Pivot-Based Bilingual Dictionary Extraction from Multiple Dictionary Resources

Mairidan Wushouer[1], Donghui Lin[1], Toru Ishida[1], and Katsutoshi Hirayama[2]

[1] Department of Social Informatics, Kyoto University
[2] Graduate School of Maritime Sciences, Kobe University
mardan@ai.soc.i.kyoto-u.ac.jp, lindh,ishida@i.kyoto-u.ac.jp,
hirayama@maritime.kobe-u.ac.jp

Abstract. High quality bilingual dictionaries are rarely available for lower-density language pairs, especially for those that are closely related. Using a third language as a pivot to link two other languages is a well-known solution, and usually requires only two input bilingual dictionaries to automatically induce the new one. This approach, however, produces many incorrect translation pairs because the dictionary entries are normally are not transitive due to polysemy and the ambiguous words in the pivot language. Utilizing the complete structures of the input bilingual dictionaries positively influences the result since dropped meanings can be countered. Moreover, an additional input dictionary may provide more complete information for calculating the semantic distance between word senses which is key to suppressing wrong sense matches. This paper proposes an extended constraint optimization model to inducing new dictionaries of closely related languages from multiple input dictionaries, and its formalization based on Integer Linear Programming. Evaluations indicated that the proposal not only outperforms the baseline method, but also shows improvements in performance and scalability as more dictionaries are utilized.

Keywords: Bilingual Dictionary Induction, Pseudo-Boolean Optimization, Constraint Satisfaction.

1 Introduction

Bilingual dictionaries are essential for many tasks in Natural Language Processing (NLP), such as machine translation [1] and cross-lingual information retrieval [2]. However, high quality dictionaries remain sparse, dated, or simply unavailable for less-resourced language pairs such as many Turkic languages. Therefore, researchers have investigated the issue of automatic creation. Solutions include the induction of a dictionary from a large-scale parallel corpora [3]. More recently, the use of comparable corpora has drawn increasing attention [4–6] since the Internet era has made monolingual data readily available while parallel corpora remain scarce.

When it comes to less-resource demanding approaches, pivot-based induction, which creates a new dictionary from two others, is fundamental, since a popular language used as a bridge between two inadequately-resourced language

D.-N. Pham and S.-B. Park (Eds.): PRICAI 2014, LNAI 8862, pp. 221–234, 2014.

pairs. A pair of words which have common translation in the pivot language are considered to be a correct translation pair. However, such an implementation yields an extremely noisy dictionary containing incorrect translation pairs as the lexicons are generally intransitive. This intransitivity stems from polysemy and ambiguous words in the pivot language. In order to solve this issue of divergence, previous studies attempted to select correct translation pairs by using semantic distances from the structures of the input dictionaries [7][8] or by using additional resources such as part of speech [9], WordNet [10], comparable corpora [11] and descriptions in the dictionary entities [12].

There has been growing interest in using the constraint optimization problem formalism for ideally describing and solving many problems in NLP and Web Service Composition [13][14][8], because these problems involve combinatorial issues that can be represented by a set of variables connected by constraints. In this paper, we propose an enhanced constraint optimization model for pivot-based dictionary induction where multiple bilingual dictionaries of a pivot language can be utilized for result enhancement by making use of the most complete parts of each input dictionary in measuring semantic distances. More precisely, we try to obtain semantic distance by constraining the types of connection in the structures of the multiple input dictionaries. Furthermore, Instances of this induction with at least two input dictionaries are represented by graphs, to which weighted edges are added to represent missing translations. In this context, a graph is modeled as an optimization problem where we maximize the coverage of the output dictionaries by adding highly probable missing edges. This optimization problem is formulated within the Pseudo-Boolean optimization framework[15] (0-1 Integer Linear Programming, or 0-1 ILP).

We designed a tool to implement the proposal by integrating an ILP solver and evaluated our approach by constructing a group of dictionaries of Turkic languages. The evaluation revealed the efficiency of our proposal.

2 Related Work

A very early attempt to create new dictionaries from existing dictionaries was made by Tanaka [7], who used a pivot language. They used Inverse Consultation (IC) to tackle lexical intransitivity divergence. IC tries to measure the intersection of two pivot word sets: the set of pivot translations of a word w in language A, and the set of pivot translations of each word in language C, candidates for a translation of w. The number of elements in the intersection indicates the nearness of the original word to its candidate. Variations of this method have also been proposed [10][16][17][18], where extra language resources are utilized such as WordNet. A weakness of the IC method is that it relies on synonymous words to identify correct translations, which may result in low precision if synonym-poor or heavily incomplete dictionaries are used. This weakness has been addressed to some degree in a recent work[8] that takes account of data incompleteness; it reported slightly improved precision without requiring an additional resource.

