Infomagnetism and Sentence Generation

Jason L. Hutchens

I. Introduction

LEXICAL attraction was introduced by Doug Beeferman, Adam Berger and John Lafferty in a 1997 conference paper [1], and was picked up by Deniz Yuret in his 1998 thesis which focused on using lexical attraction to discover a link-grammar for English utterances [2]. Searching for `"lexical attraction"` on Google reveals links to these papers, and more.

Lexical attraction is, however, nothing more than the pointwise mutual information between two quantifiable entities [3]. The concept of mutual information has long been used as a measure of the correlation between the two random variables due to the fact that it reflects the reduction of uncertainty of one random variable due to the knowledge of the other.

II. Definitions

Definition 1: The *mutual information* $I(X;Y)$ between two random variables X and Y is defined as the relative entropy between the product of the marginal probability distributions and the joint probability distribution. That is,

$$I(X;Y) = \sum_{a \in X} \sum_{b \in Y} P(a,b) \log \frac{P(a,b)}{P(a)P(b)} \tag{1}$$

Definition 2: The *pointwise mutual information* $J(a;b)$ between two quantifiable entities a and b is defined as the change in information received when we observe one of the entities in the knowledge that the other has occurred. That is,

$$J(a;b) = \log \frac{P(a,b)}{P(a)P(b)} \tag{2}$$

Remark 1: Application of Baye's rule allows us to rewrite this in a way which emphasises that the pointwise mutual information measures a change in information. Note that the measure is unbounded and can take on both positive and negative values. Note that a and b are interchangeable here as $J(a;b)$ is a symmetrical measure. That is,

$$J(a;b) = \log P(a|b) - \log P(a) \tag{3}$$

Remark 2: We choose to use the term *infomagnetism* as an equivalent terminology for the pointwise mutual information, chosen both for its compactness and for its suggestion that the pointwise mutual information between two entities has the properties of a force.

Remark 3: The mutual information between random variables may also be rewritten as the expected value of the infomagnetism, calculated as a weighted average of the infomagnetism between all possible combination of values of those random variables. This makes it obvious that the infomagnetism is a measure of information (or surprise) while the mutual information is a measure of entropy (or uncertainty), as entropy is always the expected value of information. That is,

$$I(X;Y) = \sum_{a \in X} \sum_{b \in Y} P(a,b) J(a;b) \tag{4}$$

III. Example

Let us now think about infomagnetism intuitively. Imagine that we observe two events, one following the other. Examples of such events might be pressing a button and observing a light turning on a moment later. If the light turns on reliably following a button press, and doesn't turn on at any other time, then the two events are strongly correlated, and this will be reflected in a large positive value of the infomagnetism. We might say that a strong attractive force exists between the two events.

If, however, pressing the button has no noticeable effect on the behaviour of the light, then the two events are independent, and the infomagnetism will be close to zero. We could say that no force exists between the two events.

In the third scenario, pressing the button may cause the light to turn on less frequently. In such a situation the infomagnetism between the two events will be negative, indicative of a repulsive force between them.

IV. Problems

In order to calculate the infomagnetism between two events we need to estimate two probabilities. Probability estimation is usually performed by observation, and sparsity of the data is a very real problem (basically, we may not observe sufficient occurrences of the events to warrant an accurate probability estimate). However, all data-driven approaches suffer from this drawback, and instead of throwing the baby out with the bathwater we should concentrate our efforts on developing techniques which lessen the effect of this problem. Probably the best way of doing this is by averaging out a large number of observations. This is the method we use to generate language-like utterances using the infomagnetism model.

V. Language

We may quantify the words in a sentence as atomic information units and calculate the infomagnetic forces acting between them. Doing so allows us to take long-distance dependencies into account, resulting in more realistic generated sentences than possible with traditional Markov models.

Our experimental sentence generator generated a sentence from a template (always delimited by the special ”sentence boundary” symbol) containing zero or more seed words in the order that they should appear in the final sentence. Generation progressed by postulating a new word in all possible positions in the template and selecting the word and position which results in the greatest increase in the average infomagnetism over the entire sentence (calculated by summing the individual infomagnetisms and normalising with respect to the sentence length). This procedure is illustrated in figure 1.

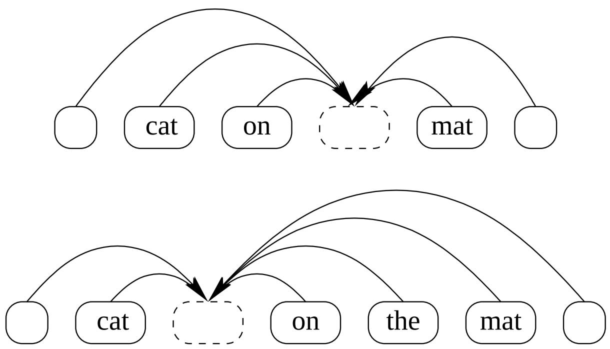

Fig. 1. One step of the generation process in which a proposed word and position are inserted into the template, resulting in a new source of infomagnetism.

To test the infomagnetic language model we inferred frequency statistics from a 174 sentence portion of an online encyclopaedia in order to estimate the probabilities necessary to the calculation of the infomagnetic forces. A portion of the training data used is shown below, with the words which were used as seed words in subsequent generation tests highlighted in boldface.

- Botany is the science of **plant** life.
- An omnivore is an animal that eats both **plant** and animal matter.
- A **plant** is a living organism which does not have the ability to move, and does not have sensory organs or digestive organs.
- A **plant** is a living organism of the vegetable kingdom.
- An artery is a vessel that conveys **blood** from the heart.
- The heart is the muscle in vertebrates which pumps **blood** around the body.
- The kidney is a gland used for filtering urine from the **blood**.
- A vein carries venous **blood** from the body to the heart.
- The Achilles Tendon is the tendon which connects the heel with the calf of the leg, and is the principal extensor of the **foot**.
- The ankle is the joint connecting the **foot** with the leg

We then performed various generations using all possible combinations of the three words plant, blood and foot as seed words in the template. As can be seen from the results below, although none of these three words ever appeared within the same training sentence, the model did a good job of generating sentences which contained all three of them. This performance encourages us that infomagnetic technology should be incorporated into our learning system.

- A **plant** is the **blood** around the **foot**
- A **plant** is a word formed from the **foot** with the **blood**
- The kidney is the **blood** around the science of **plant** is the **foot**
- The dinosaurs were a vessel that conveys **blood** from the **foot** with the organ of **plant** life
- The ankle is the joint connecting the **foot** with the muscle in vertebrates which pumps **blood** around the science of **plant** and animal
- The eye is the **foot** with the fleshy edges of **plant** is the **blood**

VI. Future

We would like to continue our experiments with the infomagnetic language model. In particular, we would like to include this model in HALthree, as we have high expectations of its performance and it seems to satisfy many of the requirements currently being formulated by Walter. We would also like to perform some experiments with this model, exploring whether it is reasonable to include nonlexical units as atoms which may also exert forces. For example, we have thought about introducing atoms which represent "concepts" in order to allow utterances to have a wider contextual effect than they do at present. Finally, I would personally like to make a MegaHAL-style chatterbot using this technology, potentially making it available for download from the Ai website, and possible making an online version available as well.

References

[1] Doug Beeferman, Adam Berger, and John Lafferty, "A model of lexical attraction and repulsion," in *Proceedings of the ACL-EACL '97 Joint Conference*, 1997.
[2] Deniz Yuret, *Discovery of Linguistic Relations Using Lexical Attraction*, Ph.D. thesis, MIT, 1998.
[3] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.