Understanding Happiness by Using a Crowd-sourced Database with Natural Language Processing

Yixia Wang
STUDENT NUMBER: 2027916

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

Giacomo Spigler
Sharon Ong

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
June 2020

Acknowledgement

First, I would like to express my gratitude to Dr. Giacomo Spigler for giving me the chance to write my thesis in the field of Natural Language Processing. Without his persistent help and supervision, the goal of this project would not have been realized.

In addition, I would like to thank my mom, who has been supporting me for whatever decision I have made, including coming to Tilburg University to further my study.

My special thanks also go to my friends Jinfan, Shuo, Fei, Xiangyu, Sophie, Qian, Ruonan, Chamanika and Angelo, who have offered their generous and unconditional help whenever I needed during the whole process.

This journey at Tilburg University has been bittersweet, but from all my professors, Emmanuel Keuleers, Sharon Ong and Maryam Alimardani, I have learned how important it is to be a man of integrity.

Understanding Happiness by Using a Crowd-sourced Database with Natural Language Processing

Yixia Wang

We conduct three studies by utilizing a crowd-sourced database called Happy DB Database, which consists of more than 100,000 descriptions of happy moments written by workers in Amazon's Machine Turk. Firstly, we apply a state-of-art word embedding algorithm BERT to transform all happy moments to context-sensitive representations to learn two critical concepts of happiness, agency and sociality. Our performance is better than that of the existing publications. Next, We study the association between alcohol consumption and happiness and our result suggests that both alcohol consumption and abstaining from drinking can lead to happiness. However, no association is found between gender and different drinking patterns in terms of breeding happiness. We also delve into happiness that results from interpersonal relationships and a significant association is found between gender and interpersonal relationships. Interestingly, men are more likely to get happy from the moments related to interpersonal connections, compared to their female counterparts.

1. Introduction

1.1 Motivation

Since the outbreak of COVID-19 pandemic, ways of human communication have been changed substantially. Instead of conventional face-to-face interaction, the new norms of contacts are largely constrained to texts, voicemails and video calls via the Internet, with textual messages taking up major proportions. Thus, huge amounts of textual data are being generated at an unprecedented pace.

The field of deciphering human natural language data is natural language processing(NLP). Tasks in this turf such as machine translation, speech recognition, and product recommendation, have vastly improved over the last few years. At the core of these new language processing technologies are language models which transform massive amounts of written or spoken human language into multi-dimensional vector representations of words or sequences, which are then used as input representations in complex artificial intelligence tasks. These models have fused research in many areas, including positive psychology.

Positive psychology as a stream of psychology studies positive human functioning and aspects that contribute the most to a well-lived and fulfilling life, both individually and at a social level(Seligman and Csikszentmihalyi 2014). Positive psychology deals with three issues: positive emotions, positive individual traits and positive institutions(Seligman and Csikszentmihalyi 2014). This thesis lays emphasis on positive emotions, precisely, happiness and factors that evolve around this emotion. Note that the terms 'happiness' and 'well-being' are used interchangeably(Delle Fave et al. 2011). Seligman and others(Seligman 2004; Peterson, Seligman et al. 2004) presented the directions to happiness framework, which comprise three different pathways to happiness; pleasure, engagement and meaning. Empirical research explicates that people experiencing high levels of all these three orientations have the highest life satisfaction compared with those who have low levels in these three aspects, but that meaning and engagement are the critical predictors to happiness in comparison to pleasure(Peterson, Park, and Seligman 2005; Vella-Brodrick, Park, and Peterson 2009). Other theories related to happiness cover a broad range of topics including 'the biological, personal, relational, cultural and a global dimensions of life'(Delle Fave et al. 2011). For example, Seligman and Peterson(2014) claimed that relationships are essential in breeding positive emotions, whether they are work-related, familial, romantic, or platonic. As Christopher Peterson(2004) puts it simply, "Other people matter." It is typical that most positive things always take place in the presence of other people. However, due to previous constraints on samples and relative monotonousness of demographic characteristics of participants, further research is yet to benefit from the recently emerged data collection method of crowd-sourcing. This thesis bridges the gap by utilizing a crowd-sourced dataset initiated in Amazon's Mechanical Turk and called Happy DB Database(Asai et al. 2018) where each participant gives accounts of their happy moments in the past 24 hours or 3 months. This dataset contains almost 100,000 descriptions of happy moments and by deciphering them, more useful insights could be added to the filed of positive psychology.

1.2 Research Questions and Objectives

The thesis consists of three studies that explore various aspects of happiness with different research questions and objective raised.

Study 1 focuses on two concepts of happiness, sociality and agency. Sociality in this context means feeling happy in the presence of others or alone, while 'agency' denotes the subject of the registered happy moment being participants themselves or others. These two concepts have been explored in a CL-Aff shared competition, where 'sociality' and 'agency' were labeled as an extension to the original Happy DB database and two corresponding classification tasks were presented(Jaidka et al. 2019). Participating teams explored diverse methods. However, when transforming sentences into vector representations, none has used a state-of-art word embedding algorithm called BERT which stands for bidirectional encoder representations from transformers (Devlin et al. 2018). BERT is a real game changer in the word embedding algorithms community. Therefore, this study uses BERT as an important weapon to complete the task and then see if the performance is better than that of the existing publications. Research questions are framed as follows:
- Which state-of-art NLP approaches could be used to classify the above-mentioned factors of sociality and agency?
- How do such approaches perform compared with baseline models and existing methods in publications?

Study 2 explores association between alcohol and happiness. Four research questions are stated:
- Does alcohol bring happiness?

- Does reducing alcohol consumption bring happiness?
- Does drinking in the presence of others bring happiness more often than drinking alone?
- Is there any gender difference when it comes to the different drinking behaviors?

Study 3 inquires into gender difference in terms of happiness from interpersonal relationships. We are interested in the research question:
- Is there any gender difference when it comes to happiness from interpersonal relationships?

1.3 Structure of the Thesis

This thesis delves into exploring different aspects of happiness by utilizing natural language processing toolkits. It is split into seven main chapters. This **introduction(Chapter 1)** presents three studies that revolve around happiness and gives a summary on the structure of the whole thesis. It is followed by **literature review(Chapter 2)**. Theories of natural language processing and various word embedding algorithms are discussed in this chapter. In addition, research on how sociality and agency are suggested in the dataset is introduced. Also, empirical studies on association between happiness and concepts like alcohol and interpersonal relationships are presented. However, as these existing works measure happiness mostly by subjective ratings of participants in questionnaires, we think subjective ratings lose some credits where brief accounts of happiness used in this thesis come in. **Chapters 3, 4 and 5 articulate on methods and results of three separate studies revolving around the core topic of happiness.** In **Chapter 3**, we conduct **two classification tasks in order to learn the concepts of 'sociality' and 'agency'** in happiness. We explore LSTM accompanied with the state-of-art word embedding algorithm BERT, which is also a part of main contributions of this thesis. We also employ traditional machine learning algorithms such as Random Forest and Logistic Regression as baseline models. **In Chapter 4**, **association between happiness and alcohol** is examined. In **Chapter 5, gender difference in how interpersonal relationships inspire happiness** is burrowed. This is followed by **Chapter 6** which contains an **overall discussion**. Limitations and potential research questions are also discussed. **Chapter 7** contains **conclusion**.

In addition, we also run another two tasks related to this dataset. In the first task, we dig into the association between age and source of happiness. However, since we apply identical methods as we do in Study 2 and Study 3, so we see no need to channel a single chapter for it. As a result, the whole case is appended in the supplementary material Appendix 7. The second additional task done to get a better picture of the dataset is entity extraction. We experiment with the total dataset to answer a question: Which places make people happy the most? We use an entity extractor called KOKO(Wang et al. 2018) for this task and results are added in Appendix 7.

2. Related Work

Three subsections construct the whole part of literature review. Research questions in the first study are related to natural language processing(NLP), which is introduced in Section2.1. At the core of NLP lie the language models, so well-known word representation methods, like bag of words, bag of words with TF-IDF transformation, GloVe and BERT, are also presented in this part. Happiness is the core issue we want to delve into, so the Section2.2 presents all previous works that have contributed most to this topic. Since the dataset we use was crowdsourced online, so we also give a rundown on crowdsourcing, highlighting its data reliability and quality in Section2.3.

2.1 Natural Language Processing

Natural language processing(NLP) is the technology that lies in the subset of artificial intelligence(AI), computer science, and linguistics(Collobert et al. 2011). It uses algorithms to digest what human beings say and generates appropriate outputs according to different tasks given. In a word, natural language processing covers all communications between a computer and a human by the use of written or spoken languages. NLP can be broken down into several AI tasks.

First of all, a machine powered by NLP has to disambiguate what a human says and her/his intent. This process is called Natural Language Understanding, which can be done by utilizing a set of technologies as follows:
- Syntactic analysis. Syntax is the rules that govern the structure of a sentence in any given language. Syntactic analysis refers to the process of parsing a string into its constituents which results in a parse tree that shows syntactic relation between all parts(Rindflesch and Fiszman 2003).
- Entity extraction. This tool searches for entities like a place, person, organization or event and determines how important these entities are for understanding human languages(Collobert et al. 2011).
- Semantic analysis. Semantic analysis of a corpus is related to the structures and occurrences of a language, ranging from the level of words to the level of paragraphs(Hofmann 2013). It understands the idea of what is written in a particular text and also judges whether the two syntactic structures derive similar meanings or not.
- Sentiment analysis. This technology identifies how humans feel in terms of their mood, emotion and opinions(Hofmann 2013).

Now that the machine understands both the meaning and the intent, NLP has many applications afterward.
- Machine translation. Text or speech from one language are converted to another by utilizing neural networks(Bahdanau, Cho, and Bengio 2014).
- Text generation. Using this technology, structured data is transformed back into natural language(McKeown 1992). Practical applications of this include automated customer reports or chatbots which can serve as customer service consultants.
- Document summarization. This tool identifies essential information of a text(Nenkova and McKeown 2012).

Apart from what has been mentioned above, NLP also contains other well-known tasks such as sentence classification and speech recognition as well since NLP is not limited to text.

The remaining part of this section focuses on how NLP technologies transform large amounts of unstructured human natural languages into vector representations that are machine-readable.

2.1.1 Word Representations. In order for a machine to do NLP downstream tasks, we need to first translate human natural languages into machine-readable digital forms that enable effective incorporation of knowledge carriers of words. These forms are real-valued vector representations of words where both syntactic and semantic meanings obtained from large amounts of unlabeled corpus are embedded at varying levels according to different translating methods. They are very powerful tools extensively used in NLP tasks, including information retrieval(Manning, Raghavan, and Schütze 2008), question answering(Kumar et al. 2016), semantic analysis(Hofmann 2013), machine translation(Bahdanau, Cho, and Bengio 2014) and dependency parsing(Kübler, McDonald, and Nivre 2009).

Research on word representations methods has been fruitful. In this thesis, we mainly discuss three models and elaborate how they are connected to one another to fuel those NLP downstream tasks.

The first one is count-based models with no context information embedded, where words in texts are mapped into a count matrix. We take Bag of Words(Wallach 2006) and Bag of words with TF-IDF transformation(Ramos et al. 2003) as examples in this thesis. These models, though preserve partial information of original texts, still face scalability challenges. To address this problem, count-based models that are trained on unsupervised learning tasks like GloVe(Ramos et al. 2003) are presented with dimensionality reduction as a step to reduce sparsity. Also, these models shift perspective from mere counts of words to architecture construction based on optimization of a certain objective(Bengio 2009). Note that when these vector representations arise from representations on hidden layers of multi-layer neural network models, they are called word embeddings. In order to compare their performances with the models mentioned above, tasks have been done: Baroni et al.2014 compared the models using synonym detection, which was similar to a task done by (Landauer and Dumais 1997), concept categorization(Almuhareb 2006; Baroni, Murphy, Barbu, and Poesio 2010; Baroni, Evert, and Lenci 2008). Pawel et al.(2017) also compared their performances on a large dataset of semantic priming, word associative norms, and so on.

However, despite huge progress in task performance, models like GloVe, according to Pawel, Emmanuel and Marc(2017), still lack psychological plausibility because they fail to mimic the human cognitive system in that they take in data in a conditional but not incremental manner. Another drawback is the conflation deficiency of word vectors, which was first identified by the works of Schütze(1998)and an example is shown in Figure 1. Conflating all meanings of a word in a single representation means that the models are probably unable to mine semantic richness of homonyms, namely, words with multiple distinct definitions. Research has been done to show that word embedding models like GloVe and word2vec fail to efficiently encapsulate different meanings of a word, even when these meanings are provided in training corpus(Yaghoobzadeh and Schütze 2016).

