
<tr>
<td>A smidgen</td>
<td>0.15</td>
<td>2.06</td>
<td>Quite a bit</td>
<td>4.18</td>
<td>1.81</td>
</tr>
<tr>
<td>Considerable amount</td>
<td>7.23</td>
<td>1.98</td>
<td>Sizeable</td>
<td>6.00</td>
<td>1.96</td>
</tr>
<tr>
<td>Extreme amount</td>
<td>9.95</td>
<td>2.06</td>
<td>Small</td>
<td>1.70</td>
<td>1.80</td>
</tr>
<tr>
<td>Fair amount</td>
<td>5.50</td>
<td>1.63</td>
<td>Some</td>
<td>3.31</td>
<td>1.94</td>
</tr>
<tr>
<td>Good amount</td>
<td>6.47</td>
<td>1.74</td>
<td>Some to moderate</td>
<td>4.70</td>
<td>1.30</td>
</tr>
<tr>
<td>High amount</td>
<td>9.00</td>
<td>2.20</td>
<td>Somewhat small</td>
<td>1.78</td>
<td>1.83</td>
</tr>
<tr>
<td>Huge amount</td>
<td>9.85</td>
<td>2.24</td>
<td>Substantial amount</td>
<td>7.63</td>
<td>2.01</td>
</tr>
<tr>
<td>Humongous amount</td>
<td>9.95</td>
<td>2.39</td>
<td>Teeny-weeny</td>
<td>0.15</td>
<td>3.08</td>
</tr>
<tr>
<td>Large</td>
<td>8.35</td>
<td>1.50</td>
<td>Tiny</td>
<td>0.15</td>
<td>2.01</td>
</tr>
<tr>
<td>Little</td>
<td>0.85</td>
<td>2.31</td>
<td>Very high amount</td>
<td>9.90</td>
<td>1.74</td>
</tr>
<tr>
<td>Low amount</td>
<td>1.20</td>
<td>1.93</td>
<td>Very large</td>
<td>9.95</td>
<td>2.43</td>
</tr>
<tr>
<td>Maximum amount</td>
<td>9.99</td>
<td>0.91</td>
<td>Very little</td>
<td>0.15</td>
<td>1.80</td>
</tr>
<tr>
<td>Medium</td>
<td>5.03</td>
<td>1.10</td>
<td>Very sizeable</td>
<td>8.13</td>
<td>2.29</td>
</tr>
<tr>
<td>Moderate amount</td>
<td>5.13</td>
<td>1.38</td>
<td>Very small</td>
<td>0.20</td>
<td>1.66</td>
</tr>
</tbody>
</table>

IV. HOW THE FUZZY SETS WERE COMPARED

In order to identify which fuzzy sets from the dictionary best match the sets from the inverted pendulum application we first transfer the dictionary to each of the three domains in the inverted pendulum system. We then identify which sets best match the exist sets.

A. Transferring the Dictionary

Fuzzy transfer learning [54] suggests how fuzzy sets from one context can be transferred to another. In the simplest terms this mean translating and scaling the parameters of membership function from one domain to another. In this paper, we only use Gaussian membership functions. As such the transfer of our dictionary to any other domain becomes a problem of translating and scaling values of c and scaling values of σ. Translating the centre of our Gaussian sets from one domain to another is trivial. We calculate the distance between the centre of each domain and move all values of c by that distance. The domain of the dictionary is [0, 10] which has a centre of 5. The first input domain in the inverted pendulum application is [−40, 40] which has a centre of 0. As such we must translate all values of c in the dictionary fuzzy sets by −5. Next the values of c and σ must be scaled. To do this we calculate a scaling factor as the width of the new domain over the width of the old domain. Again looking as the first input domain of the inverted pendulum application ([−40, 40]) we see it has width of 80 whilst the dictionary ([0, 10]) has a width of 10. This gives 80/10 so we have a scaling factor of 8. Every value of c and sigma must now be multiplied by this scaling factor. Consider the example set from the dictionary medium with parameters c = 5.03 and σ = 1.10. In order to transfer these parameters to the domain [−40, 40] we first translate the value of c by −5 giving 0.03. We then scale both c and σ by 8 giving a values of c = 0.24 and σ = 8.80. Applying the same procedure to Maximum amount (c = 9.90, σ = 0.91 yields values of c = 39.20 and σ = 7.82. The calculation of the new values of c and sigma is given formally in equations 3 and 4.

