Through The Eyes of A Poet: Classical Poetry Recommendation with Visual Input on Social Media

Daniel (Yue) Zhang, Bo Ni, Qiyu Zhi, Thomas Plummer, Qi Li, Hao Zheng, Qingkai Zeng, Yang Zhang, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame, IN, USA
{yzhang40, bni, tplummer, qzhi, qli8, hzheng3, qzeng, yzhang42, dwang5}@nd.edu

***Abstract*—With the increasing popularity of portable devices with cameras (e.g., smartphones and tablets) and ubiquitous Internet connectivity, travelers can share their instant experience during the travel by posting photos they took to social media platforms. In this paper, we present a new image-driven poetry recommender system that takes a traveler's photo as input and recommends classical poems that can enrich the photo with aesthetically pleasing quotes from the poems. Three critical challenges exist to solve this new problem: i) how to extract the implicit artistic conception embedded in both poems and images? ii) How to identify the salient objects in the image without knowing the creator's intent? iii) How to accommodate the diverse user perceptions of the image and make a diversified poetry recommendation? The proposed iPoemRec system jointly addresses the above challenges by developing heterogeneous information network and neural embedding techniques. Evaluation results from real-world datasets and a user study demonstrate that our system can recommend highly relevant classical poems for a given photo and receive significantly higher user ratings compared to the state-of-the-art baselines.**

I. Introduction

With the prevalence of portable devices with cameras and ubiquitous networks, people can take photos and post them immediately on social media to share the adventures in their lives [1], [2], [3]. Figure 1 shows that a traveller is taking a photo of an astonishing waterfall and looking for a poetic descriptor of the photo. However, searching such poetic descriptors online can be a challenging and time consuming task if the poem has not been in the traveller's mind. In this paper, we develop a novel recommender system, iPoemRec, for automatic recommendations of classical poetry using visual input. We choose classical poetry because it is an important asset of the cultural heritage and contains a rich set of quotes that touch many aspects of our lives [4]. Such a recommender system will also help today's social media users to appreciate and inherit the beauty of ancient culture in an entertaining manner.

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08...$15.00
http://doi.org/10.1145/3341161.3343708

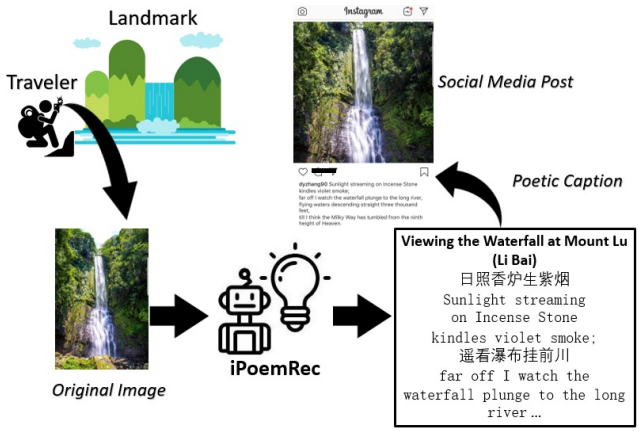


Figure 1: An Example of iPoemRec

The problem that iPoemRec addresses is more challenging than the cross-domain information retrieval task where images are used to search for relevant texts [5], [6]. The reason is that we need to explicitly explore the *artistic conception* of both the poems and images. Such artistic conception is often implicit and viewed differently by users with diverse cultural backgrounds [7]. The current solutions in recommender systems and information retrieval cannot be used to solve our problem because they do not explicitly explore the artistic conception expressed by the poems or images. We summarize the key technical challenges of iPoemRec below.

Implicit Artistic Conception: The artistic conception of a poem or image often refers to the emotions or topics that the creators intend to express in an implicit manner. In particular, a unique feature of classical poetry is the common employment of symbolism and metaphor in the poems. For example, in Chinese classical poetry, the authors often use "red beans" to refer to "lovesickness", and "climbing" to refer to "ambition". Similarly, the viewers have various "feelings" about an image that are affected by the objects, activities, and colors of the image. Existing solutions in text-image matching mainly focus on the *object relevance*, which recommend texts based on the similarities of the objects in an image and words in a poem [8]. These solutions can easily lead to wrong recommendations as illustrated in Figure 2. In this example, the image expresses the feeling of "loneliness". Both Poem A and B share the same feeling and therefore are considered as matching poems, even

though Poem B does not contain any objects described in the image[1]. In contrast, Poem C, while having a matching object to the image (i.e., “bird"), is a wrong recommendation since it expresses a different feeling of “happiness".

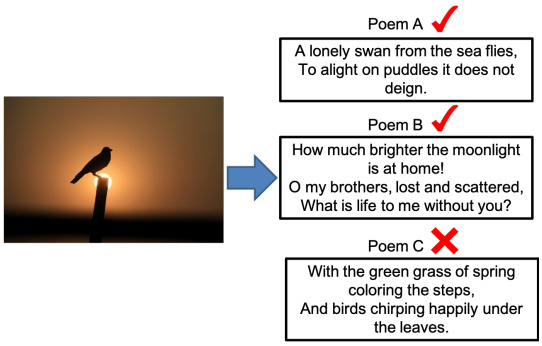


Figure 2: Poetry Recommendation Example

Poetic Visual Entity Identification: It is also challenging to correctly identify the salient objects in the image without knowing the creator’s intent *a priori*. For example, the focus of the image in Figure 2 is the bird and sunset, not the tree branch. The system can generate irrelevant poems by missing the bird and sunset from the picture but instead focusing on the tree branch. Moreover, we also need to rank the objects in the image based on their “poetic" value. In the above example, the sunset is more important than the bird since it directly relates to the “loneliness" feeling of the image. Ranking bird over sunset may mislead the recommender system to recommend inappropriate poems like Poem C in Figure 2.