In order to tackle the dual problems, state-of-art contextualized word embedding algorithms are designed, including BERT(Devlin et al. 2018),ELMo(Peters et al. 2018) and GPT-2(Radford et al. 2019). They imitate human cognitive systems of taking in information in continuous manner. More semantic richness is also mined. In the section2.1.5

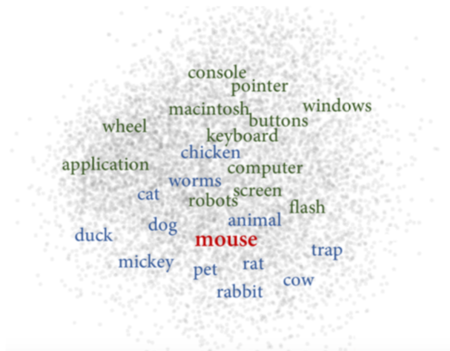


Figure 1: An illustration of meaning conflation deficiency from the work of Jose Camacho-Collados(Camacho-Collados, Pilehvar, and Navigli 2016). In this 2D visualization by t-SNE, as mouse has two meanings, the word rat is close to computer and screen with which it does not share any semantic interrelation. Source:Adapted from Camacho-Collados, J., Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research, 63, 743-788.

we articulate on BERT which is pre-trained on two unsupervised tasks: sentence reconstruction and next sentence prediction.

2.1.2 Bag of Words. Bag of words(BOW) is a simple but popular text representation approach(Wallach 2006). As the term manifests itself, a text is represented as a bag of numbers, disregarding its word order and grammatical rules. These numbers could be characterized in many ways, but the most well-known one is frequency, namely, the quantity of times a term appears in the text.

We give one example to show the core idea behind BOW. We have two simple text documents. Sentence 1) Mike eats an apple. Sentence 2)Lily drinks a cup of apple juice. We first build a vocabulary from all unique words in these two sentences. The vocabulary consists of 10 unique words: 'Mike', 'eats', 'an', 'apple', 'Lily', 'drinks', 'a', 'cup', 'of', 'juice'. Then, we quantize the occurrences of these words in each sentencein Table 1.

	Mike	eats	an	apple	Lily	drinks	a	cup	of	juice	Length of the sentence
Sentence 1	1	1	1	1	0	0	0	0	0	0	4
Sentence 2	0	0	0	1	1	1	1	1	1	1	7

Table 1: The occurrences of words in each sentence.

The objective is to transform each document of unstructured text into a vector that can be used as input for NLP downstream tasks like text classification. If we take only numbers from the table, two vectors are formed. Sentence 1: [1,1,1,1,0,0,0,0,0,0]. Sentence 2: [0,0,0,1,1,1,1,1,1,1]. These two vectors are the new representations of said sentences. In this way, we have a consistent way of extracting features from any document, ready for use in modeling.

Although simplicity and consistency characterize this method, there are salient downsides to it. Since modeling sentences using BOW ignores word order, there would be a mismatch of meaning and even loss of critical information(Landauer et al. 1997). For example, 'This is bad' and 'Is this bad' would end up having an identical representation. Another factor is the dimensionality issue(Landauer et al. 1997). It can be seen clearly that the length of the document vector is equal to the number of known words in all documents. If we have two documents that in total contain 3000 words, then each representation would have 3000 dimensions. If the list goes to millions of words, then the result is even more catastrophic. Besides, if there are many uncommon words in a document, then it definitely leads to a vector full of zeros, which is called a sparse vector(Landauer et al. 1997).

2.1.3 Bag of Words with TF-IDF transformation. In a document, domain-specific words are regarded as delivering more information than common words such as 'the'. A problem with scoring word frequency is that highly frequent words dominate in vectors with larger scores while in fact, they are not as 'informative' as those domain-specific words which appear far less. In order to give more weights to important words, one approach is to rescore each word by how often it appears in all documents so that the scores for common words that are probably frequent throughout all documents, like 'a', are penalized. This method is called Term Frequency-Inverse Document Frequency(TF-IDF)(Ramos et al. 2003) where Term Frequency represents a scoring of the frequency of a word in the current document and Inverse Document Frequency means a scoring of how rare the word is across all documents. TF measures how frequently a term, t, appears in a document, d.

$$tf_{t,d} = \frac{n_{t,d}}{Number\ of\ terms\ in\ the\ document} \tag{1}$$

Where n denotes the number of times the term 't' appears in the document 'd.' Hence, each pair of term and document has its TF score. Computing TF alone is not enough for determining the importance of a term. IDF comes in and measures how rare a term is across all documents. TF measures how frequently a term, t, appears in a document, d.

$$idf(t, D) = \log \frac{|D|}{1 + |\{d \in D : t \in d\}|} \tag{2}$$

Where the numerator D denotes our document space, so |D| is the size of the space. The denominator refers to the total number of times in which term t appears in all document. Note that no matter how many times a term appears in a document, it will still be counted as 1. In addition, the plus 1 is used for the purpose of avoiding zero division(Ramos et al. 2003).

Then by combining TF and IDF we can calculate TF-IDF score for each word in the whole corpus. By this way, words that are distinct and potentially contains more useful information in a given document are highlighted with larger scores.

$$(tf\_idf)_{t,d} = tf_{t,d} * idf_t \tag{3}$$

We extends the previous example to illustrate the key point behind BOW with TF-IDF transformation. According to 1 and 2, the TF values and the IDF values for all terms in Sentence 1 and Sentence 2 are calculated and shown in Table 2. Then we calculate TF-IDF values for all terms in the documents based on 3.

Term	Sentence 1 (counts)	Sentence 2 (counts)	TF Sentence 1	TF Sentence 2	IDF	TF-IDF Sentence 1	TF-IDF Sentence 2
Mike	1	0	1/4	0	0.69	0.17	0
eats	1	0	1/4	0	0.69	0.17	0
an	1	0	1/4	0	0.69	0.17	0
apple	1	1	1/4	1/7	0.00	0	0
Lily	0	1	0	1/7	0.69	0	0.10
drinks	0	1	0	1/7	0.69	0	0.10
a	0	1	0	1/7	0.69	0	0.10
cup	0	1	0	1/7	0.69	0	0.10
of	0	1	0	1/7	0.69	0	0.10
juice	0	1	0	1/7	0.69	0	0.10

Table 2: TF, IDF values and TF-IDF values for all terms in the documents.

We have now obtained TF-IDF scores for our documents and new BOW representations with TF-IDF transformation which are more informative. Sentence 1:[0.17,0.17,0.17,0,0,0,0,0,0,0]. Sentence2: [0,0,0,0,0.10,0.10,0.10,0.10,0.10,0.10]. However, as Bag of Words with TF-IDF transformation rescales representations produced by using Bag of Words, it still retains problems like sparse representation. Besides, it still could not capture any semantic information in that it ignores orders of words as well. That is where the following models which are based on optimization of a certain task come in.

2.1.4 Global Vectors of Word Representations(GloVe). GloVe(Pennington, Socher, and Manning 2014) which stands for Global Vectors is a prominent word embedding architecture. It leverages both local context information of words and global statistics(word co-occurrence) to obtain word vectors. Interestingly, word2ve(Mikolov et al. 2013) is popular for using local statistics. And the idea of utilizing global statistics to gain semantic relationships goes back to Latent Semantic Analysis(Landauer and Dumais 1997). The key idea underlying this method is a simple observation that it is potential to derive the meaning of a word from its co-occurrence possibilities with other words(Pennington, Socher, and Manning 2014). In the example provided by Pennington(2014) in Table 3, given two words from a 6-million word corpus, e.g. ice and steam, if a probe word k is relevant to ice but different from steam, such as the case when k is 'solid', then P(k | ice)/P(k | steam) is larger than 1, very high. If the probe word k is relevant to steam but different from ice, such as the case when k is 'gas', then P(k | ice)/P(k | steam) is less than 1, very small. When k is relevant or irrelevant to both words, then P(k | ice)/P(k | steam) is close to 1.

GloVe's ability of capturing semantic information is revealed in finding nearest neighbors. Euclidean distance(Danielsson 1980) and cosine similarity(Nguyen and Bai 2010) measure how closely a word is relevant to other one, thus providing an effective clue to the semantic similarity of the corresponding words. Through the metric of Euclidean distance, some of the seven nearest neighbors of the word 'frog' even went beyond

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

Table 3: The behavior of probability and ratio for various words. Source: Adapted from the original paper of GloVe. Pennington, J., Socher, R., Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

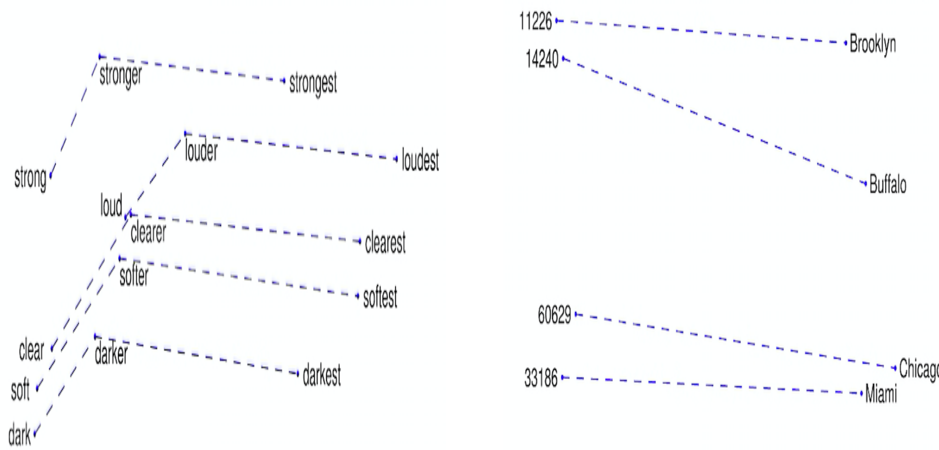
(a) litoria (b) leptodactylidae (c) eleutherodactylus

Figure 2: Rare words that go beyond average people's vocabulary but are revealed by GloVe. Source: Adapted from https://www.pinterest.com/pin/459719074447127550/.

average human's knowledge of taxonomy(Pennington, Socher, and Manning 2014). The seven nearest words are(see Figure 2): frogs, toad, litoria, leptodactylidae, rana, lizard, and eleutherodactylus.

Euclidean distance and cosine similarity produce a single number to quantify how a word is related to other one. This simplicity may cause problems because two word vectors always unveil more complicated relationships than can be captured by a scalar(Pennington, Socher, and Manning 2014). For instance, although words 'man' and 'woman' are considered similar in that they describe human beings, they are often regarded as being opposite since they represent two completely different gender types. In order to disclose more relationships between words, like the difference between 'woman' and 'man', an intuitive approach is capturing vector difference. GloVe is designed in such a way that vector differences encapsulate as much as possible the meaning identified by the juxtaposition of the said words(Pennington, Socher, and Manning 2014). This is shown in the visualization of linear substructures in Figure 3. The underlying concept that tells slow apart from slower and slowest may be equivalently identified by other word pairs, such as strong, stronger and strongest or short, shorter and shortest. 2D visualization of these words demonstrates the relatedness of all said words. The property is also observable through a set of other visualizations.

Despite its prominence, GloVe still has two major limitations. It disregards the fact that there exist homonyms, words with multiple meanings. For example, orange would have the same representation in 'The color is orange' and 'This is an orange'.Conflating



(a) original-comparative-superlative (b) zipcode-city

Figure 3: Visualizations of vector differences of word pairs that go beyond average people's vocabulary but are revealed by GloVe. Source: Adapted from Pennington, J., Socher, R., Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

all meanings into a single vector may have adverse impact on semantic modeling, resulting in loss of critical information(Schütze 2016). According to Pawel, Emmanuel and Marc(2017), GloVe also lacks psychological plausibility because it fails to mimic human cognitive system in that it takes in data in a conditional but not incremental manner. And so, contextualized word embeddings are born.