\[ c' = c + \frac{(d^n_e - d^n_s) - (d^n_o - d^n_e)}{2} \times \frac{d^n_o - d^n_s}{d^n_e - d^n_s} \]  
\[ \sigma' = \sigma \times \frac{d^n_o - d^n_s}{d^n_e - d^n_s} \]

where c’ is the value of c after the transfer, d^n_s and d^n_e are the start and end of the new (destination) domain and d^n_o and d^n_e are the start and end of original (source) domain.

This process yields a collection fuzzy sets transferred from the dictionary with its somewhat arbitrary domain of [0, 10] to the three domains in the inert pendulum application. When considering the values obtained from this process, it becomes clear whilst the values of c have transferred across perfectly sensibly, there is something of an issue with the values of σ. The values of σ given in the transferred dictionaries are significantly higher than those in the inverted pendulum system. Whilst we leave a discussion of the reasons for this to Section VI we now propose an additional step in the transfer...
A similarity measure is given below:
comparing the Gaussian membership functions. The Jaccard
is a good measure of closeness in terms of centre and width when
selected. We use the Jaccard similarity measure \([55]\) as it gives
pendulum system with each set from one of our dictionaries.
we simply compute the similarity of each set in the inverted
\[\sigma\]
dictionaries, with and without scaled values of
pendulum application with a set from our two respective
sections.

B. Selecting Sets from the Dictionary

Our goal is to replace each set in the existing inverted
pendulum application with a set from our two respective
dictionaries, with and without scaled values of \(\sigma\). To do this,
we simply compute the similarity of each set in the inverted
pendulum system with each set from one of our dictionaries.
The set from the dictionary with the highest similarity is
selected. We use the Jaccard similarity measure \([55]\) as it gives
a good measure of closeness in terms of centre and width when
comparing the Gaussian membership functions. The Jaccard
similarity measure is given below:
\[
J_s(A, B) = \frac{p(A \cap B)}{p(A \cup B)} = \frac{\int_x \mu_{A \cap B}(x) dx}{\int_x \mu_{A \cup B}(x) dx}
\]  

(5)

where \(A\) and \(B\) and fuzzy sets and \(p(A \cap B)\) and \(p(A \cup B)\)
are the cardinalities of \(A \cap B\) and \(A \cup B\) respectively. Table
III gives a complete list of all membership functions used in
this paper.

V. RESULTS

In this study, a Java simulation is developed for the ex-
periments exploring the interpretability of a fuzzy system
for controlling an inverted pendulum on a cart as a case
study. Runge Kutta method is implemented for solving the
differential equations in Eqs. 1 and 2 to illustrate the system
response to the controller outputs. The simulation embeds
a Java library, referred to as Juzzy for modeling the fuzzy
controller making fuzzy logic calculations \([56]\). The inverted
pendulum system is designed based on the parameter settings
provided as follows: \(M = 1.0kg\), \(p = 0.1kg\), \(l = 1m\),
g = 9.8m/sec\(^2\), initial \(\Theta = -30^\circ\) and \(\dot{\Theta} = 4^\circ/sec\).

Four different fuzzy controllers are implemented to compare
the system response:
- \(T\) is our reference fuzzy system which uses triangular
  membership functions \([41]\).
- \(G\) uses Gaussian membership functions that are converted
  from \(T\).
- \(S\) uses Gaussian membership functions that are chosen
  from the dictionary using the standard fuzzy transfer
  method.
- \(SN\) similarly uses Gaussian membership functions that
  are chosen from the dictionary with the standard nor-
  malised fuzzy transfer method.