Diversified User Perception: we observe that users of the poetry recommender system are often subjective and may have diversified perceptions of the same image. For example, some people may feel “loneliness" by looking at the image of Figure 2 while others may feel “sadness" or “peaceful". Therefore, if a recommender system always recommends the same poem for a given image, it would fail to meet the diversified perceptions of its users. The goal of our system is to explore different perceptions of an image and recommend a diverse set of poems with the artistic conceptions that are all relevant to the image.

The iPoemRec system jointly addresses the above challenges and provides effective classical poetry recommendations for given visual inputs. In particular, to address the implicit artistic conception challenge, iPoemRec explicitly models the sentiments, metaphors, and themes in poems and images by developing a conception-aware heterogeneous information network. To address the poetic visual entity identification challenge, iPoemRec develops a poetic visual analyzer module to extract the poetic objects that clearly represent the artistic conception of an image. Finally, iPoemRec develops a new meta-path embedding based recommendation method to address the diversified user perception challenge. To the best of our knowledge, iPoemRec is the first poetry recommender system given visual inputs, which brings value of classical poetry into the modern image sharing social media. Evaluation results on real-world datasets and a user study demonstrate that iPoemRec significantly outperforms the state-of-the-art recommender systems by recommending more relevant and enjoyable poems to its users.

[1]The poems are translated from the original classical Chinese poems.

II. Related Work

A. Information Retrieval with Visual Input

Existing information retrieval (IR) solutions using visual inputs focus on image search engines, which discover images that are visually similar to the input image [9], [10], [11], [12]. The iPeomRec scheme solves a new information retrieval problem where an image is used to search for a set of relevant poems. There exist a few cross-modality information retrieval solutions that are relevant to our work. For example, Gu *et al.* developed a textual-visual retrieval tool that leverages neural embedding techniques to jointly transform text documents and images into a homogeneous feature domain [5]. Niu *et al.* adopted a tree-structured Long Short-Term Memory (LSTM) neural network scheme to learn the hierarchical relations between phrases and visual objects [6]. Commercial systems such as Google Image have also shown to be effective in identifying the landmarks and events related to an image. These methods cannot solve our problem because they do not consider the implicit sentiment and theme of the images and poems, which are particularly important in our poetry recommendation problem.

B. Image Description Generation

Our work is related to the literature on image caption generation. For example, recent progress on image captioning has made it possible to generate statements to describe images [13], [14]. These models either use a single sentence or a few keywords to describe visual content. Krause *et al.* improved existing image captioning models by developing an image-to-paragraph tool that produces a coherent and detailed story for an image [15]. A few recent solutions further add poetic value to the image descriptors. For example, Xu *et al.* developed a poem generation tool that can write a poem using memory-based neural networks [4]. A recent work from Microsoft’s XiaoIce can generate poems that match the adjectives (e.g., beautiful, magnificent) expressed by the image [16]. However, none of these solutions can detect the artistic conception of a picture, which is particularly difficult due to the usage of symbolism and metaphor to express the sentiment and theme. In contrast, iPoemRec can identify the artistic conception in a picture by constructing a novel conception-aware heterogeneous information network.

C. Information Network in Recommender System

Information network is often used to facilitate information retrieval by providing valuable knowledge about users, the items to recommend, and their relations [17], [18], [19], [20]. For example, Oramas *et al.* developed a music recommender system that uses information network to enrich the description of music items with semantic information (e.g., genre, artist information) [21]. Zhang *et al.* developed a content-free recommender system to identify the copyright infringing videos

on Youtube by exploring the network constructed from user's comments [22]. Wang *et al.* developed an explainable recommender system by leveraging the sequential dependencies within the paths of an information network to reason about the underlying rationale of a user-item preference [23]. In our work, we present a new conception-aware information network that models the relationships among sentiments, metaphors, themes, and the phrases in a poem. We designed a new network embedding approach to fully explore the conception-aware information network for classical poetry recommendation.

III. Problem Formulation

In this section, we formally define the problem of recommending classical poetry given the image input. In particular, we consider a poetry recommendation application that takes an input image I from a user. The recommended poems are selected from a corpus containing a set of classical poems: P. For each poem $P \in P$, we assume it is associated with an *artistic conception*, which is defined below.

DEFINITION 1. ***Artistic Conception****: it conveys the sentiments and themes beyond the wordings of the poem [7].*

In this paper, we focus on two particular aspects of artistic conception: *sentiment* and *theme*. The sentiment represents the emotion (e.g., happiness, loneliness, anger) that the authors express in the poem and the theme is the topic of the poem (e.g., love, travel, festival, farewell). Such artistic conception also exists in images (e.g., an image with a guy holding a bouquet of roses represents a theme of "love", and a picture of an abandoned house represents a sentiment of "sadness").