2.1.5 Bidirectional Encoder Representations from Transformers(BERT). Replacing static representations with context-driven dynamic word embeddings has resulted in a significant improvement in almost every NLP downstream tasks. Research(Ethayarajh, Duvenaud, and Hirst 2019) shows that on average less than 5% of the variance in a word's dynamic representation produced by contextualized word embedding methods can be explained by a static embedding. This suggests that contextualized word embeddings do not assign a simple vector to represent a word sense, because otherwise, the proportion of variance would be much higher. The principal reason behind their rich representation is that in all layers of these context-oriented algorithms, the representations of all words are anisotropic, which means they would not be distributed throughout but instead occupy a narrow cone(Ethayarajh, Duvenaud, and Hirst 2019). Among all contextualized models we focus on BERT in this thesis. BERT stands for Bi-directional Encoded Representations from Transformers(Vaswani et al. 2017). As the name suggests, it is a Transformers-based model, unlike GPT-2(Radford et al. 2019) whose architecture is based on LSTM. In a sequence-sequence task, the longer computational path is likely to introduce more error as every step of computation in deep neural networks involves an approximation that leads primarily to error(Tripathi et al. 2019).That is where self-attention works well. Transformers, developed by Google, are an encoder-decoder architecture model that employs self-attention mechanisms(Vaswani et al. 2017). Layers of self-attention wrap a more complete picture

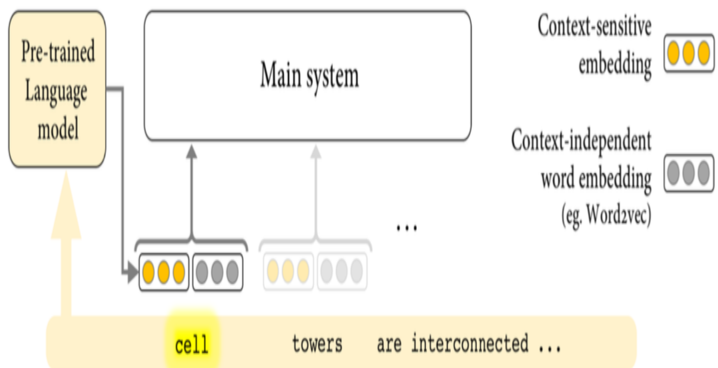


Figure 4: A general illustration of context-based word embeddings and how they are incorporated in NLP models. Unlike above-mentioned methods (i.e. BOW and GloVe) which produce static representations, context-based approaches generate dynamic word embeddings since there is a pre-trained language model component that analyzes the context of the target word(cell in the figure). Source: Retrieved from Camacho-Collados, J., Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research, 63, 743-788.

of the whole sequence and forward it to the decoder all at once. The encoder is bi-directional since self-attention can attend to tokens on both the left and right. However, the decoder is uni-directional, because the decoder is not allowed to attend to the right of the current token. Therefore, the decoder constrains the attention mechanism by masking the tokens to the right, and thus the whole Transformers are considered 'uni-directional', which inherently limits context inferring(Vaswani et al. 2017).

BERT improves on Transformers in a simple way. It only uses encoders, discarding the decoders. And the whole BERT is trained on a dual-task: Masked Language Modeling(MLM) and Next Sentence Prediction(NSP)(Devlin et al. 2018). The goal of the task is to minimize the combined loss of these two strategies.

In the MLM, when pre-training the model by using a large corpus of sentences, researchers(Devlin et al. 2018) masked 15% of the words and their model was trained to predict these words by digesting known context information(see Figure5). As the model was trained to predict the hidden words, it learned to produce a powerful intricate representations of the words.

In NSP, pairs of sentences were fed into the model, which learned to predict whether the second sentence is the ensuing sentence in the original pairs(Devlin et al. 2018). In the architecture, three steps were taken. Firstly, at the beginning of the first sentence, a [CLS] token was added and a [SEP] token was added at the end of each sentence. In this way, the model knew where each sentence began and ended. Secondly, a sentence embedding that indicated the order in the pair was appended(see Figure6).

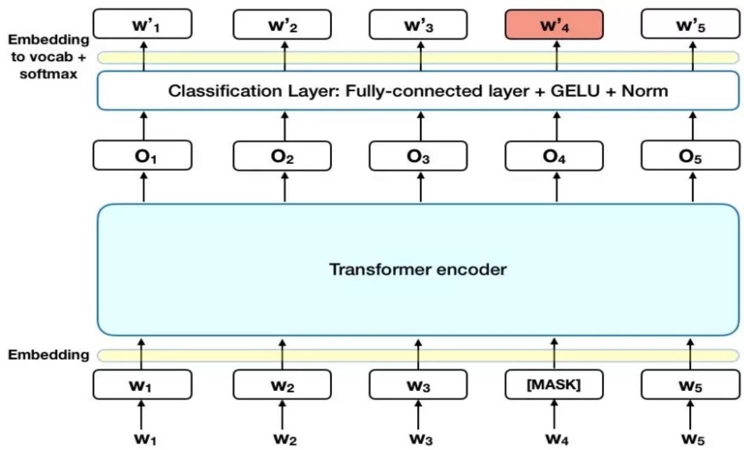


Figure 5: An illustration of BERT MLM architecture(Devlin et al. 2018). The fourth word in the input is masked. The words go through an embedding layer and an encoder. There is a classification layer on top of the encoder. Then the output vectors from the classification layer multiply the embedding matrix to form a vocabulary dimension. Lastly, the probability of each word in the vocabulary is calculated by applying activation function of softmax(Bridle 1990). Source: Adapted from Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Lastly, an extra embedding was inserted to indicate the position of each token in the sequence(Devlin et al. 2018). To predict if the subsequent sentence was in the original pair, the entire input went through the model. By applying a classification layer, the output of the [CLS] token was transformed into a 2*1 vector. Then the probability of IsNextSequence was calculated through the activation function of softmax.

In addition to BERT, the context-based word embedding community harbors other outstanding candidates like ULMFiT(Howard and Ruder 2018), ELMo(Peters et al. 2018), and GPT-2(Radford et al. 2019). All these technologies propel the whole field of NLP with an unprecedented engine power.

2.2 Happiness

Happiness has been conceptualized in many ways. Existing theories related to happiness cover a broad range of topics including 'the biological, personal, relational, cultural and global dimensions of life'(Delle Fave et al. 2011). Research depicted psychological well-being as a cognitive state including positive self-awareness and self-evaluation(Ingram and Wisnicki 1988; Epstein and Connors 1992; Kitayama, Markus, and Matsumoto 1995; Martin and Rubin 1995; Veenhoven 2010). Some other researchers characterized happiness as a general emotional state which is not constrained by any specific events(Diener and Lucas 2000; Kitayama, Markus, and Kurokawa 2000). Seligman and others (Seligman 2004; Peterson, Seligman et al. 2004) presented the directions to happiness framework, which comprise three different pathways to happiness; pleasure, engagement and meaning. Others emphasized that happiness is associated

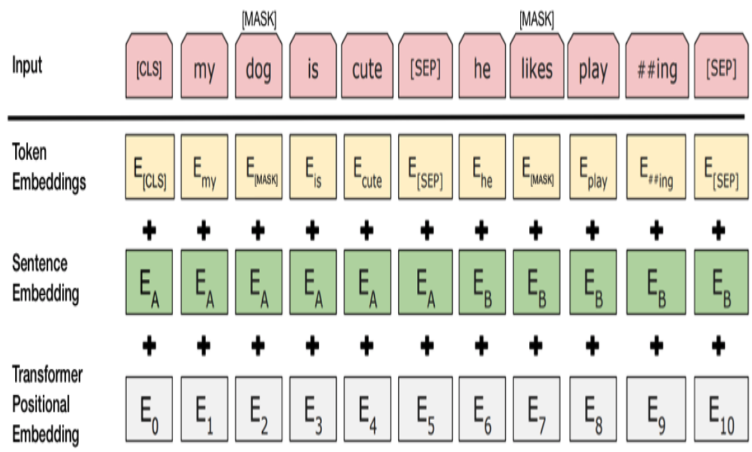


Figure 6: An illustration of three embeddings before being fed into the BERT NSP architecture. Source: Adapted from Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

with a healthy physical state(Goldberg and Hillier 1979). Also, quality and meaningful relationships with others are regarded as a critical access to happiness(Hills and Argyle 2001; Roothman, Kirsten, and Wissing 2003; Ryff, Singer, and Dienberg Love 2004; Grant, Christianson, and Price 2007). From these miscellaneous definitions of happiness, it can be seen clearly that happiness evolves around personal development in all variables ranging from physical health to cognitive flexibility. It also exists when relationships with others achieved a harmonious state.

2.2.1 Happiness and Alcohol. Low happiness is revealed in alcohol abuse amidst college students(Diulio et al. 2014). Though Murphy et al.(2005) claimed that there is no clear association between alcohol consumption and wellbeing indexes, a U-shaped relationship is discovered(Levy and Sheflin 1983). Precisely, they reported a negative linear relationship between drinking behaviors and happiness. However, in their work, heavy drinkers even record higher happiness than mild drinkers. Later, disputing this finding, Ventegodt(1995) made a hypothesis about an inverse U pattern, with mild drinkers being happier. More recently, an inverse U-shaped relationship among women and a J-shaped relationship among men were found by Massin and Kopp whereas other studies showed a linear relationship among women(Caldwell, Rodgers, Jorm, Christensen, Jacomb, Korten, and Lynskey 2002; Alati, Kinner, Najman, Fowler, Watt, and Green 2004; Zhan, Shaboltas, Skochilov, Kozlov, Krasnoselskikh, and Abdala 2012). However, little works are found in terms of depicting how both drinking alcohol and reducing alcohol can lead to happiness.

2.2.2 Happiness and Interpersonal Relationships. A strong, deep and close bond between two or more people refers to interpersonal relationships. The range spans from relations with neighbors to family connections. Association between happiness and interpersonal relationships has been explored. One concept of eudaemonic wellbeing is measured with six dimensions(Ryff 1989) and they are autonomy, self-acceptance, personal growth, environmental mastery, purpose in life and positive relations with others. These are based on the autonomy of an individual himself/ herself, except 'positive relations with others'. Researchers found that positive relations with others breed positive psychology by contributing directly to it or by indirectly buffering stress resulting from negative life events(Dean, Kolody, and Wood 1990; Wills, Blechman, and McNamara 1996; Liu, He, and Yan 2014). However, Markus et al.(1991) claimed that there is a cultural difference in terms of how interpersonal relations arouse happiness. People from individual-driven cultural contexts are tending to seek happiness from an autonomous agency. On the contrary, researchers(Uchida, Norasakkunkit, and Kitayama 2004) found that individuals influenced by East Asian cultural contexts are motivated to pursue happiness derived from healthy connections with others. In terms of gender difference, different social roles based on gender impact massively on individual's mental health(Rosenfield 2000). In addition, traditional gender roles affect happiness significantly(Kimmel 2012; Coltrane and Adams 2001). Also, Bach and others(Crose, Nicholas, Gobble, and Frank 1992; Bach, Telliez, Leke, and Libert 2000) showed that women are more accustomed to connections with others than men are. However, except for all the mentioned studies, there is still more room to fill in regarding studies on the association between gender and interpersonal relationships in terms of breeding happiness. We take the chance to contribute to this field.

2.2.3 Conclusion of Literature Review on Happiness. In this chapter we reviewed some papers that studied happiness, its association with alcohol consumption, as well as with interpersonal relationships. Although these studies generated fruitful results, they still face one minor limitation: the monotony of quantitative analysis. Precisely, all these papers measure happiness by having participants rate their subjective score for their happiness(Delle Fave et al. 2011). Participants are repeatedly asked to fill in surveys such as the General Social Survey and the World Value Survey. On the one hand, it proves that these methods gain wide recognition and validity. We can compare our result to that of studies utilizing these surveys. However, as Christopher and Hickinbottom(2008) put it, it necessitates the risk of assuming that people worldwide harbor the same definition of happiness based on the theories of western philosophers and psychologists. If new insight is yet to gain from positive psychology, qualitative assessment methods such as open-ended questions should also be embedded in the analysis. Participants should be allowed and guided to give more descriptions and comments on how they feel and what they define as happiness. Specifically, through qualitative analyses, many can be investigated, such as a personalized definition of happiness and prioritized lists of valuable things that lead to happiness. The goal of quantitive analyses is basically to explore relations among levels of happiness. Antonella et al.(2011) believed that in order to better define the contribution of different dimensions of happens and their interplay in breeding well-being, a combination of qualitative and quantitative analyses should be exploited. The thesis also takes the same stance and conduct both qualitative and quantitive analyses.

2.3 Crowd-sourced Data

Crowdsourcing is a model that leverages the collective intelligence via the Internet for problem-solving and task visualization. Jeff(2006a) first coined this word in an article. In some of the first attempts to form a concept for the term, crowdsourcing was defined as the practice initiated by a company or the sort to outsource its task, in the manner of an open call and via the Internet, to a larger network of people(Howe 2006b;

Brabham 2008; Estellés-Arolas and González-Ladrón-De-Guevara 2012; Littmann and Suomela 2014). Later, Estellés-Arolas and González-Ladrón-de-Guevara reviewed and summarized relevant literature and listed ten essential characteristics and nine typologies of crowdsourcing(Estellés-Arolas and González-Ladrón-De-Guevara 2012; Estellés-Arolas, Navarro-Giner, and González-Ladrón-de Guevara 2015). Since then, crowdsourcing has taken off with the popularization of smartphones, social media and AI.