Runge Kutta based simulations are run for 12 seconds
with 0.1 seconds time interval with the initial conditions
given above. The system responses (\(\Theta\) and \(x(position)\)) and
controller output (\(F\)orce) are illustrated in Fig. 5(a), (b) and
c, respectively. \(T\) and \(G\) use symmetric fuzzy sets which is a
standard approach in almost all control problems. For example,
the fuzzy set NB has \(c = 25\) while PB has \(c = -25\), where
ZO has \(c = 0\) in \(G\). This phenomena leads to the behavior as
observed in Fig. 5(a) for \(\Theta\) and \(F\) settling at 0 for \(T\) and \(G\).
However, those values are nonzero for \(S\) and \(SN\). This is due
to the fact that human use of words does not necessarily create
symmetric fuzzy sets. The results show that normalisation is
indeed needed in order to capture the domain characteristics
as \(SN\) provides more sensible control of the pendulum as
compared to \(N\).

Root mean square error (RMSE) values (see Eq. 6) for the
controller output and system responses are provided in Table
IV.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^*)^2}
\]  

(6)

where \(n\) is the number of samples (taken at each time
interval), \(y_i\) is \(F\), \(\Theta\), or \(x\) for the \(i^{th}\) sample, and \(y_i^* = 0\).

Although they are both interpretable fuzzy sets, the per-
formance of \(SN\) is better than \(S\) in terms of RMSE for the
controller output and system responses of \(F\), \(\Theta\) and \(x\).
Moreover, \(SN\) employs the least force to control the inverted
pendulum generating an RMSE of 0.7601 when compared to
\(T\), \(G\) and \(S\).

VI. CONCLUSION

Interpretability is one of the main advantages of using a
fuzzy system. However, that advantage disappears in some
applications, for example, while solving control problems
using a fuzzy system, since synthetic fuzzy sets are used.
In this initial study, we presented an approach to make a
fuzzy system more interpretable by making the fuzzy sets
more 'meaningful' based on a dictionary and fuzzy transfer
learning. We applied our approach to the well-known problem
of inverted pendulum obtaining encouraging results. We will
investigate the performance of the proposed approach on other
real world problems and look into the trade-off between
interpretability and performance.

The results from this study do present some issues when
attempting to use fuzzy sets in the form of words elicited
from humans in engineering applications.

Engineering application often require (at least output fuzzy
sets) to have some exact central datum, in the case of the
inverted pendulum a value of zero. Fuzzy sets elicited from

TABLE III: Membership Functions of all Fuzzy Sets used in this Paper.

<table>
<thead>
<tr>
<th>Label</th>
<th>Original Parameters</th>
<th>Gaussian Version</th>
<th>Standard Dictionary</th>
<th>Normalised Dictionary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s</td>
<td>m</td>
<td>e</td>
<td>c</td>
</tr>
<tr>
<td>NVB</td>
<td>-40</td>
<td>-40</td>
<td>-25</td>
<td>-40</td>
</tr>
<tr>
<td>NB</td>
<td>-40</td>
<td>-25</td>
<td>-10</td>
<td>-25</td>
</tr>
<tr>
<td>N</td>
<td>-2</td>
<td>-10</td>
<td>0</td>
<td>-10</td>
</tr>
<tr>
<td>ZO</td>
<td>-5</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>PB</td>
<td>10</td>
<td>25</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>PVB</td>
<td>25</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

FIG. 5: Controller output: (a) Force, system response: (b) Θ and (c) position (x).

TABLE IV: RMSE of controller output and system responses.

<table>
<thead>
<tr>
<th>Controller</th>
<th>RMSE of F</th>
<th>RMSE of Θ</th>
<th>RMSE of x</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.9125</td>
<td>4.9020</td>
<td>17.2812</td>
</tr>
<tr>
<td>G</td>
<td>0.8442</td>
<td>4.8627</td>
<td>13.7589</td>
</tr>
<tr>
<td>S</td>
<td>4.9141</td>
<td>22.6699</td>
<td>132.9342</td>
</tr>
<tr>
<td>SN</td>
<td>0.7601</td>
<td>5.3423</td>
<td>4.0910</td>
</tr>
</tbody>
</table>

VII. ACKNOWLEDGEMENTS

We would like to acknowledge the work of Jerry Mendel and his students in the area of interval fuzzy sets which informed, and gave labels to the dictionaries used in this paper.

REFERENCES