A key characteristic in classical poetry is the use of *metaphor*, which often implicitly refers to both themes and sentiments. For example, in classical Chinese poetry, the term "spring breeze (春风)" is often metaphorical to a sentiment of "happy", the term "old tree (老树)" is often metaphorical to a sentiment of "sad", and "a pair of mandarin ducks (鸳鸯)" is metaphorical to the theme "love". To extract the artistic conception from poetry, metaphor must be explicitly considered. For example, recommending a poem about the old tree to a photo of singing bird on a tree leads to a wrong recommendation by ignoring the metaphor used in the poem. In contrast, if the system recommends a poem about love to a user who took a picture of mandarin ducks, the user will feel "an element of surprise" and appreciate the recommendation of a matching theme. The goal of this paper is to recommend the poems that are most relevant to the artistic conception of a given input image. The relevance is defined by three aspects:

DEFINITION 2. ***Object Consistency****: the objects the image describes should match the objects in the recommended poems.*

DEFINITION 3. ***Theme Consistency****: the theme in an image should match the theme expressed by the recommended poem.*

DEFINITION 4. ***Sentiment Consistency****: the sentiment of the image should match the sentiment expressed by the recommended poem.*

We use $OC_{I,P}$, $TC_{I,P}$, $SC_{I,P}$ to denote the object, theme, and sentiment consistency between an input image I and a poem P, respectively.

The objective of poetry recommendation is to output a recommendation list of poems that satisfy the above three consistencies for a given image. More specifically, the output of iPoemRec is a rank list of recommended poems $RL = \{\tilde{P}_1, \tilde{P}_2, ..., \tilde{P}_K\}$, where K is the size of the recommendation list. We set K to 1 in applications where only the best poem is recommended. We formally define the objective of the iPoemRec as:

$$\arg\max_{RL} \sum_{\tilde{P} \in RL}^{\tilde{P} \in RL} (OC_{I,\tilde{P}} + TC_{I,\tilde{P}} + SC_{I,\tilde{P}}) \quad (1)$$

IV. Solution

In this section, we present the iPoemRec system (Figure 3) to address our poetry recommendation problem. It consists of three components: 1) a Conception-aware Heterogeneous Information Network (CaHIN) module that models the semantic relationships among the sentiments, themes, objects, and metaphors in both images and classical poems; 2) a Poetic Visual Analyzer (PVA) module that extracts salient objects and their descriptors to capture the artistic conception of an input image; and 3) a Semantic Enriched Meta Path Ranking (SEMPR) module that uses the CaHIN to learn the latent representation of an image and a poem for effective recommendation. We discuss each component in details below.

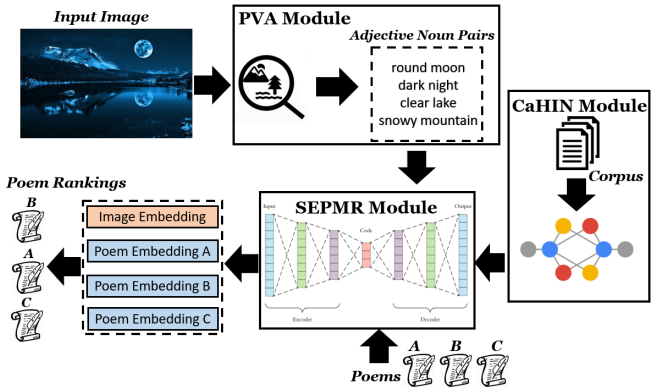


Figure 3: iPoemRec Overview

A. Conception-aware Heterogeneous Information Network (CaHIN) Module

We first define the *Conception-aware Heterogeneous Information Network* as an *undirected* graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ where $\mathbf{V}$ is the set of entities and $\mathbf{E}$ is the set of links between entities. In particular, we consider three types of entities.

Sentiment entity ($\mathbf{V}_S$): a sentiment entity denotes the emotion (e.g., happiness, sadness, anger) of a poem.

Theme entity ($\mathbf{V}_T$): a theme entity denotes the topic (e.g., romance/love, homesickness, ambition), event (e.g., travel, study), or context (e.g., spring, winter) of a poem.

Adjective-Noun Pair (ANP) entity ($\mathbf{V}_A$): an ANP entity is the combination of an adjective and a noun (e.g., “high mountain", “lonely bird" and “bright moon").

We observe that ANP entities often serve as metaphor for sentiment and theme of a poem (e.g., “high mountain" is metaphor for “ambition"). There also exists relationship between sentiment and theme entities. For example, a theme of “lovesickness" often implies sentiments of “sadness" or “loneliness". We explicitly model the metaphorical relations among these three entities in CaHIN. In particular, we define links between two entities to model the metaphorical relationship between them. A link $e_{v,v'}$ *2* **E** denotes entity v^0 is related to entity v. Each link is associated with a *weight*, $w_{v,v'}$, that represents the strength of the relationship between the connected entities. We define five types of links in CaHIN below:

ANP-ANP link (AA): a link representing the similarity between two ANPs.

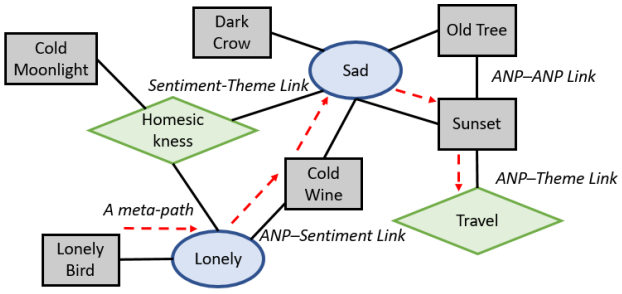
ANP-Sentiment link (AS): a link representing the relevance between an ANP and a sentiment entity.