Crowd-sourced data have the advantages of increased scalability, accelerated processes, reduced operational costs and increased consumer engagement. Crowdsourcing also fills knowledge gaps because it boosts the ability to access people who have certain skills(Estellés-Arolas, Navarro-Giner, and González-Ladrón-de Guevara 2015). All merits being listed, crowd-sourced data have fueled the academic turf massively, especially with the availability of platforms such as Amazon's Mechanical Turk. For tasks like mass data entry, user studies and cross-cultural studies, this platform provides an ideal solution(Buhrmester, Kwang, and Gosling 2016). The major reason is that it has a well-functioning participant compensation system, including a fairly large participant pool and a streamlined process of participation. Research has also been done comparing this way of massive data collection and traditional ways of participant recruitment. Results(Buhrmester, Kwang, and Gosling 2016) show that the site has necessary elements for successful data collection compared to other ways like traditional samples.

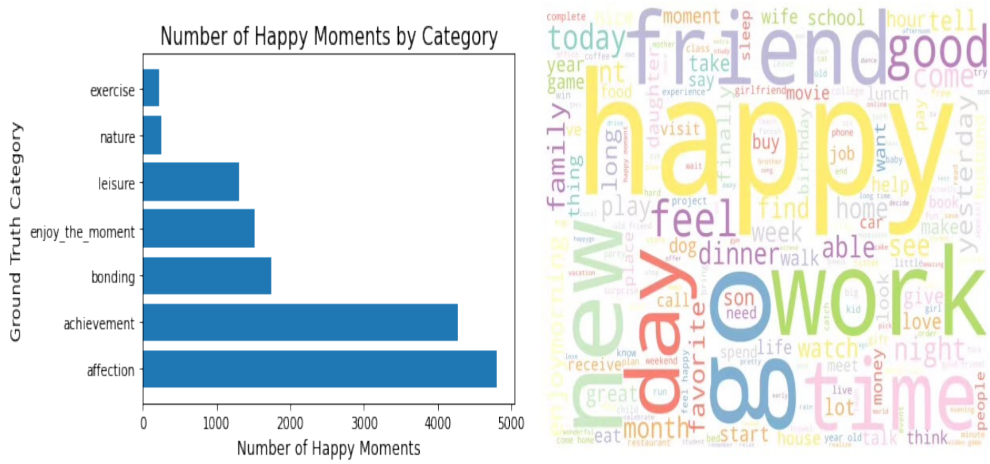
3. Study 1 Sociality and Agency in Happiness

In the previous section, we have listed all existing literature works that lead us to our research interests. From this chapter on, **we will be presenting our proposed methods and results of three separate studies from Chapter 3 to Chapter 5**. Chapter 3 introduces two core concepts of happiness, which are sociality and agency, by conducting two classification tasks. In the Chapter 4, association between happiness and alcohol is burrowed. Chapter 5 presents our work on how interpersonal relationships affect happiness and the attendant gender difference.

In Study 1, BERT-based LSTM is used to learn the concepts of sociality and agency. As stated previously, sociality in this context distinguishes feeling happy in the presence of others from feeling happy when the participants are alone, while 'agency' denotes the subject of the registered happy moment being the participants themselves or others. Precisely, there are two classification tasks where 'sociality' and 'agency' are learned separately in two independent settings.

3.1 Happy DB Database

The dataset we use throughout the thesis is an Amazon's Mechanical Turk crowd-sourced dataset called Happy DB database(Asai et al. 2018). 10,841 participants give brief descriptions of their happy moments in the past 24 hours or 3 months. They are asked to answer two questions: "What made you happy in the last 24 hours? What made you happy in the last 3 months?"(Asai et al. 2018). There are 5445 female(mean age= 34.44, SD= 11.87), accounting for 50.22% of the total population, while 5331 males(mean age= 31.12, SD= 9.43) represents 48.98%. 85 participants have no information on gender. All participants are from 100 different countries and districts, including the United States of America, India, Canada, etc.. Altogether, this database consists of



(a) Seven categories of happiness.

(b) Wordcloud of keywords that make women happy.

Figure 7: Visualization of the different categories of happiness(Asai et al. 2018) and the keywords that women feel happy about.

100,525 events of happy moments. The team(Asai et al. 2018) has already corrected some spelling mistakes and removed vacuous descriptions. Among all events(N= 100,525), 14,125 are annotated with seven labels of 'ground-truth-category' depicting the category of these events. The seven labels(see Figure 7) are distributed as follows: affection(4810), achievement(4276), bonding(750), enjoy-the-moment(1514),leisure (1306),nature(252), exercise(217). More detailed demographic information is shown in Figure 7 and in the supplementary material.

Note that each study requires different subsets of information from the original database. Therefore, we conditioned selection of subsets according to specific needs of each study.

3.2 Dataset Description for Study 1

For Study 1, there is already a corpus called CL-Aff corpus(Jaidka et al. 2019) where 'sociality' and 'agency' are annotated as an extension to the original happy DB database, so we took advantage of this. The corpus information is as follows:
- Labeled training set (N = 10,560): Labels that identify 'agency' of the participant and 'sociality' of the event are attached to the 10,560 subsamples of the original HappyDB corpus. Table4 shows the components of the labeled training data: agency ('yes'=7,796; 'no'= 2,764) and sociality ('yes'=5,625; 'no'= 4,935).
- Test set: (N = 17,215): Subsamples that contain also these labels and demographic information of authors are given.

For our models, the labeled training data are split into eighty percent training set (8,448 samples) and twenty percent validation set(2,112 samples).

Table 4: Distributions of all labels.

		Sociality		
		yes	no	Sum
Agency	yes	3554	4242	7796
	no	2071	693	2764
Sum		5625	4935	10560

3.3 Models

3.3.1 Baseline Models. We develop eight baseline models by applying traditional machine learning algorithms, including Support Vector Machine (SVM), Random Forest, Logistic Regression (Log Reg) and Naive Bayes. Each of them is implemented with two sets of word embedding algorithms and they are: Bag of Words(BOW) and Bag of Words with a TF - IDF transformation(BOW tf-idf). Default parameters in Scikit-learn are used to train all baseline models. We report binary classification accuracy of all these models in the result session. In addition to these machine learning approaches, LSTM companied with GloVe is also used as a benchmark to which our model is compared. Reasons of using GloVe are two: 1) Many of the existing publications(Jaidka et al. 2019) reported their highest accuracy from an architecture equipped with this word embedding algorithm. 2) GloVe is an unsupervised learning algorithm for distributed word representation, whose performance we want to compare to that of BERT. Training of GloVe is performed on aggregated global co-occurrence of word pairs from a corpus, and its representations display interesting linear substructures of the word vector space. We use the standard 100 dimensional GloVe embeddings trained on 840B word tokens. As to LSTM, We give more information in 3.3.2. Binary classification accuracy, area under curve(auc) and F1 scores of CloVe-Based LSTM are reported in the result session.

3.3.2 LSTM with BERT. Neural networks are great techniques and have brought breakthrough in applications like image recognition. However, there are two limitations that obstruct their performance in sequential data, such as time series and textual data. First, data in the whole dataset need to be of fixated length for a neural network to process it, while in real life sentences are mostly of different lengths. Compressing pre-trained vectors or padding them into a single representation would risk loss of information. Also, there is no memory associated with these models. Accordingly, neural networks are not a supreme option for processing sequential data. That is where Recurrent Neural Network(RNN) comes in(Mikolov et al. 2010). RNN addresses the said problems by taking in inputs of various lengths and introduces a hidden cell serving as a kind of memory. Then, LSTM, one variant of RNN, extends that idea by including a short-term component and also a long-term memory component while retaining the characteristics of varying-length inputs(Hochreiter and Schmidhuber 1997). Therefore, LSTM is a powerful tool for sequential data where there is dependence of one word on the ones that have preceded it. Since invented, LSTM has documented outstanding performance in tasks such as text generation, text synthesis and sentiment analysis, to name a few.

Parameters. As these two classification tasks are all related to sentences, we use one-layer LSTM connected with one dropout layer and on top of them one output layer. During the training, we set the unit size to 60. The hidden layer uses the default argument of Tanh(Karlik and Olgac 2011) as an activation function and the output layer uses sigmoid(Karlik and Olgac 2011) as an activation function. We use mini-batch gradient descent with size 32 and the Adam optimizer(Kingma and Ba 2014) is employed with a learning rate of 0.1. Early stopping(Prechelt 1998) is utilized to monitor whether there are signs of overfitting. If so, in order to deal with it, orthogonal initializer is used with a gain of 1.0. We also add a dropout layer(Srivastava 2013) with dropout rate being 0.2. Other methods of reducing overfitting have been experimented but no increase in accuracy is witnessed so we will only discuss them in chapter x of Discussion.

BERT. There are 24 types of BERT models, including BERT-Tiny, BERT-Small, and BERT-Large. We use the released BERT-Base model(Uncased: 12-layer Transformer, 768 dimensions, 12-heads, 110M parameters)(Devlin et al. 2018). All hyper-parameters remain as default values. For example, max sequence length is set as 128. We use padding and truncating skills to fix the length of each description to 128. So, for each input text, BERT outputs a tensor of shape (128, 768) with one vector per token. Out of 12 layers, We summed the last four layers as a pooling strategy to obtain a fixed representation for each happy moment description.

3.3.3 Technologies. We shortly brief the technologies we have applied in implementation of all models. These libraries and packages have supported the coding part.

Tensorflow. Tensorflow, an open source library for numerical computation, is applied in model implementation in this thesis. This framework contains a set of traditional machine learning algorithms and neural network models(Abadi et al. 2016). It eases the whole process of acquiring data, reshaping data, training models, producing classification results and all other refinement along the way. Tensorflow can train and run deep neural networks to finish AI-related tasks of various sorts, such as image recognition, text classification and word embedding generation(Abadi et al. 2016). Another merit of Tensorflow is that it supports prediction generation with the same models used for training(Abadi et al. 2016).

SciPy. This is a collection of open source software containing domain-specific toolboxes and various algorithms for numeric computation(Virtanen et al. 2020). This ecosystem is equipped with powerful functions for data management and high-performance computation. It is used throughout the coding part of the thesis.

NumPy. The essential package for numerical computation(Walt, Colbert, and Varoquaux 2011). By calling this package, high-dimensional arrays and matrix data structures are defined. Basic operations could also be run on them efficiently. NumPy arrays outstrip Python lists for it takes significantly less memory to process them.

pandas. This is a flexible, fast and user-friendly tool for data manipulation, including data wrangling and analysis(McKinney et al. 2011). To clean, transform, manipulate and analyze data of different kinds, pandas is a major game changer.

scikit-learn. Scikit-learn is a popular easy-to-use package that covers most standard machine-learning tasks, like clustering, regression and classification(Pedregosa et al. 2011). It provides a slew of common algorithms with consistent interface and scales to most data problems.

Models	Accuracy	
	Sociality	Agency
Logistic Regression(BOW-tfidf)	88.90	78.11
Logistic Regression(BOW)	89.77	**80.65**
Linear SVM(BOW-tfidf)	**90.49**	78.34
Linear SVM(BOW)	90.01	76.51
Random Forest(BOW-tfidf)	56.95	70.61
Random Forest(BOW)	56.91	70.60
Naive Bayes(BOW-tfidf)	70.80	46.11
Naive Bayes(BOW)	67.90	45.78

Table 5: Best performing baseline models on Sociality and Agency classification respectively.

Hyper-parameter	**Value**
Batch Size	**32**, 64, 128, 256
Numbers of Layers	**1**, 2
Units	32, 48, **60**, 128, 256
Learning Rate	0.001, 0.003, **0.1**
Droupout Rate	**0.2**, 0.4, 0.5
Optimizer	**Adam**, Rmsprop
Input Representation	GloVe, **BERT**

Table 6: The configurations of LSTMs are shown in this table. Batch sizes were iterated in steps of 32, 128 and 256. We experiment with three different learning rates and five units. The performance of a one-layer LSTM and two-layer LSTM is evaluated separately. Three dropout rates are applied in reducing overfitting. The results of two word embedding technologies are also compared. Values constructing the best performance are shown in bold.

3.4 Results

Table 5 demonstrates the accuracy of baseline models on the test set for both classification on sociality and agency. Compared to other models, the linear SVM with BOW as the word representation method performs well on the first task, achieving a score of 90.49%. In terms of classification on agency, Logistic Regression outperforms others, reaching 80.65% of accuracy. As the labels of agency are less balanced, most models perform less satisfying in this task. However, Random Forest shows a different pattern where its score is higher in the agency classification that that in the sociality classification.