With the assumption that more pivot languages could provide extra information for better evaluating the semantic distance of cross-lingual word pairs, one method uses multiple input dictionaries [19]. This idea is similar to IC method, but its use of multiple pivot languages eliminates its dependency on synonym-rich input dictionaries to some extent. However, the new problem is the need to find suitable multiple input dictionaries. Another approach to creating a dictionary without using pivot language are extraction from parallel-corpora[20] and monolingual corpora [4][21]. This approach usually demands large amount of data in order to produce a high quality dictionary. For closely related-language pairs, using spelling similarity as an additional heuristic is also common although its result largely depend on language pairs closeness [22].

Large sets of language resources such as parallel corpora or monolingual corpora are available for some language pairs, so that many existing methods for creating dictionary can take advantage of them in acquiring word sense and evaluating, with high precise, the semantic distance between cross-lingual word pairs. However, we consider that basic methods that work with the inherent structures of input dictionaries must be explored because it is useful for inadequately-resourced languages. In our work we focus on creating dictionaries from existing dictionaries, since there still many world languages that lack useful languages resources. However, Our approach, modeling complete structures of multiple input dictionaries which is an extension of[8], can handle the incompleteness of input dictionaries to some extent, and performs better in the case of closely related languages. Also, introducing more exiting dictionaries with the same pivot language into the induction problem instances is a new solution and is promising for quality improvement.

3 Using a Pivot Language

Using a pivot language to connect two others is commonly used in research on machine translation [23], and service computing [24]. Let l be a language and $D_{l_i-l_j}$ be a dictionary (non-directional) of two languages, l_i and l_j. In the context of dictionary creation, the basic idea of using a pivot language is to induce a new dictionary $D_{l_1-l_2}$ from existing $D_{l_1-l_0}$ and $D_{l_0-l_2}$ by extracting pairs of words in l_1 and l_2 languages that have the same translation in pivot language l_0. However, since word sense is not considered, $D_{l_1-l_2}$ is likely include many incorrect translation pairs. Take Uyghur-English-Kazakh as an example. The English word *tear* is the translation of Uyghur word *yash*, but only in the sense of liquid from the eyes. Further translating *tear* into Kazakh yields both the correct translation *jash* and an incorrect one, *jirtiw* (to rip). The reason for this happing is that the pivot word is polysemous or ambiguous, a common situation in existing dictionaries. It is challenging to identify such incorrect translations, because, unfortunately, most dictionaries lack adequate sense information in their entries. Therefore, it is not possible to map translation equivalents according to their senses. Most previous studies try to guide this mapping by using the semantic distances extracted from the dictionaries themselves or external resources.

Merging $D_{l_1-l_0}$ and $D_{l_0-l_2}$ via l_0 forms a big graph whose vertex is a word and edge is the indication of having common meaning between two endpoint words.

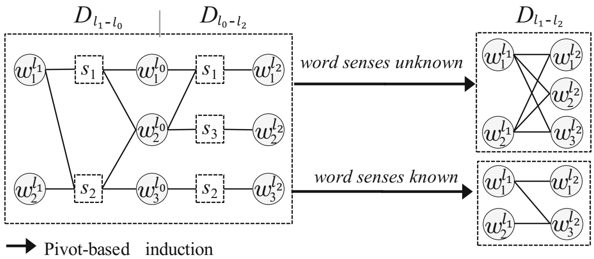

Fig. 1. Illustration of Pivot-based bilingual dictionary induction and ambiguity problem: incorrect translation pairs cannot be detected when word sense is unknown

Such graph has at least one connected component[1], which is called *transgraph* and is defined as follows [19][8]:

Definition 1. A *transgraph is defined as an undirected graph $G = \{V, E\}$, in which vertex $w_i^l \in V$ is a word in a language l, and an edge $e(w_i^{l_1}, w_j^{l_2}) \in E$, represents the belief that $w_i^{l_1}$ and $w_j^{l_2}$ have at least one meaning in common.* We also use $V^l \subset V$ to denote the set of words in language l.

4 Proposal

4.1 One-to-one Mapping of Lexicons

The lexicons of closely related languages (intra-family languages) are similar. Moreover, they share a significant number of cognates – words that are derived from same origin and are similar in both spelling and meaning (e.g. neveu [Fr.] and nephew [Eng.]). Most cognate pairs have direct translations[2], meaning that they correspond one-to-one. A classical lexicostatistical study of 15 Turkic languages[3] indicated that cognate pairs shared among its members scale from 44% to 94% of their lexicons, and majority of non-cognates tend to be noun.

Taking account such facts, an assumption of one-to-one mapping in lexicons of intra-family languages has been adopted by some studies on bilingual dictionary creation[8] [26][27]. The common assertion is that that if l_1 and l_2 are intra-family, for any $w_i^{l_1}$ there exists a unique $w_j^{l_2}$, such that they have exactly the same meaning(s). Such a pair is called one-to-one pair; the term one-to-one pair candidate refers a pair of words which have yet to be confirmed as a one-to-one pair. In this work, we also utilize this one-to-one assumption as a constraint for the output dictionary.