ANP-Theme link (AT): a link representing the relevance between an ANP and a theme entity.

Theme-Theme link (TT): a link representing the similarity between two theme entities.

Sentiment-Theme link (ST): a link representing the relevance between a sentiment entity and a theme entity.

We do not define sentiment-sentiment link above because we assume there often exists only one major sentiment for a given image or a poem. An example of CaHIN is shown in Figure 4. In this example, we not only can find the explicit relationship between connected entities (e.g., “Old Tree" and “Sad") but can also find the implicit relationship by performing graph traversals (referred to as meta-pahs). For example, an implicit link between the “Sad" entity and “Lonely" entity are identified by the meta-path shown in the dash line. We explain how to use the CaHIN to recommend poems in Section IV-C.



The rectangular blocks are ANP entities; the diamond blocks are theme entities; and the oval blocks are sentiment entities.

Figure 4: CaHIN Example

CaHIN can significantly facilitate the detection of the artistic conception related entities in both poems and images. For example, a poem with a “sad" sentiment and “old tree" ANP can traverse the graph and find the “lonely" sentiment and “sunset" are also related to the poem. Similarly, images can use the visual ANPs defined in the next subsection (e.g., “dark crow") to traverse the graph to find implicit artistic conception of the image (e.g., “sad" sentiment and “homesickness" theme).

The CaHIN is built from a corpus of classical poems, which is discussed in Section V. To extract the entities of CaHIN, we collect a set of popular sentiments and themes from the poems. The sentiments are classified into 3 predominant categories in classical Chinese poetry: i) “happy", “peaceful", and “inspirational" that represent positive emotions; ii) “angry", “lonely", and “sad" that belong to negative feelings; and iii) a “neutral" sentiment using crowdsourcing. We identify a set of 13 themes from the poems in the corpus (i.e., “romance/love, war, ambition, homesickness, lovesickness, separating, friendship, spring, summer, autumn, winter, education, travel"). The ANPs are extracted using Google Natural Language API. In particular, we first perform Part-of-Speech (POS) tagging and then extract the nouns and phrases that match the <adjective, noun> pattern.

To construct the links in CaHIN, we first derive the link weights as follows. For an AA link, the weight is defined as the text similarity score between two ANPs using Word2Vec [24]. For AS, AT, TT, and ST links, the weight is defined as the probability that the two entities co-appear in a poem. All above link weights are normalized to a (0, 1] scale. Then, we create a link between two entities only if the link weight is higher than a threshold W_{thres} to reduce the density of the graph (for algorithm efficiency concern). The assignment of W_{thres} is discussed in Section V.

B. Poetic Visual Analyzer (PVA) Module

The CaHIN allows us to query for enriched sentiments, themes and APNs of a poem or an image through the graph traversal. However, the query cannot be performed for a raw image, which is simply a composition of pixels. In this section, we develop a PVA module to extract the “visual ANPs" from the image. The visual ANP is defined as an object in the image and its descriptor. Examples of visual ANPs are “magnificent waterfall", “misty mountain", and “dark crow".

To extract these ANPs, we first use the state-of-the-art object detector YOLO V3 [25] to detect the objects in an image. To find the descriptor of each object, we use a Convolutional neural network pre-trained on ImageNet [26] and fine-tuned on SentiBank [27], a large-scale dataset that is specifically designed to extract descriptors of objects. In the above ANP extraction process, we also consider the relevance and poetic value of the objects. Consider a picture with a man holding a bouquet of roses near a car. The most salient object in this example is the “rose" rather than the object "car", which can give confusing information to the recommender system. To identify the salient APNs, PVA ranks the APN based on its “metaphorical degree", which is defined below. APNs with highest metaphorical degrees are used.

DEFINITION 5. ***metaphorical degree of an ANP*** *(δ): the sum of the weights of all AS and AT links connected to the ANP in* **G**.

The intuition of the above definition is that, if a phrase is strongly related to sentiments and themes, it may contain more artistic value than ordinary APNs that has no such connection. Note that the above process may suffer from the problem of "out of vocabulary issue" where some ANPs in an image may not appear in the CaHIN. We address this issue through *stemming* and *transformation* processes. The stemming process is a common natural language processing technique that reduces derived words to their word stem (e.g., from "raven" to "bird" or from "carp" to "fish"). The transformation process replaces the noun or descriptors of the ANP with the most similar ANP entity in CaHIN (e.g., from "black raven" to "dark crow"). After the stemming and transformation processes, the visual ANPs of images are consistent to the ANPs in CaHIN.

C. Semantic Enriched Meta Path Ranking (SEMPR) Module

The goal of the SEMPR module is to 1) learn the representations of an image and a poem that best capture their artistic conception; 2) perform effective poetry recommendation based on the similarities of the learned representations.

1) Diversity-aware Random Walk Meta-Path Extraction: As mentioned in the CaHIN module, the meta-paths allow iPoemRec to identify implicit relationships between entities. By traversing the graph with many meta-paths, the CaHIN can identify a collection of heterogeneous entities that are relevant to a poem/image. We use the set of meta-paths as the "representation" of a poem/image.