Table 6 shows the configurations of LSTM we have trailed. In order to achieve a better outcome, we iterated batch size in steps of 32, 64 128 and 256. We experimented with both one-layer LSMT and two layers. Adding up a layer consumes more time to train the model and a quicker sign of overfitting occurred. Learning rate was tested at 0.001,

Models	Accuracy	F1 Score	AUC
ELMo + LSTM(publication)	85.00%	90.00%	None
GloVe + LSTM	83.70%	88.55%	89.41%
BERT + LSTM	**86.42**%	**90.42**%	**91.41**%

Table 7: Comparison of a one-layer LSTM with Different Word Embeddings on Agency Classification

Models	Accuracy	F1 Score	AUC
ELMo + LSTM(publication)	92.00%	93.00%	None
GloVe + LSTM	90.14%	90.89%	95.70%
BERT + LSTM	**93.00%**	**93.49%**	**97.11%**

Table 8: Performance of a one-layer LSTM with Different Word Embeddings on Sociality Classification

0.003 and 0.1. We also tried different unit sizes. In Table **??**, we have highlighted the hyper-parameter value that contributes to the best performer in both tasks.
We used three sets of evaluation methods(accuracy, area under curve and F1 score) to compare three models, the ELMo, CloVe and BERT trained models on these two tasks of classification. Note that we retrieved the accuracy and the F1 score of the ELMo-based model from the publication(Jaidka et al. 2019) as it was the best performer in the previously stated CL-Aff competition. The results presented in both Table 7 and Table 8 suggest that the proposed model with BERT outstrips the other two in both tasks and in all evaluation methods. It achieves 86.42% accuracy, 90.42% F1 score and 91.41% area under curve(auc) in the agency classification task. As to the sociality task, as the samples are more balanced, the result is better. Our model with BERT hit 97.11% in auc. Its F1 score is also the highest, boasting 93.49%. The model also achieved 93% in accuracy, outperforming the previous best model(EMLo+LSTM).

4. Study 2 Does Alcohol Bring Happiness?

In this chapter we probe into happiness related to alcohol consumption. If alcohol consumption can be seen as a positive factor in peoples' life experiences of happiness, one would expect descriptions related to drinking behaviors in the Happy DB database. Indeed, a group of participants, like one whose working id(wid) is 10,835 in Figure 1, recorded their happy moments which involved alcohol consumption.

4.1 Methods

As previously stated, data collection was done in the form of filling open-ended questions. The application of this type of qualitative assessment method can a breakdown of participants' beliefs, views and perceptions in their own special terms, in contrast to researchers' programmed definitions and categories which could possibly impose restrictions on participants' ability to give away more information(Delle Fave et al. 2011).

Question: *What made you happy in the last 24 hours?*

id (10835): Rewarding myself with Skinny Pete's chicken wings and beer for dinner.
id (2330): This evening I was able to try a rare beer that I have been waiting to have for a while.
id (2278): I had an alcoholic drink.

Figure 8: Snippets of happy moments related to alcohol consumption from Happy DB Database.

4.1.1 Data Extraction. From the original Happy DB Database, we removed some samples whose token size was more than 128 as this was the maximum length of tokens we configured in BERT. After filtering the dataset, we finally managed to utilize 99,953 samples, instead of the original 100,252 samples. These samples are written by 10,830 participants(male: N= 5,526, 51.02%; female: N= 5,304, 48.97%). As the majority of the original accounts were retained, we reckoned that this sample size still held efficient information. Out of these 99,953 samples, we extracted subsets of data based on a list of alcohol-related keywords. The list which was constructed by using both KOKO and cosine distances between word embeddings is added in the supplementary materials[7]. After data extraction, out of total 99,953 moments, we have obtained 730 samples which are related to drinking alcohol.

4.1.2 Annotation. In order to know whether sociality accounts for a critical factor in alcohol consumption, we followed the format of the CL-Aff task(Jaidka et al. 2019) to annotate all sentences related to alcohol along this dimension of sociality[9]. Paulhas' conceptualization of self-presentation(Paulhus and Trapnell 2008) includes a factor of 'communion'. Sociality synthesizes with this factor, revealed in writing as the reference to any activity practiced with or in the company of others.

Three annotators were requested to label all moments in the same time span. Final results were defined as long as at least two out of the three annotators agreed on one choice. Any question, doubt or incongruence regarding decision-making along annotation was discussed and resolved by common consent between researchers and annotators.

4.1.3 Participants and Measures. 569 participants report at least one happy moment related to alcohol, with males(N=340) accounting for 59.75% and females(N=229) 40.25%. The mean age of the participants is 31.56(SD= 9.01). The data are presented as counts with percentages or means with standard deviations(SD). The comparisons between gender and alcohol consumption are made by the chi-squared test when appropriate. We also provide a visualization with t-SNE to gain more insight of the data.

4.2 Results

As it is shown in Figure 10, 569 participants record at least one happy moment related to alcohol consumption, making up 5.25% of the number of total participants(N=

Sociality: Does this moment involve other people other than the author? YES/ NO

Examples of sentences which involve other people (Answer is YES):

- " Going out to drink with my friends made me happy."
- "Visited with friends whom were celebrating their wedding anniversary, we brought them a nice a bottle of wine to enjoy."

Examples of sentences which do not involve other people (Answer is NO):

- "I hit my alcohol limit yesterday with Bacardi."
- "I went to a wine a tasting."

Figure 9: An illustration of the examples given to annotators during labelling. Note that we presented original descriptions to annotators although some contain grammatical errors like the misuse of 'were' shown in the figure.

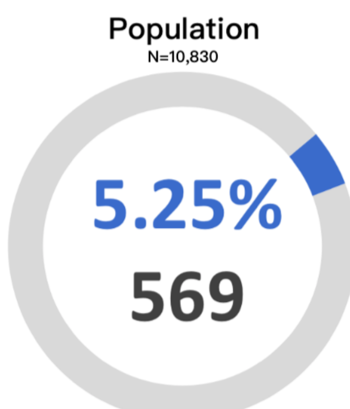


Figure 10: The proportion of participants who recorded at least one happy moment related to alcohol consumption.

10,830). Altogether, these participants report 730 happy moments relating to alcohol, accounting for 0.73% of the total moments in this study(N= 99,953). Interestingly, Figure 11 suggests that among them, 0.67% (N=671 moments) depict a scene where drinking alcohol brings happiness while 0.06% (N=59 moments) are from alcohol abstainers demonstrating how reducing or stopping alcohol consumption has resulted in happiness to their lives. Some descriptions are shown in Figure 12.

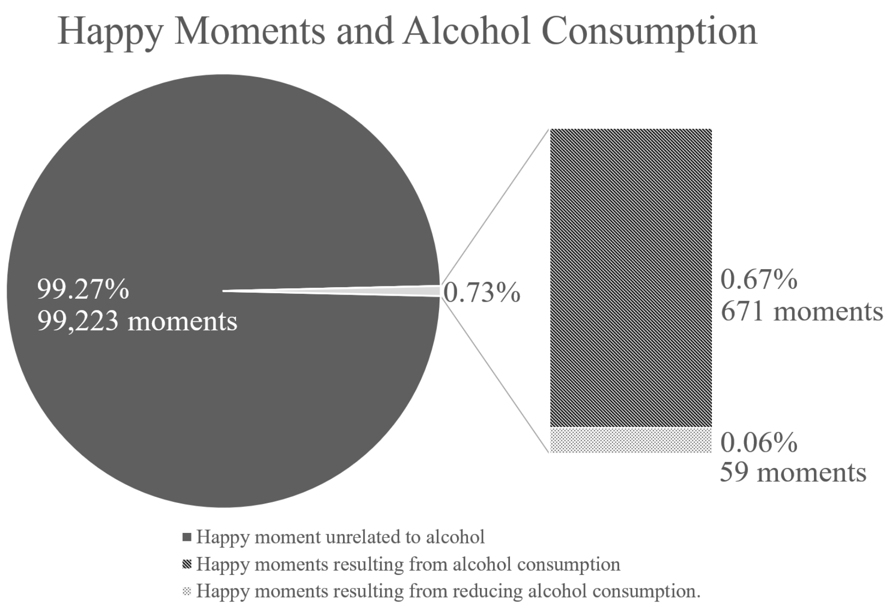


Figure 11: The proportion of happy moments that are related to alcohol consumption.

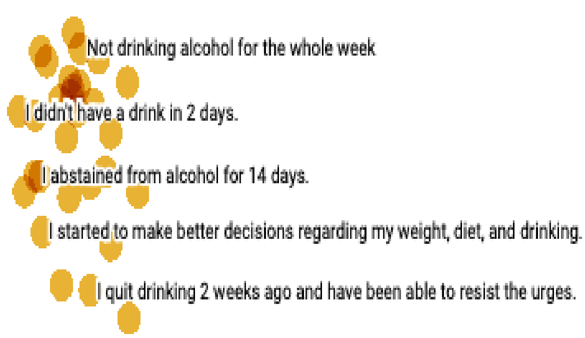


Figure 12: An illustration of 2-dimensional visualization of how reducing or stopping alcohol consumption can result in happiness. These descriptions are clustered together as we ran a t-SNE visualization with cosine similarity as the distance metric.

The result in Table 9 suggests that drinking in the company of others brings happiness more often than drinking alone, as 71.24% accounts(N= 478) depict a scene of drinking with others, while 28.76% happy moments(N= 193) occur when participants drink alone. As to gender difference, males report more cases of both drinking with others and drinking alone. 62.76%(N= 300) of happy moments of drinking in the company of others are reported by males, while the equivalent figure for females is only 37.24%(N=

178). Likewise, with regard to drinking alone, the percentage of moments recorded by males is higher than that of females, with 59.07% and 40.93% respectively.

Drinking Behaviors and Gender

	Number of Happy Moments of Drinking Alone	Number of Happy Moments of Drinking with Others	Sum.
Female	**79**	**178**	**257**
%	40.93%	37.24%	-
Male	**114**	**300**	**414**
%	59.07%	62.76%	-
Sum.	**193**	**478**	**671**
%	100.00%	100.00%	-

Table 9: Gender differences in drinking behaviors.

We also did Pearson's chi-squared test(Greenwood and Nikulin 1996) to see if there is a significant association between gender and drinking behavior when it comes to achieving happiness. There are six assumptions(Greenwood and Nikulin 1996) to satisfy before running a chi-squared test as follows: - The data in the cells should be frequencies rather than percentages or some other transformation of the data.
- All the categories of the variables are mutually exclusive.
- Each subject may contribute data to one and only one cell in the 2.
- The study groups must be independent.
- There are 2 variables and both are usually measured at the nominal level.
- At least 80% of the cells hold at least 5 values or more and no cell should have a value of less than 1.

Our study conforms to all assumptions so we can make hypotheses. The null hypothesis (H0) and the alternative hypothesis (H1) of the test are expressed in the following way:
H0: Drinking behavior is independent of gender; no association between these two variables exists.
H1: Drinking behaviors not independent of gender; an association between these two variables exists.

From the result ($\chi 2(2)>$ = 0.645, p-value = 0.421),it can be seen that since the p-value is greater than our chosen significance level ($\alpha = 0.05$), we do not reject the null hypothesis. Rather, we conclude that there is not enough evidence to suggest an association between gender and different drinking behaviors. Therefore, based on the result, we state the following: No association was found between gender and drinking behavior.

5. Study 3 Gender Difference in Happiness Derived from Interpersonal Relationship

In this chapter we burrow how happiness arises from four types of interpersonal relationships: romantic relationship, friendship, family and working relationship. Many existing literature works have studied the association between happiness and these interpersonal relationships, but few of them are equipped with the subjective descriptions of happy moments, which can obviously give us more insight. It can be seen from Figure 13, participants documented their happy moments derived from various types of bonding.

Question: *What made you happy in the last 24 hours?*

id (2): I was happy when my son got 90% marks in his examination.
id (14): A hot kiss with my girl friend last night made my day.
id(5652): I saw two close friends that I haven't seen for a couple months.
id (1783): My co-worker started playing a Carley Rae Jepson song from her phone while ringing out customers.

Figure 13: Snippets of happy moments related to the four types of interpersonal relationships from Happy DB Database.

5.1 Methods

5.1.1 Data Extraction. In order to explore whether all happy moments cover a wide range of life domains and activities, a list of keywords with highest occurrences in the original dataset was curated and these words were then grouped and saved into the nine files by Asai et al.(2018). We take advantage of this by manually selecting keywords from the files of 'people' and 'family'(see Figure 14). Then we curated four lists of keywords for each bonding type in interpersonal relationships: romantic relationship, friendship, family and working relationship.

5.1.2 Missing Value Removal. Based on the four lists of keywords, we extracted 41,601 moments from all samples(N=99,953). We also labeled them according to their corresponding type of interpersonal relationships. However, some missing values were found in terms of information about age, gender, marital status, and parenthood related to their writers. As these information is of crucial importance for analysis and the number of samples with missing values is negligible(N=164), so we removed those with missing values and ended up with 41,437 samples.