We should note that mappings between languages are nondeterministic[4], as words can have multiple translations, so the one-to-one assumption based on

[1] http://en.wikipedia.org/wiki/Connected_component_(graph_theory)

[2] A direct translation [25] is an association between two words where this two words are indeed mutual translations.

[3] http://turkic-languages.scienceontheweb.net

deterministic mapping may be too strong for the general case. However, it is considered reasonable for the case of intra-family languages. Moreover, having a one-to-one assumption has been evaluated in the literatures. For example, Melamed et al. [26] made a similar assumption for any kind of language pair when they tried to create a word-to-word model of translation. They claimed that such a constraint strengthens the explanatory power of their model in comparison to the IBM models, and also helps them to avoid indirect associations, a major source of errors in translation models. Koehn et al. [27] also used similar assumption when they extracted an English-German dictionary from monolingual corpora. In conclusion, the goal making this one-to-one assumption is not to create dictionaries with word-to-word mapping but accurate dictionaries of closely related languages (such as Turkic languages), so that the result might be of use in translation systems like Apertium [4], which uses dictionaries and manual rules to translate between related languages.

This one-to-one assumption is realized as constraints which are imposed on pair ($w_i^{l_1}$, $w_j^{l_2}$) in a *transgraph*; all constraints must be satisfied before the pair is recognized as a one-to-one pair. Next section details the constraints.

4.2 Defining Constraints

In a *transgraph*, the one-to-one assumption can be realized with two constraints, one of which demands symmetric connection of a one-to-one pair while the another guarantees its uniqueness[8]. In addition, selecting candidates of one-to-one pairs also can be seen as constraint, and is independently defined in this paper. As the initial step of pivot-based techniques, the possible translation pairs are selected to generate a noisy $D_{l_1-l_2}$ based on the structures of the input dictionaries. In this dictionary, whether word pair $(w_i^{l_1}, w_j^{l_2})$ is a one-to-one pair candidate is decided by the following constraint.

Candidate Existence: A pair of words, $w_i^{l_1}$ and $w_j^{l_2}$, in a *transgraph*, is a one-to-one pair candidate iff they are connected via at least one pivot word. That is, a word pair is taken to be a candidate and subjected to further evaluation only if they share at least one translation in the pivot language.

Moreover, A one-to-one pair is a pair of words that carry exactly same meanings. This defines another constraint as follows:

Symmetry Constraint: Given a pair of words, $w_i^{l_1}$ and $w_j^{l_2}$, if they are a one-to-one pair, then they should be symmetrically connected through pivot word(s). In other words, a one-to-one pair share the same translations in pivot language; the number of edges between $w_i^{l_1}$ and pivot words should equal the number of edges between $w_j^{l_2}$ and pivot words. For example, in Fig. 1, if $(w_1^{l_1}, w_1^{l_2})$ is a one-to-one pair, then 6 edges: $e(w_1^{l_1}, w_1^{l_0})$, $e(w_1^{l_1}, w_2^{l_0})$, $e(w_1^{l_1}, w_3^{l_0})$, $e(w_1^{l_2}, w_1^{l_0})$, $e(w_1^{l_2}, w_2^{l_0})$ and $e(w_1^{l_2}, w_3^{l_0})$ exist in the *transgraph*, where $e(w_1^{l_2}, w_3^{l_0})$ is not present. Such an edge considered to be possibly missing, meaning that the corresponding translation might not have been included in the input dictionary when it was built. Note that it is possible that synonymous words in a language can share exactly same

[4] An open-source machine translation platform at `http://www.apertium.org`

meaning(s) and can be used as alternates in the translation. Such synonymous words apparently can be one-to-one translation equivalent to the same word, but since the goal is to find a single equivalent under the one-to-one assumption, the following constraint prevents the selection of multiple equivalents.

Uniqueness Constraint: Given a pair of words, $w_i^{l_1}$ and $w_j^{l_2}$, in a *transgraph*, if they are a one-to-one pair, then they should be unique, such that all other candidates involving $w_i^{l_1}$ or $w_j^{l_2}$ are not one-to-one pairs. For example, in Fig. 1, if $(w_1^{l_1}, w_1^{l_2})$ is a one-to-one pair, then we assert that $(w_1^{l_1}, w_2^{l_2})$, $(w_1^{l_1}, w_3^{l_2})$ and $(w_2^{l_1}, w_1^{l_2})$ are not one-to-one pairs.