In particular, we traverse the CaHIN starting from each ANP of an image, and collect all relevant entities (including sentiment, theme, and ANP entities) along the paths. We use random walks, a commonly used technique for traverse and extract information of a graph [28]. Each walk randomly traverses a graph and visits at most N entities (including the APN entity that the random walk starts with). The parameter N controls the *diversity* of the recommendation as it affects how many sentiments/themes will be explored for an image. To avoid retrieving irrelevant entities when N is large, we calculate the accumulative link weight (ALW), that is the multiplication of all link weights traveled so far, and stop the random walk if $ALW \quad W_{thres}$. For each APN, the random walk is performed multiple times to ensure all relevant entities are collected. All the visited entities during a random walk constitutes a meta-path, denoted as Ω. For each image/poem, we extract M meta-paths via random walks to ensure all relevant entities are collected.

To match an image to a poem, we represent each poem as a collection of meta-paths as well. The meta-paths are collected by performing the same random walk starting from the ANP, sentiment, and theme entities of each poem. The intuition is that we would like to collect all entities that are relevant to the poem's artistic conception as well as the APNs that are relevant to the objects in the poem. The meta-path representation allows both the images and poems to be presented as a collection of heterogeneous entities, making it possible to compare their similarity for recommendation.

2) Meta-path Embedding via Stacked Autoencoder: We define the meta-path collection for an image as $\mathbf{\Omega^I}$ 2 $\mathbb{R}^{M \ N}$, where M is the total number of meta-paths used to represent image I, and N is the maximum length of the path. Similarly, we define $\mathbf{\Omega^P}$ 2 $\mathbb{R}^{M \ N}$ as the meta-path collection of a poem P. Given the two meta-path collections $\mathbf{\Omega^I}$ and $\mathbf{\Omega^P}$, the problem is to find the similarity between them - the poems with the highest similarity of the input image will be recommended. However, it is not a trivial task to find the similarity between the meta-paths of the image and poem considering the potential high dimensions of $\mathbf{\Omega^I}$ and $\mathbf{\Omega^P}$ and the heterogeneous entities in the meta-path collections.

To address these challenges, we map the meta-path collections of both images and poems into a homogeneous latent feature subspace with much lower dimensions. In particular, we embed each meta-path using the auto-encoding technique [29]. It consists of an *encoder* that maps an input vector $\mathbf{X}$ into a latent subspace $\mathbf{Z}$ and a *decoder* that uses the latent representation $\mathbf{Z}$ to recover the original input. In iPoemRec, we develop a 6-layer stacked autoencoder (denoted as SAE). The representation of the l^{th} layer of the SAE_{em} is defined as $\mathbf{Z}^l = f(\mathbf{W}^l \ \mathbf{X}^l + \mathbf{b}^l)$, where $f(.)$ is a rectified linear unit (ReLU) activation function. $\mathbf{W}^l$ and $\mathbf{b}^l$ are weighting factor and bias of the l^{th} layer. $\mathbf{X}^l$ denotes the input to the l^{th} layer which is the latent feature of the previous layer (i.e., $\mathbf{Z}^{l \ 1}$).

To train the autoencoder and derive the latent representation for Ω, we define a customized feature reconstruction loss function L to explicitly consider the heterogeneity of different types of entities in the meta-paths.

$$L = jj(\Omega \quad \widehat{\Omega}) \quad \alpha jj_2^2 \tag{2}$$

where denotes the Hadamard product and $\widehat{\Omega}$ is the reconstructed feature vector. $\alpha_{m,n}$ is defined as:

$$\alpha = \begin{cases} \lambda_S, \ \Omega_{m,n} \ 2 \ \mathbf{V_S}, \ 1 \quad m \quad M,1 \quad n \quad N \\ \lambda_T, \ \Omega_{m,n} \ 2 \ \mathbf{V_T}, \ 1 \quad m \quad M,1 \quad n \quad N \\ \delta_{m,n}, \ \Omega_{m,n} \ 2 \ \mathbf{V_A}, \ 1 \quad m \quad M,1 \quad n \quad N \\ 1, \ otherwise \end{cases} \tag{3}$$

where $\delta_{m,n}$ is the metaphorical degree of an ANP entity. $\lambda_S > 1$ and $\lambda_T > 1$ are tunable parameters for the model. The $\delta_{m,n}$, λ_S, and λ_T factors assign different penalties to the reconstruction error of ANP, sentiment, and theme entities respectively. By tuning $\delta_{m,n}$, λ_S and λ_T, the SAE can be guided to emphasize more on one entity over another.

The stacked autoencoders are trained by minimizing L via layer-wise pre-training [30]. The encoded results of $\mathbf{\Omega^I_m}$ and $\mathbf{\Omega^P_m}$ are denoted as $\mathbf{Z^I_m}$, and $\mathbf{Z^P_m}$ respectively. After training the SAE, we can learn the latent representation of a poem and an image by aggregating the meta-path embeddings. The image and poems are represented in the same latent subspace after the above neural embedding process. We then use the cosine distance to measure the similarity between image embedding and poem embedding and recommend the poems with the highest cosine similarity [24]. The iPeomRec algorithm is summarized in Algorithm 1.