5.1.3 Participants and Measures. 9,341 participants recorded happy moments arising from connections with others, with males(N= 4,496) accounting for 48.13% and females(N= 4,845) 51.86%. The mean age of the participants is 32.95(SD= 10.81). The data are presented as counts with percentages or means with standard deviations(SD). The comparisons between gender and interpersonal relationships are made by using Pearson's chi-squared test when appropriate. We also provide visualization with t-SNE to get more insight into data in Figure 16 and Figure 17.

5.2 Results

Out of 99,953 descriptions, 41.46%(N= 41,437) involve or directly relate to interpersonal relationships. Among them, 52.54%(N= 21,768) are written by males, while 47.46% are by females. From the perspective of demographics, out of all 5,526 males who participated in the survey, 4,845 have reported at least one happy moment that has

Topics	% of Sentences in Topic	Size of Keywords List
people	46.0	478
family	26.4	423
food	16.2	1073
work	14.5	115
entertainment	8.8	156
exercise	8.4	558
shopping	8.4	35
school	5.5	47
pets	4.5	149
none	20.3	N/A

Figure 14: The team(Asai et al. 2018) provided a table that gives a breakdown of the distribution of the nine topics, with the rest being labeled 'none'. We can see that these nine topics cover almost 80 percent of the total corpus. However, note that a happy moment may be grouped into multiple topics. For example, "running with my pet" is related to both "pet" and "exercise". Another example would be 'my son got an A'. It is related to both 'family' and 'school'. Source: Retrieved from Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, Xu, Y. (2018). Happydb: A corpus of 100,000 crowdsourced happy moments. arXiv preprint arXiv:1801.07746.

occurred due to others, accounting for 87.68% of the total number of males. The equivalent figure for women is 84.77%. 4,496 women out of 5,304 have logged their happy moments relating to positive connections with others at least once.

Is there any difference by gender in terms of the distribution of four types of relationships? Family matters the most for both genders as the largest number of recorded happy moments for males and females are all connected to their family members. Women have written down 12,251 moments when they felt happy spending time with their family, which account for 62.30% of the total number of moments related to these four relationships. Although men have also logged 11,496 moments, taking up more than half(52.81%), this figure is still almost 10% lower than that of women. In contrast, men have recorded 30.77% of accounts(N= 6,698) associated with their friends, while women have noted down 4,341 moments, representing only 22.07%. The rest two types of bonding, romantic relationships and working relationships, collectively account for around 15% for both men and women. Precisely, 12.29% of happy moments(N= 2,416) took place when women shared their time with their significant partner while males documented a similar figure of 11.85%(N= 2,580). Minimum descriptions are found when both males and females felt happy due to their workmates, at less than 5%.

To see whether the gender difference stated above was statistically sound, we ran a Pearson's chi-squared test. As we mentioned in section 4.2, there are six assumptions to

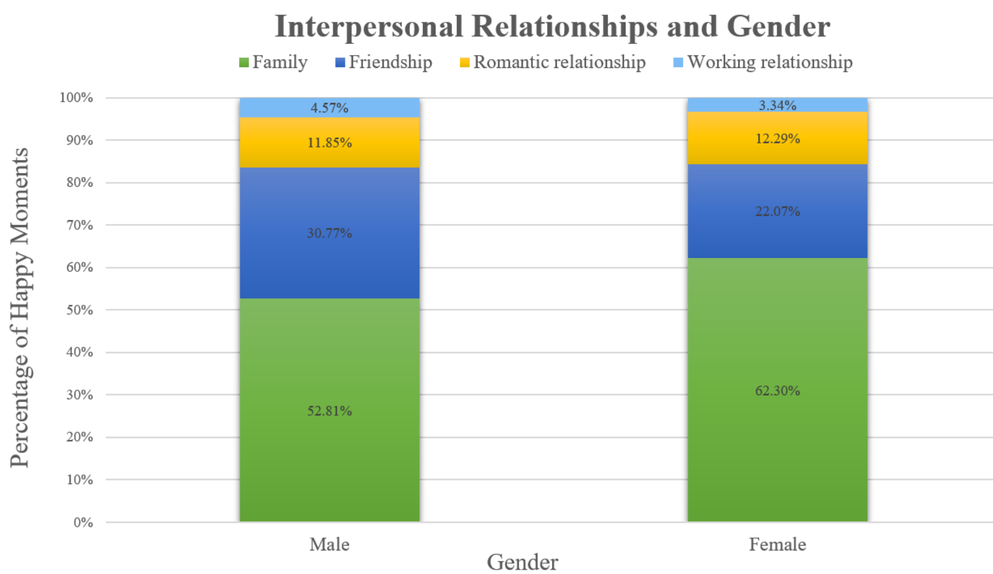


Figure 15: The bar graph of happy moments derived from four interpersonal relationships by gender. Both males and females write most happy moments related to their family members. Also, working relationship brings the least number of happy moments for both.

satisfy before running a chi-squared test and this study also satisfies all assumptions so we make the following hypotheses:
H0: Interpersonal relationship is independent of gender; no association between these two variables exists.
H1: Interpersonal relationship is not independent of gender; an association between these two variables exists.

With three degrees of freedom, the result is ($\chi 2(2)>$ = 495.969, p= 0.000). It can be seen that since the p-value is lower than our chosen significance level (α = 0.05), there is enough evidence to suggest a strong association between gender and interpersonal relationships. Therefore, based on the result, we state the following: a significant association was found between gender and interpersonal relationships.

6. Discussion

This thesis holds some congruent results with those of previous findings documented in the literature on happiness and also adds new insights to associations between variables that evolve around happiness in three studies. In short, our contribution to the current field is three-fold: 1) In terms of state-of-art algorithm implementation, we apply an outstanding word embedding algorithm BERT to transform all happy moments to semantic-sensitive representations to learn the two critical concepts of happiness, agency and sociality. Our performance is better than that of the existing publications. 2) In terms of assessment method, we leverage the crowd-sourced qualitative data on

Figure 16: 2D Visualization of happy moments related to the family. All sentences related to interpersonal relationships boast more than 40,000, so translating them into points in a 2-dimensional space could not impart easily understood information. But we zoomed in and found interesting clusters. For example, in this visualization, we spot on the sentence 'My older daughter keeps patting my younger daughter's head', a very loving description of family time. We highlighted the eleven moments that are closest to this sentence, measured by using cosine similarity(Nguyen and Bai 2010). We can see that as BERT is a context-sensitive embedding method, the resulting clustered sentences do share the common feature that most of them are related to family members. Keywords here are children, grandmother, moms, granddaughters, father, family and daughter.

which we conduct both quantitative assessment(i.e. vectors) and qualitative analysis. 3) Our result suggests that both alcohol consumption and abstaining from drinking can lead to happiness. These are useful aspects for us to gain more access to happiness.

6.1 BERT-based LSTM in Happiness Classification

We begin with discussion over Study 1 where two classification challenges are addressed. As stated before, these two challenges constituted the CL-Aff Shared Task 2019(Jaidka et al. 2019). Teams that participated in the competition experimented with both supervised learning and semi-supervised learning methods. Most teams used Convolutional Neural Network(CNN) or RNN and its variants like LSTM. The team from ASU hypothesized that a small set of word pairs were vital for representing the nature of sociality/agency of these happy moments(Saxon et al. 2019). They presented a Word Pair Convolutional Model, where they utilized Convolutional Neural Networks(CNN) for prediction on the test data. Three other teams also leveraged similar models based on CNN(Sun, Yang, Chi, Zhang, and Lin 2019; Claeser 2019; Xin and Inkpen 2019). Two teams applied RNN or its variants of LSTM or Bi-LSTM for this task(Rajendran, Zhang, and Abdul-Mageed 2019; Syed, Indurthi, Shah, Gupta,

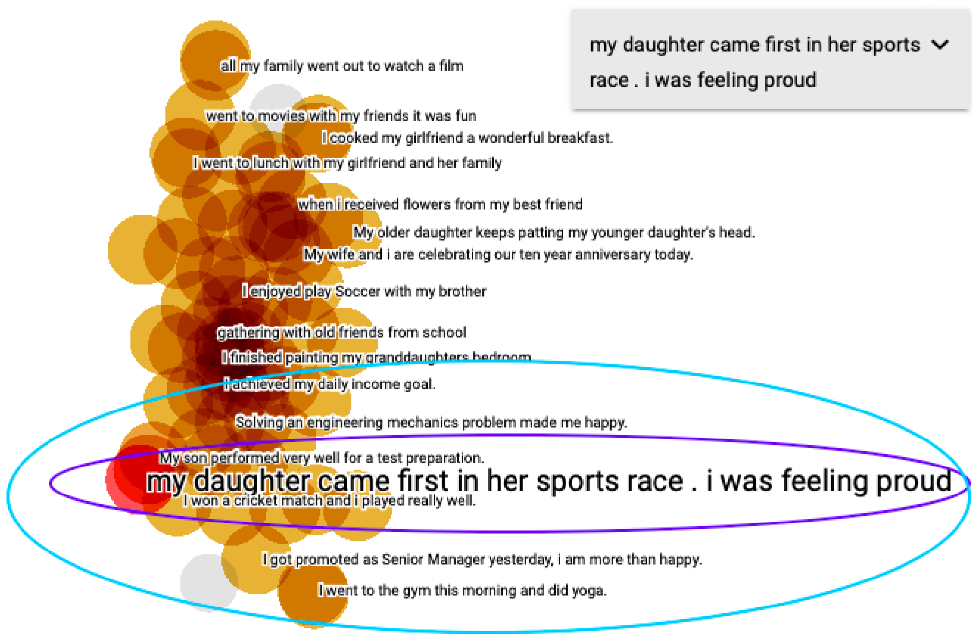


Figure 17: 2D Visualization of happy moments related to the family members and achievement. Zooming out a bit of the last visualization, we can easily see that BERT captured the feature of 'family' and 'goal achieving'. For example, we circle some sentences around the moment: 'My daughter came first in her sports race. I was in feeling proud.' The inner circle shows a similar situation where the participant's child also achieved good performance. Sentences within the outer circle depict the scenes of achievements of goals.

and Varma 2019). Apart from these, semi-supervised learning settings were designed by incorporating autoencoders(Bae, Cheong, and Song 2019) or k-means clustering scheme(Torres and Vaca 2019). Various embedding algorithms were also used in this task including word2vec(Mikolov et al. 2013)CloVe(Pennington, Socher, and Manning 2014), ELMo(Peters et al. 2018) and word embeddings pre-trained on WikiText-103 corpus(Torres and Vaca 2019). Among all these models, Elmo-based LSTM presented by UBC(Rajendran, Zhang, and Abdul-Mageed 2019) achieved the highest accuracy on test sets for both prediction on agency(85%) and sociality(92%). As it is widely known that LSTM is a holy grail for sequential data processing, from this comparison it can also be clearly seen the power of a context-sensitive word embedding algorithm like ELMo. However, our BERT-based LSTM outstrips all above-mentioned solutions, achieving 86.42% in agency prediction and 93% in sociality prediction on test sets.

To optimize the model, three modes of approaches have been taken. First, we use a grid search for hyper-parameters in LSTM. We choose different sizes for the hidden state vectors and test a dimension of 32, 48, 60, 128, 256 and 512. The grid search implements the batch size of 32, 64, 128 and 256. implementation included. Also, the learning rate is set to 0.001, 0.003 and 0.1. The dropout rate is iterated at 0.2, 0.4 and 0.5. We test all these in another two-layer LSTM as well.

During the whole process, we apply early stopping to monitor the learning trajectory with a patience of 5 minutes. It turns out that overfitting is a critical issue in these two classification tasks. Even the best performing models stop training after around mere 15 epochs of training, with a tendency of accuracy drop in validation set. As the architecture gets more complex, the epoch times reduce, so as the validation accuracy. The major reason could be lack of data in these tasks. Although we have over 100,000 samples but only marginally more than 10,000 are burrowed in the competition. Also, the complexity of bigger architectures for such a limited amount of textual data would backlash. Since the dataset is already fixated, we use a lighter model where each hyper-parameter is kept at a reasonably small level to deal with the overfitting problem. This works and the validation accuracy increases. The best result encapsulates a batch size of 32, a learning rate of 0.1, a hidden size of 60, a dropout rate of 0.2 and an optimizer of Adam. These parameters are further experimented for comparing word embeddings of GloVe and BERT.