4.3 Missing Edges

In most cases of induction, input dictionaries may be incomplete so that the result would be negatively affected. For example, 1) a pivot word is missing so that some translation pair for $D_{l_1-l_2}$ cannot be identified, (2) a non-pivot word is missing, or (3) a pair $(w_i^{l_1}, w_j^{l_0})$ or $(w_i^{l_2}, w_j^{l_1})$ is missing (an edge is missing in a *transgraph*). Apparently, first two cases cannot be resolved without additional resources. The third one, however, is vital because any possible missing edge would have been part of a symmetric connection between $w_i^{l_1}$ and $w_j^{l_2}$ which is key to detecting $(w_i^{l_1}, w_j^{l_2})$ as a one-to-one pair. This could harm the quality of induction. However, missing edges are hard to avoid since the input dictionaries are usually independently created, and their completeness is seldom guaranteed. However, adding missing edges into the *transgraph* makes it complete, and, thus, makes induction more accurate. We assign a weight to missing edge e, indicating the likelihood of it being missed. Therefore, a weight matrix needs to be generated for all the one-to-one pair candidates in order to have a full list of possible missing edges, each with a weight representing its chance of being missed.

There are different possible ways to obtain such weights[28], especially if external resources such as parallel corpora are available. In this work, a simple statistical method [29][8] is used for the sake of simplicity. However, one can extend the calculation by adopting different formula or even using external knowledge to gain more accurate weights.

4.4 Utilizing Additional Dictionary

When there is a third dictionary $d_{l_3-l_0}$ available in addition to $d_{l_1-l_0}$ and $d_{l_2-l_0}$, where l_1, l_2 and l_3 are closely related languages, adding it to existing *transgraph*s may introduce more complete semantic information that could ultimately boost the accuracy of induction result (multiple output dictionaries). This is because the number of meanings of a given pivot word in each input dictionary might depend on its completeness. Therefore, taking advantage of the most compete part of each dictionary is reasonable. In this work, we conclude the effect of an additional dictionary to the induction with only two input dictionaries as follows:

1. A more accurate weight of a possibly missing edge can be obtained by taking the maximum of the weights from each from each combination of input dictionaries. For example, in *transgraph-a* in Fig.2, edges $e(w_1^{l_0}, w_2^{l_2})$ and $e(w_2^{l_0}, w_1^{l_2})$ have the same weight (=0.50), so that it is impossible to select one of the $(w_1^{l_1}, w_1^{l_2})$ and $(w_1^{l_1}, w_2^{l_2})$ as a one-to-one pair with higher confidence. But when *transgraph-b* is formed due to the additional input dictionary $d_{l_3-l_0}$, the weights of the two edges can be recalculated for each pair-combination of the three input dictionaries. In this example, $(w_1^{l_1}, w_1^{l_2})$ secures a higher value (=0.66), which is then propagated to the *transgraph-a*.
2. A new constraint – one-to-one pairs among intra-family language pairs must be consistent – needs to be imposed, which might contribute to the accuracy of the output dictionaries. For example, given three words of three intra-family languages: w^{l_i} and w^{l_j} and w^{l_k}, which can form three word pairs. If any two of these three pairs are one-to-one pairs, then the third must also be one-to-one pair. This can prevent the false associations during the optimization process to some extent.

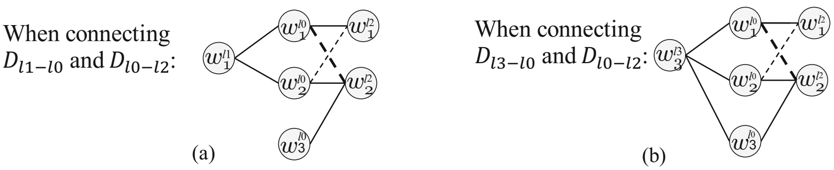

Fig. 2. Difference weights can be obtained for an edge from different combinations of input dictionaries; for the same edge $e(w_1^{l_0}, w_2^{l_2})$ two different weights are obtained

4.5 Optimization Model

We allow a possible missing edge to be added to the *transgraph* if it has non-zero weight of having been missed. If it is added, then a certain cost (equals to 1-weight) is to be paid. The process of extracting one-to-one pairs from a *transgraph* is defined as an optimization problem; the objective is to extract as many one-to-one pairs as possible while minimizing the cost of edge addition, where cost is defined as the chance that an edge does not exist (or turns out to be not missing).

Let variables $x, e \in \{0, 1\}$ denote a one-to-one pair candidate and an edge in the *transgraph*, respectively:

- $x_{w_i^{l_1} w_j^{l_2}}$, representing word pair $(w_i^{l_1}, w_j^{l_2})$, takes 1 if it is one-to-one pair; 0 otherwise.
- $e_{w_i^{l_k} w_j^{l_0}}$, representing edge $(w_i^{l_k}, w_j^{l_0})$, where $k \geq 1$, takes 1 if it must exist; 0 otherwise.
- $\omega_{w_i^{l_k} w_j^{l_0}}$, representing the weight of edge $(w_i^{l_k}, w_j^{l_0})$, whose domain is $[0, 1]$.