Algorithm 1 iPoemRec Scheme

Input: an image I, poems $\mathcal{P}$, recommendation list size K
Output: recommended poems $P_1, P_2, ..., P_K$
for all $P \in \mathcal{P}$ **do** ▷ **Offline Phase**
  extract sentiment, theme, and ANPs of P
end for
construct CaHIN from poetry corpus
train stacked auto encoder SAE for path embedding
for all $P \in \mathcal{P}$ **do**
  derive poem embedding $E(P)$
end for
extract ANPs from I ▷ **Online Phase**
initialize Z =[], RL = []
for all $ANP \in ANPs$ **do**
  while ANP $\notin$ CaHIN **do**
    convert ANP using stemming and transformation
  end while
  Extract meta-path Ω, calculate meta-path embedding Z', $Z \leftarrow Z'$
end for
Derive image embedding E_I from Z
for all $P \in \mathcal{P}$ **do**
  calculate sim(E_P,E_I), $RL \leftarrow sim(E_P, E_I)$
end for
Return top K largest element in RL

V. Evaluation

In this section, we present an extensive evaluation of our iPoemRec scheme. We first discuss the evaluation setup and baselines for comparison. We then present the evaluation results using real-world case studies. The results show that iPoemRec achieves significant performance gains in terms of relevance and user ratings of the recommended poems compared to the state-of-the-art baselines.

A. Experiment Setup

We use a dataset of classical Chinese poetry from Tang Dynasty which consists of a total of 870 poems [31]. We select four representative baselines from recent literature.

Word2Vec: a poem recommender that uses Word2Vec model to recommend Chinese poetry based on a list of keywords as input [8]. We use YOLO V3 object detector to extract these keywords.

Image2Caption: an image caption generation scheme that generates a short text descriptor for an image [14]. The text descriptor is then used to recommend poems based on the text similarity [24].

SCAN: a matching algorithm that develops a Stacked Cross Attention to perform semantic matching of objects in texts and images [13]. The poems are ranked based on the semantic matching score.

SentiBank: an image sentiment analyzer that generates sentiments from an image [27]. We match the generated sentiments with the sentiments of the poems to perform recommendation.

Please note that none of the above baseline address the exact problem of iPoemRec. Therefore, we have to modify these algorithms in order to compare with our scheme. We use a tuning set of 10 images and 50 poems with ground truth labels to tune the model parameters. We set N = 5, M = 50, λ_S = 2.5, λ_T = 1.2, and W_{thres} = 0.3 for our system.

Our evaluation consists of two stages. The first stage is a *crowdsourcing based evaluation* and the second is a *real world user study*. We elaborate these experiments below.

B. Crowdsourcing-based Evaluation

In the the crowdsourcing-based evaluation, we collect a test set of 100 landscape images from Instagram and use them as inputs to all compared schemes. The recommended list of poems together with each image are submitted as surveys with incentives to a Chinese crowdsourcing platform Wenjuan Xing [32] to obtain the user feedback. In our experiment, each image is recommended with 5 poems. The recommended poems of each image are sent to 3 different users to obtain the *relevance* and *rating* score of the recommended poems. The relevance score is a binary response denoting whether the user believes the recommended poem is relevant (score of 1) or irrelevant (score of 0). The rating score is the user's overall judgement of the recommended poem, which is represented as a integer value from 0 to 5, with 5 representing the best recommendation and 0 representing the worst.

1) Results - Relevance: The relevance of a retrieved document to an input sample is often evaluated using the Precision@K metric defined below.

DEFINITION 6. ***Precision (Pre@K)****: the percentage of relevant poems in the top K recommendation list. In particular,* $Pre@K = \frac{\{Recommended\ Poems\} \setminus \{Relevant\ Poems\}}{K}$.

Note that we did not include the $Recall@K$ that is often used in IR literature because it requires the labeling of all the relevant poems to each image, which causes excessive labeling cost [33], [34] and generates a non-trivial amount of tedious work on the crowdsourcing platform. The results are presented in Figure 5. We can observe that iPoemRec consistently outperforms all baselines. In particular, iPoemRec achieved 8%, 9%, 11%, 16%, and 19% performance gain compared to best-performing baseline when K changes from 1 to 5, respectively. This suggests that the poems recommended by iPoemRec are more relevant than the ones recommended by the baseline systems. We also observe that the performance gain of iPoemRec becomes more significant when K is large. This is because the baseline methods based on the object similarity (i.e., Word2Vec, Image2Caption, SCAN) can recommend poems with mismatched sentiments or themes for a given image. SentiBank, purely focused on the sentiment similarity, can lead to recommendations with mismatched objects and themes. In contrast, iPoemRec is artistic conception aware and holistically considers the object, sentiment and theme similarities between images and poems in its recommendation.

2) Results - Overall Rating: We then evaluate the performance of overall ratings of all compared schemes. We use two popular metric *Mean Overall Rating (MOR@K)* and *Normalized Discounted Cumulative Gain (NDCG@K)* in recommender systems. We present the MOR@K results in Figures 6. We can observe that iPoemRec achieves significantly higher ratings than the baselines. We attribute the performance gain to the explicit modeling of artistic conception in our system,

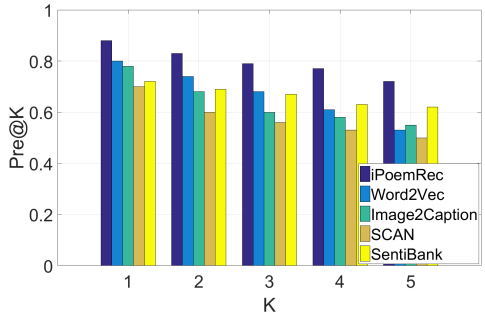


Figure 5: Precision@K for all Schemes

which ensures the recommended poems have a high poetic value. The diversity of the recommended poems also help to satisfy users' diverse perceptions of the images. In contrast, baselines often provide mundane recommendations based on a single criteria - either maximizing the similarity between objects or the similarity between sentiments. The results of NDCG@K in Figure 7 have demonstrated better consistency in the order of the recommended list of iPoemRec. Note that we do not show $NDCG@1$ because the NDCG always has a score of 1 if only one item is recommended.