Then, as (Devlin et al. 2018) claims, we do not have to focus on a computer with incredibly high-computational capability and more complex architecture to get a good outcome for a NLP downstream task, because an exceptional word representation of human language can do the work. Therefore, we test two sets of word embedding methods, GloVe and BERT. The reason for using GloVe is that many teams in the competition applied it in their model and the outcomes were satisfying. GloVe is also sensitive to context so we think it is qualified to be used to decipher all sentences. BERT is the core weapon of this thesis, so we use it for classification in this study of classification and also employ it for aiding the visualization in the other two studies. The reason is that BERT claims to be deeply contextualized(Devlin et al. 2018) compared to other contextualized models like GloVe, ELMo and ULFOM. These models are unidirectional, because when they were pre-trained, they took use of either the left or the right context words for prediction of the target word. However, BERT utilizes both sides, thus being equipped with more comprehensive ability to decipher all these sentences. For example(Devlin et al. 2018), I made a bank deposit. During vectorizing the word 'bank', other models would decipher it based on 'I made a' if they are trained by using left-sided context windows, or based on 'deposit' if they are trained on the right-sided context windows. In contrast, BERT uses both 'I made a' and 'deposit', thus eradicating the curse of meaning conflation. Our result serves as an evidence for this. With a very simple architecture, BERT can deliver a state-of-art show.

Thirdly, we experiment with some other techniques like adding penalties for weight size to the loss function and using an orthogonal kernel initializer. For adding regularization, we test with both L1 regularizer where the sum of the absolute weights is used and L2 regularizer where the sum of the squared weights is applied. Future work to improve the performance includes grid searching along order of magnitude for regularizers and combining different regularization. But all of them fail to increase our accuracy. The reason could be that we did not grid search different L1 and L2 values to a bigger extent so no optimal configuration was found with this method. In contrast, the orthogonal initialization works well and has increased the validation accuracy by 2% in both tasks. The initial weight matrix is chosen as a random orthogonal matrix in the orthogonal initializer. Therefore, even after repeated matrix manipulation there is no chance of exploding or vanishing since the eigenvalues of an orthogonal matrix have absolute

1(Kingma and Ba 2014). Our model benefits substantially from this property because once there is gradient vanish, training would be stuck in a near standstill where there is no backpropogation of information.

6.2 Alcohol consumption, Reducing Alcohol Consumption and Happiness

Alcohol can bring happiness, as previous literature suggest(Caldwell et al. 2002; Alati et al. 2004; Zhan et al. 2012). However, our finding suggests that not only alcohol consumption would lead to happiness that is worth recording from the perspective of the participants, but also reducing or stopping drinking alcohol can arose happiness. Apart from this, different drinking patterns are found among all participants: drink alone or drinking with others. That drinking alone leads to happiness is backed up by the theory of social inhibition and is also in turn a proof of it. Social inhibition mainly means avoidance of social interaction. A high level of social inhibition is possibly associated with lower possibility of disapproval from others(Denollet 2005). For behaviors like the use of alcohol, drinking alone would reduce anxiety that otherwise would arise when people who surround you have a different living habits for alcohol consumption. Drinking with others, evidenced by the most of partners, is more likely to brings happiness than drinking alone. Sociality has already been already proved as a significant predictor of happiness(Paulhus and Trapnell 2008). Compounded with alcohol consumption, sociality is catalyzed, thus resulting in a slew of happy moments from the participants where they and others have fun drinking together. Also, this finding may be partially caused by the fact that most of participants fall in the age range between 18 to 35 years old. This age span is engaged in a personal motive of seeking both isolation versus intimacy. Roothman(2003) posited that they tend to seek intimacy more. However, there is no dependence between gender and different drinking behaviors in terms of happiness associated with alcohol consumption.

6.3 Interpersonal relationships, Gender and Happiness

A significant association is found in gender and happiness from interpersonal relationships. Our findings suggest that men are more likely to be happy from interpersonal connections with others, disputing the claims of Bach and others(Crose, Nicholas, Gobble, and Frank 1992; Bach, Telliez, Leke, and Libert 2000). They found that women are more attuned to interpersonal relationships. This difference may arise from two sets of assessment methods. Our methods, which more focus on comments and opinions, are quite different from all previously mentioned studies, including those research that burrowed the association between alcohol and happiness. They usually rate happiness with an individual's subjective evaluation or with the General Social Survey and the World Value Survey without having participants giving a deeper explanation. This method results in quantized levels of happiness that are used in a slew of quantitive analyses. However, our method is to investigate a large dataset of crowd-sourced descriptions. There is no better way between these two because quantized data could be used for further modeling while accounts and comments could unveil more views from the perspectives of the participants. However, they all have limitations. Christopher and Hickinbottom(2008) claimed that surveys with mere subjective ratings on the concepts provided by researchers necessitate the risk of assuming that people worldwide harbor the same definition of happiness based on the theories of western philosophers and psychologists. Future research should consider merging and encapsulating these two methods for a deeper and comprehensive understanding of human emotions.

6.4 Limitation and Future Research Work

In the first study of classification tasks, although the performance of the proposed model excels those in the current publication. However, it still faces a serious problem of overfitting. Future work can focused on resolving this issue by applying another layer of attention mask on top of LSTMs. In the second and third tasks, although we have mined context-rich descriptions that aid our understanding of participants' mental well-being, we still lack variables that could be used to quantify their happiness. Future research posted online for crowdsourcing should combine recording event with a subjective rating of emotion, which could lead to more quality data collection. Also, comparative study of happiness across different cultural contexts should take more advantage of the currently thriving crowdsourcing so more samples could be done to decipher the cultural influence on human happiness.

7. Conclusion

We have done three studies by utilizing a crowd-sourced database called Happy DB Database. We apply a state-of-art word embedding algorithm BERT to transform all happy moments to context-sensitive representations to learn two critical concepts of happiness, agency and sociality. Our performance is better than that of the existing publications. We study the association between alcohol consumption and happiness and Our result suggests that both alcohol consumption and abstaining from drinking can lead to happiness. However, no association is found between gender and different drinking patterns. We delve into happiness that results from interpersonal relationships and significant association is found between gender and interpersonal relationships. Interestingly, men are more likely to get happy from the moments related to interpersonal connections. Another contribution of this thesis is in the assessment method, as we leverage the crowd-sourced qualitative data on which we conduct both quantitative assessment(i.e. vectors) and qualitative analysis.

References

Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283.

Alati, Rosa, Stuart Kinner, Jake M Najman, Greg Fowler, Kerrianne Watt, and David Green. 2004. Gender differences in the relationships between alcohol, tobacco and mental health in patients attending an emergency department. *Alcohol and Alcoholism*, 39(5):463–469.

Almuhareb, Abdulrahman. 2006. *Attributes in lexical acquisition*. Ph.D. thesis, University of Essex.

Asai, Akari, Sara Evensen, Behzad Golshan, Alon Halevy, Vivian Li, Andrei Lopatenko, Daniela Stepanov, Yoshihiko Suhara, Wang-Chiew Tan, and Yinzhan Xu. 2018. Happydb: A corpus of 100,000 crowdsourced happy moments. *arXiv preprint arXiv:1801.07746*.

Bach, Véronique, Frédéric Telliez, André Leke, and Jean-Pierre Libert. 2000. Gender-related sleep differences in neonates in thermoneutral and cool environments. *Journal of Sleep Research*, 9(3):249–254.

Bae, Byung-Chull, Yun-Gyung Cheong, and Youngrok Song. 2019. Modeling happiness using one-class autoencoder.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baroni, M, S Evert, and A Lenci. 2008. Esslli 2008 workshop on distributional lexical semantics. *Hamburg, Germany: Association for Logic, Language and Information*.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 238–247.

Baroni, Marco, Brian Murphy, Eduard Barbu, and Massimo Poesio. 2010. Strudel: A corpus-based semantic model based on properties and types. *Cognitive science*, 34(2):222–254.

Bengio, Yoshua. 2009. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127.

Brabham, Daren C. 2008. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1):75–90.

Bridle, John S. 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing*. Springer, pages 227–236.

Buhrmester, Michael, Tracy Kwang, and Samuel D Gosling. 2016. Amazon's mechanical turk: A new source of inexpensive, yet high-quality data?

Caldwell, Tanya M, Bryan Rodgers, Anthony F Jorm, Helen Christensen, Patricia A Jacomb, Ailsa E Korten, and Michael T Lynskey. 2002. Patterns of association between alcohol consumption and symptoms of depression and anxiety in young adults. *Addiction*, 97(5):583–594.

Camacho-Collados, José, Mohammad Taher Pilehvar, and Roberto Navigli. 2016. Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence*, 240:36–64.

Christopher, John Chambers and Sarah Hickinbottom. 2008. Positive psychology, ethnocentrism, and the disguised ideology of individualism. *Theory & psychology*, 18(5):563–589.

Claeser, Daniel. 2019. Affective content classification using convolutional neural networks. In *AffCon@ AAAI*.

Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(Aug):2493–2537.

Coltrane, Scott and Michele Adams. 2001. Men's family work. *Working families: The transformation of the American home*, pages 72–99.

Crose, Royda, Donald R Nicholas, David C Gobble, and Beth Frank. 1992. Gender and wellness: A multidimensional systems model for counseling. *Journal of Counseling & Development*, 71(2):149–156.

Danielsson, Per-Erik. 1980. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248.

Dean, Alfred, Bohdan Kolody, and Patricia Wood. 1990. Effects of social support from various sources on depression in elderly persons. *Journal of Health and Social Behavior*, pages 148–161.

Delle Fave, Antonella, Ingrid Brdar, Teresa Freire, Dianne Vella-Brodrick, and Marié P Wissing. 2011. The eudaimonic and hedonic components of happiness: Qualitative and quantitative findings. *Social indicators research*, 100(2):185–207.

Denollet, Johan. 2005. Ds14: standard assessment of negative affectivity, social inhibition, and type d personality. *Psychosomatic medicine*, 67(1):89–97.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Diener, Ed and Richard E Lucas. 2000. Subjective emotional well-being. *Handbook of emotions*, 2:325–337.

Diulio, Andrea R, Ian Cero, Tracy K Witte, and Christopher J Correia. 2014. Alcohol-related problems and life satisfaction predict motivation to change among mandated college students. *Addictive behaviors*, 39(4):811–817.

Epstein, Joyce L and Lori J Connors. 1992. School and family partnerships. *Practitioner*, 18(4):n4.

Estellés-Arolas, Enrique and Fernando González-Ladrón-De-Guevara. 2012. Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2):189–200.

Estellés-Arolas, Enrique, Raúl Navarro-Giner, and Fernando González-Ladrón-de Guevara. 2015. Crowdsourcing fundamentals: Definition and typology. In *Advances in crowdsourcing*. Springer, pages 33–48.

Ethayarajh, Kawin, David Duvenaud, and Graeme Hirst. 2019. Understanding undesirable word embedding associations. *arXiv preprint arXiv:1908.06361*.

Goldberg, David P and Valerie F Hillier. 1979. A scaled version of the general health questionnaire. *Psychological medicine*, 9(1):139–145.

Grant, Adam M, Marlys K Christianson, and Richard H Price. 2007. Happiness, health, or relationships? managerial practices and employee well-being tradeoffs. *Academy of management perspectives*, 21(3):51–63.

Greenwood, Priscilla E and Michael S Nikulin. 1996. *A guide to chi-squared testing*, volume 280. John Wiley & Sons.

Hills, Peter and Michael Argyle. 2001. Emotional stability as a major dimension of happiness. *Personality and individual differences*, 31(8):1357–1364.

Hochreiter, Sepp and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hofmann, Thomas. 2013. Probabilistic latent semantic analysis. *arXiv preprint arXiv:1301.6705*.

Howard, Jeremy and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.

Howe, Jeff. 2006a. Crowdsourcing: A definition.

Howe, Jeff. 2006b. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4.

Ingram, Rick E and Kathleen S Wisnicki. 1988. Assessment of positive automatic cognition. *Journal of consulting and clinical psychology*, 56(6):898.

Jaidka, Kokil, Saran Mumick, Niyati Chhaya, and Lyle Ungar. 2019. The cl-aff happiness shared task: Results and key insights.

Karlik, Bekir and A Vehbi Olgac. 2011. Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122.

Kimmel, Michael S. 2012. *Gender of Desire, The: Essays on Male Sexuality*. Suny Press.

Kingma, Diederik P and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kitayama, Shinobu, Hazel Rose Markus, and Masaru Kurokawa. 2000. Culture, emotion, and well-being: Good feelings in japan and the united states. *Cognition & Emotion*, 14(1):93–124.

Kitayama, Shinobu, Hazel Rose Markus, and Hisaya Matsumoto. 1995. Culture, self, and emotion: A cultural perspective on" self-conscious" emotions.

Kübler, Sandra, Ryan McDonald, and Joakim Nivre. 2009. Dependency parsing. *Synthesis lectures on human language technologies*, 1(1):1–127.

Kumar, Ankit, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing. In *International conference on machine learning*, pages 1378–1387.

Landauer, Thomas K and Susan T Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.