- X, the set of x variables representing the one-to-one pair candidate space of the *transgraph*.
- $E^{l_i l_0}$, representing the edge space of the *transgraph* for l_i and pivot language l_0, whose domain is $[0, 1]$.

The objective function can be formulated as follows:

$$\Omega = \mu_1[\sum_{x_{w^{l_1} w^{l_2}} \in X} x_{w^{l_1} w^{l_2}}] - \mu_2[\sum_{e_{w^{l_i} w^{l_0}} \in E^{l_i l_0}} (1 - \omega_{w^{l_i} w^{l_0}}) \cdot e_{w^{l_i} w^{l_0}}] \tag{1}$$

where the first segment corresponds to the objective to maximize the coverage of output dictionary, while the latter grantees minimization of the cost of edge addition; their subtraction normalizes the multi-objectives into a single maximization. Moreover, coefficients μ_1 and μ_2 can be used to control precision and recall of extracted one-to-one pairs to some extent. However, in this work, we consider only the case that they are equally treated ($\mu_1 = \mu_2 = 0.5$)

With this objective function in mind, the dictionary induction problem S can be formulated as

$$S = \text{argmax } \Omega \tag{2}$$

which subjects to *Symmetry* and *Uniqueness* constraints. We use an optimization solver to generate the optimally correct one-to-one pair set. In the next section, we will describe how we formalize this problem within the 0-1 ILP framework [5], and use a state-of-the-art solver to generate one-to-one pairs as the output dictionaries.

5 0-1 ILP-Based Modeling

5.1 Preliminaries

The Pseudo-Boolean Optimization (PBO) problem, also known as 0-1 ILP, is an of Boolean Satisfiability where constraints can be any linear inequality with integer coefficients (also known as Pseudo-Boolean constraints, or just PB) defined over the set of problem variables. The objective in PBO is to find an assignment to problem variables such that all problem constraints are satisfied and the value of a linear objective function is optimized. A Pseudo-Boolean (PB) constraint is defined over a finite set of Boolean variables x_i and has the form $\sum_i \omega_i x_i \triangleright k$ where ω_i (called weights) and k are integers, $\triangleright$ is one of the following classical relational operations $=$, $>$, $<$, $\geq$ or $\leq$, and $1 \leq i \leq 1$, where n is the number of variables in the PB constraint.

[5] For an overview and example of Integer Linear Programming, refer to [30].

5.2 Modeling

As for the *Symmetry* Constraint: For any $(w_i^{l_1}, w_j^{l_2})$ where $w_i^{l_1}$ and $w_j^{l_2}$ have $V_i^{l_0}$ and $V_j^{l_0}$ preexisting meanings in a *transgraph*, respectively. If it is a one-to-one pair, then either $w_i^{l_1}$ or $w_j^{l_2}$ have the same meanings $V_{i,j}^{l_0} = V_i^{l_0} \cup V_j^{l_0}$. This is expressed by the following inequality for given one-to-one pair candidate $(w_i^{l_1}, w_j^{l_2})$:

$$\sum_{w_k^{l_0} \in V_{i,j}^{l_0}} e_{w_i^{l_1} w_k^{l_0}} + \sum_{w_k^{l_0} \in V_{i,j}^{l_0}} e_{w_j^{l_2} w_k^{l_0}} - 2|V_{i,j}^{l_0}| \cdot x_{w_i^{l_1}, w_j^{l_2}} \geq 0 \tag{3}$$

For example, to $(w_1^{l_1}, w_1^{l_2})$ in Fig. 1, following PB constraint is needed:

$$e_{w_1^{l_1} w_1^{l_0}} + e_{w_1^{l_1} w_2^{l_0}} + e_{w_1^{l_1} w_3^{l_0}} + e_{w_1^{l_2} w_1^{l_0}} + e_{w_1^{l_2} w_2^{l_0}} + e_{w_1^{l_2} w_3^{l_0}} - 6x_{w_1^{l_1}, w_1^{l_2}} \geq 0$$

As for the *Uniqueness* constraint: For any set of one-to-one pair candidates $R^{l_i l_j}$ which commonly share a word w^{l_i} or w^{l_j}, at most one of them is one-to-one pair. This can be expressed by following PB constraint:

Given a set of pairs where all items include a common word $w_i^{l_1}$ or $w_j^{l_2}$; let X' denote the variable set corresponding to these pairs. The Uniqueness constraint is written as follows:

$$\sum_{x_{w^{l_i} w^{l_j}} \in X'} x_{w^{l_i} w^{l_j}} \leq 1 \tag{4}$$

Therefore, extracting one-to-one pairs from a given *transgraph* can be transformed into a PB optimization problem as follows:

$$\text{maximize} \quad \sum_{x_{w^{l_1} w^{l_2}} \in X} x_{w^{l_1} w^{l_2}} - \sum_{e_{w^{l_i} w^{l_0}} \in E^{l_i l_0}} (1 - \omega_{w^{l_i} w^{l_0}}) \cdot e_{w^{l_i} w^{l_0}}$$

subjected to

1) For any one-to-one pair candidate $(w_i^{l_1}, w_j^{l_2})$:

$$\sum_{w_k^{l_0} \in V_{i,j}^{l_0}} e_{w_i^{l_1} w_k^{l_0}} + \sum_{w_k^{l_0} \in V_{i,j}^{l_0}} e_{w_j^{l_2} w_k^{l_0}} - 2|V_{i,j}^{l_0}| \cdot x_{w_i^{l_1}, w_j^{l_2})} \geq 0$$

2) For any set of one-to-one pair candidates:

$$\sum_{x_{w^{l_i} w^{l_j}} \in X'} x_{w^{l_i} w^{l_j}} \leq 1$$

If more than two input dictionaries are involved, we need to extend the *Uniqueness* constraint to keep one-to-one pairs not only unique but also consistent across the intra-family languages. In other word, the words of a one-to-one pair share a same one-to-one equivalent in a third language which is also intra-family. For this reason, it is necessary to add a new PB constraint to any set of three one-to-one pair candidates $\{(w^{l_i}, w^{l_j}), (w^{l_i}, w^{l_k}), (w^{l_j}, w^{l_k})\}$ which consist of three distinct words $\{w^{l_i}, w^{l_j}, w^{l_k}\}$, $i \neq j \neq k$, such that any two of them cannot be seen as the one-to-one pairs if the third one is not a one-to-one pair. Formally, $x_{w^{l_i}w^{l_j}} + x_{w^{l_i}w^{l_k}} + x_{w^{l_j}w^{l_k}} \neq 2$. Unfortunately, PBO does not allow the $\neq$ relation[15]. Therefore, it needs to be translated into an equally valid constraint. This can usually be done by introducing new indicator variable $b \in \{0, 1\}$ as follows:

$$\begin{aligned} &x_{w^{l_i}w^{l_j}} + x_{w^{l_i}w^{l_k}} + x_{w^{l_j}w^{l_k}} - 3b \geq 0 \\ &3b - (x_{w^{l_i}w^{l_j}} + x_{w^{l_i}w^{l_k}} + x_{w^{l_j}w^{l_k}}) \geq -1 \end{aligned} \tag{5}$$

6 Experiment

We implemented the proposal using IBM Cplex[6] as it is an ILP solver commonly used by the ILP community. With this tool, we evaluated our approach by creating new dictionaries from $D_{zh\text{-}ug}$, $D_{zh\text{-}kk}$ and $D_{zh\text{-}kg}$, where *ug* (Uyghur), *kk* (Kazakh) and *kg* (Kyrgyz) are Turkic languages, while *zh* (Chinese) belongs to the Sino-Tibetan language family. Table 1 details these three input directions. Notice that *zh* words whose translation are not available in all three languages have been excluded from the experiment because the proposal does not apply to such cases. Moreover, the number of available *ug*, *kk* and *kg* words are different which indicates that the output dictionaries could have different size. However, we assume that they will have similar precision and recall in evaluating the performance of our proposal.

Table 1. Details of input dictionaries in the experiment

Dictionary	*zh* words	*ug* / *kk* / *kg* words	Pairs
$D_{zh\text{-}ug}$	28, 806	44,400	76,501
$D_{zh\text{-}kk}$	28, 806	61, 000	143,515
$D_{zh\text{-}kg}$	28, 806	27,351	40,381

6.1 Experiment Settings

Total 664 *transgraphs* are formed by merging $D_{zh\text{-}ug}$, $D_{zh\text{-}kk}$ and $D_{zh\text{-}kg}$, from which we selected smaller ones which involve 2289 *ug*, 3264 *kk* and 1634 *kg* words as samples. The evaluation is conducted in two phases with manual determination of precision and recall.

[6] http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/

1. Evaluating the performance of induction in the case of two input dictionaries with comparison to a baseline method: Three input dictionaries were paired into three groups, and each group independently processed to extract a new dictionary. In this phase, we compared our proposal to a baseline method, IC (Inverse Consultation)[7]. Let's denote the output dictionaries produced in this phase as $\mathbb{D}_1 = \{D_{ug-kk}, D_{ug-kg}, D_{kk-kg}\}$.
2. Evaluating the proposal of using more than two input dictionaries: In this phase, we created a new set of dictionaries $\mathbb{D}_2 = \{D_{ug-kk}, D_{ug-kg}, D_{kk-kg}\}$ by processing the same input dictionaries as a single optimization problem. By doing this, we observed what effect the use of an additional input dictionary has on the quality of the output dictionaries.