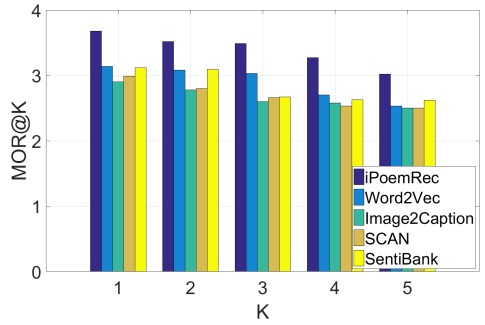


Figure 6: MOR@K for all Schemes

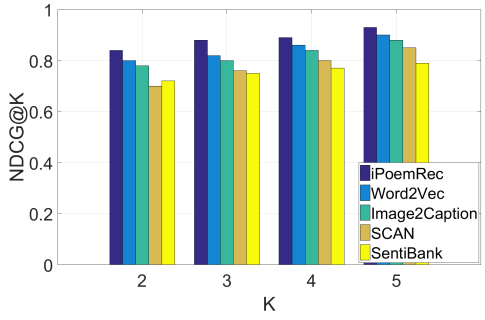


Figure 7: NDCG@K for all Schemes

C. Online User Study Evaluation

We perform an online user study to further evaluate the performance of iPoemRec and baselines. The reason for doing this case study on real world users is multi-fold: 1) it gives more insight of the real-world experience of users who are interested in using the application; 2) the images uploaded by these users are provided by themselves (instead of the chosen test set), which can help evaluate the performance of systems on diverse inputs; 3) these users are not provided with any incentives (except for the free usage of our app) and are expected to provide more genuine feedback. In our experiment, we develop a set of public web tools using Flask to perform poetry recommendation to users. Each web tool uses one of the compared schemes as the backend recommendation algorithm and we posted the web tool links to social media sites (Facebook, Twitter, Instagram) to attract online users to test them and provide feedback (i.e., relevance and rating scores). In this experiment, we only recommend 3 poems because we found real-world users could quickly become impatient if we recommend too many poems for them to rate.

We have collected a total of 297 responses from anonymized users (59 responses on average for each scheme). The results are presented in Table I. We observe iPoemRec continues to outperform baselines in all compared metrics. It improves the best-performing baseline by 8.5% and 13.2% in terms of Pre@1 and Pre@3 respectively. The performance gains are more significant compared the ones from the crowdsourcing-based evaluation. The reason is the users upload their own pictures containing objects that are not commonly seen in the classical Chinese poetry (e.g., portraits, cars), which lead to the degraded performance of object matching based schemes such as Word2Vec. In contrast, iPoemRec is robust to diverse user inputs as its PVA module extracts only the salient objects in the image for recommendation. The overall rating score of iPoemRec continues to outperform baselines as well. This indicates the our scheme did bring aesthetic satisfaction and appreciation to users in a real-world application scenario.

Table I: Evaluation Results for an Online User Study

	Pre@1	Pre@3	MOR@1	MOR@3	NDCG@3
iPoemRec	**0.881**	**0.855**	**3.513**	**3.421**	**0.816**
Word2Vec	0.796	0.722	3.074	2.907	0.778
Image2Caption	0.785	0.714	2.696	2.536	0.750
SCAN	0.714	0.698	2.444	2.127	0.698
SentiBank	0.792	0.708	3.145	3.021	0.729

D. Generality Analysis

Finally, we evaluate the performance of iPoemRec when it is applied to classical poetry in English culture to study the generality of our model. We use a Shakespeare dataset which contains 140 sonnets [35]. The results are reported in in Table II. While iPoemRec is designed for the classical poetry with rich artistic conception (e.g., Chinese poetry), it also performs well for English classical poetry and outperforms all the compared baselines. We observe that the performance of all schemes improve in this experiment. We attribute this observation to the fact that the artistic conception in English poetry is often more explicit than Chinese poetry. In particular, the classical Chinese poetry heavily employs metaphors and symbolism to deliver its artistic conception [36], which is much more difficult to model.

VI. Conclusion

In this paper, we develop iPoemRec, the first classical poetry recommender system using visual inputs. The iPoemRec

Table II: Evaluation Results for Classical English Poetry

	Pre@1	Pre@3	MOR@1	MOR@3	NDCG@3
iPoemRec	**0.894**	**0.848**	**3.470**	**3.212**	**0.864**
Word2Vec	0.846	0.807	3.211	3.019	0.827
Image2Caption	0.805	0.787	2.914	2.702	0.787
SCAN	0.791	0.729	2.812	2.396	0.750
SentiBank	0.800	0.781	3.254	3.127	0.763

system can explicitly model the artistic conception of poems and images and effectively recommend poetry that matches the sentiment and theme of the image. Using real-world datasets and a user study, we have demonstrated that iPoemRec can recommend classical poems to users with high relevance and receive significantly higher user ratings than the state-of-the-art baselines. We envision iPoemRec will also enable many interesting and useful applications in the future (e.g., image-based poetry search engines, smart classical poetry education apps for students).