Landauer, Thomas K, Darrell Laham, Bob Rehder, and Missy E Schreiner. 1997. How well can passage meaning be derived without using word order? a comparison of latent semantic analysis and humans. In *Proceedings of the 19th annual meeting of the Cognitive Science Society*, pages 412–417.
Levy, David and Neil Sheflin. 1983. New evidence on controlling alcohol use through price. *Journal of studies on alcohol*, 44(6):929–937.
Littmann, Mark and Todd Suomela. 2014. Crowdsourcing, the great meteor storm of 1833, and the founding of meteor science. *Endeavour*, 38(2):130–138.
Liu, EK, WQ He, and CR Yan. 2014. ‘white revolution’to ‘white pollution’—agricultural plastic film mulch in china. *Environmental Research Letters*, 9(9):091001.
Mandera, Paweł, Emmanuel Keuleers, and Marc Brysbaert. 2017. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92:57–78.
Manning, Christopher D, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge university press.
Markus, Hazel R and Shinobu Kitayama. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, 98(2):224.
Martin, Matthew M and Rebecca B Rubin. 1995. A new measure of cognitive flexibility. *Psychological reports*, 76(2):623–626.
McKeown, Kathleen. 1992. *Text generation*. Cambridge University Press.
McKinney, Wes et al. 2011. pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9).
Mikolov, Tomáš, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*.
Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
Murphy, Christopher M, Jamie Winters, Timothy J O'Farrell, William Fals-Stewart, and Marie Murphy. 2005. Alcohol consumption and intimate partner violence by alcoholic men: comparing violent and nonviolent conflicts. *Psychology of addictive behaviors*, 19(1):35.
Nenkova, Ani and Kathleen McKeown. 2012. A survey of text summarization techniques. In *Mining text data*. Springer, pages 43–76.
Nguyen, Hieu V and Li Bai. 2010. Cosine similarity metric learning for face verification. In *Asian conference on computer vision*, pages 709–720, Springer.
Paulhus, Delroy L and Paul D Trapnell. 2008. Self-presentation of personality: An agency-communion framework.
Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.
Pennington, Jeffrey, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
Peters, Matthew E, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
Peterson, Christopher, Nansook Park, and Martin EP Seligman. 2005. Orientations to happiness and life satisfaction: The full life versus the empty life. *Journal of happiness studies*, 6(1):25–41.
Peterson, Christopher, Martin EP Seligman, et al. 2004. *Character strengths and virtues: A handbook and classification*, volume 1. Oxford University Press.
Prechelt, Lutz. 1998. Early stopping-but when? In *Neural Networks: Tricks of the trade*. Springer, pages 55–69.
Radford, Alec, Jeffrey Wu, Dario Amodei, Daniela Amodei, Jack Clark, Miles Brundage, and Ilya Sutskever. 2019. Better language models and their implications. *OpenAI Blog https://openai.com/blog/better-language-models*.
Rajendran, Arun, Chiyu Zhang, and Muhammad Abdul-Mageed. 2019. Happy together: Learning and understanding appraisal from natural language. *arXiv preprint arXiv:1906.03677*.
Ramos, Juan et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 133–142,

Piscataway, NJ.
Rindflesch, Thomas C and Marcelo Fiszman. 2003. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of biomedical informatics*, 36(6):462–477.
Roothman, Brett, Doret K Kirsten, and Marié P Wissing. 2003. Gender differences in aspects of psychological well-being. *South African journal of psychology*, 33(4):212–218.
Rosenfield, Sarah. 2000. Gender and dimensions of the self: Implications for internalizing and externalizing behavior.
Ryff, Carol D. 1989. Happiness is everything, or is it? explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, 57(6):1069.
Ryff, Carol D, Burton H Singer, and Gayle Dienberg Love. 2004. Positive health: connecting well–being with biology. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1449):1383–1394.
Saxon, Michael, Samarth Bhandari, Lewis Ruskin, and Gabrielle Honda. 2019. Word pair convolutional model for happy moment classification.
Schutze, Andreas, James Y Jeong, Steven E Babayan, Jaeyoung Park, Gary S Selwyn, and Robert F Hicks. 1998. The atmospheric-pressure plasma jet: a review and comparison to other plasma sources. *IEEE transactions on plasma science*, 26(6):1685–1694.
Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*.
Seligman, Martin EP. 2004. *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. Simon and Schuster.
Seligman, Martin EP and Mihaly Csikszentmihalyi. 2014. Positive psychology: An introduction. In *Flow and the foundations of positive psychology*. Springer, pages 279–298.
Srivastava, Nitish. 2013. Improving neural networks with dropout. *University of Toronto*, 182(566):7.
Sun, Baohua, Lin Yang, Catherine Chi, Wenhan Zhang, and Michael Lin. 2019. Squared english word: A method of generating glyph to use super characters for sentiment analysis. *arXiv preprint arXiv:1902.02160*.
Syed, Bakhtiyar, Vijaysaradhi Indurthi, K Shah, M Gupta, and V Varma. 2019. Ingredients for happiness: Modeling constructs via semi-supervised content driven inductive transfer learning. In *Proceedings of the 2nd Workshop on Affective Content Analysis@ AAAI (AffCon2019), Honolulu, Hawaii (January 2019)*.
Torres, Johnny and Carmen Vaca. 2019. Cl-aff deep semisupervised clustering.
Tripathi, Suraj, Chirag Singh, Abhay Kumar, Chandan Pandey, and Nishant Jain. 2019. Bidirectional transformer based multi-task learning for natural language understanding. In *International Conference on Applications of Natural Language to Information Systems*, pages 54–65, Springer.
Uchida, Yukiko, Vinai Norasakkunkit, and Shinobu Kitayama. 2004. Cultural constructions of happiness: theory and emprical evidence. *Journal of happiness studies*, 5(3):223–239.
Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
Veenhoven, Ruut. 2010. Greater happiness for a greater number. *Journal of happiness studies*, 11(5):605–629.
Vella-Brodrick, Dianne A, Nansook Park, and Christopher Peterson. 2009. Three ways to be happy: Pleasure, engagement, and meaning—findings from australian and us samples. *Social Indicators Research*, 90(2):165–179.
Ventegodt, S. 1995. Liskvalitet i danmark [quality of life in denmark]. *Results from a population survey. Copenhagen: Forskningscentrets Forlag*.
Virtanen, Pauli, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272.
Wallach, Hanna M. 2006. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*, pages 977–984.
Walt, Stéfan van der, S Chris Colbert, and Gael Varoquaux. 2011. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30.

Wang, Xiaolan, Jiyu Komiya, Yoshihiko Suhara, Aaron Feng, Behzad Golshan, Alon Halevy, and Wang-Chiew Tan. 2018. Koko: a system for scalable semantic querying of text. *Proceedings of the VLDB Endowment*, 11(12).

Wills, Thomas Ashby, Elaine A Blechman, and Grace McNamara. 1996. Family support, coping, and competence. *Stress, coping, and resiliency in children and families*, pages 107–133.

Xin, Weizhao and Diana Inkpen. 2019. Happiness ingredients detection using multi-task deep learning.

Yaghoobzadeh, Yadollah and Hinrich Schütze. 2016. Corpus-level fine-grained entity typing using contextual information. *arXiv preprint arXiv:1606.07901*.

Zhan, Weihai, Alla V Shaboltas, Roman V Skochilov, Andrei P Kozlov, Tatiana V Krasnoselskikh, and Nadia Abdala. 2012. Gender differences in the relationship between alcohol use and depressive symptoms in st. petersburg, russia. *Journal of addiction research & therapy*, 3(2).

Appendix A: Study 4 Age and Sources of Happiness

Study 4 explores association between age and seven different sources of happiness. Research questions in this study are as follows: - Is there any age difference in terms of source of happiness?

In this study we experiment with a sub-dataset of the original Happy DB database to see what is the main source of happiness for participants at different ages. In the sub-dataset provided by the same team who has conducted the original research(Asai et al. 2018), 14,066 random moments were selected . The team clustered these moments into seven groups and labeled them as seven sources of happiness, which include affection, achievement, bonding, enjoy-the-moment, leisure, nature and exercise. Two major steps were taken to prepare the dataset for further analysis. One step is to remove values that are not legitimate for the task, such as extremely large values as 223 years old. Another preprocessing step is the creation of age bins. There are misleading information given by participants concerning their age and two actions have been conducted as follows.

Extreme values removal. Samples containing weird age information were omitted. Ages such as 2, 3, 233 and 227 are what we do not reckon as logical. It is intellectually challenging for young babies aged 2 or 3 to finish all open-ended questions with technological gadgets and the worlds' documented longest lifespan is xx years old. Accordingly, we removed all samples with said ages. Non-related descriptions removal: One participant's answer to the questions concerning age was 'prefer not to say' and the very sample was dismissed.

Missing values removal. There are some samples including missing values in terms of age of the participants. As age is a critical factor in this study and also the number of missing values is negligible(N=12), so these samples were erased. In addition, we divide all ages which range from 18 to 88 into four bins: [18,30),[30,45),[45-65),[65-89). Each moment was annotated with corresponding label according to age information of its author. The four age bins are young, young-mid, mid and old.

Result. After data preprocessing, 13,951 moments are retained. Among them, 7,757 are recorded by young people, 4,849 are written by participants labeled as young-mid. 1,182 are from the middle-aged authors. The rest 163 moments are noted down by participants labeled as old. There are 5,460 participants in this sub-dataset, with males(N=2823) accounting for 51.70 and females(N=2637) 48.30%. The mean age of the participants is 32.55(SD = 10.18). The data are presented as counts with percentages or means with standard deviations(SD). As we are interested in whether there is any dependence between age bins and different source of happiness, we conduct Pearson's chi-squared test since the study abides by all the assumptions mentioned in the section4.2. The result shows an significant association was found between age and source of happiness (χ2(2)> =171.664, p =0.000, degrees of freedom=18).

Appendix B: Extracting the places that make people happy the most

We use KOKO for this task. KOKO is a Python3-based tool for language extraction(Wang et al. 2018). Users begins with a KOKO file and write in queries to search for needed information. KOKO advances information extraction in that it exploits multi-indexing scheme. Precisely, users can write queries based on conditions on the surface of texts and also on the structure of sentences. Also, as KOKO aggregates evince from the whole document, it achieves more refined extraction results(Wang et al. 2018).

```
extract "NPs" x from "happydb1.txt" if
        ("in" x {1.0}) or
        ("at" x {1.0})
with threshold 0.0
excluding(str(x) matches "(I|i|me|we|us|he|him|she|her|they|them|it)")
excluding(str(x) matches "(order|while|mail|lunch)")
excluding(str(x) matches ".*(time|noon|afternoon|morning|night|months|weeks|years|hours|
days|minutes|seconds)")
excluding(str(x) matches ".*(pm|am|AM|PM|p.m.|a.m.|month|while)")
excluding(str(x) matches ".*(month|year|winter|summer|April|January|February|March|May|
June|August|September|October|November|December|front|end)")
```

Figure 1: KOKO file that contains inquiries for places.

Places

Nr.	Places	Entity count	Threshold
1	work	1752	1.000000
2	school	155	1.000000
3	my house	62	1.000000
4	home	187	1.000000
5	the world	111	1.000000
6	bed	76	1.000000
7	Mumbai	26	1.000000
8	my pocket	23	1.000000
9	the park	42	1.000000
10	the gym	21	1.000000

Figure 2: A table of ten places that make people happy the most. We ran the KOKO extractor in the whole data file and printed the 10 places with the highest frequencies in people's accounts of happy moments.

Topics	% of Sentences in Topic	Size of Keywords List
people	46.0	478
family	26.4	423
food	16.2	1073
work	14.5	115
entertainment	8.8	156
exercise	8.4	558
shopping	8.4	35
school	5.5	47
pets	4.5	149
none	20.3	N/A

Figure 1: Nine frequently mentioned topics curated in order to get keywords of each topic. Source: Retrieved from Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, A., ... Xu, Y. (2018). Happydb: A corpus of 100,000 crowdsourced happy moments. arXiv preprint arXiv:1801.07746.

Appendix C: Some basic statistics of the original Happy DB Database

Appendix D: Some basic statistics of the original Happy DB Database

```
[ ]  alcohol = ['Alcohol',
     'beers'
     ,'Wine'
     ,'Drink'
     ,'Beverage'
     ,'Beer'
     ,'Cider'
     ,'drank'
     ,'Distilled'
     ,'drinks'
     ,'drink'
     ,'Rectified'
     ,'Spirits '
     ,'spirits'
     ,'Mead '
     ,'Fermented tea'
     ,'Whisky'
     ,'Sparkling wine'
     ,'Liquor '
     ,'Liqueur'
     ,'Cocktail'
     ,'Tequila'
     ,'Gin'
     ]
```

Figure 1: Keywords of alcohol. We compile this list of keywords by using the entity extractor KOKO.