6.2 Result and Analysis

In the first phase of evaluation, we randomly selected 3×100 sample pairs from newly created $D_{ug-kk} \in \mathbb{D}_1$, and asked a bilingual human to judge whether they are indeed correctly mapped as one-to-one. The results were about precision of 80% and 69% recall are achieved when we assume that the size of the one-to-one space is equal to the maximum of numbers of unique ug and kk words. However, it is not reasonable to directly compare these numbers with one in related works and reach a conclusion on the efficiency of the proposal, since the experimental language pairs and resources chosen in each similar research are not quite the same. In response, we processed the same dataset with the IC method, because it is a well-known approach to creating new dictionary from only two input dictionaries without additional resources and heuristics, hence often used as a baseline method [28][8]. As a result, the proposal yields about 10% higher precision with similar recall 72% of IC.

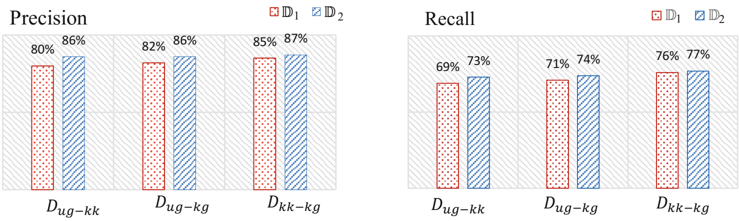

Fig. 3. Precision and recall comparison for the cases of two and three input dictionaries.

In the second phase, we conducted a human evaluation on samples from six output dictionaries in $\mathbb{D}_1$ and $\mathbb{D}_2$. As the details show in Fig.3 and Table 2, both precision and recall were slightly improved when three input dictionaries were processed as a single problem. Although the degree of the improvements varies from one language pair to another, an improvement was achieved in every case. On average, 4%, 2.6% and 4% gains in precision, recall and F_1-measure are achieved, respectively, which prove the efficiency of the proposal in utilizing more input dictionaries, although more experiments with different language pairs and dictionaries and a deeper analysis are essential for a precise conclusion. We attribute these improvements to efficient utilization of most complete parts of each input dictionaries.

Table 2. Details of results for the case of two and three input dictionaries

Input Dictionary	# of 1-to-1 pairs		Precision			Recall			F_1-measure		
	$\mathbb{D}_1$	$\mathbb{D}_2$	$\mathbb{D}_1$	$\mathbb{D}_2$	+/-	$\mathbb{D}_1$	$\mathbb{D}_2$	+/-	$\mathbb{D}_1$	$\mathbb{D}_2$	$\mathbb{D}_2$ over $\mathbb{D}_1$
D_{ug-kk}	1973	1954	80%	86%	+6%	69%	73%	+4%	76%	80%	+6%
D_{ug-kg}	1415	1414	82%	86%	+4%	71%	74%	+3%	76%	80%	+4%
D_{kk-kg}	1465	1457	85%	87%	+2%	76%	77%	+1%	80%	82%	+2%

7 Conclusion

Automatic creation of bilingual dictionaries has always been challenging because many language pairs lack any really useful language resources like a parallel corpora or even comparable corpora. To provide an efficient method for low-resourced language pairs by making use of available bilingual dictionary resources which are possibly incomplete, we presented an extended constraint optimization approach to pivot-based dictionary induction, where new dictionaries of closely related languages are induced from multiple input dictionaries using a distant language as a pivot. Our approach allows the utilization of as many as possible existing dictionaries for improving output performance by taking advantage of most complete part of each dictionary. In this approach, the lexical intransitivity divergence which stems from polysemy and ambiguous words in pivot language is approached by modeling instance of induction as an optimization problem[8], where the new dictionaries are produced as optimal solutions of the problem. Moreover, our proposal considers dropped meanings in the dictionaries to some extent and so efficiently handles low quality and incomplete input dictionaries. An experiment showed the feasibility of our proposal in practice. However, we note the following points: (1) There is a potential of including spelling as additional information; (2) More comparisons are expected to find whether the method can indeed rely purely on dictionary structure and still outperform the methods that utilize cheap external resources such as monolingual data; (3) Applying the proposal to extra-family language pairs is also promising and should be explored.

Acknowledgment. This research was partially supported by Service Science, Solutions and Foundation Integrated Research Program from JST RISTEX, and a Grant-in-Aid for Scientific Research (S) (24220002) from Japan Society for the Promotion of Science.