ACKNOWLEDGMENT

This research is supported in part by the National Science Foundation under Grant No. CNS-1831669, CBET-1637251, Army Research Office under Grant W911NF-17-1-0409. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "The age of social sensing," *Computer*, vol. 52, no. 1, pp. 36–45, 2019.

[2] D. Zhang, N. Vance, and D. Wang, "When social sensing meets edge computing: Vision and challenges," *arXiv preprint arXiv:1905.07528*, 2019.

[3] D. Wang, T. Abdelzaher, and L. Kaplan, *Social sensing: building reliable systems on unreliable data*. Morgan Kaufmann, 2015.

[4] L. Xu, L. Jiang, C. Qin, Z. Wang, and D. Du, "How images inspire poems: Generating classical chinese poetry from images with memory networks," in *32 AAAI Conference on Artificial Intelligence*, 2018.

[5] J. Gu, J. Cai, S. R. Joty, L. Niu, and G. Wang, "Look, imagine and match: Improving textual-visual cross-modal retrieval with generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7181–7189.

[6] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Hierarchical multimodal lstm for dense visual-semantic embedding," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[7] L. Liang and R. Chen, "Realization of figure-ground in tang poems and its effect on artistic conception," *Journal of Foreign Languages*, 2008.

[8] Z. Liu and H. Huang, "Research on personalized recommendation of ancient poetry based on word2vec model," 2018.

[9] Y. Jing and S. Baluja, "Visualrank: Applying pagerank to large-scale image search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1877–1890, 2008.

[10] C. Frankel, M. J. Swain, and V. Athitsos, "Webseer: An image search engine for the world wide web," 1996.

[11] H. Neven Sr and H. Neven, "Image-based search engine for mobile phones with camera," Jul. 21 2009, uS Patent 7,565,139.

[12] D. Y. Zhang, Y. Zhang, Q. Li, T. Plummer, and D. Wang, "Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications," in *Distributed Computing Systems (ICDCS), 2019 IEEE 39th International Conference on*. IEEE, 2019.

[13] K.-H. Lee, X. Chen, G. Hua, H. Hu, and X. He, "Stacked cross attention for image-text matching," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 201–216.

[14] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, "Image captioning with semantic attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4651–4659.

[15] J. Krause, J. Johnson, R. Krishna, and L. Fei-Fei, "A hierarchical approach for generating descriptive image paragraphs," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2017.

[16] W.-F. Cheng, C.-C. Wu, R. Song, J. Fu, X. Xie, and J.-Y. Nie, "Image inspired poetry generation in xiaoice," *arXiv:1808.03090*, 2018.

[17] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Twenty-Eighth AAAI conference on artificial intelligence*, 2014.

[18] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1026–1037, 2013.

[19] J. Marshall and D. Wang, "Mood-sensitive truth discovery for reliable recommendation systems in social sensing," in *Proceedings of International Conference on Recommender Systems (Recsys)*. ACM, 2016, pp. 167–174.

[20] D. Wang, M. T. Al Amin, T. Abdelzaher, D. Roth, C. R. Voss, L. M. Kaplan, S. Tratz, J. Laoudi, and D. Briesch, "Provenance-assisted classification in social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 624–637, 2014.

[21] S. Oramas, V. C. Ostuni, T. D. Noia, X. Serra, and E. D. Sciascio, "Sound and music recommendation with knowledge graphs," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, p. 21, 2017.

[22] D. Y. Zhang, Q. Li, H. Tong, J. Badilla, Y. Zhang, and D. Wang, "Crowdsourcing-based copyright infringement detection in live video streams," in *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2018*, 2018.

[23] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua, "Explainable reasoning over knowledge graphs for recommendation," *arXiv preprint arXiv:1811.04540*, 2018.

[24] T. Kenter and M. De Rijke, "Short text similarity with word embeddings," in *Proceedings of the 24th ACM international on conference on information and knowledge management*. ACM, 2015, pp. 1411–1420.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[27] D. Borth, T. Chen, R. Ji, and S.-F. Chang, "Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 459–460.

[28] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016.

[29] D. Y. Zhang, L. Shang, B. Geng, S. Lai, K. Li, H. Zhu, M. T. Amin, and D. Wang, "Fauxbuster: A content-free fauxtography detector using social media comments," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 891–900.

[30] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in neural information processing systems*, 2007, pp. 153–160.

[31] "Chinese classical poetry dataset," http://www.shigeku.org/xlib/lingshidao/hanshi/index.html, accessed: 2019-04-23.

[32] "Wenjuan xing," https://www.wjx.cn/, accessed: 2019-04-23.

[33] T. Saracevic, "Evaluation of evaluation in information retrieval," in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. Citeseer, 1995.

[34] D. Y. Zhang, D. Wang, H. Zheng, X. Mu, Q. Li, and Y. Zhang, "Large-scale point-of-interest category prediction using natural language processing models," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 1027–1032.

[35] "shakespeare sonnets dataset," https://www.opensourceshakespeare.org/views/sonnets/sonnets.php, accessed: 2019-04-23.

[36] C. C.-c. Sun, "Mimesis and xing: Two modes of viewing reality comparing english and chinese poetry," *Comparative Literature Studies*, vol. 43, no. 3, pp. 326–354, 2006.