References

1. Brown, P.F., Cocke, J., Della Pietra, S.A., Jelinek, F., Lafferty, J.D., Mercer, R.L., Roossin, P.S.: A statistical approach to machine translation. Computational Linguistics 16(2), 79–85 (1990)
2. Nie, J.-Y., Simard, M., Isabelle, P., Durand, R.: Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the web. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 74–81. ACM (1999)
3. Wu, D., Xia, X.: Learning an English-Chinese lexicon from a parallel corpus. In: Proceedings of the First Conference of the Association for Machine Translation in the Americas, pp. 206–213. Citeseer (1994)
4. Dou, Q., Knight, K.: Large scale decipherment for out-of-domain machine translation. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 266–275. Association for Computational Linguistics (2012)
5. Yu, K., Tsujii, J.: Bilingual dictionary extraction from wikipedia. In: Proceedings of Machine Translation Summit XII, pp. 379–386 (2009)
6. Haghighi, A., Liang, P., Berg-Kirkpatrick, T., Klein, D.: Learning bilingual lexicons from monolingual corpora. In: Proceedings of ACL 2008: HLT, pp. 771–779 (2008)
7. Tanaka, K., Umemura, K.: Construction of a bilingual dictionary intermediated by a third language. In: Proceedings of the 15th Conference on Computational Linguistics, COLING 1994, vol. 1, pp. 297–303. Association for Computational Linguistics, Stroudsburg (1994)
8. Wushouer, M., Lin, D., Ishida, T., Hirayama, K.: Bilingual dictionary induction as an optimization problem
9. Bond, F., Ogura, K.: Combining linguistic resources to create a machine-tractable Japanese-Malay dictionary. Language Resources and Evaluation 42(2), 127–136 (2008)
10. István, V., Shoichi, Y.: Bilingual dictionary generation for low-resourced language pairs. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, vol. 2, pp. 862–870. Association for Computational Linguistics (2009)
11. Kaji, H., Tamamura, S., Erdenebat, D.: Automatic construction of a Japanese-Chinese dictionary via English. In: LREC, vol. 2008, pp. 699–706 (2008)
12. Sjobergh, J.: Creating a free digital Japanese-Swedish lexicon. In: Proceedings of PACLING, pp. 296–300. Citeseer (2005)
13. Matsuno, J., Ishida, T.: Constraint optimization approach to context based word selection. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, vol. 3, pp. 1846–1851. AAAI Press (2011)
14. Ben Hassine, A., Matsubara, S., Ishida, T.: A constraint-based approach to horizontal web service composition. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 130–143. Springer, Heidelberg (2006)
15. Barth, P., Stadtwald, I.: A davis-putnam based enumeration algorithm for linear pseudo-boolean optimization (1995)
16. Saralegi, X., Manterola, I., Vicente, I.S.: Analyzing methods for improving precision of pivot based bilingual dictionaries. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 846–856. Association for Computational Linguistics (2011)

17. Kaji, H., Aizono, T.: Extracting word correspondences from bilingual corpora based on word co-occurrences information. In: Proceedings of the 16th Conference on Computational Linguistics, vol. 1, pp. 23–28. Association for Computational Linguistics (1996)
18. Bond, F., Yamazaki, T., Sulong, R.B., Okura, K.: Design and construction of a machine–tractable Japanese-Malay lexicon. In: Annual Meeting of the Association for Natural Language Processing, vol. 7, p. 1 (2001)
19. Soderland, S., Etzioni, O., Weld, D.S., Skinner, M., Bilmes, J., et al.: Compiling a massive, multilingual dictionary via probabilistic inference. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, vol. 1, pp. 262–270. Association for Computational Linguistics (2009)
20. Nerima, L., Wehrli, E.: Generating bilingual dictionaries by transitivity. In: LREC (2008)
21. Shezaf, D., Rappoport, A.: Bilingual lexicon generation using non-aligned signatures. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 98–107. Association for Computational Linguistics (2010)
22. Schulz, S., Markó, K., Sbrissia, E., Nohama, P., Hahn, U.: Cognate mapping: A heuristic strategy for the semi-supervised acquisition of a Spanish lexicon from a Portuguese seed lexicon. In: Proceedings of the 20th International Conference on Computational Linguistics, p. 813. Association for Computational Linguistics (2004)
23. Tanaka, R., Murakami, Y., Ishida, T.: Context-based approach for pivot translation services. In: IJCAI, pp. 1555–1561 (2009)
24. Ishida, T.: The Language Grid. Springer (2011)
25. Dan Melamed, I.: Models of translational equivalence among words. Computational Linguistics 26(2), 221–249 (2000)
26. Dan Melamed, I.: A word-to-word model of translational equivalence. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, pp. 490–497. Association for Computational Linguistics (1997)
27. Koehn, P., Knight, K.: Learning a translation lexicon from monolingual corpora. In: Proceedings of the ACL 2002 Workshop on Unsupervised Lexical Acquisition, vol. 9, pp. 9–16. Association for Computational Linguistics (2002)
28. Shezaf, D., Rappoport, A.: Bilingual lexicon generation using non-aligned signatures. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 98–107. Association for Computational Linguistics (2010)
29. Nakov, P., Ng, H.T.: Improving statistical machine translation for a resource-poor language using related resource-rich languages. Journal of Artificial Intelligence Research 44(1), 179–222 (2012)
30. Schrijver, A.: Theory of linear and integer programming. John Wiley & Sons (1998)